Role of the N-Terminal Tail of Metal-Transporting P_{1B}-type ATPases from Genome-Wide Analysis and Molecular Dynamics Simulations

Shailesh Sharma[†] and Antonio Rosato*[,†,‡]

Magnetic Resonance Center (CERM) - University of Florence, Via L. Sacconi 6, 50019 Sesto Fiorentino, Italy, and Department of Chemistry - University of Florence, Via della Lastruccia 3, 50019 Sesto Fiorentino, Italy

Received July 9, 2008

Copper is an essential trace metal but can be potentially toxic *in vivo*. Consequently, its intracellular concentration and distribution is tightly controlled. A widespread system involved in maintaining copper homeostasis in a variety of prokaryotic and eukaryotic organisms involves a P-type ATPase, which pumps the metal ion from the cytosol at the expense of ATP hydrolysis. Copper-transporting ATPases of this kind are often associated with a small soluble metal-transporter (metallochaperone). In this work, we investigated the occurrence and properties of the ATPases and, partly, of their partner metallochaperones. We found that the latter proteins are typically encoded in organisms containing also ATPases of the subtypes 1B-1 or 1B-2. These subtypes have a characteristically extended N-terminal cytoplasmic tail that contains multiple metal-binding domains (MBDs), which can receive the metal ion from the metallochaperone. We observed a significant variability in the number and spacing in sequence of the MBDs. On the basis of molecular dynamics simulations, we proposed that the MBDs could be quite free to reorient with respect to one another. The relative conformational freedom increased rapidly with the length of the linker between the MBDs. Also based on available experimental studies, these data suggested that the reciprocal mobility of the MBDs is instrumental to permit the tuning of the selectivity and/or affinity of the ATPase for the substrate as well as to modulate the enzymatic activity of the system. We additionally detected a small but significant number of instances in which a metallochaperone is likely to interact directly with the transmembrane domain of P-type ATPases lacking cytoplasmic MBDs.

INTRODUCTION

For many organisms copper is an essential metal, because of its role as a cofactor in a variety of enzymes and electron carriers.[1] In eukaryotes, these copper-binding proteins are localized in various cellular compartments (such as the cytosol or the mitochondrion) or can be extracellular.[1,2] In prokaryotic organisms, copper-binding proteins are mainly periplasmic, in Gram-negative organisms, or associated to the plasma membrane.[1,2] A notable exception is observed in the case of photosynthetic prokaryotes that contain copper proteins in thylakoids.[3] Notwithstanding its crucial role for the correct functioning of cells, copper can be potentially toxic in vivo.[1] Copper toxicity is indeed at the basis of e.g. the use of this metal as a parasiticide in agriculture. These features made it necessary for living organisms to develop mechanisms that take care of copper uptake and transport to the appropriate subcellular locations as well as of removal of excess intracellular copper.[4] For the many prokaryotic organisms that do not use copper within their cellular processes, only copper removal is a relevant aspect.[2]

Among the various biochemical solutions occurring in Nature to address the above-mentioned needs, one that is particularly widespread involves the use of two protein partners, a soluble small (ca. 70 amino acids) copper(I) binding protein (called a metallochaperone) and an ATPase that can transport copper(I) ions across membranes at the expenses of ATP hydrolysis.[4−6] Copper(I)-transporting ATPases are of the so-called P-type, i.e. they catalyze reactions proceeding through a covalent phosphorylated "P" intermediate.[6] Based on their structural organization, and in particular on the number and position of transmembrane segments, P-type ATPases can be further separated into subgroups, with proteins of the P_1-subgroup being responsible for the transport of heavy metals, such as Cd^{2+}, Zn^{2+}, Pb^{2+}, Co^{2+}, Cu^{2+}, Ag^{+}, and Cu^{+}.[7] The H^{+},K^{+} ATPase is instead an example of P_2-type ATPase. Phylogenetic analyses have shown that the P_1-subgroup encompasses also some relatively uncommon bacterial ATPases that feature an organization in multiple protein subunits and are involved in potassium transport.[8] The latter form the so-called P_{1A} subsubgroup, whereas the ATPases transporting heavy metals are dubbed P_{1B}.[9] The aforementioned combination of a small, soluble copper(I)-transporter, operating in the cell cytosol, and of an enzyme that actively catalyzes the translocation of the metal ions allows cells to remove copper(I) ions from the cytosol and either pump them outside the cell or into intracellular organelles, depending on the localization of the ATPase.[6,10,11]

All P-type ATPases share a basic "core" architecture (Figure 1),[8] comprising a hydrophilic region protruding into the cytosol, which contains the phosphorylation and ATP-binding sites and a smaller cytosolic region (sometimes called the Actuator or A-domain), which has a regulatory function

* Corresponding author phone: +39 055 4574267; fax: +39 055 4574253; e-mail: rosato@cerm.unifi.it.
† Magnetic Resonance Center (CERM).
‡ Department of Chemistry.

10.1021/ci8002304 CCC: $40.75
Published on Web 12/18/2008

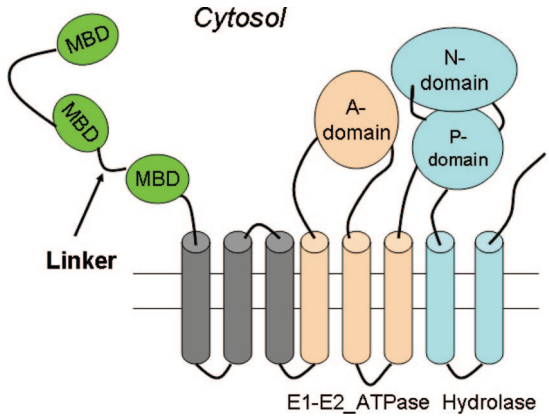

Figure 1. Scheme of the architecture of P_{1B}-type ATPases and correspondence to the Pfam domains used for complete proteome analysis. An ATPase with three metal-binding domains (MBDs) is shown. Each MBD corresponds to a single HMA Pfam domain (green). The transmembrane and cytosolic regions map to two different Pfam domains: the Pfam domain E1-E2_ATPase (light orange) contains three transmembrane helices and the Actuator (A-) soluble domain; the Pfam domain Hydrolase (light blue) contains the ATP-binding site (in the N-domain) and the phosphorylation site (in the P-domain) as well as the two most C-terminal helices. The gray transmembrane helices do not belong to a Pfam domain, and their position is different in different ATPase classes.[7,8] The regions corresponding to the E1-E2_ATPase and Hydrolase Pfam domains are present in all P-type ATPases. The linker region on which the present study focuses is highlighted by an arrow. Relative region sizes are not to scale.

and is required for the phosphatase step of the catalytic cycle (dephosphorylation of the intermediate formed during ATP hydrolysis[12]). These cytosolic parts of the polypeptide chain are connected by a number of transmembrane helices, which are involved in the formation of an intramembranous channel, and whose organization, as mentioned, leads to the definition of the P_1 and P_2 subgroups.[7] In addition to the core structure, the distinguishing feature of P_{1B}-type copper(I)-transporting ATPases, which are the focus of this work, is their long N-terminal tail, which contains a variable (between one and six) number of 70-aa independently folded domains (Figure 1).[5] The fold of these domains and of the soluble metallochaperone is alike and consists of a ferredoxin-like $\beta\alpha\beta\beta\alpha\beta$ fold. Each domain harbors a conserved sequence motif CXXC, through which it can bind one equivalent of copper(I).[13] The motif is often preceded by a methionine in position -2 (i.e., MXCXXC), which however is not involved in copper(I) coordination. In humans, there are two relevant ATPases, namely ATP7A and ATP7B, also known as the Menkes (MNK) and Wilson (WND) disease proteins, respectively. Many studies are available for these two systems that demonstrate that, even *in vivo*, the presence of either the intact fifth or intact sixth metal-binding domain (i.e., the two closest to the transmembrane domain) is sufficient to support the activity of the protein, including intracellular trafficking, at levels normal or close to normal.[14−18]

The stretches of aminoacidic sequence linking the folded domains in the N-terminal tail are poorly structured.[19,20] Notably, the length of such linker regions is very variable,[5] both for linkers connecting different domains within the same protein or for linkers between corresponding domains in different proteins, and ranges from three to several tens of amino acids. In the systems for which detailed experimental studies on the structure at the atomic level of the N-terminal tail are available,[20−24] the last two domains are connected by a relatively small number of amino acids (less than ten, to be compared to a few tens of amino acids linking the most N-terminal domains). The features of the flexible N-terminal tail and its structural plasticity are important for the understanding of the overall functioning of P_{1B} ATPases. Indeed, there is a substantial body of evidence that the interaction between the various regions of the enzyme as well as with the metallochaperone affect significantly its activity and, for the mammalian enzymes, the balance between onward and backward protein trafficking.[11,25] A possible role of the N-terminal tail could be that of modulating the ATPase activity through metal-dependent interactions with the ATP-binding domain and the A domain of the enzyme.[26]

In the present work, we aimed at furthering our understanding of the role of the linker region through molecular dynamics simulations of systems having different linker length and through a bioinformatic analysis of the spacing occurring between corresponding domains in various ATPases with multidomain cytoplasmic tails.

METHODS

Sequence Analysis. We used the SMART (http://smart.embl-heidelberg.de/) database[27] to identify proteins having a domain architecture similar to that of the yeast copper-transporting ATPase *Ccc*2, i.e. containing at least two soluble metal-binding domains in addition to all other domains characteristic of P_{1B}-type ATPases (date of access: June 2008). Incomplete (i.e., containing only protein fragments) sequences were discarded. The SMART database contains protein sequences from Swiss-Prot and spTrembl databases as well as Ensembl proteomes.[27] Archaeal, bacterial, and eukaryotic organisms were investigated. We retrieved more than 400 proteins and extracted from the latter the sequences of the two metal-binding domains closest to the transmembrane region. These sequences were aligned to check the conservation of the metal-binding CXXC motif. The proteins in which one or both of the two domains lacked the motif were removed. Then we separated the domain pairs on the basis of the length of the polypeptide region linking the two domains by a locally written program, which exploited the definition of domain boundaries of SMART. We selected a few different representative systems and performed molecular dynamics simulations on their apo and holo-forms: *Bacillus subtilis*; the human Menkes's disease protein (MNK); the human Wilson's disease protein (WND); *Deinococcus geothermalis*; and *Brucella abortus* domains whose interdomain linker regions comprise three, seven, seven, eleven, and thirty-five amino acids, respectively.

Among these five, only for *B. subtilis* and WND the structure of the two-domain construct is in the PDB (entries 1P6T[23] and 2EW9,[22] respectively). Structural models for the other sequences were thus built using the program Modeller with standard parameters on the basis of these two available structures. Because of the choice of the template structures, in all MD calculations the two domains were close in space at the beginning of the simulation. Several models were generated for each sequence, and each of them was then visually inspected. Models without apparent defects were

ranked on the basis of their stereochemical quality and energy; the best model was used as input for molecular dynamics simulations. For each sequence, only the model with the best stability in the first hundreds of picoseconds of the trajectory (after equilibration) was retained.

We used the Pfam[28] domains HMA, E1_E2_ATPase, and Hydrolase, which are contained in P-type ATPases and metallochaperones (HMA only), to investigate the occurrence of these proteins in completely sequenced prokaryotic proteomes. Note that the E1-E2 ATPase domain is specific of P-type ATPases, whereas the Hydrolase domain is not specific but is required for phosphatase activity. The program HMMER[29] was used for this purpose, with an E-value threshold of 10^{-5} (i.e., only domains with an E-value better than 10^{-5} were retained for analysis). Proteome sequences were retrieved from the Ref_Seq database[30] (date of access: June 2008). We used this approach, which is similar to what was done in other studies by our and other laboratories,[31−34] to obtain a more detailed view in a data set of protein sequences more restricted than SMART. The present data set however had the advantage of including only complete proteomes and therefore allowed us to perform meaningful comparisons among the ATPase and metallochaperone content of different organisms. P-type ATPases were identified by the simultaneous presence of the Hydrolase and E1_E2_ATPase domains (which define their common basic "core" architecture[8]); the additional presence of one or more HMA domains identified P_{1B-1} or P_{1B-2} ATPases (Figure 1).

Molecular Dynamic Simulations (MD). AMBER 8.0[35] was used to make individual simulations both in apo and in holo forms. Holo forms were built from the corresponding apo forms by adding a copper(I) ion in between the two Sγ atoms of the cysteines of the motif CXXC. To do so, the side chains of the cysteines were properly preoriented by restraining the distance between the Sγ atoms and minimizing the structure of the apoprotein. After insertion of the copper(I) ion, the Sγ−Cu−Sγ angle was loosely restrained at 180 degrees for a short time during the MD. Parameters for the metal site were taken from refs 36 and 37. The SHAKE algorithm was used to maintain bond lengths fixed, permitting the use of a time-step of 1.5 fs. The protein was solvated using TIP3P water and a 10 Å buffering distance between the edges of the box and the protein. Initially, we minimized the energy of each system in two stages. In the first stage, we minimized the water molecules while holding the protein and counterions fixed, in order to relax solute−solvent contacts. In the second stage, we minimized the complete system. MD were then performed using periodic boundary conditions, at constant pressure (1 atm) and temperature (298 K), for six nanoseconds. For each trajectory, we analyzed a portion of four nanoseconds after the system had equilibrated. rms versus time graphs are shown in Figure S1.

RESULTS

Separation of MBDs and Its Effect on the System Properties. We retrieved the sequences of P_{1B}-type ATPases that contained at least two metal-binding domains (MBDs) from both the SMART database, which contains sequences from organisms in all domains of Life, and completely sequenced prokaryotic proteomes, in the latter case using the releant HMMs provided by the Pfam database. This resulted in, respectively, 320 and 290 sequences. The length of the linker between the two MBDs closest to the transmembrane region (highlighted by an arrow in Figure 1) was determined using the domain definitions of either the SMART or the Pfam database. A summary of the data is given in Table 1. As shown in Figure 2, the computed linker lengths featured a significant variability. The most likely separation between the two MBDs taken into account was three amino acids, which was equal or very close to the first quartile of the distribution (Table 1). Nevertheless, the third quartile was 16−17 amino acids, implying that one-quarter of the sequences (i.e., 70−80) in the ensembles examined had linker regions with a length exceeding this value. The computed separation did not depend on the total number of domains in the ATPase (not shown).

Table 1. Distribution of the Lengths of the Linker Connecting the Two MBDs Closest to the Transmembrane Domain in the Pfam and SMART Data Sets of P_{1B}-type ATPases

	Pfam	SMART
mean	15 ± 22	16 ± 24
first quartile	4	3
third quartile	17	16
95th percentile	49	70
minimum value	1	3
maximum value	144	179

We then verified whether, independently of their separation, two consecutive MBDs formed a tight unit thanks to energetically favorable interdomain interactions. To evaluate this hypothesis we built structural models of two-MBD units with different linker lengths and subjected them to molecular dynamics simulations. During MD, individual MBDs remained stable with backbone rms deviations, after equilibration, in the range 1−2 Å. On the other hand, when considering their relative position, we observed larger rearrangements with respect to the initial orientation the longer the linker length (Supporting Information Figure S1). These rearrangements did not depend on the modeling procedure, as the MD trajectories starting from the model

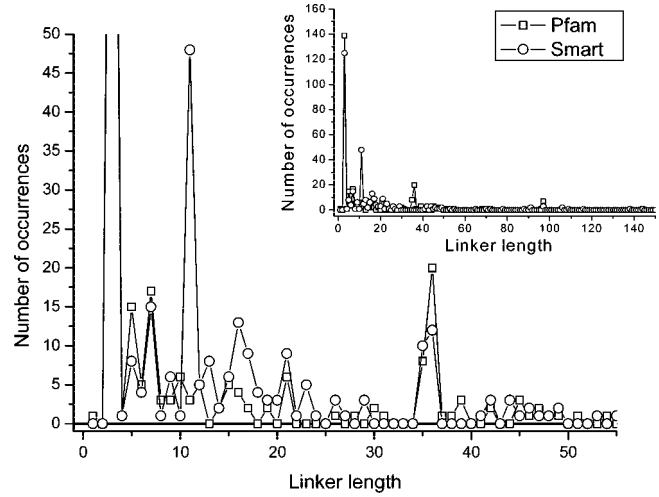

Figure 2. Number of proteins with a given length of the linker connecting the two MBDs closest to the transmembrane domain. Open squares: proteins from complete proteomes retrieved using the Pfam domains. Open circles: proteins from the SMART data set. Inset: full graph. The *R* correlation coefficient between the two data sets is 0.91.

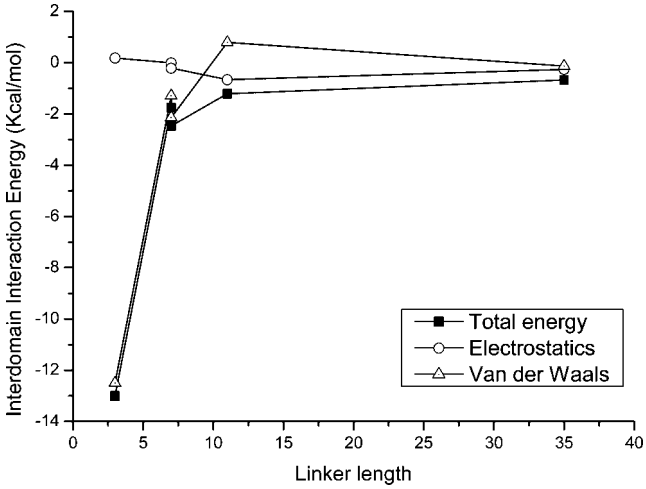

Figure 3. Interdomain interaction energy (separated into electrostatics, van der Waals, and total energy, including also hydrogen bonding contributions).

structure of the MNK protein feature rmsd values as large as (for the apoprotein) or even smaller than (for the copper-protein) those observed for the simulation of the WLN protein, which has the same linker length of MNK and started from an experimental structure (Supporting Information Figure S1). We evaluated the energies of interdomain interaction along the various trajectories and computed for each system the average energy of the apo- and copper-protein in the 4 ns production trajectory. The results are shown in Figure 3 as a function of the separation in sequence of the two MBDs. It can be readily observed that the average interdomain interaction energy falls sharply with increasing linker length.

Distribution of Metallochaperones and ATPases in Prokaryotic Organisms. In this work, we investigated selected members of a subclass of P-type ATPases, namely copper(I)-transporting P_{1B}-type ATPases. These ATPases catalyze the transport of copper(I) ions across biomembranes at the expenses of ATP hydrolysis. The catalytic cycle involves the formation of a covalent phosphorylated "P" intermediate, hence the label "P-type".[6] P_1-type ATPases are characterized by a common organization of their transmembrane segments that is distinct from that of P_2-type ATPases,[7] and the P_{1B} subtype is in particular responsible for the transport of heavy metals (Cd^{2+}, Zn^{2+}, Pb^{2+}, Co^{2+}, Cu^{2+}, Ag^{+}, Cu^{+}).[8,9] The entire group of P_{1B}-type ATPases can be further split in various subgroups, which have specific structural and sequence features throughout the entire polypeptide that are linked to their metal specificity.[38,39] The presence of (most frequently) N-terminal, cytoplasmic MBDs typically harboring a CXXC metal binding pattern is common to P_{1B}-type ATPases transporting Cu^{+} (subgroup 1B-1), Zn^{2+}, Cd^{2+}, and Pb^{2+} (subgroup 1B-2).[38−40] The MBDs themselves contain sequence features that help in discriminating the above metals.[41] ATPases containing multiple MBDs normally transport copper(I) and exceptionally transport zinc(II).[39]

It is known that copper(I)-transporting ATPases from eukaryotes tend to have two or more MBDs.[5] We investigated how common these are in prokaryotic systems, by scanning the complete proteome sequences of as many as 594 organisms using the Pfam HMM's representing the HMA (corresponding to the soluble MBD), Hydrolase, and E1_E2_ATPase domains. P-type ATPases were identified by the simultaneous presence of the latter two domains (which define their common basic "core" architecture,[8] see Figure 1); proteins were instead classified as P_{1B-1} or P_{1B-2} if they additionally contained one or more HMA domains. At the present level of investigation, the P_{1B-1} and P_{1B-2} subtypes cannot be distinguished; therefore, from now on we will refer to the ATPases of P_{1B-1} or P_{1B-2} type as $P_{1B-1,2}$-type. Hereafter, the figures referring to P-type ATPases exclude $P_{1B-1,2}$-type ATPases. The results obtained are summarized in Table 2.

Table 2. Summary of the Results of the Analysis on 594 Fully Sequenced Prokaryotic Genomes Using Pfam Domains

	number of proteins	number of organisms (percentage of the 594 organisms analyzed)
Individual Proteins		
metallochaperone	521	335 (56.4%)
P-type ATPase (excluding $P_{1B-1,2}$ type)	1625	462 (77.8%)
$P_{1B-1,2}$ ATPase	826	468 (78.8%)
none	na	50 (8.4%)
Combinations of Proteins		
metallochaperone and $P_{1B-1,2}$ATPase	1072 (499 + 573)	315 (53.0%)
metallochaperone and not $P_{1B-1,2}$ATPase	22	20 (3.4%)
$P_{1B-1,2}$ATPase and not metallochaperone	253	153 (25.8%)
not $P_{1B-1,2}$ATPase and not metallochaperone	na	106 (17.8%)
P-type ATPase and $P_{1B-1,2}$ ATPase	2082	386 (65.0%)
P-type ATPase and not $P_{1B-1,2}$ ATPase	240	76 (12.8%)
$P_{1B-1,2}$ ATPase and not P-type ATPase	129	82 (13.8%)

Sixty-five percent of the prokaryotic organisms analyzed simultaneously contained one or more $P_{1B-1,2}$-type ATPases and one or more other P-type ATPases. In addition, 82 (14%) organisms encoded only $P_{1B-1,2}$-type ATPases, yielding a total of 468 (79%) organisms that encoded at least one $P_{1B-1,2}$-type ATPase. In all the organisms analyzed, we detected only three instances of proteins containing the E1-E2 ATPase but not the Hydrolase domain, which is not specific of P-type ATPases but is nevertheless required for function. When counting individual proteins, we detected as many as 826 $P_{1B-1,2}$-type ATPases. Most of these organisms encoded one or two $P_{1B-1,2}$ ATPases; three was also relatively common. *Haloarcula marismortui* encoded nine different such ATPases, which is probably related to its halophilic lifestyle. Bacterial $P_{1B-1,2}$-type ATPases had between one and four MBDs (Table 3). Fourteen organisms encoded a single ATPase with four MBDs, with the separation between the two MBDs closest to the transmbrane domain ranging between less than 10 and 30 amino acids. In *Ralstonia metallidurans* and *P. lavamentivorans*, the four MBDs did not contain the canonical CXXC metal-binding pattern, which was replaced by CXXEE. For *R. metallidurans* this was implied with the substrate being lead(II) rather than copper(I), and the gene was called PbrA.[42] In the proteins with three MBDs, all of them had the canonical CXXC pattern and featured a quite variable spacing. For example, in the ATPases from *Yersinia* species the first and second

Table 3. Number of $P_{1B-1,2}$ ATPases with the Indicated Number of HMA Domains in 594 Fully Sequenced Prokaryotic Genomes

number of HMA domains	number of $P_{1B-1,2}$ ATPases	percentage over the total number of $P_{1B-1,2}$ ATPases
1	540	65.4%
2	245	29.7%
3	28	3.4%
4	13	1.6%

MBDs were closely spaced, whereas the third MBD was relatively distant (in sequence) from the second as well as from the transmembrane domain. Finally, 244 bacterial $P_{1B-1,2}$-type ATPases had two MBDs. *R. metallidurans* had two such ATPases, one with the canonical patterns and the other with the CXXEE patterns in the both MBDs. Proteins similar to the latter, i.e. having two MBDs with the CXXEE pattern, are found in a variety of organisms (e.g., *Klebsiella pneumoniae*, *Shewanella frigidimarina*, etc.). Instances of proteins in which the two MBDs have different patterns also exist, not only in *R. metallidurans*[42] but also in distant organisms such as *Aeropyrum pernix*.

We analyzed also the distribution of the proteic partners of copper(I)-transporting ATPses, i.e. metallochaperones. We detected putative metallochaperones, assigned as such on the basis of their sequence containing only MBDs and no other known domains, in 335 organisms (Table 2). Of these, only 20 did not contain any $P_{1B-1,2}$-type ATPase. On the other hand, the majority of the organisms lacking a metallochaperone encoded one or more $P_{1B-1,2}$-type ATPases. Of the metallochaperones potentially without a partner $P_{1B-1,2}$-type ATPase, 10 were from as many *Lactobacilli* species and were all adjacent in the genome to a P-type ATPase lacking any MBD. The same was true for *Leuconostoc mesenteroides*. In *Thermoplasma acidophilum* and *Thermoplasma volcanium* the only metallochaperone encoded was instead next to a mercuric reductase (MerA) and thus presumably had Hg^{2+} as its target (i.e., was part of an operon containing the *merA* and *merP* genes;[43] MerP features a single HMA domain). In *Thermofilum pendens* the gene encoding the metallochaperone was next to one encoding a protein containing a rubrerythrin domain, which binds iron. Other instances of organisms containing a metallochaperone but not $P_{1B-1,2}$-type ATPase appeared as sequencing errors, where the ATPase sequence was interrupted or abnormal in some way.

DISCUSSION

P_{1B}-type ATPases of subgroups 1 and 2, which transport respectively Cu^+ or Zn^{2+}, Cd^{2+}, Pb^{2+} ion, contain between one and six MBDs.[5] Prokaryotic ATPases contain up to four MBDs (Table 3). The separation in protein sequence between the two cytoplasmic metal-binding domains (MBDs) is highly variable (Figure 2). In structurally characterized two-domain systems (*B. subtilis* CopA[23,24] and the human protein WND[22]), it has been proposed that the two MBDs that are closest to the transmembrane region domains form relatively tight units, possibly owing to their short linkers. In more complex multidomain constructs, such as a three-domain construct from the MNK protein, the two terminal MBDs (domains 5 and 6) are more rigidly connected to one another than domain 5 is connected to the preceding MBD (domain 4).[20] A study of the entire N-terminal tail of MNK is also available, where it is readily observed that within any pair of consecutive domains besides 5 and 6 the two MBDs are nearly completely free to reorient with respect to one another.[21]

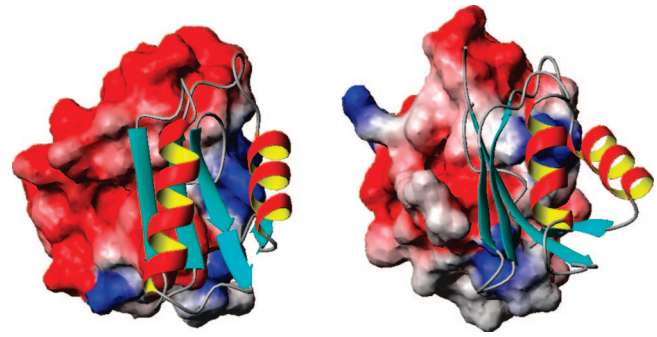

Figure 4. Surface electrostatic potential of the two MBDs of the WND protein (based on the experimental 2EW9 structure). Left: surface of the first domain show, with the second domains in ribbon representation. Right: vice versa. The two panels are interchanged by a rotation of 180° along the vertical axis. The electrostatic potential was calculated with the program MOLMOL.[51]

The presence of a relatively short linker region between the two MBDs closest to the trans-membrane part of the enzyme was proposed to be instrumental to maintain a fixed relative orientation of their respective metal-binding sites.[21] This would avoid interactions between them in the presence of the metal substrate. In *Anabaena* AztA the interaction between a pair of MBDs can result in enhanced selectivity in a zinc(II)-transport system, where other divalent cations such as lead(II) or cadmium(II) form a stable metal-bridged intermediate involving the two MBDs that inhibits the ATPase.[44] Within this frame it is notable that a small number of copper-transporting ATPases feature one MBD at the N-terminal part of the trans-membrane region and one at the C-terminal part, which are therefore truly independent units. These are the ATPases from *Archeoglobus fulgidus*,[45] *Bacteroides fragilis*, and *Troponema denticola*. A model of the structure of the *A. fulgidus* ATPase is available,[26] where however the C-terminal MBD has been removed and thus no information on its possible interaction with the N-terminal MBD.

In principle, the length of the linker separating two consecutive MBDs is not sufficient to establish whether they form a tight unit or not. Indeed, there could exist interdomain interactions that are sufficiently stable (from the thermodynamic point of view) to hold the two together regardless of the length of the linker. However, MD simulations showed that the two domains display increasing freedom of relative reorientation with increasing separation in sequence and, in parallel, less significant energies of interdomain interaction (Figure 3). Therefore, the linker effectively uncouples the two MBDs. These conclusions applied to both the apo and holo forms, possibly to a larger extent in the latter than in the former even though metalation did not trigger significant changes within a single MBD, besides the region of the metal-binding site.[46] Indeed, the regions of interdomain contact at the surfaces of the two MBDs appear in general to be poorly optimized for interaction. For example, Figure 4 shows the electrostatic potential at the interfaces of the two MBDs of the WND protein, based on the experimental 2EW9 structure: it can be noted that regions with similar

electrostatic potential are in proximity (at the top and in the center of the interfacial regions). The above MD data are in agreement with the experimental NMR relaxation data for the two-domain systems from *B. subtilis* CopA and from WND, which show that the linker region is essentially as rigid as the rest of the protein in the former[23] but more mobile in the latter.[22] This confirms that even the short increase in linker length from the CopA to the WND protein (from three to five residues) is sufficient to appreciably increase the relative freedom of the two MBDs. Analogously, relaxation data for the three-domain construct of MNK already mentioned, which contains the fourth, fifth and sixth MBDs, demonstrate the flexibility of all linker regions.[20] The linker between the fourth and fifth MBDs is nearly 50 amino acids in length and displays significantly higher flexibility than the linker between the fifth and sixth MBDs, which comprises only seven amino acids. The order parameters for bond vectors in the linker residues extracted from our simulations fell sharply from around 0.9 to around 0.5 when going from the shortest to the longest linker. In summary, the sequence analysis data (Table 1 and Figure 2) and the MD simulations concur to establish that even the MBD pairs closest to the transmembrane domain do not need to have a fixed orientation with respect to one another. Relatively short linker lengths already permit an appreciable degree of relative conformational freedom; systems where the linker length is sufficient to allow the two MBDs to bring their metal-binding sites at short distance are fairly common (Table 1).

As mentioned, P_{1B}-type ATPases are enzymes that transport heavy metals across biomembranes and play a key role in the homeostasis and the mechanisms of biotolerance of these metals.[38−40] Here we focused on $P_{1B-1,2}$-type ATPases, which feature N-terminal, cytoplasmic MBDs. ATPases containing multiple MBDs normally transport copper(I) or, less commonly, divalent cations such as zinc(II).[39] The common metal-binding pattern for these systems is CXXC; we also detected proteins with CXXEE patterns that are presumably involved in the transport of divalent cations, such as lead(II). In some instances, proteins with two MBDs had a different metal-binding pattern in each of them. These systems were common to phylogenetically distant organisms. Notably, extensive horizontal gene transfer of $P_{1B-1,2}$-type ATPases has been recently reported among bacteria isolated from subsurface soils contaminated by metals, involving not only proteobacteria but also actinobacteria and firmicutes.[47]

We carried out an extensive investigation of the distribution of P-type ATPases and of their partner metallochaperones in prokaryotes (Table 2). Metallochaperones always occurred as single-MBD proteins, with the five detected exceptions appearing again as sequencing errors, introducing breaks within $P_{1B-1,2}$-type ATPase sequences. They are known to be involved in the transport of either copper(I) or mercury(II).[43] In the currently accepted model, the copper(I) ion is bound by the soluble metallochaperone and then transferred to the metal-binding site in the transmembrane region of the ATPase via its soluble MBDs, to be eventually removed from the cytosol (Figure 5, pathway I). An analogous pathway is followed for mercury(II) detoxification. In some organisms it is actually possible to find both copper(I) chaperones and mercury(II) chaperones, having respectively a $P_{1B-1,2}$-type ATPase and a protein of the *mer* operon as their partners. Notably, MerA contains a MBD

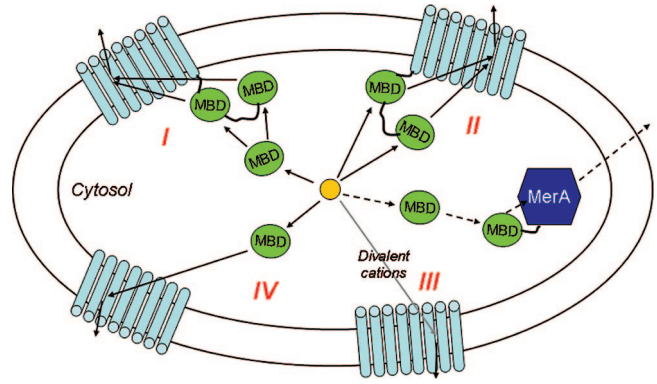

Figure 5. Possible pathways for the export of a metal ion from the cytosol of a prokaryotic cell by a P-type ATPase. The metal ion is represented by the orange sphere at the center of the figure. Only the transmembrane helices and, for $P_{1B-1,2}$ ATPases, the cytosolic MBDs are shown. Metallochaperones are represented by a single MBD. The various pathways are numbered with roman numerals (clockwise order). Pathways I, II and IV are relevant for copper(I), whereas pathway III is presumably relevant only for some divalent cations.[39] The mercury(II) detoxification pathway is also shown for completeness, with dashed arrows.

domain as well, which is the domain for interaction with the Hg^{2+} chaperone[43] (Figure 5). In *Streptococcus gordonii* we identified an operon with a MerA containing two MBD domains and a MerP homologue. Metallochaperones were detected only in proteomes encoding also at least one $P_{1B-1,2}$-type ATPase, with only 22 exceptions (Table 2). These included 10 *Lactobacilli*, where the metallochaperone was in the same operon of a P-type ATPase lacking any MBD. However, in 75 organisms we detected a larger number of metallochaperones than $P_{1B-1,2}$-type ATPases. In fact, we observed the occurrence of metallochaperones in the same operon of P-type ATPases not belonging to the 1B-1 or 1B-2 subgroups also in the proteome of organisms that did encode $P_{1B-1,2}$-type ATPases. This finding suggests that the ATPase does not strictly need to receive its metal substrate through a cytosolic domain first (Figure 5, pathway IV). Indeed, for CopA from *A. fulgidus* it has been recently suggested that the metallochaperone may interact directly with the metal-binding site in the transmembrane region, without an intermediate step involving the MBDs of the ATPase.[45,48] This is reminiscent of the proposed mechanism of function of the SERCA Ca^{2+}-ATPase, which is an archetype for P-type ATPases, where calcium(II) is directly bound to the trans-membrane site[49] (Figure 5, pathway III). In other instances, metallochaperones had as their protein partner a MerA homologue. Finally, the observation that a large number of organisms encode $P_{1B-1,2}$-type ATPase sequences but not metallochaperones (Table 2) strongly suggests that these enzymes can sequester their metal substrate directly in the cytoplasmic space, either from a small-molecule complex or from an unidentified metallochaperone (Figure 5, pathway II).

CONCLUDING REMARKS

We investigated the occurrence and properties of $P_{1B-1,2}$-type ATPases and, partly, of their partner metallochaperones. $P_{1B-1,2}$-type ATPases may contain multiple MBDs in the N-terminal cytoplasmic part of their sequence. This, together with the observation that the number of these domains tends

to be higher in more complex organisms, has stimulated numerous studies of their overall properties, function, and specialization. In particular, it has been shown that in the human MNK and WND proteins, which contain six MBDs, the two domains closest in sequence to the transmembrane part of the protein have a different role than the other four.[10] Here we demonstrated through sequence analysis across a large data set of organisms and molecular dynamics that in the majority of organisms these two domains are structurally independent and can reorient one with respect to the other. Therefore, it appears very likely that they can exert their function independently. The MBDs have a role in regulating the ATPase activity through interactions with the other protein domains.[26] The presently observed relative flexibility can thus be instrumental in optimizing the regulation of the activity in multi-MBD ATPases.

Metallochaperones are single-MBD proteins that typically deliver copper(I) ions to partner $P_{1B-1,2}$-type ATPases or mercury(II) ions in the mercury detoxification system. However, also when involved in the latter process, metallochaperones are detected nearly exclusively in organisms encoding $P_{1B-1,2}$-type ATPases as well, possibly indicating that they evolved originally to interact with the ATPases and then adapted to scavenge also mercury(II). We also described several hints suggesting that metallochaperones might interact also with P-type ATPases lacking MBDs (Pathway IV in Figure 5). Conversely, and much more commonly, there are several $P_{1B-1,2}$-type ATPases that likely function in the absence of a partner metallochaperone (Pathway II). This is typically the case of ATPases transporting divalent cations.[39] For these systems, in addition to modulating the overall enzymatic activity, the reciprocal mobility of MBDs could be important to tune the selectivity and/or the affinity for the substrate, which could be the metal ion complexed to either an organic molecule[50] or an unidentified cytoplasmic metallochaperone.

ACKNOWLEDGMENT

We thank prof. Ivano Bertini for useful discussions. MIUR is acknowledged for financial support (projects PRIN, prot. 2007M5MWM9, and FIRB, prot. RBLA032ZM7).

Supporting Information Available: Backbone rmsd vs time plots for all MD simulations (Figure S1). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Linder, M. C. *Biochemistry of Copper*; Plenum Press: New York, 1991; pp 1−13.
(2) Andreini, C.; Banci, L.; Bertini, I.; Rosato, A. Occurrence of copper through the three domains of life: a bioinformatic approach. *J. Proteome Res.* **2008**, *1*, 209–216.
(3) Cavet, J. S.; Borrelly, G. P.; Robinson, N. J. Zn, Cu and Co in cyanobacteria: selective control of metal availability. *FEMS Microbiol. Rev.* **2003**, *27*, 165–181.
(4) O'Halloran, T. V.; Culotta, V. C. Metallochaperones: An Intracellular Shuttle Service for Metal Ions. *J. Biol. Chem.* **2000**, *275*, 25057–25060.
(5) Arnesano, F.; Banci, L.; Bertini, I.; Ciofi-Baffoni, S.; Molteni, E.; Huffman, D. L.; O'Halloran, T. V. Metallochaperones and metal transporting ATPases: a comparative analysis of sequences and structures. *Genome Res.* **2002**, *12*, 255–271.
(6) Singleton, C.; Le Brun, N. E. Atx1-like chaperones and their cognate P-type ATPases: copper-binding and transfer. *Biometals* **2007**, *20*, 275–289.
(7) Lutsenko, S.; Kaplan, J. H. Organization of P-type ATPases: significance of structural diversity. *Biochemistry* **1995**, *34*, 15607–15613.
(8) Moller, J. V.; Juul, B.; le Maire, M. Structural organization, ion transport, and energy transduction of P-type ATPases. *Biochim. Biophys. Acta* **1996**, *1286*, 1–51.
(9) Axelsen, K. B.; Palmgren, M. G. Evolution of substrate specificities in the P-type ATPase superfamily. *J. Mol. Evol.* **1998**, *46*, 84–101.
(10) Lutsenko, S.; Barnes, N. L.; Bartee, M. Y.; Dmitriev, O. Y. Function and regulation of human copper-transporting ATPases. *Physiol. Rev.* **2007**, *87*, 1011–1046.
(11) La Fontaine, S.; Mercer, J. F. Trafficking of the copper-ATPases, ATP7A and ATP7B: role in copper homeostasis. *Arch. Biochem. Biophys.* **2007**, *463*, 149–167.
(12) Clausen, J. D.; Vilsen, B.; McIntosh, D. B.; Einholm, A. P.; Andersen, J. P. Glutamate-183 in the conserved TGES motif of domain A of sarcoplasmic reticulum Ca2+-ATPase assists in catalysis of E2/E2P partial reactions. *Proc. Natl. Acad. Sci. U.S.A.* **2004**, *101*, 2776–2781.
(13) Lutsenko, S.; Petrukhin, K.; Cooper, M. J.; Gilliam, C. T.; Kaplan, J. H. N-terminal domains of human copper-transporting adenosine triphosphatases (the Wilson's and Menkes disease proteins) bind copper selectively in vivo and in vitro with stoichiometry of one copper per metal-binding repeat. *J. Biol. Chem.* **1997**, *272*, 18939–18944.
(14) Payne, A. S.; Gitlin, J. D. Functional expression of the Menkes disease protein reveals common biochemical mechanisms among the copper-transporting P-type ATPase. *J. Biol. Chem.* **1998**, *273*, 3765–3770.
(15) Huster, D.; Lutsenko, S. The distinct roles of the N-terminal copper-binding sites in regulation of catalytic activity of the Wilson's disease protein. *J. Biol. Chem.* **2003**, *2789*, 32212–32218.
(16) Cater, M. A.; Forbes, J. R.; La Fontaine, S.; Cox, D.; Mercer, J. F. Intracellular trafficking of the human Wilson protein: the role of the six N-terminal metal-binding sites. *Biochem. J.* **2004**, *380*, 805–813.
(17) Strausak, D.; La Fontaine, S.; Hill, J.; Firth, S. D.; Lockhart, P. J.; Mercer, J. F. The role of GMXCXXC metal binding sites in the copper-induced redistribution of the Menkes protein. *J. Biol. Chem.* **1999**, *274*, 11170–11177.
(18) Voskoboinik, I.; Strausak, D.; Greenough, M.; Brooks, H.; Petris, M.; Smith, S.; Mercer, J. F.; Camakaris, J. Functional analysis of the N-terminal CXXC metal-binding motifs in the human menkes copper-transporting P-type ATPase expressed in cultured mammalian cells. *J. Biol. Chem.* **1999**, *274*, 22008–22012.
(19) Walker, J. M.; Huster, D.; Ralle, M.; Morgan, C. T.; Blackburn, N. J.; Lutsenko, S. The N-terminal metal-binding site 2 of the Wilson's disease protein plays a key role in the transfer of copper from Atox1. *J. Biol. Chem.* **2004**, *279*, 15376–15384.
(20) Banci, L.; Bertini, I.; Cantini, F.; Chasapis, C.; Hadjiliadis, N.; Rosato, A. A NMR study of the interaction of a three-domain construct of ATP7A with copper(I) and copper(I)-HAH1: the interplay of domains. *J. Biol. Chem.* **2005**, *280*, 38259–38263.
(21) Banci, L.; Bertini, I.; Cantini, F.; Della Malva, N.; Migliardi, M.; Rosato, A. The different intermolecular interactions of the soluble copper-binding domains of the Menkes protein, ATP7A. *J. Biol. Chem.* **2007**, *282*, 23140–23146.
(22) Achila, D.; Banci, L.; Bertini, I.; Bunce, J.; Ciofi-Baffoni, S.; Huffman, D. L. Structure of human Wilson protein domains 5 and 6 and their interplay with domain 4 and the copper chaperone HAH1 in copper uptake. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 5729–5734.
(23) Banci, L.; Bertini, I.; Ciofi-Baffoni, S.; Gonnelli, L.; Su, X. C. Structural basis for the function of the N terminal domain of the ATPase CopA from *Bacillus subtilis*. *J. Biol. Chem.* **2003**, *278*, 50506–50513.
(24) Singleton, C.; Banci, L.; Ciofi-Baffoni, S.; Tenori, L.; Kihlken, M. A.; Boetzel, R.; Le Brun, N. E. Structure an Cu(I)-binding properties of the N-terminal soluble domains of *Bacillus subtilis* CopA. *Biochem. J.* **2008**, *411*, 571–579.
(25) Lutsenko, S.; LeShane, E. S.; Shinde, U. Biochemical basis of regulation of human copper-transporting ATPases. *Arch. Biochem. Biophys.* **2007**, *463*, 134–148.
(26) Wu, C. C.; Rice, W. J.; Stokes, D. L. Structure of a copper pump suggests a regulatory role for its metal-binding domain. *Structure* **2008**, *16*, 976–985.
(27) Letunic, I.; Copley, R. R.; Pils, B.; Pinkert, S.; Schultz, J.; Bork, P. SMART 5: domains in the context of genomes and networks. *Nucleic Acids Res.* **2005**, *34*, D257-D260.
(28) Finn, R. D.; Mistry, J.; Schuster-Bockler, B.; Griffiths-Jones, S.; Hollich, V.; Lassmann, T.; Moxon, S.; Marshall, M.; Khanna, A.; Durbin, R.; Eddy, S. R.; Sonnhammer, E. L.; Bateman, A. Pfam: clans, web tools and services. *Nucleic Acids Res.* **2006**, *34*, D247-D251.
(29) Eddy, S. R. Profile hidden Markov models. *Bioinformatics* **1998**, *14*, 755–763.
(30) Pruitt, K. D.; Tatusova, T.; Maglott, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **2007**, *35*, D61-D65.

(31) Gabaldon, T.; Huynen, M. A. Prediction of protein function and pathways in the genome era. *Cell Mol. Life Sci.* **2004**, *61*, 930–944.
(32) Zambelli, B.; Musiani, F.; Savini, M.; Tucker, P.; Ciurli, S. Biochemical studies on *Mycobacterium tuberculosis* UreG and comparative modeling reveal structural and functional conservation among the bacterial UreG family. *Biochemistry* **2007**, *46*, 3171–3182.
(33) Lee, S. W.; Mitchell, D. A.; Markley, A. L.; Hensler, M. E.; Gonzalez, D.; Wohlrab, A.; Dorrestein, P. C.; Nizet, V.; Dixon, J. E. Discovery of a widely distributed toxin biosynthetic gene cluster. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5879–5884.
(34) Bertini, I.; Cavallaro, G.; Rosato, A. Cytochrome c: occurrence and functions. *Chem. Rev.* **2006**, *106*, 90–115.
(35) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Merz, K. M. ; Wang, B.; Pearlman, D. A.; Crowley, M.; Brozell, S.; Tsui, V.; Gohlke, H.; Mongan, J.; Hornak, V.; Cui, G.; Beroza, P.; Schafmeister, C. E.; Caldwell, J. W.; Ross, W. S.; Kollman, P. A. *AMBER 8. (8.0)*;University of California: San Francisco, CA, 2004.
(36) Fuchs, J. F.; Nedev, H.; Poger, D.; Ferrand, M.; Brenner, V.; Dognon, J. P.; Crouzy, S. New model potentials for sulfur-copper(I) and sulfur-mercury(II) interactions in proteins: from ab initio to molecular dynamics. *J. Comput. Chem.* **2006**, *27*, 837–856.
(37) Poger, D.; Fuchs, J. F.; Nedev, H.; Ferrand, M.; Crouzy, S. Molecular dynamics study of the metallochaperone Hah1 in its apo and Cu(I)-loaded states: role of the conserved residue M10. *FEBS Lett.* **2005**, *579*, 5287–5292.
(38) Argüello, J. M. Identification of ion-selectivity determinants in heavy-metal transport P1B-type ATPases. *J. Membr. Biol.* **2003**, *195*, 93–108.
(39) Argüello, J. M.; Eren, E.; Gonzalez-Guerrero, M. The structure and function of heavy metal transport P1B-ATPases. *Biometals* **2007**, *20*, 233–248.
(40) Solioz, M.; Vulpe, C. CPx-type ATPases: a class of P-type ATPases that pump heavy metals. *Trends Biochem. Sci.* **1996**, *21*, 237–241.
(41) Banci, L.; Bertini, I.; Ciofi-Baffoni, S.; Su, X. C.; Miras, R.; Bal, N.; Mintz, E.; Catty, P.; Shokes, J. E.; Scott, R. A. Structural basis for metal binding specificity: the N-terminal cadmium binding domain of the P1-type ATPase CadA. *J. Mol. Biol.* **2006**, *356*, 638–650.
(42) Mergeay, M.; Monchy, S.; Vallaeys, T.; Auquier, V.; Benotmane, A.; Bertin, P.; Taghavi, S.; Dunn, J.; van der, L. D.; Wattiez, R. Ralstonia metallidurans, a bacterium specifically adapted to toxic metals: towards a catalogue of metal-responsive genes. *FEMS Microbiol. Rev.* **2003**, *27*, 385–410.
(43) Brown, N. L.; Shih, Y. C.; Leang, C.; Glendinning, K. J.; Hobman, J. L.; Wilson, J. R. Mercury transport and resistance. *Biochem. Soc. Trans.* **2002**, *30*, 715–718.
(44) Liu, T.; Reyes-Caballero, H.; Li, C.; Scott, R. A.; Giedroc, D. P. Multiple metal binding domains enhance the Zn(II) selectivity of the divalent metal ion transporter AztA. *Biochemistry* **2007**, *46*, 11057–11068.
(45) Mandal, A. K.; Argüello, J. M. Functional roles of metal binding domains of the Archaeoglobus fulgidus Cu(+)-ATPase CopA. *Biochemistry* **2003**, *42*, 11040–11047.
(46) Rodriguez-Granillo, A.; Wittung-Stafshede, P. Structure and dynamics of Cu(I) binding in copper chaperones Atox1 and CopZ: a computer simulation study. *J. Phys. Chem. B* **2008**, *112*, 4583–4593.
(47) Martinez, R. J.; Wang, Y.; Raimondo, M. A.; Coombs, J. M.; Barkay, T.; Sobecky, P. A. Horizontal gene transfer of PIB-type ATPases among bacteria isolated from radionuclide- and metal-contaminated subsurface soils. *Appl. Environ. Microbiol.* **2006**, *72*, 3111–3118.
(48) Gonzalez-Guerrero, M.; Argüello, J. M. Mechanism of Cu+-transporting ATPases: soluble Cu+ chaperones directly transfer Cu+ to transmembrane transport sites. *Proc. Natl. Acad. Sci. U.S.A.* **2008**, *105*, 5992–5997.
(49) Toyoshima, C.; Nakasako, M.; Nomura, H.; Ogawa, H. Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 A resolution. *Nature* **2000**, *405*, 647–655.
(50) Cobine, P. A.; Ojeda, L. D.; Rigby, K. M.; Winge, D. R. Yeast contain a non-proteinaceous pool of copper in the mitochondrial matrix. *J. Biol. Chem.* **2004**, *279*, 14447–14455.
(51) Koradi, R.; Billeter, M.; Wüthrich, K. MOLMOL: a program for display and analysis of macromolecular structure. *J. Mol. Graphics* **1996**, *14*, 51–55.

CI8002304