

University of Nebraska - Lincoln

DigitalCommons@University of Nebraska - Lincoln

---

Vadim Gladyshev Publications

Biochemistry, Department of

---

April 2008

## Molybdoproteomes and evolution of molybdenum utilization

Yan Zhang

University of Nebraska-Lincoln, yzhang3@unl.edu

Vadim N. Gladyshev

University of Nebraska-Lincoln, vgladyshev@rics.bwh.harvard.edu

Follow this and additional works at: <https://digitalcommons.unl.edu/biochemgladyshev>



Part of the [Biochemistry, Biophysics, and Structural Biology Commons](#)

---

Zhang, Yan and Gladyshev, Vadim N., "Molybdoproteomes and evolution of molybdenum utilization" (2008). *Vadim Gladyshev Publications*. 78.

<https://digitalcommons.unl.edu/biochemgladyshev/78>

This Article is brought to you for free and open access by the Biochemistry, Department of at DigitalCommons@University of Nebraska - Lincoln. It has been accepted for inclusion in Vadim Gladyshev Publications by an authorized administrator of DigitalCommons@University of Nebraska - Lincoln.

# Molybdoproteomes and evolution of molybdenum utilization

Yan Zhang and Vadim N. Gladyshev\*

Department of Biochemistry, University of Nebraska–Lincoln, Lincoln, NE 685880664

\*Corresponding author — tel 402 472-4948, fax 402 472-7842, email [vgladyshev1@unl.edu](mailto:vgladyshev1@unl.edu)

## Abstract

The trace element molybdenum (Mo) is utilized in many life forms, where it is a key component of several enzymes involved in nitrogen, sulfur, and carbon metabolism. With the exception of nitrogenase, Mo is bound in proteins to a pterin, thus forming the molybdenum cofactor (Moco) at the catalytic sites of molybdoenzymes. Although a number of molybdoenzymes are well characterized structurally and functionally, evolutionary analyses of Mo utilization are limited. Here, we carried out comparative genomic and phylogenetic analyses to examine occurrence and evolution of Mo utilization in bacteria, archaea, and eukaryotes at the level of (i) Mo transport and Moco utilization trait, and (ii) Mo-dependent enzymes. Our results revealed that most prokaryotes and all higher eukaryotes utilize Mo, whereas many unicellular eukaryotes, including parasites and most yeasts, lost the ability to use this metal. In addition, eukaryotes have fewer molybdoenzyme families than prokaryotes. Dimethylsulfoxide reductase (DMSOR) and sulfite oxidase (SO) families were the most widespread molybdoenzymes in prokaryotes and eukaryotes, respectively. A distant group of the ModABC transport system, was predicted in the hyperthermophilic archaeon *Pyrobaculum*. ModE-type regulation of Mo uptake occurred in less than 30% of Moco-utilizing organisms. A link between Mo and selenocysteine utilization in prokaryotes was also identified wherein the selenocysteine trait was largely a subset of the Mo trait, presumably due to formate dehydrogenase, a Mo- and selenium-containing protein. Finally, analysis of environmental conditions and organisms that do or do not depend on Mo revealed that host-associated organisms and organisms with low GC content tend to show reduced Mo utilization. Overall, our data provide new insights into Mo utilization and show the wide occurrence, yet limited use, of this metal in individual organisms in all three domains of life.

**Keywords:** molybdenum, molybdopterin, molybdoenzyme, comparative genomics, evolution

## Introduction

The trace element molybdenum (Mo) occurs in a wide variety of metalloenzymes in both prokaryotes and eukaryotes, where it forms part of active sites of these enzymes.<sup>1–3</sup> Except for iron-Mo cofactor (FeMoco) in nitrogenase,<sup>4</sup> Mo is complexed by pterin molecules, thereby generating the molybdenum cofactor (Moco or molybdopterin, MPT) in Mo-dependent enzymes (molybdoenzymes).<sup>5–7</sup> Some microorganisms are able to utilize tungsten (W) that is also coordinated by MPT.<sup>8</sup> As a result, the term Moco refers to the utilization of both metals.

Moco-containing enzymes catalyze important redox reactions in the global carbon, nitrogen, and sulfur cycles.<sup>2</sup> More than 50 Mo-enzymes, mostly of bacterial origin, have been previously identified.<sup>2, 3, 9</sup> On the basis of sequence comparison and spectroscopic properties, these Moco-containing enzymes are divided into four

families: sulfite oxidase (SO), xanthine oxidase (XO), dimethylsulfoxide reductase (DMSOR), and aldehyde:ferredoxin oxidoreductase (AOR).<sup>10, 11</sup> Each family is further divided into different subfamilies based on the use of their specific substrates. For example, the DMSOR family also includes trimethylamine-N-oxide reductase, biotin sulfoxide reductase, nitrate reductase (dissimilatory), formate dehydrogenase, and arsenite oxidase. All four of these families can be detected in prokaryotes, however, only two families (SO and XO), containing four subfamilies, occur in eukaryotes. The SO family includes nitrate reductase (NR) and SO, whereas the XO family is represented by xanthine dehydrogenase (XDH) and aldehyde oxidase (AO). These enzymes are typical for essentially all Mo-utilizing eukaryotes analyzed thus far. Recently, two additional Moco-binding enzymes were reported: pyridoxal oxidase and nicotinate hydroxylase, which were exclusively found in *Drosophila melanogaster* and *Aspergillus nidulans*, respectively.<sup>7</sup>

Functions of molybdoenzymes depend on additional gene products that transport molybdate anions into cells and synthesize and assemble Moco. In bacteria, high-affinity molybdate ABC transporters (ModABC, products of *modABC* genes) have been described that consist of ModA (molybdate-binding protein), ModB (membrane integral channel protein) and ModC (cytoplasmic ATPase).<sup>7, 12, 13</sup> In addition, a new class of the Mo/W transport system (WtpABC) and a highly specific tungstate ABC transporter (TupABC) have been reported.<sup>14, 15</sup> Although both transporter systems exhibited low sequence similarity to ModABC transporters, they showed different anion affinity than ModA. TupA specifically binds tungstate, whereas WtpA has a higher affinity for tungstate than ModA and its affinity for molybdate is similar to that of ModA.<sup>14, 15</sup> In contrast to bacteria, eukaryotic molybdate transport is poorly understood, but recent studies in *Arabidopsis thaliana* suggested the occurrence of a high-affinity molybdate transport system, MOT1.<sup>16</sup>

In *E. coli*, the *modABC* operon is regulated by a repressor protein, ModE, which also controls the transcription of genes coding for molybdopterin synthesis (*moaABCDE*), and molybdoenzymes.<sup>17-20</sup> *E. coli* ModE is composed of an N-terminal DNA-binding domain (ModE\_N, COG2005) and a C-terminal molybdate-binding domain.<sup>17, 18, 21</sup> The C-terminal domain contains a tandem repeat of the MPT-binding protein (Mop, COG3585; also referred to as Di-Mop domain).<sup>18</sup> The ModABC-ModE systems are widespread in prokaryotes, but not ubiquitous.<sup>22-25</sup> Variations of ModE-like proteins were also observed in other Moco-utilizing organisms.<sup>25, 26</sup> On the other hand, regulation of WtpABC and TupABC transporters is unclear.

In organisms studied thus far with respect to Moco utilization (e.g., bacteria, plants, fungi, and mammals), this cofactor is synthesized by a conserved multistep biosynthetic pathway.<sup>7</sup> The first model of Moco biosynthesis was derived from studies in *E. coli*.<sup>6</sup> In this organism, the proteins required for biosynthesis and regulation of the pterin cofactor are encoded by the *moa-mog* operon.<sup>27, 28</sup> The *moa* and *moe* operons are responsible for biosynthesis of the mononucleotide form of pterin cofactor, and the *mob* operon encodes pterin guanine dinucleotide synthase that adds GMP to the Mo-complexed pterin cofactor. Functions of other operons linked to Mo utilization are unclear. In eukaryotes, six gene products catalyze Moco biosynthesis that have been studied in plants (Cnx1-3, Cnx5-7),<sup>28</sup> fungi,<sup>29</sup> and humans.<sup>30-32</sup> Although these proteins are homologous to their counterparts in bacteria, not all of the eukaryotic Moco biosynthesis machinery could functionally complement the corresponding bacterial mutant strains. Different nomenclature has been used in humans and plants,<sup>30</sup> and in this work we use the plant nomenclature to refer to the eukaryotic Moco synthetic genes.

In recent years, the complete genomes of many organisms from the three domains of life became available. It is now possible to examine occurrence and evolution of numerous biochemical pathways that an organism utilizes, including metal utilization. Several comparative and functional genomic analyses have been carried out for different trace elements.<sup>33-38</sup> However, a comprehensive investigation of either Moco biosynthesis systems or Mo-containing enzymes has not been performed.

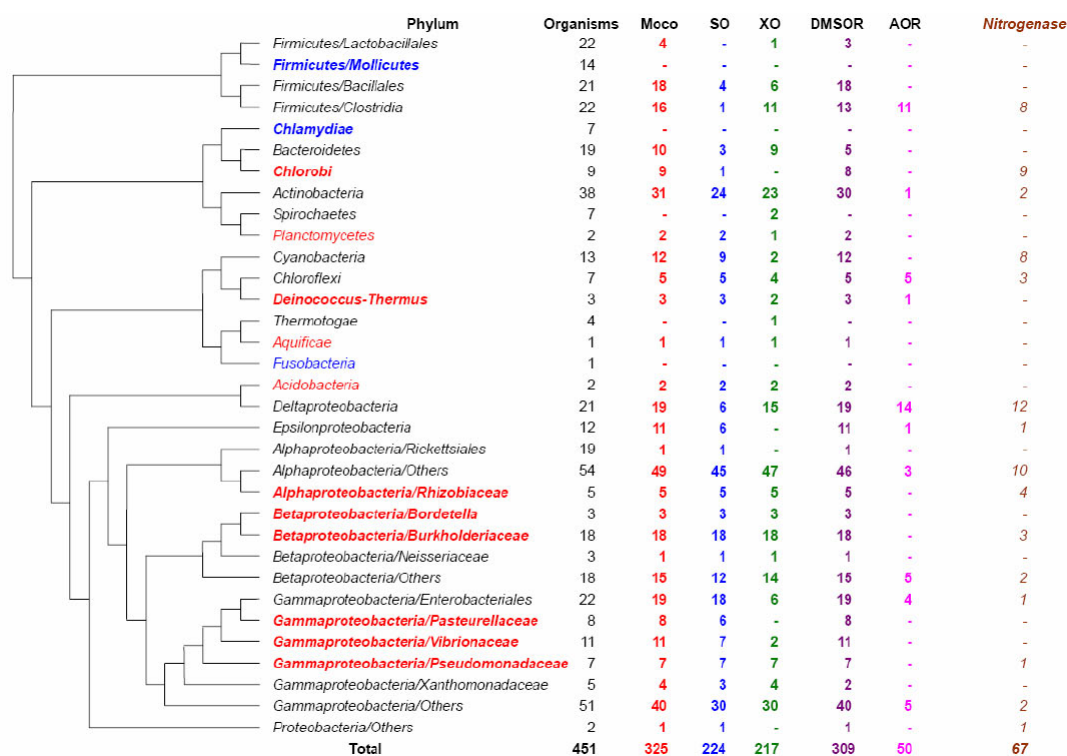
In this study, we used comparative genomic analyses to better understand Mo utilization in various life forms. Our data showed a widespread utilization of Mo in all three domains of life and revealed that evolutionary changes in Mo utilization can be influenced by various factors. Our results also highlight complexity of regulation of the Mo/W uptake systems. Moreover, the relationship between Mo and selenium (Se) utilization in prokaryotes suggests a possibility that Se utilization may be dependent on Mo. These studies reveal widespread utilization of Mo in various life forms and its limited use in individual organisms, and are important for understanding the evolution of both Mo utilization trait and molybdoenzymes.

## Results

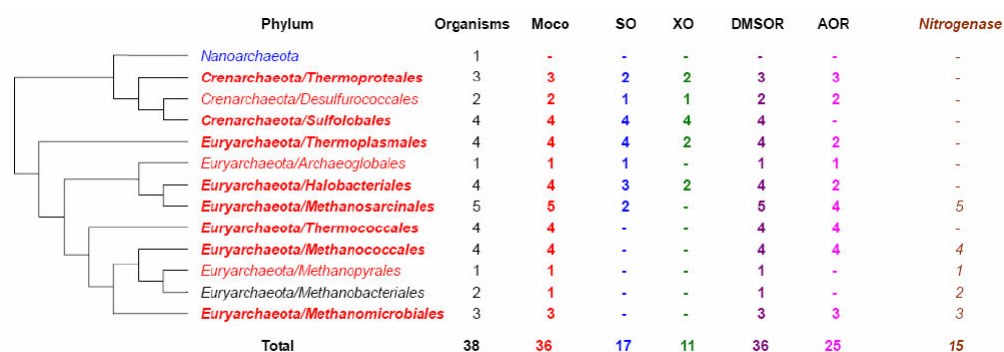
### *Occurrence of Mo utilization in prokaryotes and eukaryotes*

Analysis of prokaryotic genomes revealed a wide distribution of genes encoding Moco biosynthesis pathway and Mo-containing proteins (a complete list is in Table S1). Almost all organisms were found either to possess both Moco biosynthesis proteins and known molybdoenzymes or lack them, suggesting a very good correspondence between the occurrence of Moco biosynthesis trait and Moco-dependent enzymes. In total, 325 (~72.1%) bacterial organisms were found to utilize Moco. Figure 1 shows the distribution of Moco biosynthesis trait and Moco-containing protein families in different bacterial taxa based on a highly resolved phylogenetic tree of life.<sup>39</sup>

Except for the phyla containing few sequenced genomes (<3, for example, *Planctomycetes*, *Aquificae*, and *Acidobacteria*), Mo was found to be utilized by almost all bacterial phyla. All sequenced organisms in *Chlorobi*, *Deinococcus-Thermus*, *Alphaproteobacteria/Rhizobiaceae*, *Betaproteobacteria/Bordetella*, *Betaproteobacteria/Burkholderiaceae*, *Gammaproteobacteria/Pasteurellaceae*, *Gammaproteobacteria/Vibrionaceae*, and *Gammaproteobacteria/Pseudomonadaceae*, as well as the majority of *Cyanobacteria* (92.3%), *Epsilonproteobacteria* (91.7%), *Deltaproteobacteria* (90.5%), *Gammaproteobacteria/Enterobacteriales* (86.4%), and many other bacterial subdivisions utilize Moco. In contrast, neither Moco biosynthesis trait components nor Moco-containing proteins were detected in *Firmicutes/Mollicutes*, and *Chlamydiae*. It should be noted that we found orphan XO homologs in five com-



**Figure 1.** Occurrence of Moco biosynthesis pathway and molybdoenzymes in bacteria. The tree is based on the bacterial part of a highly resolved phylogenetic tree of life.<sup>27</sup> Moco, molybdopterin cofactor biosynthesis pathway; SO, sulfite oxidase; XO, xanthine oxidase; DMSOR, dimethylsulfoxide reductase; AOR, aldehyde:ferredoxin oxidoreductase. Phyla in which none of the organisms possess Moco biosynthesis pathway are shown in blue (if containing at least 3 organisms, shown in bold and blue). Phyla in which all organisms possess the Moco biosynthesis pathway are shown in red (if containing at least 3 organisms, shown in bold and red).



**Figure 2.** Occurrence of Mo utilization and Moco-containing proteins in archaea. Moco, molybdopterin cofactor biosynthesis pathway; SO, sulfite oxidase; XO, xanthine oxidase; DMSOR, dimethylsulfoxide reductase; AOR, aldehyde:ferredoxin oxidoreductase. Phyla in which none of the organisms possess the Moco biosynthesis pathway are shown in blue (if containing at least 3 organisms, shown in bold and blue). Phyla in which all organisms possess Moco biosynthesis pathway are shown in red (if containing at least 3 organisms, shown in bold and red).

pletely sequenced organisms belonging to *Deltaproteobacteria*, *Firmicutes/Clostridia*, *Spirochaetes*, and *Thermotogae*, which lack genes for either Mo/W transporters or known Moco biosynthesis trait components (see Table S1). This observation suggests either that there may be an unknown Mo utilization pathway in these organisms (unlikely scenario) or that they use other proteins that functionally replace XO and other molybdoproteins. It is also possible that the functions carried out by molybdo-

proteins are dispensable in these organisms. Nevertheless, the wide distribution of Moco utilization observed in the present study suggests that, in addition to several metal ions utilized by all or most organisms, e.g., iron, zinc, and magnesium, Mo also shows widespread occurrence in bacteria.

An even wider Mo utilization was observed in archaea (Figure 2). About 95% of sequenced archaeal organisms were found to utilize Moco. Thus, it appears

Phylum	Organisms	Moco	SO	XO
<i>Diplomonadida</i>	1	-	-	-
<i>Parabasalidea</i>	1	-	-	1
<i>Kinetoplastida</i>	4	-	-	-
<b><i>Stramenopiles</i></b>	4	<b>4</b>	<b>4</b>	<b>4</b>
<b><i>Alveolata/Apicomplexa</i></b>	12	-	-	-
<i>Alveolata/Ciliophora</i>	2	-	-	-
<i>Rhodophyta</i>	1	<b>1</b>	<b>1</b>	-
<i>Viridiplantae/Chlorophyta</i>	2	<b>2</b>	<b>2</b>	<b>1</b>
<b><i>Viridiplantae/Streptophyta</i></b>	3	<b>3</b>	<b>3</b>	<b>3</b>
<i>Dictyosteliida</i>	1	<b>1</b>	<b>1</b>	<b>1</b>
<b><i>Entamoebidae</i></b>	3	-	-	-
<b><i>Fungi/Ascomycota/Pezizomycotina</i></b>	22	<b>22</b>	<b>22</b>	<b>22</b>
<b><i>Fungi/Ascomycota/Saccharomycotina</i></b>	22	-	-	-
<i>Fungi/Ascomycota/Schizosaccharomycetes</i>	2	-	-	-
<i>Fungi/Basidiomycota</i>	8	<b>3</b>	<b>3</b>	<b>1</b>
<i>Fungi/Microsporidia</i>	1	-	-	-
<i>Fungi/Zygomycota</i>	1	<b>1</b>	<b>1</b>	-
<b><i>Metazoa/Pseudocoelomata/Nematoda</i></b>	3	<b>3</b>	<b>3</b>	<b>3</b>
<i>Metazoa/Coelomata/Others</i>	2	<b>2</b>	<b>2</b>	<b>2</b>
<b><i>Metazoa/Coelomata/Arthropoda (Insects)</i></b>	17	<b>17</b>	<b>17</b>	<b>17</b>
<b><i>Metazoa/Coelomata/Chordata/Others (Fish, Frog)</i></b>	9	<b>9</b>	<b>9</b>	<b>9</b>
<b><i>Metazoa/Coelomata/Chordata/Mammals</i></b>	21	<b>21</b>	<b>21</b>	<b>21</b>
<b>Total</b>	<b>142</b>	<b>89</b>	<b>89</b>	<b>85</b>

**Figure 3.** Occurrence of Mo utilization and Moco-containing proteins in eukaryotes. Moco, molybdopterin cofactor biosynthesis pathway; SO, sulfite oxidase; XO, xanthine oxidase. Phyla in which none of the organisms possess the Moco biosynthesis pathway are shown in blue (if containing at least 3 organisms, shown in bold and blue). Phyla in which all organisms possess the Moco biosynthesis pathway are shown in red (if containing at least 3 organisms, shown in bold and red).

that Mo utilization is an ancient and essential trait which is common to essentially all species in this domain of life as well as in bacteria.

In eukaryotes, the only known use of Mo is Moco. Our analysis identified 89 (62.7%) Mo-dependent organisms (Figure 3, details are shown in Table S1). All animals, land plants, algae, stramenopiles (including diatoms and oomycetes), and certain fungi (all *Pezizomycotina* and some *Basidiomycota*) possess Moco biosynthesis genes and known molybdoenzymes. However, MOT1 molybdate transporter<sup>16</sup> was only found in one third of Mo-utilizing eukaryotes, which are land plants, green algae, peizizomycotina, and stramenopiles. In contrast, all parasites (14.8%, including *Alveolata/Apicomplexa*, *Entamoebidae*, *Kinetoplastida*, *Parabasalidea*, and *Diplomonadida*), yeasts (21.1%, including *Saccharomycotina* and *Schizosaccharomycetes*), and free-living ciliates (1.4%, *Alveolata/Ciliophora*) lack Mo biosynthesis proteins, molybdoenzymes, and MOT1 transporters. Since Mo utilization is widespread in all three domains of life, it appears that many protozoa, especially parasites, lost the ability to utilize Mo. A unique exception was the detection of an orphan XO in a parasitic flagellated protozoan, *Trichomonas vaginalis* (*Parabasalidea* phylum). Considering that its genome sequence is not fully completed, it is possible that the Mo biosynthesis proteins could correspond to unfinished sequences. Alternatively, this organism may rely on uptake of Moco from the host.

#### *Distribution and phylogeny of Mo or W transporters*

We analyzed both well-characterized Mo ABC transport system (ModABC) and two secondary systems: WtpABC and TupABC (W-specific) in prokaryotes. A summary of the distribution of these Mo/W transporter families is shown in Table 1. In bacteria, 294 organisms that account for 90.5% of Mo-utilizing bacteria possess ModABC transporter. Occurrence of the other two systems is more restricted, especially of WtpABC which was identified in only 10 organisms. The W-specific transporter TupABC was found in 85 (26.2%) Moco-utilizing organisms. In contrast, the distribution of these transporters in archaea was different. WtpABC was the most common transporter that was found in 23 (63.9%) Mo-utilizing organisms, whereas ModABC and TupABC systems showed lower occurrence (38.9% and 33.3% respectively). These data are consistent with the previous hypothesis that WtpABC is an archaeal Mo/W transporter, whereas ModABC and TupABC occur predominantly in bacteria.<sup>14</sup>

Phylogenetic analysis was used to further examine the evolutionary relationships of Mo/W transport systems in different organisms. We used ModA (periplasmic component of the ModABC transport system), WtpA (periplasmic component of the WtpABC transport system), and TupA (periplasmic component of the TupABC transport system) to build a phylogenetic tree (Figure 4). First, all orthologs of the three different fam-

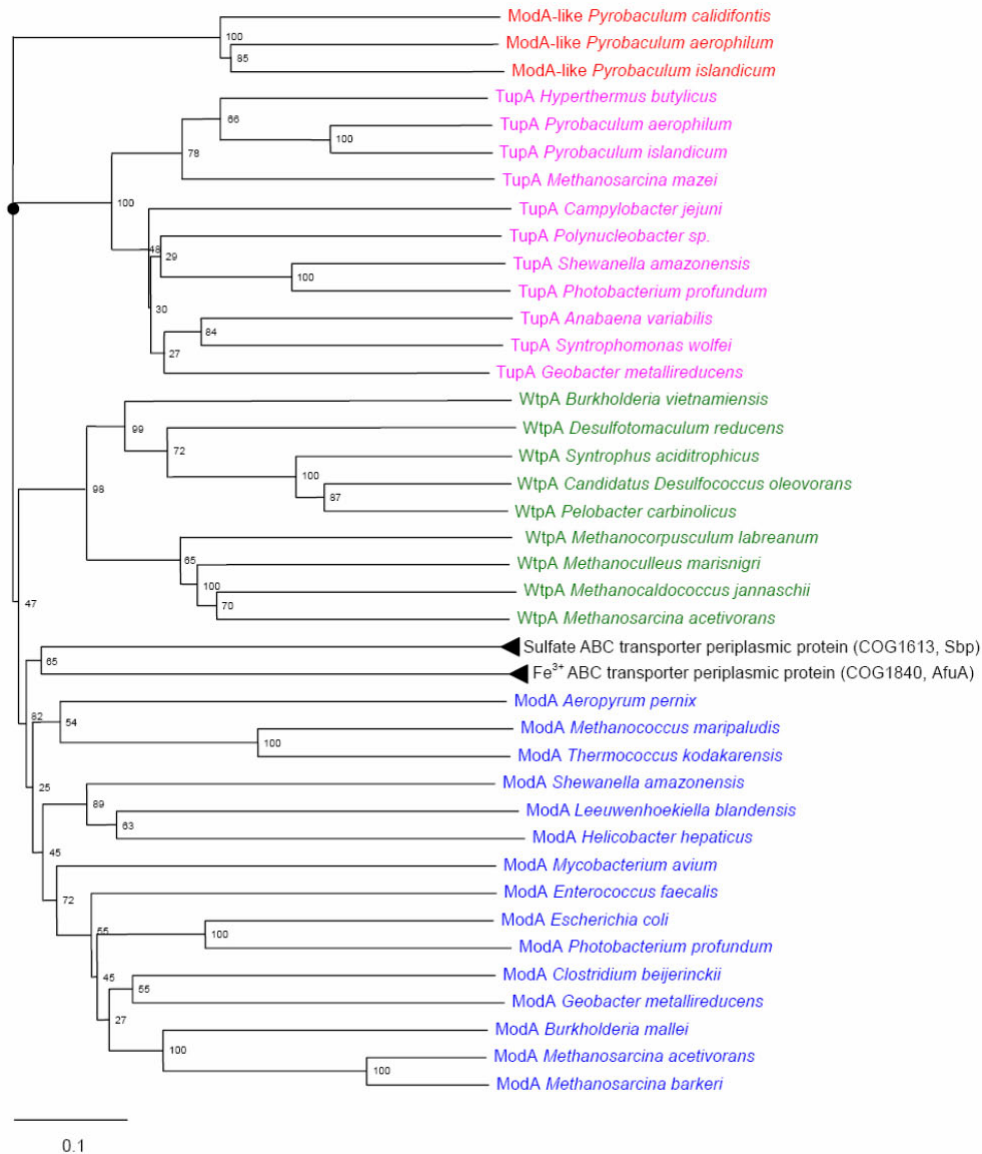
**Table 1.** Distribution of known Mo/W transporters in different prokaryotic phyla

Kingdom	Moco-utilizing organisms	ModABC transporter (Mo/W)	WtpABC transporter (Mo/W)	TupABC transporter (W)	No known Mo/W transporters
<b>Archaea (Total)</b>	<b>36</b>	<b>14</b>	<b>23</b>	<b>12</b>	<b>2</b>
<i>Crenarchaeota/Desulfurococcales</i>	2	2	-	1	-
<i>Crenarchaeota/Sulfolobales</i>	4	-	4	-	-
<i>Crenarchaeota/Thermoproteales</i>	3	-	-	3	-
<i>Euryarchaeota/Archaeoglobales</i>	1	-	1	-	-
<i>Euryarchaeota/Halobacteriales</i>	4	-	4	2	-
<i>Euryarchaeota/Methanobacteriales</i>	1	1	-	1	-
<i>Euryarchaeota/Methanococcales</i>	4	3	4	-	-
<i>Euryarchaeota/Methanomicrobiales</i>	3	2	3	3	-
<i>Euryarchaeota/Methanopyrales</i>	1	1	-	-	-
<i>Euryarchaeota/Methanosarcinales</i>	5	4	1	2	-
<i>Euryarchaeota/Thermococcales</i>	4	1	4	-	-
<i>Euryarchaeota/Thermoplasmatales</i>	4	-	2	-	2
<i>Nanoarchaeota</i>	-	-	-	-	-
<b>Bacteria (Total)</b>	<b>325</b>	<b>294</b>	<b>10</b>	<b>85</b>	<b>24</b>
<i>Actinobacteria</i>	31	31	-	-	-
<i>Bacteroidetes</i>	10	2	-	-	8
<i>Chlorobi</i>	9	9	-	-	-
<i>Chlamydiae</i>	-	-	-	-	-
<i>Cyanobacteria</i>	12	11	-	1	1
<i>Firmicutes/Bacillales</i>	18	18	-	-	-
<i>Firmicutes/Clostridia</i>	16	10	1	9	1
<i>Firmicutes/Lactobacillales</i>	4	2	-	1	2
<i>Firmicutes/Mollicutes</i>	-	-	-	-	-
<i>Acidobacteria</i>	2	2	-	-	-
<i>Aquificae</i>	1	1	-	-	-
<i>Planctomycetes</i>	2	1	-	-	1
<i>Chloroflexi</i>	5	6*	-	-	-
<i>Deinococcus-Thermus</i>	3	3	-	1	-
<i>Fusobacteria</i>	-	-	-	-	-
<i>Thermotogae</i>	-	-	-	-	-
<i>Proteobacteria/others</i>	1	1	-	-	-
<i>Alphaproteobacteria/others</i>	49	41	1	16	6
<i>Alphaproteobacteria/Rhizobiaceae</i>	5	5	-	1	-
<i>Alphaproteobacteria/Rickettsiales</i>	1	-	-	-	1
<i>Betaproteobacteria/Bordetella</i>	3	3	-	3	-
<i>Betaproteobacteria/Burkholderiaceae</i>	18	18	2	5	-
<i>Betaproteobacteria/Neisseriaceae</i>	1	1	-	-	-
<i>Betaproteobacteria/others</i>	15	14	1	7	1
<i>Deltaproteobacteria</i>	19	18	4	13	1
<i>Epsilonproteobacteria</i>	11	11	-	5	-
<i>Gammaproteobacteria/Enterobacteriales</i>	19	18	1	-	-
<i>Gammaproteobacteria/Others</i>	40	38	-	12	2
<i>Gammaproteobacteria/Pasteurellaceae</i>	8	8	-	-	-
<i>Gammaproteobacteria/Pseudomonadaceae</i>	7	7	-	-	-
<i>Gammaproteobacteria/Vibrionaceae</i>	11	11	-	11	-
<i>Gammaproteobacteria/Xanthomonadaceae</i>	4	4	-	-	-
<i>Spirochaetales</i>	-	-	-	-	-

\*: includes one organism which has nitrogenase but lacks Moco biosynthesis pathway and Moco-dependent proteins

ilies were used to generate a preliminary tree (see Materials and Methods). Representative sequences were then manually selected to condense the original tree without changing its topology. In addition, the periplasmic components of sulfate and Fe<sup>3+</sup> transporters which show low similarity to ModA were used for reference. Robustness of the phylogenetic tree was evaluated with additional programs, which showed a similar topology (see Ma-

terials and Methods and Figure S1). It should be noted that although WtpA and ModA sequences belong to the same COG (COG0725), they showed different anion affinities based on previous experimental analysis.<sup>14</sup> In the phylogenetic tree, they cluster in different branches, suggesting that they derived from a common ancestral gene and have since diverged from the parent copy by mutation and selection or drift.



**Figure 4.** Phylogenetic tree of periplasmic components of Mo/W transporters in prokaryotes. ModA-like proteins are shown in red and bold, TupA in pink, WtpA in green, and ModA in blue. Representative sequences were selected from a large number of orthologous proteins based on sequence similarities. The sulfate and Fe<sup>3+</sup> ABC transporter branches were compressed and represented by family names. The measurement of distance for the branch lengths (shown by a bar) is indicated.

Distant ModA-like proteins were identified in several *Pyrobaculum* species which are hyperthermophilic archaea. Blast-based pairwise alignment showed less than 25% similarity between these ModA-like proteins and *E. coli* ModA or *Pyrococcus furiosus* WtpA. Phylogenetic analysis also suggested they are outgroups of all known ModA proteins (Figure 4). However, they belong to the same COG (e-value 2e-17) as ModA. We further examined the genomic context of *modA*-like genes and the conservation of residues involved in molybdate binding in *E. coli* ModA (1AMF)<sup>40</sup> and tungstate binding in *Archaeoglobus fulgidus* WtpA (2ONS).<sup>41</sup> These *modA*-like genes were always located in an operon containing a complete ABC transport system in-

cluding an ABC-type permease and an ATPase component. Both components were distantly homologous to ModB and ModC, respectively (similarity <25% and e-value >0.1 based on BLAST pairwise alignment). In addition, in one organism, *Pyrobaculum islandicum*, the *modA*-like gene was located next to *modD* gene, which is present in some *modABC* operons in prokaryotes and is involved in molybdate transport (its exact function is unclear).<sup>12, 13</sup> Multiple alignment of ModA, WtpA, and ModA-like sequences revealed that two or three out of five residues involved in Mo binding in *E. coli* ModA (Ser36, Ser63, and Tyr194)<sup>40</sup> were conserved in these ModA-like sequences (Figure S2 and S3). The other two residues (Ala149 and Val176), which only provide the

backbone hydrogen to form hydrogen bonds with molybdate,<sup>40</sup> were not strictly conserved, but other amino acids may similarly provide ligands to the metal ion. These data suggest that the *Pyrobaculum* ModA-like proteins should be considered as a distant group of the ModA family. The absence of ModA-like proteins in other sequenced organisms suggests a limited distribution of this subfamily.

We also found that several completely sequenced organisms, including 2 archaea and 24 bacteria, which contained both Moco biosynthesis pathway and Moco-containing enzymes, did not possess any of the known transporters. Most of these organisms were distantly related, free-living organisms. This observation suggests that additional Mo/W uptake systems may exist. We examined genes in Moco biosynthesis operons in these organisms, however, no good candidates for new Mo/W-specific transport system could be found. It is also possible that molybdate is transported by either sulfate transport system or nonspecific anion transporter in these organisms.

MOT1 is the only known Mo transporter in eukaryotes, which was recently identified in *A. thaliana*.<sup>16</sup> In this study, we analyzed the occurrence of this transporter in sequenced eukaryotic genomes. Among 89 Mo-utilizing organisms, only 31 possess MOT1 orthologs, including *Fungi/Ascomycota/Pezizomycotina*, land plants (*Viridiplantae/Streptophyta*), green algae (*Viridiplantae/Chlorophyta*), and stramenopiles. The absence of MOT1 in all animals implied the presence of a currently unknown Mo transport system in these organisms.

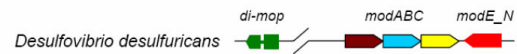
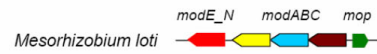
### Regulation of Mo/W transporters

In *E. coli*, the ModABC repressor, ModE, is positioned immediately upstream and transcribed divergently from the *modABC* operon (Figure 5A). However, full-length ModE orthologs were absent in many other organisms such as the Gram-positive Bacteria and *Cyanobacteria*.<sup>25</sup> In addition, various domain fusions were observed for ModE\_N or Mop, indicating complexity of ModE regulation.<sup>25</sup> Although the roles of these variants are unclear, they have been suggested to be non-functional in ModABC regulation.<sup>25</sup> In this study, we analyzed the occurrence of full-length ModE and its variants (including separate ModE\_N, Mop/Di-Mop proteins as well as their additional fusion forms) in sequenced prokaryotes. Here, only the full-length ModE orthologs were considered as true regulators of ModABC transporters. The results are shown in Table 2 (a complete distribution is shown in Table S1). Only a small portion of Moco-utilizing organisms (28.9% and 16.7% in bacteria and archaea, respectively) possessed a full-length ModE, suggesting that most prokaryotes may use additional or unspecific repressors for ModABC regulation.

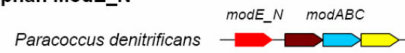
#### A. Full-length ModE (*E. coli*-type)



#### B. ModE\_N + Mop/Di-Mop



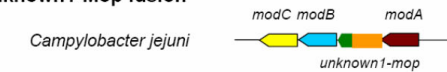
#### C. Orphan ModE\_N



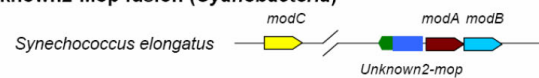
#### D. MerR-Mop fusion



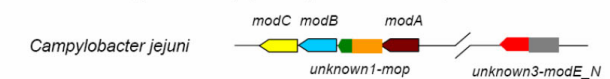
#### E. Unknown1-Mop fusion



#### F. Unknown2-Mop fusion (*Cyanobacteria*)



#### G. Unknown3-ModE\_N fusion (*Epsilonproteobacteria*)



#### H. ModE\_N-COG1910 fusion



**Figure 5.** Genomic organization of ModABC, ModE, and different ModE variants in Moco-utilizing organisms. Different genes in representative genomes are shown by indicated color schemes. (A). Full-length ModE (*E. coli*-type); (B). ModE\_N + Mop/Di-Mop; (C). Orphan ModE\_N; (D). MerR-Mop fusion; (E). Unknown1-Mop fusion; (F). *Cyanobacteria*-specific unknown2-Mop fusion; (G). *Epsilonproteobacteria*-specific Unknown3-ModE\_N fusion; (H). ModE\_N-COG1910 fusion.

In bacteria, some ModABC-containing organisms, which lack ModE, have separate ModE\_N and Mop/Di-Mop proteins or orphan ModE\_N proteins (Table 2, Figure 5B and 5C). In addition, 5 different types of domain fusions were identified for Mop (3 types, Figure 5D–5F), and ModE\_N (2 types, Figure 5G and 5H), mostly in bacteria. Analysis of genomic locations of both separate domains and fusion proteins revealed that, except for two ModE\_N fusion proteins (including a ModE\_N-COG1910 fusion protein which was suggested to regulate the transcription of formate dehydrogenase, as well as a epsilonproteobacteria-specific unknown 3-ModE\_N fusion protein which might be a transcriptional activator rather than a repressor),<sup>25</sup> genes coding for these proteins are close to or even within the *modABC* operon, suggesting functional relationship with ModABC transporters (Figure 5B–5F). Orthologs of these



Table 2. Distribution of ModE, ModE-related and other fusion proteins in prokaryotes

Phylum	Moco-utilizing organisms	ModA BC	ModE (+)	ModE (-)		
				ModE_N + Mop/Di-Mop (excluding fusion)	Orphan ModE_N (excluding fusion)	ModE_N/Mop fused with other domains
<b>Archaea (Total)</b>	<b>36</b>	<b>14</b>	<b>6</b>	<b>-</b>	<b>10</b>	<b>3 (2 also contain ModE_N)</b>
<i>Crenarchaeota/Desulfurococcales</i>	2	2	-	-	-	ModE_N-COG1910: 1
<i>Crenarchaeota/Sulfolobales</i>	4	-	-	-	4	-
<i>Crenarchaeota/Thermoproteales</i>	3	-	-	-	3	ModE_N-COG1910: 2 (all contain ModE_N)
<i>Euryarchaeota/Archaeoglobales</i>	1	-	-	-	1	-
<i>Euryarchaeota/Halobacteriales</i>	4	-	2	-	-	-
<i>Euryarchaeota/Methanobacteriales</i>	1	1	-	-	1	-
<i>Euryarchaeota/Methanococcales</i>	4	3	-	-	-	-
<i>Euryarchaeota/Methanomicrobiales</i>	3	2	-	-	-	-
<i>Euryarchaeota/Methanopyrales</i>	1	1	-	-	1	-
<i>Euryarchaeota/Methanosarcinales</i>	5	4	4	-	-	-
<i>Euryarchaeota/Thermococcales</i>	4	1	-	-	-	-
<i>Euryarchaeota/Thermoplasmatales</i>	4	-	-	-	-	-
<b>Bacteria (Total)</b>	<b>325</b>	<b>294</b>	<b>94</b>	<b>9</b>	<b>36</b>	<b>64 (33 also contain ModE or ModE_N)</b>
<i>Actinobacteria</i>	31	31	-	-	1	MerR-Mop: 17
<i>Bacteroidetes</i>	10	2	-	-	1	Unknown1-Mop: 2
<i>Chlorobi</i>	9	9	9	-	-	Unknown1-Mop: 5 (all contain ModE)
<i>Chlamydiae</i>	-	-	-	-	-	-
<i>Cyanobacteria</i>	12	11	-	-	-	Unknown2-Mop: 4
<i>Firmicutes/Bacillales</i>	18	18	-	-	-	-
<i>Firmicutes/Clostridia</i>	16	10	-	1	3	-
<i>Firmicutes/Lactobacillales</i>	4	2	-	-	2	-
<i>Firmicutes/Mollicutes</i>	-	-	-	-	-	-
<i>Acidobacteria</i>	2	2	-	-	-	MerR-Mop : 1
<i>Aquificae</i>	1	1	-	-	-	-
<i>Planctomycetes</i>	2	1	-	-	-	-
<i>Chloroflexi</i>	5	6**	-	-	-	-
<i>Deinococcus-Thermus</i>	3	3	-	-	-	-
<i>Fusobacteria</i>	-	-	-	-	-	-
<i>Thermotogae</i>	-	-	-	-	-	-
<i>Proteobacteria/others</i>	1	1	1	-	-	-
<i>Alphaproteobacteria/others</i>	49	41	7	3	14	ModE_N-COG1910: 2 (1 contains ModE_N + Mop) MerR-Mop : 1 (1 contains ModE)
<i>Alphaproteobacteria/Rhizobiaceae</i>	5	5	-	-	5	-
<i>Alphaproteobacteria/Rickettsiales</i>	1	-	-	-	-	-
<i>Betaproteobacteria/Bordetella</i>	3	3	3	-	-	ModE_N-COG1910: 3 (all contain ModE)
<i>Betaproteobacteria/Burkholderiaceae</i>	18	18	17	-	-	ModE_N-COG1910: 18 (17 contains ModE)
<i>Betaproteobacteria/Neisseriaceae</i>	1	1	-	-	-	-
<i>Betaproteobacteria/others</i>	15	14	8	1	4	ModE_N-COG1910: 7 (4 contain ModE, 2 contain ModE_N)
<i>Deltaproteobacteria</i>	19	18	5	3	3	-
<i>Epsilonproteobacteria</i>	11	11	5	-	1	Unknown1-Mop & Unknown3-ModE_N: 4
<i>Gammaproteobacteria/Enterobacteriales</i>	19	18	18	-	1	-
<i>Gammaproteobacteria/Others</i>	40	38	7	1	1	-
<i>Gammaproteobacteria/Pasteurellaceae</i>	8	8	4	-	-	-
<i>Gammaproteobacteria/Pseudomonadaceae</i>	7	7	7	-	-	ModE_N-COG1910: 1 (1 contains ModE)
<i>Gammaproteobacteria/Vibrionaceae</i>	11	11	3	-	-	-
<i>Gammaproteobacteria/Xanthomonadaceae</i>	4	4	-	-	-	-
<i>Spirochaetales</i>	-	-	-	-	-	-

ModE-like variants could be detected in several ModE-containing organisms (see Table S1). Currently, no conclusion could be made regarding the functions of these

ModE variants. One hypothesis is that separate ModE\_N and Mop/Di-Mop proteins together may have a function similar to that of full-length ModE in regulating

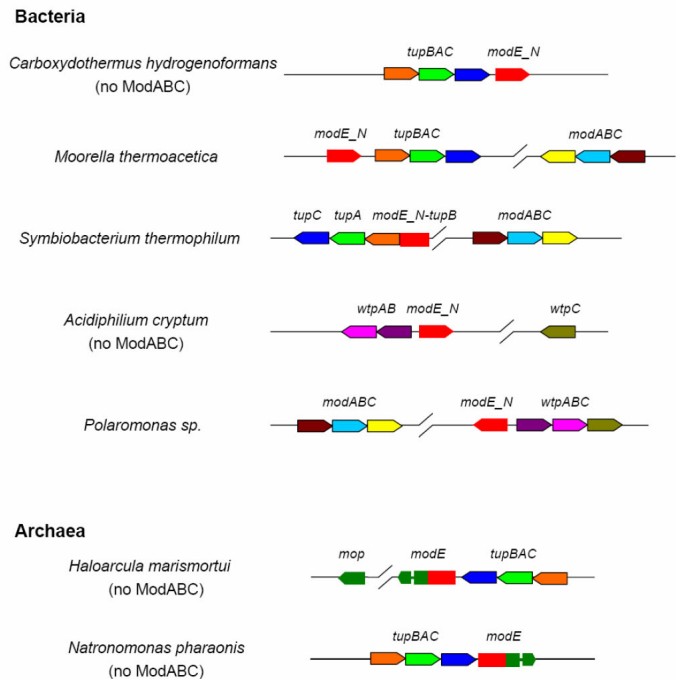
ModABC transporters (Figure 5B). The function of orphan ModE\_N is unclear. It was previously suggested that ModE\_N might be sufficient to mediate DNA binding for ModABC regulation, albeit weakly.<sup>21</sup> In addition, the MerR-Mop fusion protein identified in *Actinobacteria* could be a candidate regulator for ModABC or other Mo-related genes as this protein contains both a MerR-like transcription factor domain and Mop domain. However, the fact that almost half of ModABC-containing organisms lack both ModE and its variants suggests that new regulators are present in these organisms for ModABC regulation.

Orthologous ModE or ModE\_N sequences were also identified in several prokaryotes which lack ModABC transporters, especially in archaea where 7 out of 10 ModE\_N-containing organisms lacked ModABC transporters. We noticed that in some genomes, ModE or ModE\_N genes were located close or next to genes coding for TupABC or WtpABC transporters, suggesting that the two secondary Mo/W transporter systems may be also regulated, in some organisms, by ModE-like mechanisms (Figure 6). Further experimental verification is needed to test this possibility.

#### Occurrence of molybdoenzymes in prokaryotes and eukaryotes

Figures 1–3 also show the occurrence of different molybdoprotein families, including Moco-containing enzymes and nitrogenase, in the three domains of life. As discussed above, there was a good correspondence between occurrence of Moco biosynthesis/Mo transport components and molybdoenzymes. In bacteria, except for the AOR family (found in 50 organisms), other Moco-containing enzymes were widespread in Moco-utilizing organisms (95.1%, 68.9%, and 66.8% for DMSOR, SO, and XO, respectively). The family used by most organisms, DMSOR, was largely represented by nitrate reductase (dissimilatory) and formate dehydrogenase. Many organisms possessed 23 Moco-containing protein families and several subfamilies within these families. However, the low occurrence or absence of SO and XO families in some phyla (e.g., SO in *Firmicutes/Clostridia*, *Bacteroidetes*, *Chlorobi*, XO in *Chlorobi*, *Cyanobacteria*, *Epsilonproteobacteria*, and several *Gammaproteobacteria* clades), most of which possess the DMSOR family, suggested an independent relationship among molybdoenzymes. Only 67 organisms were found to possess nitrogenase and most of them (~97%) utilized Moco.

In archaea, members of the DMSOR family were found in all Mo-utilizing organisms. In contrast to bacteria, the AOR family was found in 69.4% of Moco-utilizing organisms, whose occurrence was much higher than that of SO and XO families (47.2% and 30.6% respectively). Nitrogenase was only present in methanogenic archaea, but in all of them.



**Figure 6.** Genomic organization of ModE, ModE\_N, and secondary Mo/W transporters in some Moco-utilizing organisms. Different genes in representative genomes are shown by indicated color schemes.

In contrast to prokaryotes, eukaryotes had only two molybdoenzyme families: SO and XO. All organisms which possessed the Mo utilization trait had SO family and 95.5% had XO family. All animals (*Metazoa*), land plants, stramenopiles, and pezizomycotina had both molybdoenzymes. Interestingly, no Moco utilization trait was detected in yeast *Saccharomycotina*. It was reported that *Saccharomyces cerevisiae* does not contain molybdoenzymes.<sup>7</sup> However, it has also been reported that some other yeasts, such as *Candida nitratophila*, *Pichia anomala*, and *Pichia angusta*, utilize Mo-containing assimilatory NR.<sup>42–44</sup> In the present study, we could not detect homologs of such NR in sequenced yeast genomes, including *Candida albicans*, *Candida glabrata*, *Candida tropicalis*, and *Pichia guilliermondii*. The absence of both Moco biosynthesis pathway and assimilatory NR strongly suggested the loss of Mo utilization in most yeast species.

#### A general evolutionary model of Mo utilization in the three domains of life

Based on the findings discussed above, it is possible to infer a general model of Mo utilization in the three domains of life. Considering that the common role of various Moco-binding proteins is to catalyze important redox reactions in the global carbon, nitrogen, and sulfur cycles, it is not surprising that Moco is essential for most organisms. However, some organisms or even complete clades may have evolved alternative mechanisms for

such reactions due to the loss of both Moco biosynthesis pathway and Moco-containing enzymes.

Of the four large molybdoenzyme families which include more than 50 subfamilies in prokaryotes, only SO and XO (including NR, SO, XDH, and AO subfamilies) span all three domains of life. If a protein family has representatives in all domains of life, it is thought that it was present in the last universal common ancestor.<sup>45</sup> Therefore, we speculate that SO and XO families evolved in the common ancestor. The other two molybdoenzyme families, DMSOR and AOR, show a more limited occurrence and are detected only in prokaryotes.

In most phyla of prokaryotes, most organisms retained the Mo utilization trait although some organisms lost it. In order to investigate the contribution of horizontal gene transfer (HGT) to Moco utilization in these organisms, we analyzed the phylogeny of both Moco biosynthesis enzymes and Moco-binding proteins, but could not identify a single HGT event for the complete Mo utilization trait (including both Moco biosynthesis pathway and the corresponding molybdoenzymes) in distantly related organisms (data not shown). This observation is consistent with the idea that HGT is unlikely to play a significant role for acquisition of Moco utilization because genes involved in Moco biosynthesis are located in several operons, some of which are typically scattered throughout the genomes. On the other hand, a complete loss of Moco utilization trait was observed in two distantly related phyla: *Firmicutes/Mollicutes* and *Chlamydiae*. The fact that their sister phyla (such as *Bacillales* and *Clostridia* for *Mollicutes*) commonly utilize Moco suggests that the loss of Moco utilization trait happened independently in the early ancestors of the two clades. All sequenced organisms in the two phyla were host-associated organisms, and it is possible that they exploit the Moco-binding proteins of the host. In several other evolutionarily distant lineages, such as *Firmicutes/Lactobacillales* and *Alphaproteobacteria/Rickettsiales*, very few organisms are able to use Moco. Phylogenetic analysis of the Mo utilization trait in these few organisms (as described above) did not support a HGT event from other species. Therefore, we inferred that Moco was used in the ancestors of *Firmicutes/Lactobacillales* and *Alphaproteobacteria/Rickettsiales* and was later independently lost. In addition, the loss of molybdoenzymes should accompany the loss of the Moco biosynthesis pathway. However, in *Spirochaetes* and *Thermotogae* which completely lost Moco biosynthesis pathways, XO homologs were detected. It is unclear whether these orphan XO homologs could still use Mo as a cofactor.

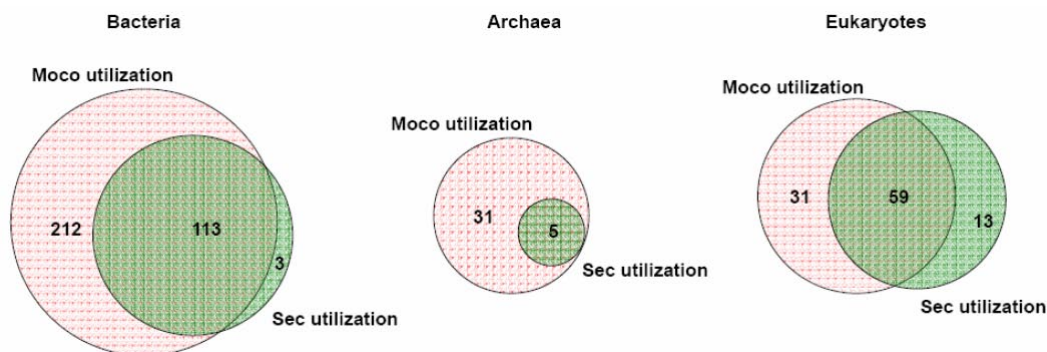
Similar trends were observed in eukaryotes. Most phyla (including all animals) inherited the Moco utilization trait from the universal ancestor of all eukaryotes, whereas certain lineages including all parasites appeared to have lost it. An interesting case was observed in fungi. All sequenced pezizomycotina contained both

Moco biosynthesis trait and the four eukaryotic molybdoenzyme subfamilies. In contrast, only a small number of yeasts possessed Mo-dependent NR which is the only reported molybdoenzyme in these organisms. *S. cerevisiae*, *Schizosaccharomyces pombe*, and all other sequenced yeasts lost the ability to use this trace element. Considering the difficulty of acquisition of the whole Mo utilization trait from distant species in eukaryotes, we suggest that the common ancestor of yeasts (including *Saccharomycotina* and *Schizosaccharomycetes*) utilized Mo as cofactor, at least for NR. However, this trait was later lost. The fact that Mo-containing NR is absent in sister species of Mo-utilizing yeasts (e.g., it is present in *Candida nitratophila* but absent in *Candida albicans* and *Candida glabrata*) suggests a recent loss event. NR catalyzes the reduction of nitrate to nitrite which is only present in autotrophic organisms such as plants, algae, and fungi.<sup>2,3</sup> The absence of Mo-dependent NR in most yeast species suggests either that Mo-dependent reduction of nitrate to nitrite is unnecessary for these organisms or that alternative Mo-independent mechanisms have evolved.

## Discussion

Mo and W are found in the mononuclear form in the active sites of diverse enzymes in all three domains of life.<sup>46–48</sup> The active sites of these enzymes include the metal ion coordinated to pyranopterin molecules and to a variable number of other ligands, such as oxygen, sulfur, and selenium.<sup>49, 50</sup> In addition, these proteins may also have other redox cofactors, such as iron–sulfur centers, flavins, and hemes, which are involved in intramolecular and intermolecular electron transfer processes.<sup>49</sup> Much effort has previously been placed on identifying and characterizing Moco biosynthesis components and Mo-dependent enzymes in various species and groups of organisms. In contrast, occurrence and evolution of the overall Mo utilization trait remained unclear. In this study, we analyzed phylogenetic profiles and regulation of Mo uptake systems, Moco biosynthesis genes, and Mo-containing proteins to better understand evolution and current use of Mo in nature. Our data reveal patterns and properties of Mo utilization among organisms with sequenced genomes and provide new insights into understanding the dynamic evolution of Mo utilization trait in prokaryotes and eukaryotes.

The widespread distribution of the Mo utilization trait in prokaryotes suggested that this trace element could be used by essentially all prokaryotic phyla. In contrast, the absence of the Mo utilization trait in several evolutionarily distant phyla (e.g., *Firmicutes/Mollicutes* and *Chlamydiae*) implied a loss of this trait in these clades. There was a good correspondence between occurrence of the Mo biosynthesis pathway and presence of known Moco-containing protein families. However, a few exceptions wherein some organisms lacked either Moco-containing proteins or Moco biosynthesis



**Figure 7.** Distribution of Moco utilization and Sec utilization in the three domains of life. Relationships between Moco utilization and Sec utilization in archaea, bacteria, and eukarya are shown by a Venn diagram.

components, suggest the presence of additional Moco-dependent protein families or alternative Mo utilization pathways in these organisms.

Besides the classic ModABC transport system, a distant ModABC-like group was predicted in *Pyrobaculum*. Although the ModA-like proteins appeared to be an outgroup of all three known Mo/W transporters, they belong to the same COG as *E. coli* ModA. The presence of *modB*-like and *modC*-like genes (as well as *modD* gene) in the same operon implied that they form a distant group of ModABC transporters and are involved in Mo/W uptake. Orthologs of this group could only be found in *Pyrobaculum* species but not in other sister species in the same archaeal phylum. It is possible that these ModABC-like transporters evolved from an ancestral ModABC system and diverged rapidly in *Pyrobaculum*. On the other hand, MOT1, which is the only known Mo transporter in eukaryotes, was only detected in one third of Mo-utilizing organisms, suggesting that most eukaryotes (including all animals) use additional unknown transport system for Mo uptake.

We investigated ModE-related ModABC regulation in prokaryotes. Surprisingly, less than 30% of Mo-utilizing organisms possessed full-length ModE regulators. Over 70% bacteria and 80% archaea appeared not to use *E. coli*-type ModE for ModABC regulation. Orphan Mop or Di-Mop proteins are not specific for ModE-related regulation because they also occur in other proteins with distinct functions (e.g., Mop domain is present in the C-terminus of ModC, and Di-Mop domain is present in ModG which is implicated in intracellular Mo homeostasis). Although some species contain either both ModE\_N and Mop/Di-Mop proteins (which suggests a function similar to that of ModE) or orphan ModE\_N (which may mediate weak ModABC regulation), almost half of ModABC-containing organisms lacked ModE-type ModABC regulation. This finding suggests the presence of novel or unspecific pathways for molybdate uptake in these organisms. In addition, the occurrence of different fusion proteins composed of ModE\_N and Mop domains suggests the presence of more complex

regulatory networks for Mo uptake, and Moco biosynthesis and utilization. Analysis of gene neighborhoods of ModE\_N/ModE and TupABC/WtpABC transporters implied that the two secondary Mo/W transporters may be also regulated by ModE-type system in some organisms.

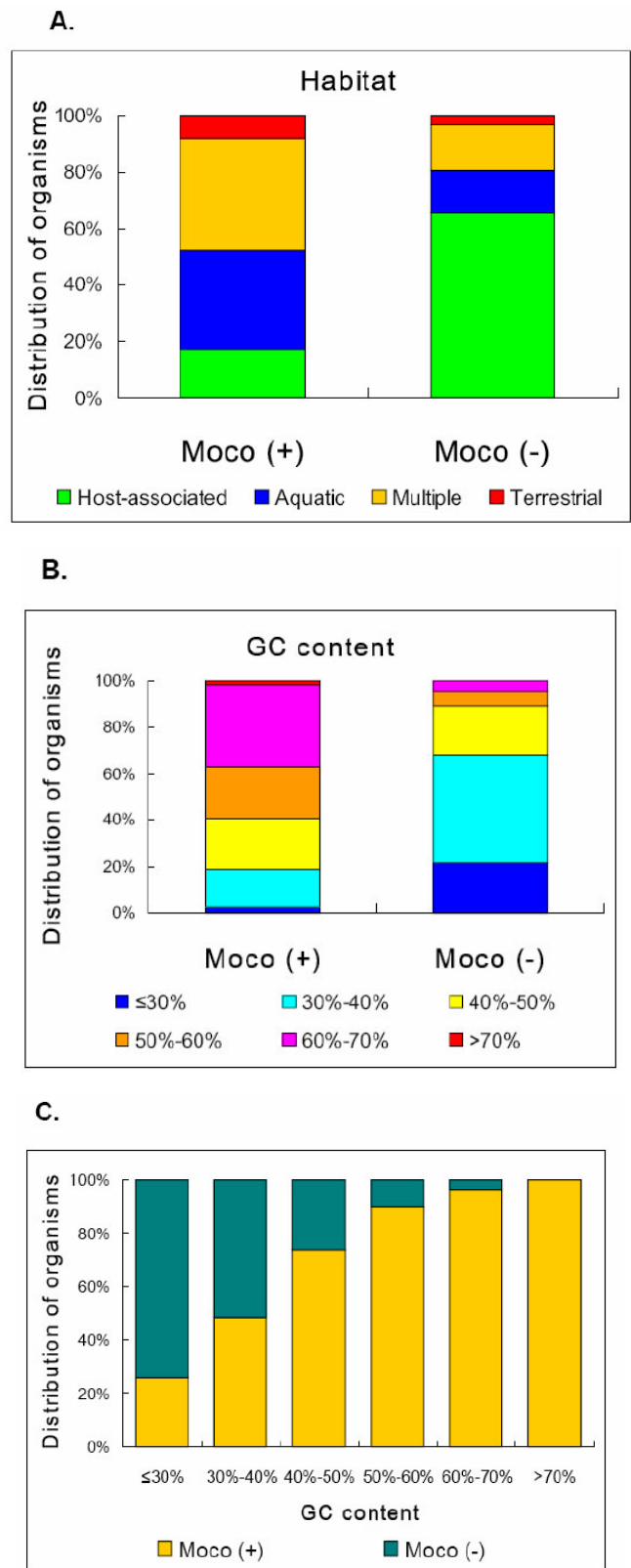
Analysis of Mo-containing proteins provided a straightforward approach to analyze the distribution and evolution of molybdoproteomes in various organisms. AOR was the first enzyme which was structurally characterized as a protein containing a Moco-type cofactor<sup>8</sup> and has been proposed to be the primary enzyme responsible for catalysis of the interconversion of aldehydes and carboxylates in archaea.<sup>51</sup> However, it is the rarest known bacterial Moco-containing protein, suggesting that AOR-dependent oxidation of aldehydes is not essential for general metabolism in most bacterial species. The other three molybdoenzyme families are much more widely distributed, especially the DMSOR family which is found in almost all Mo-utilizing bacteria and all Mo-utilizing archaea. Enzymes of the DMSOR family catalyze a variety of reactions that involve oxygen atom transfer to or from an available electron pair of a substrate or cleavage of a CH bond.<sup>2, 10, 52-55</sup> NR (disimilatory) and formate dehydrogenase are the two major members of the DMSOR family. Formate dehydrogenase alpha subunit (FdhA) is also a selenocysteine (Sec)-containing protein and might be responsible for maintaining the Sec-decoding trait in prokaryotes.<sup>56</sup> We compared the distribution of Mo- and Sec-utilizing organisms in both prokaryotes and eukaryotes and found that Sec-utilizing organisms were essentially a subset of Moco-dependent organisms in prokaryotes (Figure 7, Table S1). These data suggest that the Sec trait is dependent on the Mo utilization trait in prokaryotes because of the function of formate dehydrogenase which is not only a widespread Mo-enzyme but is also the main user of Se in prokaryotes. In addition, occurrence of the only non-Moco-containing protein, nitrogenase, was limited in both bacteria and archaea. This enzyme is used by several organisms to fix atmospheric nitrogen gas (N<sub>2</sub>).

The fact that it was found in all methanogenic archaea implied that the function of this protein is essential for these organisms.

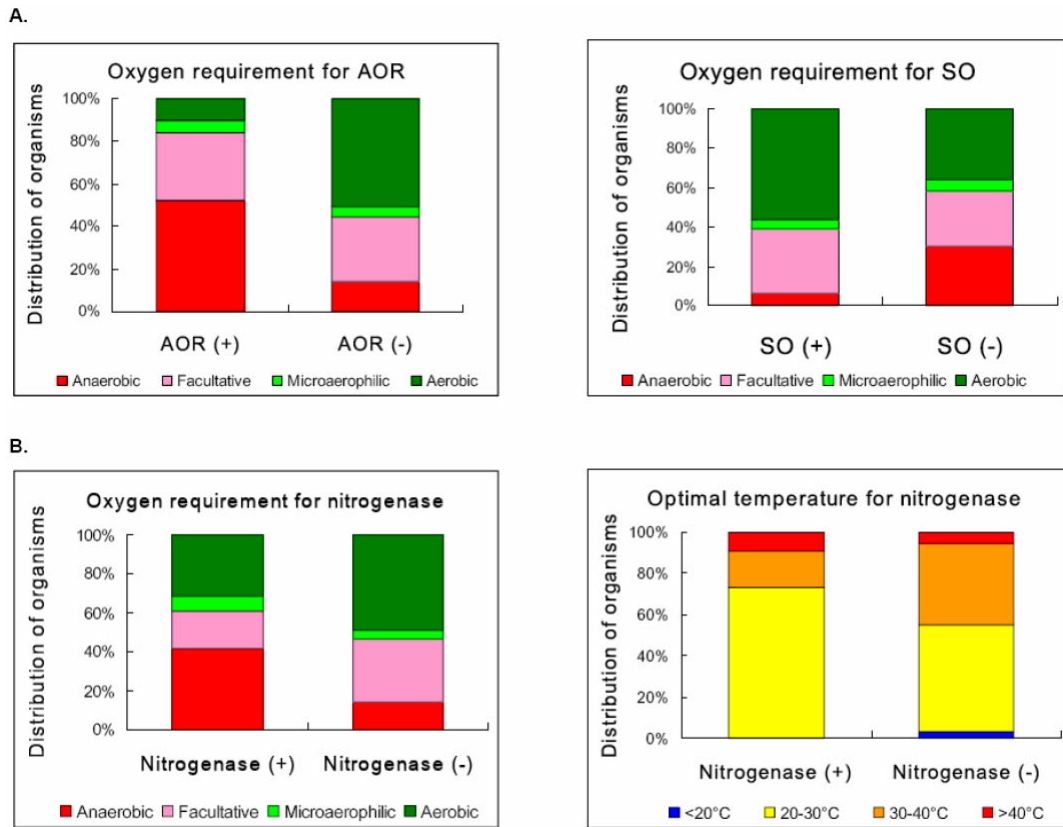
We attempted to generate a general evolutionary model of Mo utilization in the three domains of life. The Moco biosynthesis pathway and at least two molybdoenzyme families (SO and XO) were likely present in the last universal common ancestor. The Moco utilization trait is evolutionarily conserved in most prokaryotic and eukaryotic species due to the important redox reactions catalyzed by molybdoenzymes in carbon, nitrogen, and sulfur metabolism. In addition, an independent loss of the Moco utilization trait (instead of a HGT from other species) and perhaps appearance of alternative Mo-independent pathways play a role in the evolution of Mo utilization.

We hypothesized that since both Moco biosynthesis trait and molybdoenzymes were found to be present (or both were absent) in organisms, and these patterns were observed in various bacterial phyla, certain common factors (e.g., habitat) may have affected acquisition/loss of Mo utilization. To examine this possibility, we analyzed a role of environmental conditions (e.g., habitat, oxygen requirement, optimal temperature, and optimal pH) and other factors (e.g., genome size, GC content) in Mo utilization in sequenced prokaryotes. Previously, a similar strategy was used to analyze the evolution of Se in bacteria.<sup>56</sup> Figure 8 shows the distribution of organisms that possess or lack Moco utilization with respect to several such factors.

We found that the majority of bacteria that do not utilize Moco were host-associated (i.e., parasites or symbionts, Figure 8A), implying that host-associated lifestyle often leads to the loss of Mo utilization, perhaps due to limited space and resources or availability of Mo pathways of the host. This is consistent with the observation in *Firmicutes/Mollicutes* and *Chlamydiae*, all of which are host-associated and could not utilize Moco. This idea is also supported by analysis of Mo utilization in *Alphaproteobacteria/Rickettsiales*. In this phylum, only one out of 19 organisms utilized Mo (*Candidatus Pelagibacter ubique*, a marine bacterium living in ocean surface water). However, it is also the only non-host-associated organism in this clade. Our data suggested a complete loss of the Mo utilization trait in all host-related organisms in this phylum instead of a HGT into *Candidatus Pelagibacter ubique*. In addition, in many phyla, genomes of Moco-utilizing organisms had a significantly higher GC content, suggesting that the increase in GC content correlates with increased Mo utilization (Figure 8B and 8C). Organisms with low GC content (i.e., GC <40%) which lack the Moco utilization trait were found in a variety of phyla, indicating that such correlation is significant. The reason why low GC content organisms in different clades lost the Moco utilization trait is not clear. Other factors, such as oxygen requirement, gram strain, optimal



**Figure 8.** Relationship between environmental factors, properties of organisms and the Mo utilization trait. All organisms were classified into two groups: Moco (+), i.e., containing Moco utilization trait; Moco (-), i.e., lacking Moco utilization. (A) Habitat. (B) GC content. (C) A different representation of the influence of GC content.



**Figure 9.** Relationship between environmental factors, properties of organisms and different molybdoenzymes. (A) Oxygen requirement for AOR and SO. (B) Oxygen requirement and optimal temperature for nitrogenase.

temperature, and pH, did not appear to have a role in Mo utilization. In archaea, only two organisms, *Methanosphaera stadtmanae* (the only sequenced parasite in archaea) and *Nanoarchaeum equitans* (an ancient hyperthermophilic and anaerobic obligate symbiont which has a small genome<sup>57</sup> and has lost the ability to use most trace elements such as nickel, cobalt, copper, and selenium), lacked Mo utilization and both genomes had a very low GC content (27.6% and 31.6%). These data provide an additional support for our observation in bacteria. Thus, host-associated life style as well as reduced GC content seem to correlate with the loss of Mo utilization.

We also examined distribution, based on the factors discussed above, of different molybdoenzyme families, and similar trends were found. Moreover, additional features were observed for different molybdoenzymes (Figure 9). For example, organisms possessing AOR proteins favor an anaerobic environment, whereas organisms containing SO, XO, or DMSOR proteins favor aerobic conditions. Organisms containing nitrogenase favor both anaerobic and relatively warm conditions (all psychrophilic organisms did not possess nitrogenase). These data illustrate 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.

Similar investigation of molybdoenzymes in eukaryotes provided the information on Mo utilization in this domain of life. As in prokaryotes, distribution of eukaryotic Mo-containing proteins essentially matched the Moco utilization trait. However, only SO (including NR and SO) and XO (including XDH and AO) families could be detected, suggesting a much smaller molybdo-proteome in eukaryotes than in prokaryotes. Functional roles of these four subfamilies have previously been investigated in different organisms.<sup>2,3,7</sup> Besides NR, which is a key enzyme of nitrate assimilation and does not occur in animals, the other three enzymes are present in a variety of clades including unicellular organisms and animals. SO catalyzes the oxidation of sulfite to sulfate (the final step in the degradation of sulfur-containing amino acids).<sup>7</sup> XDH is a key enzyme in purine degradation and oxidizes both hypoxanthine to xanthine and xanthine to uric acid, whereas AO catalyzes the oxidation of a variety of aromatic and nonaromatic heterocycles and aldehydes and converts them to the respective carboxylic acids.<sup>7</sup> All parasites lost the ability to synthesize Moco, which is consistent with what we found in prokaryotes, suggesting that Mo utilization may have been present in the eukaryotic progenitor and became unnecessary for parasites because of reduced availability of Mo or dependence on the corresponding meta-

bolic pathways of the host. Both Mo-dependent and Mo-independent organisms were found among fungi. The recent loss of the Mo biosynthesis pathway and Mo-dependent NR in most yeasts, including *S. cerevisiae*, suggested that Mo-dependent nitrate assimilation may be unnecessary or may have been replaced by other pathways in these organisms. It is known that nitrate assimilation is one of two major biological processes by which inorganic nitrogen is converted to ammonia and hence to organic nitrogen.<sup>58</sup> Although *S. cerevisiae* lacks both Moco biosynthesis trait and NR, it contains a number of genes which convert glutamine to glutamate, providing a major source of organic nitrogen.<sup>59</sup> In addition, glutathione (GSH) stored in the yeast vacuole can serve as an alternative nitrogen source during nitrogen starvation.<sup>60</sup> It is unclear whether the ancestor of yeasts possessed other Mo-binding enzymes. However, alternative Mo-independent pathways for sulfur and carbon metabolism may have evolved in yeasts. Both Mo-dependent and Mo-independent fungi are free-living organisms and in this case we could not identify a common environmental factor which is related to Mo utilization. Thus, additional unidentified factors may have affected Mo utilization in fungi. A future challenge would be to discover these factors as well as additional features influencing Mo utilization in the three domains of life.

In conclusion, we report a comprehensive comparative genomics analysis of Mo utilization in prokaryotes and eukaryotes by examining occurrence of proteins involved in Moco biosynthesis, Mo transport, and Mo utilization (molybdoenzymes). Our data reveal a complex and dynamic evolutionary process of Mo utilization. Most bacteria and archaea utilize Mo, with the exception of parasites and organisms with low genomic GC content. A distant group of ModABC transport system was identified in *Pyrobaculum* species. Regulation of Mo uptake must be more complex than previously thought as ModE-type ModABC regulatory systems occurred only in a limited number of Moco-utilizing organisms. In contrast to the wide use of Mo in prokaryotes, the utilization of this element in eukaryotes is more restricted, both with regard to the number of organisms that depend on Mo and the number of molybdoprotein families that occur in them. Again, host-associated conditions appear to lead to the loss of Mo utilization.

## Materials and Methods

### *Genomic sequence resources*

Sequenced genomes of archaea, bacteria and eukaryotes were retrieved from NCBI website ([http://www.ncbi.nlm.nih.gov/sutils/genom\\_table.cgi](http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi)). Only one strain was used for each species (e.g., *E. coli* K12 was used as

a representative of *E. coli*). A total of 451 bacterial, 38 archaeal, and 142 eukaryotic organisms were analyzed (as of Feb. 2007).

### *Identification of Mo transporters, transporter repressors, Moco biosynthesis genes and Mo-containing enzymes*

We used several well-characterized proteins, which are Mo/W transporters or known to be involved in the Moco biosynthesis pathways as our seed sequences to search for homologs in sequence databases. In prokaryotes, products of *moa* (*moaA-moaE*), *mod* (*modABC* and *modE*) and *moe* (*moeA* and *moeB*) operons from *E. coli*, WtpABC from *P. furiosus* and TupABC from *E. acidaminophilum* were used to identify a set of primary homologous sequences using TBLASTN with an e-value <1. Iterative TBLASTN searches were then performed within each phylum, using different homologous sequences from the primary set as queries, to identify more distant homologs. In parallel, three cycles of PSIBLAST with default parameters were used for the identification of distant homologs. Orthologous proteins were defined as bidirectional best hits.<sup>61</sup> When necessary, orthologs were also confirmed by genomic location analysis or building phylogenetic trees for the corresponding protein families. Occurrence of the Moco trait was verified by the requirement for presence of most of these genes. Members of known Moco protein families as well as nitrogenase were identified using a similar approach.

In eukaryotes, we used MOT1 (a recently identified Mo-specific transporter in plants), and Cnx1-3 and Cnx5-7 from *A. thaliana* as seed sequences to detect molybdate transporter and Moco utilization in sequenced genomes. Considering uncertainty of Moco biosynthesis pathway in unicellular eukaryotes and incompleteness of some genome sequences, the presence of the Moco utilization trait was verified in these organisms by the following criteria: at least 2 orthologs of proteins involved in Moco biosynthesis and at least 1 known Mo-containing protein detected in the same organism.

### *Multiple sequence alignment and phylogenetic analysis*

To investigate distribution of organisms that utilize Mo in different phyla, we adopted a phylogenetic tree developed by Ciccarelli et al.,<sup>39</sup> which is based on concatenation of 31 orthologs occurring in 191 species with sequenced genomes. Phylogenetic trees of each Mo/W transporter systems were reconstructed by standard approaches. Sequences were aligned with CLUSTALW<sup>62</sup> using default parameters. Ambiguous alignments in highly variable (gap-rich) regions were excluded. The resulting multiple alignments were then checked for conservation of functional residues and manually edited. In addition, MUSCLE<sup>63</sup> alignment tool was used to evaluate the CLUSTALW results. Phylogenetic anal-

yses were performed using PHYLIP programs.<sup>64</sup> Pairwise distance matrices were calculated by PROTDIST to estimate the expected amino acid replacements per position. Neighbor-joining (NJ) trees were obtained with NEIGHBOR and the most parsimonious trees were determined with PROTPARS. Robustness of these trees was then evaluated by maximum likelihood (ML) analysis with PHYML<sup>65</sup> and Bayesian estimation of phylogeny with MrBayes.<sup>66</sup>

## References

- Enemark, J. H., and Young, C. G. (1993). Bioinorganic chemistry of pterin-containing molybdenum and tungsten enzymes. *Adv. Inorg. Chem.* **40**, 188.
- Hille, R. (1996). The mononuclear molybdenum enzymes. *Chem. Rev.* **96**, 2757-2816.
- Schwarz, G., and Mendel, R. R. (2006). Molybdenum cofactor biosynthesis and molybdenum enzymes. *Annu. Rev. Plant Biol.* **57**, 623-647.
- Allen, R. M., Chatterjee, R., Madden, M. S., Ludden, P. W., and Shah, V. K. (1994). Biosynthesis of the iron-molybdenum cofactor of nitrogenase. *Crit. Rev. Biotechnol.* **14**, 225-249.
- Rajagopalan, K. V. (1991) Novel aspects of the biochemistry of the molybdenum cofactor. *Adv. Enzym.* **64**, 215-290.
- Rajagopalan, K. V., and Johnson, J. L. (1992) The pterin molybdenum cofactors. *J. Biol. Chem.* **267**, 10199-10202.
- Mendel, R. R., and Bittner, F. (2006). Cell biology of molybdenum. *Biochim Biophys Acta.* **1763**, 621635.
- Chan, M. K., Mukund, S., Kletzin, A., Adams, M. W., and Rees, D. C. (1995). Structure of a hyperthermophilic tungstopterin enzyme, aldehyde ferredoxin oxidoreductase. *Science* **267**, 1463-1469.
- Sigel, A., and Sigel, H. (2002). *Molybdenum and tungsten, Their Roles in Biological Processes* (Sigel, A., and Sigel, H., eds.), pp. 130, Marcel Dekker, New York.
- Kisker, C., Schindelin, H., and Rees, D. C. (1997). Molybdenum-cofactor containing enzymes: structure and mechanism. *Annu Rev Biochem.* **66**, 233-267.
- Kisker, C., Schindelin, H., Baas, D., Retey, J., Meckenstock, R. U., and Kroneck, P. M. (1998). A structural comparison of molybdenum cofactor-containing enzymes. *FEMS Microbiol. Rev.* **22**, 503-521.
- Grunden, A. M., and Shanmugam, K. T. (1997). Molybdate transport and regulation in bacteria. *Arch. Microbiol.* **168**, 345-354.
- Pau, R. N., and Lawson, D. M. (2002). Transport, homeostasis, regulation, and binding of molybdate and Tungstate to proteins. *Met. Ions Biol. Syst.* **39**, 31-74.
- Bevers, L. E., Hagedoorn, P. L., Krijger, G. C., and Hagen, W. R. (2006). Tungsten transport protein A (WtpA) in *Pyrococcus furiosus*: the first member of a new class of tungstate and molybdate transporters. *J. Bacteriol.* **188**, 6498-6505.
- Makdessi, K., Andreessen, J. R., and Pich, A. (2001). Tungstate Uptake by a highly specific ABC transporter in *Eubacterium acidaminophilum*. *J. Biol. Chem.* **276**, 24557-24564.
- Tomatsu, H., Takano, J., Takahashi, H., Watanabe-Takahashi, A., Shibagaki, N., and Fujiwara, T. (2007). An *Arabidopsis thaliana* high-affinity molybdate transporter required for efficient uptake of molybdate from soil. *Proc. Natl. Acad. Sci. USA* (Nov 14; Epub ahead of print).
- Grunden, A. M., Ray, R. M., Rosentel, J. K., Healy, F. G., and Shanmugam, K. T. (1996). Repression of the *Escherichia coli modABCD* (molybdate transport) operon by ModE. *J. Bacteriol.* **178**, 735-744.
- Hall, D. R., Gourley, D. G., Leonard, G. A., Duke, E. M., Anderson, L. A., Boxer, D. H., and Hunter, W. N. (1999). The high-resolution crystal structure of the molybdate-dependent transcriptional regulator (ModE) from *Escherichia coli*: a novel combination of domain folds. *EMBO J.* **18**, 14351446.
- Tao, H., Hasona, A., Do, P. M., Ingram, L. O., and Shanmugam, K. T. (2005). Global gene expression analysis revealed an unsuspected *deo* operon under the control of molybdate sensor, ModE protein, in *Escherichia coli*. *Arch. Microbiol.* **184**, 225-233.
- Anderson, L. A., Palmer, T., Price, N. C., Bornemann, S., Boxer, D. H., and Pau, R. N. (1997). Characterisation of the molybdenum-responsive ModE regulatory protein and its binding to the promoter region of the *modABCD* (molybdenum transport) operon of *Escherichia coli*. *Eur. J. Biochem.* **246**, 119-126.
- McNicholas, P. M., Mazzotta, M. M., Rech, S. A., and Gunsalus, R. P. (1998) Functional dissection of the molybdate-responsive transcription regulator, ModE, from *Escherichia coli*. *J. Bacteriol.* **180**(17), 4638-4643.
- Wiethaus, J., Wirsing, A., Narberhaus, F., and Masepohl, B. (2006) Overlapping and specialized functions of the molybdenum-dependent regulators MopA and MopB in *Rhodobacter capsulatus*. *J. Bacteriol.* **188**(24), 8441-8451.
- Zahalak, M., Pratte, B., Werth, K. J., and Thiel, T. (2004). Molybdate transport and its effect on nitrogen utilization in the cyanobacterium *Anabaena variabilis* ATCC 29413. *Mol. Microbiol.* **51**, 539-549.
- Delgado, M. J., Tresierra-Ayala, A., Talbi, C., and Bedmar, E. J. (2006). Functional characterization of the *Bradyrhizobium japonicum modA* and *modB* genes involved in molybdenum transport. *Microbiology* **152**, 199-207.
- Studholme, D. J., and Pau, R. N. (2003) A DNA element recognised by the molybdenum-responsive transcription factor ModE is conserved in Proteobacteria, green sulphur bacteria, and Archaea. *BMC Microbiol.* **3**:24
- Rodionov DA, Dubchak I, Arkin A, Alm E, Gelfand MS. (2004) Reconstruction of regulatory and metabolic pathways in metal-reducing delta-proteobacteria. *Genome Biol.* **5**, R90.
- Shanmugam, K. T., Stewart, V., Gunsalus, R. P., Boxer, D. H., Cole, J. A., Chippaux, M., DeMoss, J. A., Giordano, G., Lin, E. C., and Rajagopalan, K. V. (1992). Proposed nomenclature for the genes involved in molybdenum metabolism in *Escherichia coli* and *Salmonella typhimurium*. *Mol. Microbiol.* **6**, 3452-3454.
- Mendel, R. R., and Schwarz, G. (2002). Biosynthesis and molecular biology of the molybdenum cofactor (Moco). *Met. Ions Biol. Syst.* **39**, 317-368.
- Millar, L. J., Heck, I. S., Sloan, J., Kana'n, G. J., Kinghorn, J. R., and Unkles, S. E. (2001). Deletion of the *cnxE* gene encoding the gephyrinlike protein involved in the final stages of molybdenum cofactor biosynthesis in *Aspergillus nidulans*. *Mol. Genet. Genomics* **266**, 445-453.
- Reiss, J., Cohen, N., Dorche, C., Mandel, H., Mendel, R. R., Stallmeyer, B., Zobot, M. T., and Dierks, T. (1998). Mutations in a polycistronic nuclear gene associated with molybdenum cofactor deficiency. *Nat. Genet.* **20**, 51-53.



31. Stallmeyer, B., Drugeon, G., Reiss, J., Haenni, A. L., and Mendel, R. R. (1999). Human molybdopterin synthase gene: identification of a bicistronic transcript with overlapping reading frames. *Am. J. Hum. Genet.* **64**, 698-705.
32. Stallmeyer, B., Schwarz, G., Schulze, J., Nerlich, A., Reiss, J., Kirsch, J., and Mendel, R. R. (1999). The neurotransmitter receptor-anchoring protein gephyrin reconstitutes molybdenum cofactor biosynthesis in bacteria, plants, and mammalian cells. *Proc. Natl. Acad. Sci. USA* **96**, 1333-1338.
33. Panina, E. M., Mironov, A. A., and Gelfand, M. S. (2003). Comparative genomics of bacterial zinc regulons: enhanced ion transport, pathogenesis, and rearrangement of ribosomal proteins. *Proc. Natl. Acad. Sci. USA* **100**, 9912-9917.
34. Andreini, C., Bertini, I., and Rosato, A. (2004). A hint to search for metalloproteins in gene banks. *Bioinformatics* **20**, 1373-1380.
35. Rodionov, D. A., Gelfand, M. S., Todd, J. D., Curson, A. R., and Johnston, A. W. (2006). Computational Reconstruction of Iron- and Manganese-Responsive Transcriptional Networks in alpha-Proteobacteria. *PLoS Comput. Biol.* **2**, e163.
36. Permina, E. A., Kazakov, A. E., Kalinina, O. V., and Gelfand, M. S. (2006). Comparative genomics of regulation of heavy metal resistance in Eubacteria. *BMC Microbiol.* **6**, 49.
37. Andreini, C., Banci, L., Bertini, I., & Rosato, A. (2006). Zinc through the three domains of life. *J. Proteome Res.* **5**, 3173-3178.
38. Andreini, C., Banci, L., Bertini, I., Elmi, S., and Rosato, A. (2007). Nonheme iron through the three domains of life. *Proteins* **67**, 317-324.
39. Ciccarelli, F. D., Doerks, T., von Mering, C., Creevey, C. J., Snel, B., and Bork, P. (2006) Toward automatic reconstruction of a highly resolved tree of life. *Science* **311**, 1283-1287.
40. Hu, Y., Rech, S., Gunsalus, R. P., and Rees, D. C. (1997). Crystal structure of the molybdate binding protein ModA. *Nat. Struct. Biol.* **4**, 703-707.
41. Hollenstein, K., Frei, D. C., and Locher, K. P. (2007). Structure of an ABC transporter in complex with its binding protein. *Nature* **446**, 213-216.
42. Hipkin, C. R., Kau, D. A., and Cannons, A. C. (1993). Further characterization of the assimilatory nitrate reductase from the yeast *Candida nitratophila*. *J. Gen. Microbiol.* **139**, 473-478.
43. Barbier, G. G., Joshi, R. C., Campbell, E. R., and Campbell, W. H. (2004). Purification and biochemical characterization of simplified eukaryotic nitrate reductase expressed in *Pichia pastoris*. *Protein Expr. Purif.* **37**, 61-71.
44. Fischer, K., Barbier, G. G., Hecht, H. J., Mendel, R. R., Campbell, W. H., and Schwarz, G. (2005). Structural basis of eukaryotic nitrate reduction: crystal structures of the nitrate reductase active site. *Plant Cell.* **17**, 1167-1179.
45. Doolittle, W. F. (2000). The nature of the universal ancestor and the evolution of the proteome. *Curr. Opin. Struct. Biol.* **10**, 355-358.
46. Rajagopalan, K. V. (1988). Molybdenum: an essential trace element in human nutrition. *Annu. Rev. Nutr.* **8**, 401-427.
47. Schwarz, G. (2005). Molybdenum cofactor biosynthesis and deficiency. *Cell. Mol. Life Sci.* **62**, 2792-2810.
48. Johnson, M. K., Rees, D. C., and Adams, M. W. (1996). Tungsten enzymes. *Chem. Rev.* **96**, 2817-2840.
49. Brondino, C. D., Romao, M. J., Moura, I., and Moura, J. J. (2006). Molybdenum and tungsten enzymes: the xanthine oxidase family. *Curr. Opin. Chem. Biol.* **10**, 109-114.
50. Gladyshev, V. N. (2002). Comparison of selenium-containing molybdoenzymes. *Met. Ions Biol. Syst.* **39**, 655-672.
51. Ma, K., Hutchins, A., Sung, S. J., and Adams, M. W. (1997). Pyruvate ferredoxin oxidoreductase from the hyperthermophilic archaeon, *Pyrococcus furiosus*, functions as a CoA-dependent pyruvate decarboxylase. *Proc. Natl. Acad. Sci. USA* **94**, 9608-9613.
52. Schindelin, H., Kisker, C., Hilton, J., Rajagopalan, K. V. and Rees, D. C. (1996). Crystal structure of DMSO reductase: redox-linked changes in molybdopterin coordination. *Science* **272**, 1615-1621.
53. Schneider, F., Löwe, J., Huber, R., Schindelin, H., Kisker, C., and Knäblein, J. (1996). Crystal structure of dimethyl sulfoxide reductase from *Rhodobacter capsulatus* at 1.88 Å resolution. *J. Mol. Biol.* **263**, 53-69.
54. Boyington, J. C., Gladyshev, V. N., Khangulov, S. V., Stadtman, T. C., and Sun, P. D. (1997). Crystal structure of formate dehydrogenase H: catalysis involving Mo, molybdopterin, selenocysteine, and an Fe<sub>4</sub>S<sub>4</sub> cluster. *Science* **275**, 1305-1308.
55. Dias, J. M., Than, M. E., Humm, A., Huber, R., Bourenkov, G. P., Bartunik, H. D., et al. (1999). Crystal structure of the first dissimilatory nitrate reductase at 1.9 Å solved by MAD methods. *Structure* **7**, 65-79.
56. Zhang, Y., Romero, H., Salinas, G., and Gladyshev, V. N. (2006). Dynamic evolution of selenocysteine utilization in bacteria: a balance between selenoprotein loss and evolution of selenocysteine from redox active cysteine residues. *Genome Biol.* **7**, R94.
57. Waters, E., Hohn, M. J., Ahel, I., Graham, D. E., Adams, M. D., Barnstead, M., etc. (2003) The genome of *Nanoarchaeum equitans*: insights into early archaeal evolution and derived parasitism. *Proc. Natl. Acad. Sci. USA* **100**, 12984-12988.
58. Solomonson, L. P., and Spehar, A. (1977). Model for the regulation of nitrate assimilation. *Nature* **265**, 373-375.
59. Minehart, P. L., and Magasanik, B. (1991). Sequence and expression of GLN3, a positive nitrogen regulatory gene of *Saccharomyces cerevisiae* encoding a protein with a putative zinc finger DNA-binding domain. *Mol. Cell. Biol.* **11**, 6216-6228.
60. Mehdi, K., and Penninckx, M. J. (1997). An important role for glutathione and gamma-glutamyltranspeptidase in the supply of growth requirements during nitrogen starvation of the yeast *Saccharomyces cerevisiae*. *Microbiology* **143**, 1885-1889.
61. Tatusov, R. L., Galperin, M. Y., Natale, D. A., and Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**, 33-36.
62. Higgins, D., Thompson J., Gibson T., Thompson J. D., Higgins D. G., Gibson T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673-4680.
63. Edgar, R. C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.* **32**, 1792-1797.
64. Felsenstein, J. (1989) PHYLIP Phylogeny Inference Package (Version 3.2). *Cladistics* **5**, 164-166.
65. Guindon, S., and Gascuel, O. (2003) A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* **52**, 696-704.
66. Ronquist, F., and Huelsenbeck, J. P. (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572-1574.