Published in final edited form as:

J Biol Inorg Chem. 2014 August ; 19(6): 947–960. doi:10.1007/s00775-014-1129-2.

Diversity of the metal-transporting P_{1B}-type ATPases

Aaron T. Smith, Kyle P. Smith, and Amy C. Rosenzweig

Departments of Molecular Biosciences and of Chemistry, Northwestern University, 2205 Tech Drive, Evanston, IL 60208, USA

Amy C. Rosenzweig: amyr@northwestern.edu

Abstract

The P_{1B}-ATPases are integral membrane proteins that couple ATP hydrolysis to metal cation transport. Widely distributed across all domains of life, these enzymes have been previously shown to transport copper, zinc, cobalt, and other thiophilic heavy metals. Recent data suggest that these enzymes may also be involved in nickel and/or iron transport. Here we have exploited large amounts of genomic data to examine and classify the various P_{1B}-ATPase subfamilies. Specifically, we have combined new methods of data partitioning and network visualization known as Transitivity Clustering and Protein Similarity Networks with existing biochemical data to examine properties such as length, speciation, and metal-binding motifs of the P_{1B}-ATPase subfamily sequences. These data reveal interesting relationships among the enzyme sequences of previously established subfamilies, indicate the presence of two new subfamilies, and suggest the existence of new regulatory elements in certain subfamilies. Taken together, these findings underscore the importance of P_{1B}-ATPases in homeostasis of nearly every biologically relevant transition metal and provide an updated framework for future studies.

Keywords

P-type ATPase; P_{1B}-ATPase; Metal homeostasis; Metal transport; Transitivity clustering

Introduction

The P-type ATPases are a superfamily of integral membrane proteins that utilize energy derived from ATP hydrolysis to transport ions across cellular membranes [1, 2]. Widely distributed among all kingdoms of life, these multidomain proteins mediate fundamental processes, including transport of H⁺, Na⁺/K⁺, Ca²⁺, transition metal ions, and even phospholipids [1–5]. Over the years, these proteins have been subdivided into five classes (Table 1) based on substrate specificity, sequence similarity, and overall architecture. Every P-type ATPase is composed of three conserved domains: (1) the transmembrane helix (TM) bundle that includes 6–10 TMs and provides the substrate translocation pathway; (2) the ATP-binding domain (ATPBD), which includes the nucleotide-binding domain (N-domain)

© SBIC 2014

Correspondence to: Amy C. Rosenzweig, amyr@northwestern.edu.

Electronic supplementary material The online version of this article (doi:10.1007/s00775-014-1129-2) contains supplementary material, which is available to authorized users.

and invariant and transiently phosphorylated Asp residue located in the phosphorylation domain (P-domain); and (3) the actuator domain (AD), which is believed to transmit changes in the ATPBD to the transmembrane region and to drive dephosphorylation (Fig. 1) [6, 7]. The AD and ATPBD are soluble whereas the TMs are largely hydrophobic and are imbedded in the cellular membrane.

The best biochemically and structurally studied P-type ATPases are the “hard” cation transporters such as the Na⁺/K⁺ ATPases (subclass 2), the H⁺ ATPases (subclass 3), and especially the sarcoplasmic/endoplasmic reticulum Ca²⁺ transporter (SERCA, subclass 2) [1, 2, 8]. Early biochemical studies on these ATPases were used to establish the Post-Albers catalytic cycle, which describes the P-type ATPase as alternating between two cation (or substrate) binding states (E1, E2) driven by ATP-mediated phosphorylation and dephosphorylation [9, 10]. Structural data on SERCA (and subsequently the Na⁺/K⁺ and H⁺ ATPases) revealed specific conformational changes that occur concomitantly with the E1/E2 transition [11–14]. In addition to translocating redox inactive cations such as the alkali and alkaline earth metals, the P-type ATPases are also capable of translocating the redox active d-block metals (subclass 1B), linking these ATPases to overall cellular metal homeostasis.

The maintenance of metal homeostasis is crucial for normal metabolic functions. Regulation of transition metals and their handling is of particular importance because many of these metals are redox active; unabated, Fenton-like chemistry can create reactive oxygen species that can destroy peptides, DNA, and cell walls [15, 16]. Nearly a third of all proteins contain a metal cofactor of some sort [17], and insufficiencies in, or misloading of, transition metals in particular may cause metabolic stress [17, 18]. Thus, numerous regulatory mechanisms exist to control metal uptake, delivery, storage, and efflux [18–20]. Abnormalities in these processes are associated with metabolic defects in bacteria and plants [21, 22], as well as diseases in humans. For example, anomalies of Fe uptake are linked to hereditary and neurological diseases such as hemochromatosis, Parkinson’s syndrome, and Friedrich’s ataxia [19, 23, 24], whereas aberrations in the Cu⁺-transporting P_{1B}-ATPases lead to Menkes syndrome and Wilson’s disease [23, 25, 26].

The P_{1B}-ATPases have the largest substrate scope and play an essential role in transition-metal homeostasis, although relatively little is known about these enzymes. Originally identified as copper and zinc transporters, recent biochemical studies have demonstrated a wider substrate range of the P_{1B} subfamily [27–31]. The P_{1B}-ATPases were initially divided into several subfamilies based on a combination of biochemical data and sequence similarities among several characterized enzymes [3, 5]. Specifically, differences in a conserved sequence of three amino acids located in TM4 (Fig. 2) were hypothesized to be responsible for discrimination among various metal substrates. It was then realized that various P_{1B}-type ATPases with identical TM4 identifiers were capable of translocating various metal substrates, and several lines of evidence implicate residues in TM5 and TM6 in formation of the high-affinity TM metal-binding site(s) [3, 4, 32, 33]. Additional levels of regulation derive from N- and C-terminal soluble metal-binding domains (MBDs) (Fig. 2) [3, 4]. Thus, differences in conserved residues, metal substrate specificity, and soluble MBDs have historically determined subtype designation. However, despite the ballooning

amount of genomic information that has become available recently, a systematic analysis of the P_{1B} subfamilies has not been performed in nearly a decade [3].

Here we have combined genomic data with the BLAST similarity function and new methods of data partitioning and network visualization known as Transitivity Clustering and Protein Similarity Networks [34] to analyze the P_{1B}-ATPases. In these techniques, single protein sequences or multiple similar protein sequences are represented as nodes (points) on a map, and their relative connectivity or similarity (determined by an all-by-all BLAST expectation value) is represented as a series of edges (lines) among the nodes. This technique allows comparisons over large protein sequence data sets that would otherwise be intractable. In addition, using the suite of Cytoscape plug-ins [35], we have been able to map various properties such as type of MBD, length, and specific sequence motifs to compare various subfamilies. This analysis reveals key new insights into the P_{1B}-ATPases, including the presence of new and emerging subfamilies as well as possible functional divergence among established subfamilies.

Materials and methods

Data sources

All P_{1B}-ATPase sequences were obtained from the Universal Protein Resource (UniProt) Knowledgebase and Reference Clusters (<http://www.unprot.org>) or the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). For established subfamilies (P_{1B-1} to P_{1B-4}), input sequences were obtained by standard protein–protein BLAST searches (blastp) using representative sequences as inputs (e.g. *Archaeoglobus fulgidus* CopA, *Escherichia coli* ZntA, *Thermus thermophilus* CopB). To prevent possible clustering bias, sequences with randomized expectation (E) values were obtained and used as inputs for the clustering models. For emerging subfamilies (P_{1B-5} to P_{1B-7}), input sequences were obtained in a similar manner except representative sequences for the protein–protein BLAST were determined by the presence of a conserved fourth transmembrane helix (TM4) signature motif. Sequences for clustering models were similarly randomized by E value. For every subfamily, approximately 100 unique sequences with randomized E values were used, as well as subfamily representative sequences, except for the P_{1B-7} subfamily, for which only 12 sequences are currently known (Table S1), resulting in a total of 672 base sequences. Every sequence used represents its natural sequence length; no manually truncated data were used. The presence of the invariant DKTGT phosphorylation signature sequence and several strictly conserved amino acids in the AD and ATPBD were verified for each sequence via multiple sequence alignments that were performed using JalView v. 2.7 [36] implementing the ClustalW algorithm and the Blosom62 matrix [37, 38]. The predicted presence and number of TMs were determined using the TMHMM 2.0 online server (<http://www.cbs.dtu.dk/services/TMHMM/>) [39], the TMPred online server (http://www.ch.embnet.org/software/TMPRED_form.html) [40], and the TOPCONS online server (<http://topcons.net/>) [41].

Sequence similarity network generation

Sequence similarity networks were generated as described in [34] and [35]. Briefly, nodes were generated from obtained P_{1B}-ATPase sequences (vide supra), and edges (connectivities) were generated using the all-by-all BLAST function obtained from an unmodified local install of the NCBI BLAST suite (v. 2.2.28+) after curating a custom BLAST database and importing a file with all 672 base sequences that were not prealigned and contained no gaps in FASTA format. Initial networks were generated using the Blast2Sim Cytoscape (v. 2.8.3) plug-in [35] with an initial BLAST threshold of 100, a Coverage Factor of 15, and a “Sum of all Hits” similarity function. Low levels of connectivity were removed using a numeric filter until the four historically established subfamilies began to appear, as determined visually. Optimal levels of connectivity were determined via use of the TransClust plug-in with an initial value range set from 80 to 140 and a step size of 1. The greatest F-measure of the data set was determined to appear at the Transitivity Clustering setting of 82, and this value was set and applied to the similarity network. Properties of interest (e.g. length, predicted MBD motifs, conserved TM amino acid sequences, domain, phylum, subfamily) were tabulated manually and overlaid onto the network using the Cytoscape VizMapper plug-in [35]. Properties such as length and conserved TM amino acid sequences were determined by visual inspection of each sequence. Domain and phylum were determined from taxonomy provided from the UniProt database. In some cases, the kingdom was used instead of the phylum for simplicity to minimize the number of phyla designators necessary in the figure (e.g. *Metazoa* for all animals).

The presence and types of MBDs were determined in several ways. The presence of a CXXC motif was determined by manual inspection. Domains rich in metal-binding residues (His, Cys, Met, Asp, Glu) were identified based on the presence of 20 % composition of these residues prior to the onset of the first TM (N-terminal) or after the final TM (C-terminal). The presence of a hemerythrin-like (Hr-like) domain was determined by sequence alignments to the *Acidothermus cellulolyticus* and *Sinorhizobium meliloti* P_{1B-5}-ATPase C-terminal extensions and the conservation of putative diiron site ligands as described in [42] and [30]. Predicted folds defined by various MBDs were determined using the profile hidden Markov models implemented on the HMMER database (<http://hmmer.janelia.org/>) [43] and annotated by the Pfam database (<http://pfam.sanger.ac.uk/>) [44].

Genomic contextual data

Data used to illustrate the genomic neighborhoods of the P_{1B-6}- and P_{1B-7}-ATPases were obtained from manually curated databases as detailed in [45] and [46]. Putatively encoded proteins along open-reading frames (ORFs) are represented as arrows and are sized relative to one another; their annotation is taken directly from the database unless otherwise noted. Ferric uptake regulator (Fur) binding sequences are indicated and were identified based on known Fur sequences [46] for their respective prokaryote.

Results and discussion

Domain organization

The core structure of all P_{1B}-ATPases is likely very similar. Each P_{1B}-ATPase consists of 6–8 predicted transmembrane helices, a soluble AD, and a soluble ATPBD (in which the invariant DKTGT motif that contains the transiently phosphorylated aspartate residue is located). In addition, one or multiple soluble MBDs located at the N- or C-terminus may be present (Fig. 2) [4, 5]. The largest number of strictly conserved residues across all P_{1B}-ATPases is located in the AD or ATPBD, not in the transmembrane regions. This observation is consistent with previous analyses [47, 48], and the expanded sequence alignment has allowed us to refine our designations for residues that are conserved in the P_{1B}-ATPases (Fig. 2a).

The GE loop in the AD is preserved across all P-type ATPases and is responsible for driving dephosphorylation of the Asp residue in the ATPBD [6, 7, 14, 49]. Interestingly, in >90 % of the P_{1B} sequences, a Ser or Thr residue is found directly preceding the GE loop that may be involved in hydrogen bonding within the P-domain [47]. Moreover, there is a strictly conserved Asp residue near the N-terminus of the AD whose exact function is unknown. In the ATPBD, the final Thr in the DKTGTX_T sequence (corresponding to Thr¹⁰³³ in *Homo sapiens* ATP7B, a common Wilson's disease mutation site [48, 50]) is wholly conserved throughout our P_{1B} subset; one postulated function of this residue is to hydrogen bond with another sidechain to maintain structural stability throughout the ATPBD hinge region [48]. In many cases, this partner residue in the hinge region is an Asp; however, several P_{1B} sequences have a Thr in this position, suggesting incomplete conservation of this hydrogen-bonding pair. Closer to the C-terminus of the ATPBD is a completely conserved His residue (corresponding to His¹⁰⁶⁹ in *Hs*ATP7B, another common Wilson's disease mutation site [48, 50]), which may be involved in nucleotide-protein interactions [48]. Another interesting observation is the strongly conserved S/TGD motif in the N-domain, the sidechains of which likely make direct contact with the beta and gamma phosphates and the ribose moiety of the ATP molecule [48]. These residues are strictly conserved across all P_{1B}-ATPases, suggesting a common mode of interaction with the ATP molecule across all subfamilies. Finally, there are two sequences, PXXK (unknown function) and GXGXND (defines part of the interface between the N- and P-domains [48]), that appear to be wholly conserved across all P_{1B} sequences. Thus, comparison of hundreds of P_{1B}-ATPase sequences reveals conservation of specific residues that either directly contact ATP/ADP or drive transient interdomain interactions and are strictly conserved across all subfamilies [4, 47, 48, 51–53], highlighting the common catalytic mechanism despite the family's wide substrate diversity and limited overall sequence conservation.

Subfamily classification

Our clustering data reveal the presence of numerous and distinct P_{1B} subfamilies, independent of any biochemical information. After filtering low-level edge connectivity and generating our initial set of clustering data, several strong node bundles readily appeared, suggesting distinct groups. Upon reviewing the initially clustered sequences, it was clear that these initial groups represented the four historically established subfamilies (P_{1B-1} to

P_{1B-4}) (Table 2), which are each known to possess distinct motifs in their respective protein sequences, and are known to be related at the biochemical level (*vide infra*). After further edge filtration and the application of the Transitivity Clustering method, several distinct clusters emerged that appeared to belong to both new and previously identified subfamilies (Fig. 3). We have compiled and analyzed TM4 sequences, kingdom and phylum data, polypeptide length, and MBD information and superimposed these data onto the nodes (Figs. 3, 4, 5, 6, S1). These results are summarized by subfamily below.

Established subfamilies

P_{1B-1}-ATPases—The P_{1B-1}-ATPases are the best-characterized of the P_{1B}-type subfamilies [32, 53–57], chiefly stemming from their involvement in Menkes syndrome and Wilson’s disease in humans [25, 26, 58–61]. The P_{1B-1} sequences are composed of roughly 8 TMs: 6 TMs (1–6) that constitute the core α -helical bundle as well as two additional helices (A, B) (Fig. 2) [3, 4, 53, 55]. Enzymes from the P_{1B-1} subfamily are partly classified by the complete conservation of a CPC motif in TM4 (Figs. 2b, 3, 4), as well as strong conservation of a YN(X)₄P motif in TM5 and an MXXSS motif in TM6 (Table 3) [3, 62]. Members of the P_{1B-1} subclass have been shown to transport Cu⁺ as well as Ag⁺ in organisms across all kingdoms of life (Table 2; Figs. 4b, 5) [54, 56, 63]. These enzymes are on average the longest of the P_{1B}-type ATPases (approximately 875–900 amino acids); however, this length is highly variable, ranging from 650 to 1,500 amino acids total (Fig. S1). *A. fulgidus* CopA, a model ATPase for the P_{1B-1} subclass, has been shown to bind two Cu⁺ ions in the transmembrane region: the first Cu⁺ ion is coordinated via two Cys residues (deriving from the CPC motif) and one Tyr residue, and the second Cu⁺ ion is coordinated via a novel collection of Asn, Met, and Ser residues [64]. Our data indicate that the latter three residues are nearly completely conserved, suggesting that this second TM metal-binding site may be a general phenomenon of this subclass.

Clustering data on the P_{1B-1} subfamily reveal some interesting relationships. The protein sequences derive from all three domains of life (eukaryotic, archaeal, and prokaryotic) and represent a diverse array of organisms, from the large and multicellular (e.g. *H. sapiens*, *Mus musculus*) to the microscopic and unicellular (e.g. *Saccharomyces cerevisiae*, *Helicobacter pylori*, *A. fulgidus*) (Figs. 4, 5), underscoring the importance of Cu⁺ transport across all species in nature. Interestingly, the P_{1B-1} sequences display the largest difference in the overall length of the polypeptide chain of any subfamily (Fig. S1). Perhaps unsurprising, the longest of these sequences are those that derive from various eukaryotic organisms (Figs. 5, S1). Our clustering data and sequence analyses show that the high variability in overall sequence length is a consequence not of the number of TMs or lengths of loop regions, but rather of the diverse lengths of soluble N-terminal extensions that contain one or several tandem MBDs (Figs. 6, S1). Virtually every P_{1B-1} sequence contains at least one MBD with a characteristic Cu⁺-binding CXXC-containing motif with one or several predicted ferredoxin-like folds (Fig. 6). These N-terminal extensions appear before the beginning of TMA and may also be rich (up to 30 % composition) in common metal-binding residues (Fig. 6a). Notably, CopA from *Streptococcus pneumoniae* has been recently shown to contain an N-terminal cupredoxin-like fold that is capable of binding a dicopper center, instead of the more common ferredoxin-like fold (Fig. 6b) [65]. In our data, this sequence

does not cluster strongly with other P_{1B-1} sequences and appears to be one of only two sequences lacking an N-terminal CXXC motif. The presence of this cupredoxin-like motif in SpCopA may be responsible for delivery of copper to the TM metal-binding site in this P_{1B-1}. Additionally, in at least one case, the presence of a ferredoxin-like fold on both the N- and C-termini has been identified (Fig. 6b) [66]. In some instances, these MBDs have been shown to be able to receive Cu⁺ ions from Atx1-like metallochaperones, and Cu⁺ binding by these domains alters ATPase activity as well as intracellular localization of the enzyme [54, 67, 68]. The function of these MBDs remains controversial [32, 68–72], but it would be logical for these domains to represent allosteric regulation in higher organisms, since metal homeostasis in eukaryotes is more complex than in prokaryotes, as has been previously suggested [32, 73].

Intriguingly, the Menkes syndrome (ATP7A) and Wilson's disease proteins (ATP7B) from human (*H. sapiens*), mouse (*M. musculus*), and rat (*Rattus norvegicus*) cluster strongly with other bacterial and archaeal P_{1B-1} sequences despite a strong difference in overall sequence length, rather than with other eukaryotic sequences such as the Cu⁺-transporting ATPases of similar length from yeast and other fungal organisms (e.g. *S. cerevisiae*, *Komagataella pastoris*, *Vanderwaltozyma polyspora*) (Figs. 5, 6). One notable observation is that the yeast and fungal N-terminal extensions bear a higher percentage composition (25–30 %) of metal-binding residues than those of the mammalian sequences in addition to the presence of one or several CXXC motifs (Fig. 6). Interestingly, the mammalian sequences have the largest number of predicted ferredoxin-like folds despite bearing a lower percentage of metal-binding residues in the N-terminal domains (Fig. 6). These observations suggest that ATP7A and ATP7B may be more similar to bacterial homologs than to homologs in model fungal organisms (i.e. yeast) that are frequently used to study copper transport [23].

P_{1B-2}-ATPases—The well-characterized P_{1B-2}-ATPases are involved in the maintenance of Zn²⁺ homeostasis as well as detoxification of thiophilic heavy metals such as Cd²⁺ and Pb²⁺ in bacteria and plants (Fig. 4b) (Table 2) [3, 4, 33, 74, 75]. Like the P_{1B-1}-ATPases, these enzymes are composed of roughly 8 TMs, but are slightly shorter, averaging approximately 750 amino acids in length (Figs. 2b, S1). The P_{1B-2}-ATPases contain a CPC motif in TM4 (Figs. 2, 4), and also have a strongly conserved SXP sequence in TM4, a T(X)₆QN(X)₇K motif in TM5, and a DXG(X)₇N motif in TM6 (Table 3) [3, 62]. Divergence from this trend occurs in 12 sequences in which a CSC motif aligns strongly with the TM4 CPC motif, but we speculate that these sequences may represent a distinct subfamily (vide infra). Interestingly, the previously characterized TM5 Lys residue necessary for Zn²⁺ transport in *E. coli* ZntA [76] is conserved in all but four of the sequences; however, it appears that those sequences that bear an Asn or Gln in this position instead of Lys do not cluster strongly with the other P_{1B-2}s, suggesting that this substitution is uncommon.

In contrast to the Cu⁺ transporters, mutagenesis and XAS studies indicate that only one Zn²⁺ ion interacts directly with the TM region of *E. coli* ZntA, and it has been suggested that differences in geometric preferences at a single tetrahedral binding site, as well as residues lining the TM metal-binding channel, select for zinc over copper in the high-affinity TM metal binding sites of these transporters [62, 77, 78]. Like the P_{1B-1} subfamily, the majority

of the P_{1B-2} subfamily sequences include one or several soluble MBD(s). However, these domains come in different types and may be an N-terminal extension with one or many CXXC motif(s) contained within one to three ferredoxin-like folds, a C-terminal extension rich in metal-binding amino acids, or both (Fig. 6). These metal-binding domains have been shown to coordinate both Zn²⁺ and Cd²⁺ in bacteria [74, 75], but much less is known about coordination by the His- and Cys-rich domains that are common in plants [79, 80]. It is speculated that these domains represent a regulatory element in the P_{1B-2}-ATPases [3, 81].

Clustering data on the P_{1B-2}-ATPases indicate strong similarities among members of this subfamily (Fig. 3). Like the P_{1B-1}-ATPases, the P_{1B-2} sequences derive from organisms that span the various domains of life, and many sequences of high similarity have been identified in bacterial and plant species (Fig. 5), likely owing to the dual role of this subfamily in Zn²⁺ transport as well as Cd²⁺ and Pb²⁺ detoxification [5, 22, 82]. In fact, *Arabidopsis thaliana* has at least three P_{1B-2}-ATPases (HMA2–4) of various lengths and cellular localizations, although there seems to be some redundancy [83–86]. The P_{1B-2} sequences appear to cluster strongly not based on length, but rather according to type of N- and C-terminal MBDs despite diversity among the parent phyla, with those bearing ferredoxin-like motifs, several metal-binding residues, and/or both clustering distinctly (Figs. 5, 6, S1). Interestingly, clustering data indicate that ZntA from *E. coli*, a model enzyme for this family [87], diverges strongly from the core of bacterial and archaeal enzyme sequences (Fig. 4b). Furthermore, strong clustering centered about CadA from *Staphylococcus aureus* suggests this enzyme may be more representative of the P_{1B-2} subfamily despite the paucity of information that exists on this system [75, 88].

P_{1B-3}-ATPases—The P_{1B-3}-ATPases are an interesting set of oxidation state-specific Cu²⁺ transporters [3, 4, 89]. Sequences from this subfamily may have up to 8 TMs, although they average only approximately 700 amino acids in length (Figs. 2b, S1). A unique identifier of these sequences is the strict conservation of a CPH motif in TM4 (Fig. 4). A GYN(X)₄P motif in TM5 and a MSXST motif in TM6 are also strongly conserved (Table 3) [3, 89]. Unlike the previous subtypes, the P_{1B-3}-ATPases have only been discovered in bacterial and archaeal species thus far (Fig. 5), and the best-characterized enzyme of this subfamily is CopB from *A. fulgidus*, *T. thermophilus*, and *Enterococcus hirae* [89–91]. Studies on *E. hirae* CopB originally suggested that this subfamily was involved in Cu⁺/Ag⁺ transport [90]; however, it has been shown that although members of this subfamily (specifically *A. fulgidus* and *T. thermophilus* CopB) may be partially activated by Cu⁺ and Ag⁺, tighter binding and higher activity are obtained by Cu²⁺ activation, strongly suggesting that this oxidation state of copper is the preferential substrate [90, 91]. Consistent with their role as Cu²⁺ transporters, the majority of the P_{1B-3}-ATPases have soluble N-terminal extensions that are rich in His residues, and some of these extensions have CXXC motifs (but not ferredoxin-like folds) (Figs. 2b, 6). The predominance of His residues in these domains is likely a consequence of the better hard Lewis acid-hard Lewis base match of the Cu²⁺ substrate and the nitrogen imidazole lone pair from the His sidechain. These His-rich N-terminal extensions appear to be necessary for maximal ATPase function [89], but no structural or spectroscopic data have elucidated the precise interaction of Cu²⁺ with these domains.

Clustering data on the P_{1B-3}-ATPases illustrate strong similarities among these sequences and demonstrate the least divergence of any subfamily (Fig. 3). Despite the disparate nature of the phyla represented (Fig. 5), the core similarities among these sequences are evidenced by the tight clustering data. In fact, the only discrimination in this subfamily appears to be that of length (Fig. S1). For example, the P_{1B-3} sequence from *Fibrisoma limi* is predicted to have an unusually long (approximately 200 amino acids) N-terminal extension of unknown function, whereas the P_{1B-3} sequence from *Nitrosococcus halophilus* is predicted to have an approximately 200 amino acid-long C-terminal LysM-like domain (Fig. 6b). These extensions do not appear to be rich in metal-binding residues and likely play a completely different role. The presence of a TM4 CPH motif in both of these outliers suggests that they still transport Cu²⁺.

P_{1B-4}-ATPases—Of the established subfamilies, the most recently identified and characterized are the Co²⁺-transporting P_{1B-4}-ATPases, which are some of the simplest and shortest P_{1B}-ATPases. The P_{1B-4}-ATPases average approximately 700 amino acids in length and 6 TMs, have a completely conserved SPC motif in TM4, and bear a strongly conserved HEGT motif in TM6 (Table 3) (Figs. 2b, 4, S1). Early bioinformatic and phenotypic evidence implicated these P_{1B}-ATPases in cobalt transport [92]. However, some controversy surrounded this subfamily's metal substrate until recently; various pieces of evidence had suggested involvement of these enzymes in Zn²⁺ transport [93–95]. While these early studies examined gene expression and/or culture and plant growth without studying the intact enzyme, biophysical characterization and activity assays on the purified ATPases sCoaT from *Sulfitobacter* sp. NAS-14.1 and CtpD from *Mycobacterium smegmatis* clearly demonstrated that these members of the P_{1B-4} subfamily are specific for Co²⁺ [27, 29], strongly suggesting that Co²⁺ is at least one metal substrate for the entire subfamily. Interestingly, EXAFS and activity data on sCoaT showed that the completely conserved Cys residue in the TM4 SPC motif (Fig. 2b) is necessary for optimal activity, but this residue does not ligate the Co²⁺ ion, at least in the E1-metal bound state [29]. However, it is speculated that this residue may be involved in metal efflux.

Clustering data on the P_{1B-4} subfamily indicate a strong divergence among its members (Fig. 5). The protein sequences in this subfamily are chiefly derived from bacterial organisms, although some predicted eukaryotic sequences with the TM4 SPC signature motif have been identified in *Streptophyta*. Unlike the P_{1B-1}–P_{1B-3} subfamilies, the sequence clustering here is more sporadic, suggesting less overall sequence conservation among this subfamily. Furthermore, the loose clustering data on this subfamily may suggest that there is a functional divergence in the P_{1B-4}-ATPases. Support for this argument derives from a recent study that demonstrated ATPase activation by both Co²⁺ and Ni²⁺ for *M. tuberculosis* CtpD and CtpJ [28], the results of which contrast with metal-specific stimulation demonstrated for sCoaT and MsCtpD; however, no direct interactions of either Co²⁺ and Ni²⁺ with the TM region in MtCtpD, MtCtpJ, or MsCtpD have been shown. Interestingly, while on average the P_{1B-4}-ATPases are relatively short in length, several new members of this subfamily have been identified that are predicted to contain soluble N-terminal extensions of ~ 150 amino acids prior to the onset of TM1 that are rich in common metal-binding residues, including in some cases a CXXC ferredoxin-like motif (Figs. 4, 6b). These

N-terminal extensions have not been functionally characterized, and it is not yet known whether these residues form a consensus motif similar to other known metal-binding signatures, but we speculate that these MBDs impart an added level of metal specificity and/or regulation.

Emerging subfamilies

P_{1B-5}-ATPases—The presence of the P_{1B-5} subfamily was suggested nearly a decade ago based on the identification of five P_{1B}-ATPase sequences containing a unique PCP motif in TM4 and a strongly conserved QEXXD motif in TM6 (Table 3) (Fig. 2b) [3]. We have identified several hundred sequences that maintain these characteristics. The enzymes of the P_{1B-5} subfamily average approximately 700 amino acids in length with approximately 6 TMs, similar to the P_{1B-4} subfamily. However, the identity of the subfamily's metal substrate(s) is still unknown. Interestingly, in approximately 25 % of the known P_{1B-5} sequences, a Hr-like domain is found as a predicted 100–200 amino acid C-terminal extension (Figs. 2b, 6, S1). Studies on the isolated, soluble Hr domain of the P_{1B-5}-ATPase from *A. cellulolyticus* [42] as well as the full-length P_{1B-5}-ATPase from *S. meliloti* [30] have suggested that either iron or nickel may be the biologically relevant substrate for this subclass. There is a strong sequence similarity between the P_{1B-5} Hr-like domains and the Hr-like domain of the protein FBXL5, which senses the presence of iron in mammalian systems [96–99], suggesting that the Hr-like domain may regulate P_{1B-5}-ATPase activity in response to iron availability.

Clustering data emphasize the split nature of P_{1B-5} subfamily (Fig. 3). All members include the PCP motif in TM4 (Fig. 4), and only bacterial P_{1B-5}-ATPases have been identified thus far (Fig. 5). However, the sequences appear to cluster based on the apparent length (long, medium, and short), with the longest sequences (and largest cluster) arising from those containing a Hr-like extension on their C-terminus, and the shortest sequences consisting chiefly of the minimal core structure (similar to the P_{1B-4} subfamily), with only one outlying sequence that appears to have a poly metal-binding residue motif that is not a Hr-like domain (Figs. 6, S1). Interestingly, the sequences from gram-positive firmicutes, which represent the second-largest number of unique organisms that contain P_{1B-5}-ATPases (Fig. 5b), exclusively lack the Hr-like domain. Recent phylogenetic analyses of bacterial Hr proteins have implicated firmicutes in having the largest global distribution of single-domain Hrs of any of the gram-positive bacterial phyla [100, 101]. By contrast, actinobacteria, which represent the largest number of unique P_{1B-5} sequences, but endogenously produce very few single-domain Hrs, have a C-terminal Hr-like domain in nearly 60 % of their P_{1B-5} sequences (Figs. 5b, 6a). A similar distribution occurs in the gram-negative proteobacteria, although the sample size is smaller. An examination of the genomic contexts of the firmicute, actinobacterial, and proteobacterial sequences reveals that nearly 50 % of the putative P_{1B-5} sequences lie downstream of small (50–160 amino acids) hypothetical uncharacterized proteins, many of which contain CXXC or CXXS sequences and ferredoxin-like folds or are rich in metal-binding residues such as Met, Cys, His, Asp, or Glu. In at least one instance, the putative P_{1B-5}-ATPase from *Xanthobacter autotrophicus* is directly downstream of a hypothetical rubrerythrin protein, which contains both a diiron site and a rubredoxin-like motif [45].

One hypothesis is that the Hr-like domain functions in a regulatory manner, similar to other MBDs of the P_{1B}-ATPases. In this scenario, the clustering and bioinformatic results suggest that regulation of the P_{1B-5} subfamily may be a general phenomenon and can be achieved by a separate accessory protein or by a fused C-terminal MBD. A further complication is the predicted presence of a fused cytochrome *b*₅₆₁ N-terminal extension for the P_{1B-5}-ATPase from several of the *Burkholderia* species [42]. The presence of a regulatory protein or domain in this subfamily might be consistent with an iron transport function, but further experimental data are clearly required to understand the P_{1B-5}-ATPases.

P_{1B-6}-ATPases—The presence of a P_{1B-6} subfamily has been suggested previously [3]; however, it appeared as if this subfamily originally existed as a catch-all for the few apparent P_{1B}-type ATPases with variant TM4 designators (e.g. APC, CPS). Recent BLAST searches of the P_{1B-5} subfamily revealed the presence of several P_{1B}-type ATPase sequences with poor *E* values that contain the signature DKTGT cytoplasmic phosphorylation motif, but are otherwise dissimilar from the P_{1B-5} subfamily. A BLAST search of these first unique sequences from *Desulfovibrio piger* and *D. desulfuricans* retrieved several hundred apparent P_{1B}-type ATPases with high sequence similarity. Given the large number of sequences, we have designated this subfamily the P_{1B-6}-ATPases. The protein similarity network analysis of these sequences reveals that they indeed cluster distinctly from those of the other subfamilies (Fig. 3) and all contain a unique SCA motif that aligns strongly with the TM4 motif of other P_{1B} subfamilies (Figs. 2b, 4). Present mainly in bacteria and archaea (Fig. 5), these sequences average approximately 700 amino acids in length, although they are highly variable (Fig. S1). While the average P_{1B-6} sequence appears to have approximately 6 TMs (Fig. 2b), for several sequences in this subfamily the predicted consensus TM topology of the average sequence length places the AD between TM3 and TM4, the SCA motif within TM5 (not TM4), the ATPBD between TM5 and TM6, and a completely conserved TXXNHN motif in TM6 (Table 3) (Fig. 2b). Interestingly, the predicted number of TMs in these sequences is highly variable, and some sequences have as few as 4 TMs (e.g. *Eubacterium siraeum* DSM 15702) and as many as 12 TMs (e.g. *D. vulgaris* strain Miyazaki F/DSM 19637). Should the former case be correct, this predicted topology suggests the ATPBD and its following residues would define not only the N- and P-domains, but also the C-terminus including the completely conserved TXXNHN motif (Fig. 2b), departing from the common topology of all other P_{1B}-type ATPases. Data from these prediction algorithms should be viewed with caution since these predictions are not always reliable; however, it seems likely that this subfamily may be unique among the P_{1B}-ATPases. Additionally, in several sequences the predicted topology of these enzymes places the soluble N- and C-termini opposite of the soluble AD and ATPBD, suggesting that these enzymes may function as importers rather than exporters. Although most P_{1B}-type ATPases are thought to function in metal export, a controversial role in metal import has been suggested for some systems [102–104].

A possible role for the P_{1B-6} subfamily in iron import is suggested by the genomic context of several sequences. For example, in *Rhodospirillum rubrum* (ATCC 11170), a Fur binding site is located at the beginning of an operon encoding four putative proteins: a hypothetical protein, an ABC transporter permease component, another hypothetical protein, and finally a

P_{1B-6}-ATPase (Fig. 7) [45, 46]. A similar genomic organization is present for several of the *Desulfovibrionales* P_{1B-6} sequences. Specifically, downstream of Fur binding sites in *D. vulgaris* (Hildenborough) and *D. vulgaris* (str. Miyazaki) are operons encoding two hypothetical proteins that sandwich a P_{1B-6}-ATPase (Fig. 7) [45, 46]. This operon architecture bears some resemblance to the better characterized Feo system (FeoA/B/C) that drives Fe²⁺ uptake from the environment of several bacteria [105, 106]. In this system, FeoA functions as an Fe²⁺-binding protein that likely delivers iron to FeoB, an inner membrane GTPase that functions to import Fe²⁺. Less is known about FeoC, but it is believed that FeoC functions as a transcriptional regulator and responds to bioavailable iron levels [105–107]. A similar three-component system under Fur regulation is evident in the operon architecture of the P_{1B-6}-ATPases from *Desulfovibrionales*; in fact, FeoA is predicted to lie within an ORF just upstream of the P_{1B-6}-ATPase in *D. piger* (ATCC 29098). It is tempting to speculate that these ATPases may function in a manner similar to the FeoB GTPase in Fe²⁺ import, either to substitute for a lack of a distinct FeoB protein or to complement an already functioning and complete Feo system. Some circumstantial evidence has previously implicated iron as a potential P_{1B}-type ATPase substrate, but no assays on purified ATPases were performed in these studies [108, 109]. Further characterization at the biochemical level will be necessary to decipher the possible role of the P_{1B-6}-ATPases in iron transport.

P_{1B-7}-ATPases—Finally, there is evidence for a seventh subclass, the P_{1B-7}-ATPases. These sequences, of which only 12 exist, average approximately 660 amino acids and 6 predicted TMs (Figs. 2b, S1). The P_{1B-7} sequences were originally identified based on the presence of a CSC motif in TM4 (Figs. 2b, 4), but clustering data on this emergent subtype suggest that there is a strong similarity between these sequences and those of the bacterial P_{1B-2} sequences (Figs. 3, 4). Most of the currently identified sequences derive from *S. epidermidis* and related species that are opportunistic bacteria implicated in nosocomial infections [46]; the genomic context of nearly every sequence displays an ORF encoding for the P_{1B-7}-type ATPase sandwiched by a putative and small heavy metal-associated protein that may contain a ferredoxin-like fold and a putative helix-turn-helix (HTH) transcriptional regulator (Fig. 7) [45, 46]. Several of these transcriptional regulators are predicted to be responsive to the presence of metals, such as the zinc/cobalt sensor CzcA and the arsenic/antimony sensor ArsR [110]. The putative P_{1B-7}-ATPase from *S. epidermidis* (VCU045) lies upstream of a putative HTH transcriptional regulator that bears high sequence homology (>50 % identity) to ArsR from *S. hominis* and *E. faecium*. These data suggest that the P_{1B-7}-ATPases may play a role in metal resistance as a virulence determinant for *S. epidermidis* and related species.

Unclassified subtypes

Several P_{1B}-ATPase sequences are difficult to classify. Most notably, the recently characterized Mn²⁺-transporting CtpC from *M. tuberculosis* contains a TM4 CPC motif [31]. However, clustering analyses show that this sequence and others related to it (chiefly derived from *Mycobacteria* and related species) are distinct from the P_{1B-1} and P_{1B-2} sequences that also contain a TM4 CPC motif (Figs. 3, 4). This finding correlates with biochemical data showing that these ATPases preferentially transport Mn²⁺ over Cu⁺ or

Zn²⁺; it is speculated and complemented by mutational data that distinct amino acid residues in TM6 are responsible for the selectivity of Mn²⁺ over other metals [31]. Whether these enzymes represent a distinct subclass or are evolutionary adaptations from a previously existing subclass that are necessary for the establishment of virulence in select organisms is unknown.

Conclusions

In this work, we have updated the classification of the P_{1B}-ATPases. Using genomic data, we have analyzed protein trends in the historically established Cu⁺²⁺-, Zn²⁺-, and Co²⁺-transporting P_{1B} subfamilies. Moreover, we have used new genomic and biochemical data to examine emerging subfamilies of the P_{1B}-TPases that appear to have functional roles in manganese, iron, and nickel transport. These findings suggest that the P_{1B}-ATPases are connected to maintaining cellular homeostasis of nearly every biologically relevant first-row transition metal. In addition, these data provide an updated guiding framework for future genetic, biochemical, biophysical, and enzymatic characterization.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was supported by National Institutes of Health Grants GM58518 (A.C.R.) and F32GM105339 (A.T.S.) and by the Walter S. and Lucienne Driskill Graduate Training Program in Life Sciences (K.P.S). Sequence searches utilized both database and analysis functions of the Universal Protein Resource (UniProt) Knowledgebase and Reference Clusters (<http://www.unprot.org>) and the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>).

Abbreviations

AD	Actuator domain
ATP	Adenosine-5'-triphosphate
ATPBD	ATP-binding domain
Fur	Ferric uptake regulator
Hr	Hemerythrin
MBD	Metal-binding domain
ORF	Open reading frame
SERCA	Sarcoplasmic/endoplasmic reticulum Ca ²⁺ transporter
TM	Transmembrane helix

References

1. Axelsen KB, Palmgren MG. J Mol Evol. 1998; 46:84–101. [PubMed: 9419228]
2. Lutsenko S, Kaplan JH. Biochemistry. 1995; 34:15607–15613. [PubMed: 7495787]
3. Argüello JM. J Membr Biochem. 2003; 195:93–108.

4. Argüello JM. *Biometals*. 2007; 20:233–248. [PubMed: 17219055]
5. Klein JS, Lewinson O. *Metallomics*. 2011; 3:1098–1108. [PubMed: 21901186]
6. Jorgensen PK, Håkansson KO, Karlisch SJD. *Annu Rev Physiol*. 2003; 65:817–849. [PubMed: 12524462]
7. Toyoshima C. *Biochim Biophys Acta*. 2009; 1793:941–946. [PubMed: 19010358]
8. Kühlbrandt W. *Nat Mol Cell Biol*. 2004; 5:282–295.
9. Albers RW. *Annu Rev Biochem*. 1967; 36:727–756. [PubMed: 18257736]
10. Post RL, Hegyvary C, Kume S. *J Biol Chem*. 1972; 247:6530–6540. [PubMed: 4263199]
11. Møller JV, Olesen C, Winther AM, Nissen P. *Q Rev Biophys*. 2010; 43:501–566. [PubMed: 20809990]
12. Morth JP, Pedersen BP, Toustrup-Jensen MS, Sørensen TL-M, Petersen J, Andersen JP, Vilsen B, Nissen P. *Nature*. 2007; 450:1043–1049. [PubMed: 18075585]
13. Pedersen BP, Buch-Pedersen MJ, Morth JP, Palmgren MG, Nissen P. *Nature*. 2007; 450:1111–1114. [PubMed: 18075595]
14. Toyoshima C, Mizutani T. *Nature*. 2004; 430:529–535. [PubMed: 15229613]
15. Galanis A, Karapetsas A, Sandaltzopoulos R. *Mutat Res*. 2009; 674:31–35. [PubMed: 19022395]
16. Galaris D, Evangelou A. *Crit Rev Oncol Hemat*. 2002; 42:93–103.
17. Andreini C, Bertini I, Cavallaro G, Holliday GL, Thornton JM. *J Biol Inorg Chem*. 2008; 13:1205–1218. [PubMed: 18604568]
18. Waldron KJ, Rutherford JC, Ford D, Robinson NJ. *Nature*. 2009; 460:823–830. [PubMed: 19675642]
19. Nelson N. *EMBO J*. 1999; 18:4361–4371. [PubMed: 10449402]
20. Ma Z, Jacobsen FE, Giedroc DP. *Chem Rev*. 2009; 109:4644–4681. [PubMed: 19788177]
21. Silver S. *Gene*. 1996; 179:9–19. [PubMed: 8991852]
22. Williams LE, Pittman JK, Hall JL. *Biochim Biophys Acta*. 2000; 1465:104–126. [PubMed: 10748249]
23. Askwith C, Kaplan J. *TIBS*. 1998; 23:135–138. [PubMed: 9584616]
24. Andrews NC, Levy JE. *Blood*. 1998; 92:1845–1851. [PubMed: 9731040]
25. Vulpe C, Levinson B, Whitney S, Packman S, Gitschier J. *Nat Genet*. 1993; 3:7–13. [PubMed: 8490659]
26. Bull PC, Cox DW. *Trends Genet*. 1994; 10:246–252. [PubMed: 8091505]
27. Raimunda D, Long JE, Sasseti CM, Argüello JM. *Mol Microbiol*. 2012; 84:1139–1149. [PubMed: 22591178]
28. Raimunda D, Long JE, Padilla-Benavides T, Sasseti CM, Argüello JM. *Mol Microbiol*. 2013; 91:185–197. [PubMed: 24255990]
29. Zielazinski EL, Cutsail GE, Hoffman BM, Stemmler TL, Rosenzweig AC. *Biochemistry*. 2012; 51:7891–7900. [PubMed: 22971227]
30. Zielazinski EL, González-Guerrero M, Subramaniam P, Stemmler TL, Argüello JM, Rosenzweig AC. *Metallomics*. 2013; 5:1614–1623. [PubMed: 24056637]
31. Padilla-Benavides T, Long JE, Raimunda D, Sasseti CM, Argüello JM. *J Biol Chem*. 2013; 288:11334–11347. [PubMed: 23482562]
32. Mandal AK, Argüello JM. *Biochemistry*. 2003; 42:11040–11047. [PubMed: 12974640]
33. Liu J, Dutta SJ, Stemmler AJ, Mitra B. *Biochemistry*. 2006; 45:763–772. [PubMed: 16411752]
34. Wittkop T, Emig D, Truss A, Albrecht M, Böcker S, Baumbach J. *Nat Protoc*. 2011; 6:285–295. [PubMed: 21372810]
35. Cline MS, Smoot M, Cerami E, Kuchinsky A, Landys N, Workman C, Christmas R, Avila-Campilo I, Creech M, Gross B, Hanspers K, Isserlin R, Kelley R, Killcoyne S, Lotia S, Maere S, Morris J, Ono K, Pavlovic V, Pico AR, Vailaya A, Wang P-L, Adler A, Conklin BR, Hood L, Kuiper M, Sander C, Schmulevich I, Schwikowski B, Warner GJ, Ideker T, Bader GD. *Nat Protoc*. 2007; 2:2366–2382. [PubMed: 17947979]
36. Waterhouse AM, Procter JB, Martin DMA, Clamp M, Barton GJ. *Bioinformatics*. 2009; 25:1189–1191. [PubMed: 19151095]

37. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG. *Bioinformatics*. 2007; 23:2947–2948. [PubMed: 17846036]
38. Henikoff S, Henikoff JG. *Proc Natl Acad Sci USA*. 1992; 89:10915–10919. [PubMed: 1438297]
39. Krogh A, Larsson B, von Heijne G, Sonnhammer ELL. *J Mol Biol*. 2001; 305:567–580. [PubMed: 11152613]
40. Hofmann K, Stoffel W. *Biol Chem*. 1993; 374:166.
41. Bernsel A, Viklund H, Hennerdal A, Elofsson A. *Nucl Acids Res*. 2009; 37:W465–W468. [PubMed: 19429891]
42. Traverso ME, Subramanian P, Davydov R, Hoffman BM, Stemmler TL, Rosenzweig AC. *Biochemistry*. 2010; 49:7060–7068. [PubMed: 20672819]
43. Finn RD, Clements J, Eddy SR. *Nucl Acids Res*. 2011; 39:W29–W37. [PubMed: 21593126]
44. Punta M, Coggill PC, Eberhardt RY, Mistry J, Tate J, Boursnell C, Pang N, Forslund K, Ceric G, Clements J, Heger A, Holm L, Sonnhammer ELL, Eddy SR, Bateman A, Finn RD. *Nucl Acids Res*. 2012; 40:D290–D301. [PubMed: 22127870]
45. Dehal PS, Joachimiak MP, Price MN, Bates JT, Baumohl JK, Chivian D, Friedland GD, Huang KH, Keller K, Novichkov PS, Dubchak IL, Alm EJ, Arkin AP. *Nucl Acids Res*. 2010; 38:D396–D400. [PubMed: 19906701]
46. Novichkov PS, Laikova ON, Novichkova ES, Gelfand MS, Arkin AP, Dubchak I, Rodionov DA. *Nucl Acids Res*. 2010; 38:D111–D118. [PubMed: 19884135]
47. Sazinsky MH, Agarwal S, Argüello JM, Rosenzweig AC. *Biochemistry*. 2006; 45:9949–9955. [PubMed: 16906753]
48. Sazinsky MH, Mandal AK, Argüello JM, Rosenzweig AC. *J Biol Chem*. 2006; 281:11161–11166. [PubMed: 16495228]
49. Sørensen TL-M, Møller JV, Nissen P. *Science*. 2004; 304:1672–1675. [PubMed: 15192230]
50. Cox DW, Moore SDP. *J Bioenerg Biomembr*. 2002; 34:333–338. [PubMed: 12539960]
51. Lübken M, Guldenhaupt J, Zoitner M, Deigweier K, Haebel P, Urbanek C, Scheidig AJ. *J Mol Biol*. 2007; 369:368–385. [PubMed: 17434529]
52. Jaykathan S, Roberts SA, Weichsel A, Argüello JM, McEvoy MM. *Biosci Rep*. 2012; 32:443–453. [PubMed: 22663904]
53. Gourdon P, Liu X-Y, Skjærringe T, Morth JP, Møller LB, Pederson BP, Nissen P. *Nature*. 2011; 475:59–64. [PubMed: 21716286]
54. Fan B, Rosen BP. *J Biol Chem*. 2002; 277:46987–46992. [PubMed: 12351646]
55. Melchers K, Weitznegger T, Buhmann A, Steinhilber W, Sachs G, Schäfer KP. *J Biol Chem*. 1996; 271:446–457. [PubMed: 8550601]
56. Southron JL, Basu U, Taylor GJ. *FEBS Lett*. 2004; 566:218–222. [PubMed: 15147898]
57. Lowe J, Vieyra A, Catty P, Guillain F, Mintz E, Cuillel M. *J Biol Chem*. 2004; 279:25986–25994. [PubMed: 15078884]
58. Bull PC, Thomas GR, Rommens JM, Forbes JR, Cox DW. *Nat Genet*. 1993; 5:327–337. [PubMed: 8298639]
59. DiDonato M, Sarkar B. *Biochim Biophys Acta*. 1997; 1360:3–16. [PubMed: 9061035]
60. Prohaska JR, Gybina AA. *J Nutr*. 2004; 134:1003–1006. [PubMed: 15113935]
61. Kaplan JH, Lutsenko S. *J Biol Chem*. 2009; 284:25461–25465. [PubMed: 19602511]
62. Zhitnitsky D, Lewinson O. *Mol Microbiol*. 2013; 91:777–789. [PubMed: 24350798]
63. Mandal AK, Cheung WD, Argüello JM. *J Biol Chem*. 2002; 277:7201–7208. [PubMed: 11756450]
64. González-Guerrero M, Eren E, Rawat S, Stemmler TL, Argüello JM. *J Biol Chem*. 2008; 283:29753–29759. [PubMed: 18772137]
65. Fu Y, Tsui H-CT, Bruce KE, Sham L-T, Higgins KA, Lisher JP, Kazmierczak KM, Maroney MJ, Dann CE III, Winkler ME, Giedroc DP. *Nat Chem Biol*. 2013; 9:177–183. [PubMed: 23354287]
66. Argawal S, Hong D, Desai NK, Sazinsky MH, Argüello JM, Rosenzweig AC. *Proteins*. 2010; 78:2450–2458. [PubMed: 20602459]
67. Singleton C, Le Brun NE. *Biometals*. 2007; 20:275–289. [PubMed: 17225061]

68. Padilla-Benavides T, McCann CJ, Argüello JM. *J Biol Chem*. 2013; 288:69–78. [PubMed: 23184962]
69. Voskoboinik I, Mar J, Strausak D, Camakaris J. *J Biol Chem*. 2001; 276:28620–28627. [PubMed: 11373292]
70. Voskoboinik I, Greenough M, La Fontaine S, Mercer JFB, Camakaris J. *Biochem Biophys Res Commun*. 2001; 281:966–970. [PubMed: 11237756]
71. Tsikovskii R, MacArthur BC, Lutsenko S. *J Biol Chem*. 2001; 276:2234–2242. [PubMed: 11053407]
72. Wu C-C, Rice WJ, Stokes DL. *Structure*. 2008; 16:976–985. [PubMed: 18547529]
73. Mattle D, Sitsel O, Autzen HE, Meloni G, Gourdon P, Nissen P. *J Mol Biol*. 2013; 425:2299–2308. [PubMed: 23500486]
74. Banci L, Bertini I, Ciofi-Baffoni S, Finney LA, Outten CE, O'Halloran TV. *J Mol Biol*. 2002; 323:883–897. [PubMed: 12417201]
75. Banci L, Bertini I, Ciofi-Baffoni S, Su X-C, Miras R, Bal N, Mintz E, Catty P, Shokes JE, Scott RA. *J Mol Biol*. 2006; 356:638–650. [PubMed: 16388822]
76. Okkeri J, Haltia T. *Biochim Biophys Acta*. 2006; 1757:1485–1495. [PubMed: 16890908]
77. Dutta SJ, Liu J, Stemmler AJ, Mitra B. *Biochemistry*. 2007; 46:3692–3703. [PubMed: 17326661]
78. Raimunda D, Subramaniam P, Stemmler TL, Argüello JM. *Biochim Biophys Acta*. 2012; 1818:1374–1377. [PubMed: 22387457]
79. Eren E, Kennedy DC, Maroney MJ, Argüello JM. *J Biol Chem*. 2006; 281:33881–33891. [PubMed: 16973620]
80. Eren E, González-Guerrero M, Kaufman BM, Argüello JM. *Biochemistry*. 2007; 46:7754–7764. [PubMed: 17550234]
81. Mitra B, Sharma R. *Biochemistry*. 2001; 40:7694–7699. [PubMed: 11412123]
82. Lee J, Bae H, Jeong J, Lee J-Y, Yang Y-Y, Hwang I, Martinoia E, Lee Y. *Plant Physiol*. 2003; 133:589–596. [PubMed: 14512517]
83. Morel M, Crouzet J, Gravot A, Auroy P, Leonhardt N, Vavasseur A, Richaud P. *Plant Physiol*. 2009; 149:894–904. [PubMed: 19036834]
84. Baxter I, Tchieu J, Sussman MR, Boutry M, Palmgren MG, Gribskov M, Harper JF, Axelsen KB. *Plant Physiol*. 2003; 132:618–628. [PubMed: 12805592]
85. Eren E, Argüello JM. *Plant Physiol*. 2004; 136:3712–3723. [PubMed: 15475410]
86. Hussain D, Haydon MJ, Wang Y, Wong E, Sherson SM, Young J, Camakaris J, Harper JF, Cobbett CS. *Plant Cell*. 2004; 16:1327–1339. [PubMed: 15100400]
87. Sharma R, Rensing C, Rosen BP, Mitra B. *J Biol Chem*. 2000; 275:3873–3878. [PubMed: 10660539]
88. Nucifora G, Chu L, Misra TK, Silver S. *Proc Natl Acad Sci USA*. 1989; 86:3544–3548. [PubMed: 2524829]
89. Mana-Capelli S, Mandal AK, Argüello JM. *J Biol Chem*. 2003; 278:40534–40542. [PubMed: 12876283]
90. Solioz M, Odermatt A. *J Biol Chem*. 1995; 270:9217–9221. [PubMed: 7721839]
91. Schurig-Briccio LA, Gennis RB. *J Bacteriol*. 2012; 194:4107–4113. [PubMed: 22636781]
92. Rutherford JC, Cavet JS, Robinson NJ. *J Biol Chem*. 1999; 274:25827–25832. [PubMed: 10464323]
93. Gaballa A, Helmann JD. *Mol Microbiol*. 2002; 45:997–1005. [PubMed: 12180919]
94. Kim Y-Y, Choi H, Segami S, Cho H-T, Martinoia E, Maeshima M, Lee Y. *Plant J*. 2009; 58:737–753. [PubMed: 19207208]
95. Scherer J, Nies DH. *Mol Microbiol*. 2009; 73:601–621. [PubMed: 19602147]
96. Chollangi S, Thompson JW, Ruiz JC, Gardner KH, Bruick RK. *J Biol Chem*. 2012; 287:23710–23717. [PubMed: 22648410]
97. Ruiz JC, Walker SD, Anderson SA, Eisenstein RS, Bruick RK. *J Biol Chem*. 2013; 288:552–560. [PubMed: 23135277]

98. Thompson JW, Salahudeen AA, Chollangi S, Ruiz JC, Brautigam CA, Makris TM, Lipscomb JD, Tomchick DR, Bruick RK. *J Biol Chem*. 2012; 287:7357–7365. [PubMed: 22253436]
99. Shu C, Sung MW, Stewart MD, Igumenova TI, Tan X, Li P. *ChemBioChem*. 2012; 13:788–791. [PubMed: 22492618]
100. French CE, Bell JML, Ward FB. *FEMS Microbiol Lett*. 2008; 279:131–145. [PubMed: 18081840]
101. Bailly X, Vanin S, Chabasse C, Mizuguchi K, Vinogradov SN. *BMC Evol Biol*. 2008; 8:244–254. [PubMed: 18764950]
102. Francis MS, Thomas CJ. *Microb Path*. 1997; 22:67–78.
103. Lewinson O, Lee AT, Rees DC. *Proc Natl Acad Sci*. 2009; 106:4677–4682. [PubMed: 19264958]
104. Tottey S, Rondet SAM, Borrelly GPM, Robinson PJ, Rich PR, Robinson NJ. *J Biol Chem*. 2002; 277:5490–5497. [PubMed: 11739376]
105. Kim H, Lee H, Shin D. *Biochem Biophys Res Commun*. 2012; 423:733–738. [PubMed: 22705302]
106. Naikare H, Palyada K, Panciera R, Marlow D, Stintzi A. *Infect Immun*. 2006; 74:5433–5444. [PubMed: 16988218]
107. Cartron ML, Maddocks S, Gillingham P, Craven CJ, Andrews SC. *Biometals*. 2006; 19:143–157. [PubMed: 16718600]
108. Barañano DE, Wolosker H, Bae B-I, Barrow RK, Snyder SH, Ferris CD. *J Biol Chem*. 2000; 275:15166–15173. [PubMed: 10809751]
109. McLaughlin HP, Xiao Q, Rea RB, Pi H, Casey PG, Darby T, Charbit A, Sleator RD, Joyce SA, Cowart RE, Hill C, Klebba PE, Gahan CGM. *PLoS ONE*. 2012; 7:1–12.
110. Busenlehner LS, Pennella MA, Giedroc DP. *FEMS Microbiol Rev*. 2003; 27:131–143. [PubMed: 12829264]

Fig. 1.

Domain architecture of a P₁B-ATPase. The typical P₁B-ATPase consists of approximately 6–8 transmembrane helices (TMs, *blue*) through which the cation is translocated, soluble nucleotide-binding (N) and phosphorylation (P) domains that together constitute the ATP-binding domain (ATPBD, *red*), and a soluble actuator domain (AD, *yellow*). Structure taken from PDB 3RFU [53]

Fig. 2.

a Topology of a typical P_{1B} -ATPase. Residues in the fourth transmembrane helix (TM4) have been shown to interact with the metal substrate; several lines of evidence also implicate residues in the fifth and sixth transmembrane helices (TM5, TM6) in metal coordination. Soluble N- and C-terminal metal-binding extensions (MBDs) are believed to play a regulatory role. Strictly conserved residues in the AD and ATPBD are indicated. **b** Schematic illustration of the average polypeptide composition of the P_{1B} -ATPase subfamilies. Each P_{1B} -ATPase has approximately 6–8 TMs (shown as *cylinders*). The first two accessory TMs, which are not strictly conserved across all subfamilies, are labeled “A” and “B,” whereas the next six are labeled “1–6.” For the P_{1B-6} -ATPases, the first and last TMs are outlined with dotted lines to indicate the putative lack of their conservation. The strictly conserved three-residue sequences in TM4 are thought to be important for metal recognition and binding. The presence of various N- and C-terminal soluble MBDs are indicated by rectangles, and the rough composition of these domains is indicated in *bold letters* above the *boxes*. The approximate position of the strictly conserved Asp signature sequence (DKTGT) in the ATPBD is also shown

Fig. 3.

The P_{1B} -ATPase similarity network. In this single network, one or several strongly similar protein sequences are represented as nodes (*points*) on the image, and the strength of their similarity is indicated by edges (*connective lines*). Each subfamily is color-coded, and the clear grouping of each *color* indicates that each subfamily indeed clusters strongly. Nodes corresponding to representative sequences from particular subfamilies are noted. The *pink* cluster of nodes represents a currently unclassified set of enzymes that derives chiefly from *Mycobacteria* and related species

Fig. 4.

P_{1B}-ATPase clustering data analyzed according to **a** TM4 three-residue sequence motif and **b** metal specificity as determined from biochemical data (Table 2). Metal substrates that have been shown to be translocated are Cu⁺/Ag⁺, Zn²⁺/Cd²⁺/Pb²⁺, Cu²⁺, Co²⁺, and Mn²⁺. A role for the P_{1B-5}-ATPases in Fe²⁺ binding and/or Fe²⁺/Ni²⁺ transport has been suggested

Fig. 5.

P_{1B} -ATPase clustering data analyzed according to **a** domain of life and **b** phylum. The P_{1B} -ATPases occur in all domains of life; however, sequences from every subfamily have been identified in bacteria and archaea, whereas eukaryotic sequences have only been identified from subfamilies $1B-1$, $1B-2$, and $1B-4$. **b** Clustering data of the P_{1B} -ATPases color-coded based on the phylum of the sequence's parent organism

Fig. 6.

P_{1B}-ATPase clustering data analyzed according to **a** presence and type of soluble metal-binding domains (MBDs) and **b** predicted domain type. The MBDs are classified as those with CXXC motifs (*blue*), those with several candidate metal-binding residues such as His, Cys, Met, Asp, Glu (*yellow*), and those with a hemerythrin-like (Hr) domain (*red*). The combinations of various MBDs are represented by the combination of colors (see *inset*). The presence of a predicted type and fold of domain within the N- or C-terminal extensions are noted in panel B and are color-coded according to type and number of such domain. In both panels, gray nodes represent the lack of an identified MBD

Fig. 7.

Genomic contexts of emerging subfamilies. **a** Genomic context of several P_{1B-6} -ATPases from select organisms. The presence of a ferric uptake regulator (Fur) binding site is indicated by an inverted triangle. **b** Genomic context of a representative P_{1B-7} -ATPase from *S. epidermidis*; every P_{1B-7} -ATPase that has been identified thus far displays a very similar context. Genomic information was retrieved from manually curated databases as indicated in [45] and [46]. In both figures, ORF lengths are all relative to one another

Table 1

P-type ATPase classes and their associated substrates

Class	Substrate(s)	References
1A	K ⁺	[1–3, 8]
1B	Transition and heavy metals	
2	Ca ²⁺ , Na ⁺ /K ⁺ , H ⁺ /K ⁺	
3	H ⁺	
4	Phospholipids	
5	Unknown	

Table 2

P_{1B}-ATPase subclasses and their biochemically established substrates

Subclass	Substrate(s)	References
1B-1	Cu ⁺ , Ag ⁺	[54, 55, 57–59, 63, 64]
1B-2	Zn ²⁺ , Cd ²⁺ , Pb ²⁺	[33, 81, 83, 85, 86, 88, 94]
1B-3	Cu ²⁺	[89]
1B-4	Co ²⁺	[27, 29]
1B-5	Unknown	
1B-6	Unknown	
1B-7	Unknown	

Table 3Conserved motifs in TMs 4, 5, and 6 of the P_{1B}-ATPase subclasses

Subclass	Conserved motifs			References
	TM4	TM5	TM6	
1B-1	CPC	YN(X) ₄ P	MXXSS	[3, 62]
1B-2	CPC	T(X) ₆ QN(X) ₇ K	DXG(X) ₇ N	[3, 62]
1B-3	CPH	GYN(X) ₄ P	MSXST	[3, 89]
1B-4	SPC		HEGT	[3, 29]
1B-5	PCP		QEXXD	[3]
1B-6	SCA		TXXNHN	This work
1B-7	CSC			This work