Structural models of the human copper P-type ATPases ATP7A and ATP7B

Pontus Gourdon¹, Oleg Sitsel¹, Jesper Lykkegaard Karlsen¹, Lisbeth Birk Møller² & Poul Nissen¹

¹ Centre for Membrane Pumps in Cells and Disease, Danish National Research Foundation, Aarhus University, Dept. of Molecular Biology and Genetics. Gustav Wieds Vej 10C, DK – 8000 Aarhus, Denmark

² Center for Applied Human Molecular Genetics, Kennedy Center, Gl. Landevej 7, DK – 2600 Glostrup, Denmark

* Correspondence e-mail: pn@mb.au.dk

Abstract

The human copper exporters ATP7A and ATP7B contain domains common to all P-type ATPases, as well as class-specific features such as six sequential heavy metal binding domains (HMBD1-HMBD6) and a type-specific constellation of transmembrane helices. Despite the medical significance of ATP7A and ATP7B related to Menkes’ and Wilson’s diseases, respectively, structural information has only been available for isolated, soluble domains. Here we present homology models based on the existing structures of soluble domains and the recently determined structure of the homologous LpCopA, from the bacterium Legionella pneumophila. The models and sequence analyses show that the domains and residues involved in the catalytic phosphorylation events and copper transfer are highly conserved. In addition, there are only minor differences in the core structures of the two human proteins and the bacterial template, allowing protein-specific properties to be addressed. Furthermore, the mapping of known disease-causing missense mutations indicate that among the heavy-metal binding domains, HMBD5 and HMBD6 are the most crucial for function, thus mimicking the single or dual HMBD(s) found in most copper-specific P-type ATPases. We propose a structural arrangement of the HMBDs and how they may interact with the core of the proteins to achieve auto-inhibition.
Introduction

Copper is a trace element, which is essential for survival of probably all organisms. However, while required for many metalloenzymes such as in oxidative phosphorylation, only minute amounts are freely available in cells due to high toxicity (Koch et al., 1997). Critical components of the copper homeostasis machinery in humans include sequestering glutathione and metallothioneins and metallochaperone proteins, such as ATOX1. Furthermore proteins responsible for transport across the membranes of cells and their compartments; the importers CTR1 and CTR2 as well as ATP7A and ATP7B are required for export and sub-cellular distribution (Macreadie, 2008). ATP7A and ATP7B share a high degree of sequence conservation (55.6 % identity) and provide copper to proteins in the trans-Golgi network at low or normal copper concentrations while relocating to the plasma membrane for metal export from cells at elevated copper levels. Their expression profiles are different with ATP7A being ubiquitously expressed except for the liver, while ATP7B is found mainly in this organ (Lutsenko et al., 2007).

Emphasizing their significance, malfunction of ATP7A and ATP7B is the direct cause of the rare but severe Menkes’ and Wilson’s diseases, respectively. Defects in ATP7A lead to the inability to take up dietary copper from the intestine yielding a general copper deficiency in most tissues and in turn symptoms such as neurodegeneration, kinky hair and typically early death (Tumer and Moller, 2010). In contrast, patients with Wilson’s disease have impaired copper export from their livers leading to general copper overload and e.g. Kayser-Fleischer rings, cirrhosis and Parkinsonism, but the disorder is treatable using copper chelators or supplementation of zinc to induce metallothionein expression (Ala et al., 2007). In addition to these genetic conditions, the upregulation of ATP7A and ATP7B has also been linked to resistance to some chemotherapeutic drugs, and ATP7A may even promote tumor growth (Furukawa et al., 2008; Mangala et al., 2009).

ATP7A and ATP7B are primary transporters belonging to the P_{iib} subclass of heavy metal pumps of the P-type ATPase superfamily. Common to all P-type ATPases is an invariant DKTGTXT motif, which includes an aspartate residue that undergoes autophosphorylation and dephosphorylation during the catalytic cycle (Figure 1). Their mechanism is summarized by the so-called Post-Albers scheme describing a cycle through four principal conformations including phosphoenzyme (EP) intermediates (Albers, 1967; Post et al., 1972): E1, E1P, E2P and E2, where E1 and E2 denote high and low affinity for the exported ion, associated with binding at the cytoplasmic side and release at the opposite side, respectively. Structurally this is controlled by autophosphorylation and dephosphorylation events which are catalyzed by the soluble A- (actuator), P- (phosphorylation) and N- (nucleotide) domains and coupled to the ion binding sites in the transmembrane domain (M) through tertiary interactions (Figure 1). Briefly, the N-domain contains a binding pocket for ATP which provides the phosphate moiety to the reactive aspartate in the P-domain for phosphorylation, while the conserved, surface exposed, TGE loop in the A-domain mediates dephosphorylation. The
large conformational changes of these domains are communicated to more subtle changes of the ion-binding pathway and its, one to three, ion binding site(s) in the M-domain.

Also known as CopA proteins, the copper pumps are the most prevalent of the P$_{iB}$-ATPases and characterized by eight transmembrane spanning helices, MA, MB and M1 to M6, where the A-domain constitutes the “loop” between M2 and M3 and the P- and N-domains are positioned between M4 and M5 (Figure 1). In addition, they typically contain amino- and/or carboxy-terminal heavy metal binding domains (HMBDs) each with a $\beta$/$\beta$/$\alpha$/$\beta$ ferredoxin fold harboring a solvent-exposed copper-binding CXXC motif; ATP7A and ATP7B contain six such domains in sequential order in the amino-terminus. The role of these domains is most likely autoinhibitory rather than being involved in copper transport (Huster and Lutsenko, 2003); transfer to the ATPase M-domain is believed to be achieved directly from copper(I)-chaperones such as ATOX1 with an identical fold as the HMBDs. However, ATOX1 also delivers the metal to the HMBDs and experimental data suggested that is achieved between their CXXC motifs (Walker et al., 2002).

Recently we reported the first structure of a full-length CopA enzyme determined by X-ray crystallography providing an atomic model that consists of the M-domain and the cytoplasmic A-, P- and N-domains but excluding HMBD (from now referred to as the core). The structure was derived from the Legionella pneumophila gene product Lpg1024, coined LpCopA (Gourdon et al., 2011), which was stabilized in a copper-free transition state of dephosphorylation (mimicking the E2.P$_i$ state). In short, the structure demonstrated that the P$_{iB}$-ATPases share a basic architecture with the other P-type ATPases despite a low degree of sequence identity (e.g. 15 % identity to the well-studied sarco(endo)plasmic reticulum Ca$^{2+}$-ATPase, SERCA1a, from rabbit) although with noteworthy differences. A novel finding was that the P$_{iB}$-specific transmembrane helix MB was kinked by a conserved GG motif and formed a platform at which copper presumably is to be loaded to the membrane binding sites, most likely from copper chaperones. The location of six invariant residues of CopA proteins (the cysteines in a CPC motif in M4 as well as YN and MXXXS motifs in M5 and M6, respectively) also showed resemblance to the configuration of equivalent Ca$^{2+}$ binding residues in SERCA1a, in agreement with previous studies implicating this sextet in the formation of, quite likely, two membranous ion-binding sites (Figure 1)(Gonzalez-Guerrero et al., 2008). In addition, the structure revealed a conserved carboxyl group in M2 that may serve to stimulate release from the internal sites for further transport to the extra-cellular side.

With the medical relevance of ATP7A and ATP7B there is a significant interest in detailed structural information for these proteins of the CopA family, but only structures of individual soluble domains of ATP7A (HMBD1 through HMBD6 as well as the A- and N-domains) and ATP7B (complexes of HMBDs3-4 and HMBDs5-6 as well as the N-domain) (Achila et al., 2006; Banci et al., 2005a; Banci et al., 2006; Banci et al., 2010; Banci et al., 2008; Banci et al., 2005b; Banci et al., 2004; DeSilva et al., 2005; Dmitriev et al., 2006; Gitschier et al., 1998)
exist. However, with a significant level of sequence identity between LpCopA and ATP7A/ATP7B (35.4 and 36.3 % identity to core residues 79 to 731 of LpCopA) and with no sequence gaps/inserts except for the extracytosolic loops, in the M2:A-domain linker and in the structurally characterized N-domains (see later), the conditions are met for obtaining reliable homology models of the core of the human ATP7A and ATP7B.

In this study, we have produced homology models of ATP7A and ATP7B using the existing, experimentally determined, structural information of the proteins and homologues, displaying the catalytic cores of the proteins. We describe in detail the differences between the two copper pumps and why their catalytic properties may be unique. Furthermore, the six–module HMBD is analyzed on how it may assemble and interact with the protein core and we present a model for how the entire enzyme is assembled.
Results and Discussion

The ATP7A and ATP7B core structures

As expected from sequence alignments (Figure 2), the homology models of ATP7A and ATP7B can be threaded on the bacterial LpCopA core structure with only small deviations (Figure 3). In the soluble A-, P- and N-domains responsible for auto(de)phosphorylation, amino acids critical for establishing the phosphorylation site comprise (with ATP7A/ATP7B numbering) the DKTGTXT 1044-50/1027-33, D1301/1267 and ND 1304-5/1270-71 motifs, while the TGE dephosphorylation loops are formed by 875-77/858-60 (Figure 4A). The ATP binding pocket in the N-domain is achieved by EXXSEHPL1081-8/1064-71, G1116/1099, G1118/1101, GN1184-5/1150-1 and D1256/1222, thus including the site of the most widely observed Wilson’s disease mutation, H1069Q (Caca et al., 2001) (Dmitriev et al., 2006; Tsuda and Toyoshima, 2009). In the modeled transition state of dephosphorylation, the TGE loop is positioned adjacent to the catalytic aspartate. At the phosphorylation site (Figure 4A), AlF₄⁻ interacts with multiple residues in the A- and P-domains mimicking the phosphoryl transfer of hydrolysis, most likely catalyzed by the glutamate side chain of the TGE loop. The site is stereochemically defined by coordination to a magnesium ion, which also helps to reduce electrostatic repulsion in the phosphoryl transfer reactions; P-type ATPases are indeed magnesium dependent (Skou, 1957).

The M-domains of human ATP7A and ATP7B display a high degree of similarity to LpCopA even in the extracytosolic loops between M1/M2 and M5/M6. MB, last of the two most amino-terminal and class-specific transmembrane helices, kinks at the GG-motif (727-8/710-1) and forms an amphipathic and interfacial helix with positively charged residues facing the cytoplasm (Figure 4B). With a suggested metal entry site at this platform, a putative copper pathway is maintained, including: M746/M749, E798/E781 and D935/D918 forming the suggested entry site, C1000/C983, C1002/C985, Y1365/Y1331, N1366/N1332, M1393/M1359 and S1397/S1359 at the internal membranous high-affinity sites, as well as D782/D765 (and possibly M668/668 and M1388/1354) at the putative exit site (Figure 4B). The significance of these sites is underscored by several residues in this region being targeted in both Menkes’ and Wilson’s disease mutations: the two GG-motif residues, the S761/744 and P1386/1352 residues of M1 and M6 at the exit site, as well as the T994/977 and G1005/988 residues on M4 at the CPC-motif (see Supplementary Table 2 in (Gourdon et al., 2011) and Table 1, Figure 2). The number of common Menkes’ and Wilson’s disease-causing mutations in this region is significant, considering that only 13 shared disease-causing mutations are known in total between ATP7A and ATP7B (Figure 2). In addition, the most common cause of Wilson’s disease in the Asian population, an R778L mutation, is located in close vicinity of the entry site (Mak et al., 2008).

Compared to LpCopA four notable insertions are present in ATP7A and ATP7B: in the extracytosolic loops connecting MA/MB and M3/M4 as well as in a β-turn in the A-domain and in the N-domain (Figures 2 and 3). Of these the M3/M4 insertion and the extended β-
turn in the A-domain are relatively well-conserved among vertebrate ATP7A and ATP7B ATPases yet their importance remains poorly understood. The M3/M4 loop in ATP7A has been proposed as a site for ATP7A-specific glycosylation on N975 (Liu et al., 2010). However, ATP7B is not glycosylated suggesting that the extended M3/M4 loop may also serve other purposes, for example assisting in copper release (Barry et al., 2011) (Figure 3).

ATP7A/ATP7B-specific elements

The major differences between the core structures of the human copper pumps are found in two regions (Figure 2). The connecting loop between MA and MB is significantly larger in ATP7A, consisting of 40 residues (670-710) compared to 23 residues in ATP7B (670-693). Furthermore, an extended loop of the N-domain (1124-78/1106-1142) has been shown to be flexible by NMR studies (Banci et al., 2010; Dmitriev et al., 2006) and is within reach for the A-domain in the modeled E2.P state. These two regions are obvious candidates as determinants of specific properties such as the faster phosphorylation kinetics observed for ATP7A compared to ATP7B (Tadini-Buoninsegni et al., 2010; Tsivkovskii et al., 2002) and/or the different sub-cellular localizations of the proteins (Barry et al., 2010). Disease-associated mutations such as M687V in ATP7A (Gourdon et al., 2011) and P690L, G691R and G1111D in ATP7B support that they are important for function (Figures 3B and 4B and Table 1). Indeed, biochemical experiments have implicated that the MA/MB-loop influences the rate of copper release and dephosphorylation in ATP7A (Barry et al., 2011). Similarly, for ATP7B it has been suggested that the insert is involved in allosteric regulation of a putative modulatory ATP-binding site defined by Mg\(^{2+}\) binding to E1186/1152, D1201/1167 and D1205/1171, on the surface of N-domain (Hercend et al., 2011).

Interestingly, the extracytosolic MA/MB loop harbors a putative ATP7A specific glycosylation site on the disease-targeted NMS686-8 motif (Liu et al., 2010). The motif is conserved in many vertebrate ATP7A, but not ATP7B. Similarly, ATP7B is exclusively phosphorylated on S1121 in the N-domain insertion, which is a conserved residue in many vertebrate ATP7B but not ATP7A (Pilankatta et al., 2009). Thus, it is possible that differential catalytic properties and/or sub-cellular localization patterns of ATP7A and ATP7B can be ascribed in part to different post-translational modification patterns in the two mentioned regions.

However, there are also notable differences between ATP7A and ATP7B in N- and C-terminal regions. ATP7B harbors an amino-terminal extension prior to HMBD1 that is known to be important for cellular localization (Braiterman et al., 2009), while secondary structure predictions indicate that an alpha helix is present at the carboxy-terminus of ATP7A (Figure 2) which contains several ATP7A-specific phosphorylation sites, highly conserved among vertebrates (S1463, S1466 and S1469) (Veldhuis et al., 2009). The function of another ATP7A-unique phosphorylation motif remains to be elucidated. It is located in the HMBD1/HMBD2 linker, which is extended in ATP7A compared to ATP7B. Finally, both
proteins share a common serine phosphorylation site at the C-terminus, in ATP7A being at the end of the aforementioned predicted C-terminal alpha helical structure. This motive is closely upstream of the well-known carboxy-terminal LL sorting motif. The highly conserved di-leucine (LL) motif (Francis et al., 1999), typically preceded immediately by a phosphorylation site (Liu et al., 2010), is present in the carboxy-terminus of both ATP7A and ATP7B (1488-9/1454-5) ATPases. This motif is known to be involved in re-localizing proteins from the plasma membrane (e.g. class II SPCA P-type ATPases (Vanoevelen et al., 2008)).

The HMBD6 and the core-connecting linker

We have previously shown that many of the known missense mutations that cause Menkes’ disease are positioned in the core (Gourdon et al., 2011). Now we extend this analysis for ATP7A and reveal the positions of equivalent mutations in ATP7B that are associated with Wilson’s disease in the modeled structures, respectively (Figure 2). In contrast to the X-linked recessive character of Menkes’ disease, the phenotype of Wilson’s disease depends on the integrity of two copies of ATP7B which, together with extensive polymorphisms in the gene, complicate the genetic analysis and the interpretation of the effects of Wilson mutations. To establish a robust subset of Wilson’s disease missense mutations, reduced for the bias of general polymorphism, we collected those sites that are found in at least two independent studies (Figures 3B and 4 and Table 1). First of all we find thirteen Wilson mutations from this subset that target a residue which is also affected by Menkes disease mutations: the mutations in the M-domain described earlier as well as I874/857 at the TGE loop, T1048/1031 in DKTGT motif, L1100/1083 on the surface of the N-domain as well as G1300/1266, D1301/1267, N1304/1270 and G1315/1281. The latter seven positions are all near the catalytic phosphorylation site in the P-domain (see Supplementary Table 2 in (Gourdon et al., 2011) and Table 1). The residues of the filtered subset of Wilson mutations generally cluster in the core, but some are also found in HMBD6 (G591S/D, A604P, R616Q/W and G626A) and the linker between these entities (L641S and D642H). A similar pattern emerges for additional Menkes’ mutations in ATP7A outside the core, that map to the carboxy-terminal end of HMBD6 (E628V and A629P) or the linker to the core (K633R and S637L) (Gourdon et al., 2011) (Figures 3A and 4). Unless related to impaired biogenesis, this may suggest an important role for the HMBD6 and the core-connecting residues for the protein function coinciding with an almost invariant primary structure in this region among known vertebrate ATP7A and ATP7B ATPases, which gradually diminishes towards the amino-terminus. Oppositely, the general absence of disease causing missense mutations in HMBD1-HMBD5 and in the carboxy-terminus may indicate that these are not strictly required for function, as also indicated by e.g. deletion studies of HMBD1-4 where ATP7B function is not abolished (Huster and Lutsenko, 2003; Mercer et al., 2003). Exceptions are P1413R in the carboxy-terminus of ATP7A as well as ATP7B mutations G85V in HMBD1,
L492S in HMBD5 and the C-terminal T1434M. G85V and L492S have been shown to impair interactions with ATOX1 (Kenney and Cox, 2007).

To add additional insight into this matter we performed a sequence analysis of 1599 P-type ATPases currently available in the P-type ATPase Database with the characteristic motifs for copper specificity (see Materials and Methods). Overall, more than 70% of these entries contain one or two CXXC motifs in the amino-terminus and consequently most likely one or two HMBDs (Table 2). Although it should be noted that 75 proteins were deficient of CXXC motifs N-terminal to the core (and manual inspection only suggested functional replacements like SXC, CSXXC or SXXS for 49 of these), it favors a notion that the two HMBDs closest to the core represent a conserved and important functional role in most copper transporting P-type ATPases. Re-sequencing and/or analysis of the origin genomes of the 26 entries for accessory ferredoxin folds may further show that at least one amino-terminal HMBD is strictly required for auto-inhibition in the absence of copper; in contrast carboxy-terminal HMBDs appear to optional and always in combination with a corresponding amino-terminal unit.

A short length of the linker between HMBD6 and the core appears to be a conserved feature in most class IB copper pumps, independently of the number of HMBDs present, and ranging in median length of 14-22 amino acids (Table 2). With a proximity to the membrane interface, the possible locations for HMBD6 (or the equivalent domain in proteins with fewer HMBDs) to interact with the core and exert its function are in fact quite limited. Interestingly, the surface from which the carboxy-terminal part of HMBD6 protrudes from the ferredoxin fold is positively charged, and as a consequence it may be repulsed by the positive charges on the MB platform nearby (K737/720, K740/R723, H741/724 and K742/R725) (Figure 4B). In contrast, on the opposite side of the protein, around the linker between M2 and the A-domain, there is a highly conserved surface in CopA proteins, exposing functionally important amino acids in the P-domain and adjacent regions (Gourdon et al., 2011). Keeping in mind that surface-exposed residues are generally less conserved unless they form binding interfaces (Stevens and Arkin, 2001), this region of the core constitutes a putative interaction surface for HMBD6, and hence the entire six-module HMBD entity of ATP7A and ATP7B. Indeed, the surface of this region of the P-domain is highly negatively charged and may thus attract HMBD6. Furthermore, a large portion of the surface-exposed and disease-causing missense mutations of ATP7B are found in this particular area, possibly explaining their impairing phenotype through prevention of the interaction with the HMBDs. Guided by these arguments we manually placed the structurally determined HMBD5/HMBD6 pair of ATP7B (and the modeled equivalent of ATP7B, respectively) adjacent the amino-terminus of the core and the discussed portion of the P-domain at a likely spatial position relative to the core. This suggestion corroborates with a number of experimental observations for the ATP7A, ATP7B and bacterial CopA proteins. We also suggested a similar position for the atypical HMBD of LpCopA based on additional features of the experimental electron density maps (Gourdon et al., 2011). In addition, cross-
linking experiments using bifunctional reagents to map the approximate distances between separate parts of the core and the single HMBD of Enterococcus hirae CopA (Lubben et al., 2009), as well as recent electron microscopy data for Archeoglobus fulgidus CopA (Allen et al., 2011) agree with this assignment of the HMBD location. Importantly, the position is also in agreement with the observed copper-dependent interaction of the HMBDs with the P- and N-domains of ATP7B (Tsivkovskii et al., 2001).

It is generally accepted that the HMBDs are auto-inhibitory (Barry et al., 2010). In agreement with this hypothesis, at our suggested position the HMBDs may inhibit the pump’s function by blocking the conformational changes associated with the reaction cycle. Binding and inhibition might be associated with the fully dephosphorylated E2 conformation or (an) E1 state(s) and not in the modeled transition state of dephosphorylation. This would explain why the HMBD is not clearly observed in the LpCopA crystal structure. We have attempted virtual docking procedures, including ab-initio docking of the HMBD5/HMBD6 complex using ClusPro (http://cluspro.bu.edu/login.php) and refinement of the suggested complex position using the RosettaDock server (http://rosettadock.graylab.jhu.edu/), but we reached only non-realistic solutions. The concept of inhibition through suppression of conformational changes is further supported by the observed increase in ATP affinity when copper is present (Tsivkovskii et al., 2002) and that the HMBDs of ATP7B interact more tightly with the core in the absence of copper (Tsivkovskii et al., 2001).

The six heavy-metal binding domains may arrange in pairs

With the produced homology models and the experimentally determined structures of the HMBDs, about 30 % of human ATP7A and ATP7B remain to be structurally characterized. Limited to the linkers between the HMBDs, the carboxy-termini and the few insertions in the cores, these portions are nonetheless important as they establish how the flanks, with their multiple motifs for targeting and post-translation modifications, are positioned relative to each other and the core. Despite the absence of structural data on the spatial arrangements of HMBD1-HMBD6, which is not alleviated by our homology models, some data may still be retrieved from the complex structures of HMBD5/HMBD6 of ATP7B and sequence analysis.

The short distance, eight residues, between HMBD5 and HMBD6 (555/555 to 564/564) combined with the stabilized complex structures of the HMBD5/HMBD6 complex of ATP7B (Achila et al., 2006) and the dual HMBD of Bacillus subtilis (Banci et al., 2003), suggest that HMBD5 and HMBD6 are associated as a pair. Extending on this hypothesis, it has previously been suggested based on a limited number of sequences that an even number of HMBDs frequently are observed in eukaryotic genomes (Minghetti et al., 2010), such as two in Saccharomyces cerevisiae CCC2, four in Drosophila melanogaster ATP7 and Caenorhabditis elegans CUA-1 and six in mammals. The general matter of HMBD pairs have also been discussed by others (Barry et al., 2010). Over-all our analysis with 1599 proteins displays a
somewhat different pattern (Table 2) with three and five HMBDs in 205 and 34 sequences, respectively, while for comparison 42 and 28 proteins with four or six HMBDs were found, assuming one ferredoxin fold per CXXC motif. However, conclusions are complicated by the fact that sequences with three to six CXXCs were typically associated with an unusual length between at least one pair of CXXCs possibly affecting the actual number of available HMBDs; two CXXC motifs may be present in one ferredoxin fold or one CXXC motif missing in one of two connected folds (for proteins with a significantly shorter or longer distance between HMBDs than expected). However, there is another trend that is in agreement with the notion that the dual-complexes of HMBDs are a common structural theme. Examination of the putative linker lengths between predicted HMBDs suggest that the medians of the stretches between HMBD2/3 and HMBD4/5 in proteins with six CXXCs and between HMBD2/3 in sequences with four CXXCs are significantly longer than others. For example, the inter-domain linkers between HMBD1/HMBD2, HMBD3/HMBD4 and HMBD5/HMBD6 in human ATP7B are limited to 18, 32 and 8 residues, significantly decreasing the possible structural orientations of the two folds within each putative pair and mimicking the observed complex of two HMBDs (Achila et al., 2006; Banci et al., 2003). In contrast, the median of the segments between HMBD2/HMBD3 and HMBD4/HMBD5 in vertebrates are 43 and 60 amino acids long which may allow the pairs to achieve a preferred spatial orientations relative to the core and the other HMBD pairs.

The peripheral localization of the CXXC motifs to the HMBD-to-HMBD interaction interface in the NMR structures of the dual HMBD complexes (Achila et al., 2006; Banci et al., 2003) disfavor direct copper transfer between the CXXC motifs of the putative pairs. With the termini of the dual HMBDs pointing towards the shared interacting surface in accordance with the NMR structures as well as the observed communication between CXXC motifs recorded using circular dichroism data (DiDonato et al., 2000; LeShane et al., 2010) and the proximity of some copper binding sites identified using X-ray absorption spectroscopy (Ralle et al., 2004) provide strong restraints on how the putative pairs may be structurally arranged. We propose that the six HMBD domains arranged in three pairs might appear as two sets of stacked logs with CXXC motifs at their ends, three and three (Figure 1), in accordance with e.g. the suggestion that ATOX1 interacts only with HMBD2 and HMBD4 (Achila et al., 2006). The tight and conserved connecting loop between HMBD5 and HMBD6 would enable communication between the two stacks, explaining the observed copper-induced rearrangements of the HMBD cluster (Bartee et al., 2009; DiDonato et al., 2000) and displacement of the auto-inhibitory HMBD6 from the ATPase core in the presence of copper as discussed above. An apparent contradiction with many of the studies covering copper exchange in separate HMBDs is that the ferredoxin fold remains virtually unaltered independent on if the binding site is metallated (Banci et al., 2004; Gitschier et al., 1998; Jones et al., 2003). The HMBD5/HMBD6 linker working as a hinge between the two stacks may however solve this dilemma, propagating the small changes in the CXXC-containing loops associated with copper presence or absence binding or release.
The reason that higher organisms display multiple copies of HMBDs could be that it allows more fine-tuned regulation and trafficking.

Materials and Methods

The canonical FASTA sequences of ATP7A (isoform 4, Q04656-1, 1500 residues) and ATP7B (isoform 1, P35670-1, 1465 residues) were retrieved from the UniProt database and these are hereafter simply denoted ATP7A and ATP7B. They were used separately in Modeller version 9.9 (Eswar et al., 2006) using LpCopA (pdb-id 3rfu) and the alignment in Figure 2 as input parameters for modeling. The best-scored models were selected for further processing. The unbiased scores from Modeller for ATP7A were 9591 (molpdf), -100518 (DOPE), and 1.0 (GA341) while they were 8159 (molpdf), -98036 (DOPE), and 1.0 (GA341) for ATP7B suggesting that over-all the homology models were of genuine quality. The experimentally characterized structures for available soluble domains (pdb-ids 2kij and 2kmv for ATP7A as well as 2kij and 2arf for ATP7B) were then used to replace the corresponding portions of the model, following alignment using SSM in COOT (Emsley et al., 2010), except for the linkers that were truncated from the experimentally determined structures. In compliance with all other experimentally determined P-type ATPase structures and the sequence alignment of ATP7A and LpCopA, a surface exposed helix in 2arf was reoriented to mimic the corresponding in 2kmv (residues 1196 to 1207 in ATP7A and 1163 to 1174 in ATP7B). For the major insertion in the N-domain (see later) one separate conformation from the NMR structures was selected that did not yield clashes with the generated model. To further improve the models and remove artifacts in the extracytosolic loops of ATP7A and ATP7B generated by Modeller, alanines were introduced into the LpCopA portion of the alignment such that no gaps between the template and the target proteins remained in these loops. As a consequence, the extracytosolic loops between MA/MB and M3/M4 have been generated de novo and are therefore only included to illustrate their relative sizes.

Although not present in the LpCopA structure, we have also included the experimentally determined structure of the HMBD5/HMBD6, 2ew9 (Achila et al., 2006), complex in the ATP7B model as well as a modeled equivalent in ATP7A (Figure 3). Given the short distance between the final residue of the ferredoxin fold of HMBD6 and the first residue in the LpCopA structure (corresponding to S630-H643 in ATP7A and ATP7B) as well as the limitations imposed by the adjacent hydrophobic membrane interface, the approximate position of the HMBD6 relative to the core, and with that also HMBD5, has been attempted. While the approximate spatial position of the HMBD5/HMBD6 complex is likely to be rather accurate, specific details certainly remain undetermined. The final coordinate files for the homology models have been deposited at the PUMPkin website (http://www.pumpkin.au.dk/atp7a and http://www.pumpkin.au.dk/atp7b). Figure 1 was generated using Powerpoint, Figure 2 with Jalview and ESPript 2.2 (http://espript.ibcp.fr/) while Figure 3 was made using Pymol (http://www.pymol.org).
For sequence analysis, a sequence alignment of full length class IB P-type ATPases (all containing the CPC motif in M4, YN in M5, and MXXSS in M6 – motifs associated with Cu⁺/Ag⁺ pumps (Arguello, 2003)) was established using entries found in the P-type ATPases Database (http://www.pumpkin.au.dk/pump-classifier/p-type-atpase-database/) and MUSCLE (Edgar, 2004), in accordance with our previous work (Gourdon et al., 2011). This set was extended with new vertebrate sequences available in the latest Uniprot database version and with the same motifs (Magrane and Consortium, 2011). Following removal of 21 entries for which the amino-terminus of the core relative to LpCopA could not be ascertained, a total of 1599 proteins remained. The positions of each CXXC motif of these sequences were extracted using a shell script and used to estimate the lengths of the linkers between the ferredoxin folds, assuming a typical fold size of 67 amino acids with the CXXC motif at residues 15-18.
Figures, Tables & legends
Figure 1. Overview of the human ATP7A and ATP7B ATPases. The proteins contain six heavy-metal binding domains (H1-H6, green and cyan) that may play a regulatory role, A- (yellow), P- (blue) and N- (red) domains associated with auto(de)phosphorylation and a transmembrane domain with eight membrane spanning helices (note that several helices protrude further into the intracellular side, not shown for clarity); the class specific-MA and MB (cyan) and the M1-M6 (wheat) of the general P-type ATPase core. However, in comparison to other classes of P-type ATPases, the amino-terminal fraction of the A-domain and transmembrane helices M7-M10 are absent. Most likely, copper is delivered to the proteins by chaperones, such as ATOX1 (green), with a similar fold as the HMBDs. Proposed copper binding resides in regulatory subunits, in the copper transport pathway, as well as residues involved in auto(de)phosphorylation are denoted in white, black and red, respectively. The relative position of the HMBDs represents a possible tertiary arrangement.
Figure legend Figure 2. Sequence alignment of the human copper pumps. The alignment displays the human protein ATP7A (A) and ATP7B (B) as well as LpCopA (L) and the consensus sequence for the selected ATP7A (A_c) and ATP7B (B_c) ATPases, respectively. HMBD1-HMBD5 are shown with dark green background, HMBD6 and the P_{IB}-specific TM helices MA and MB in cyan, TM helices M1-M6 in wheat, the A-domain in yellow, the P-domain in red and the N-domain in blue. Black background signifies major insertions as compared to the structurally determined LpCopA structure. Boxes indicated CXXC motifs in the HMBDs (red) as well as important residues at the catalytic phosphorylation site (orange), ATP-binding (grey), positively charged amino acids at the platform (black) and the putative copper transport pathway (brown). Underscored residues display positions that are target by missense mutations causing Menkes’ (green) and Wilson’s (purple) diseases, respectively. The experimentally determined (black) and predicted (PSIPRED (ref), maroon) secondary structure is also shown above the corresponding sequences. The figure is derived and improved from a previous sequence alignment of 1713 CopA proteins (Gourdon et al., 2011). Separately, vertebrate ATP7A (from *D. rerio*, *O. beta*, *O. niloticus*, *S. aurata*, *M. domestica*, *C. griseus*, *R. norvegicus*, *M. musculus*, *L. africana*, *B. taurus*, *E. caballus*, *O. cuniculus*, *H. sapiens*, *P. abelii*, *T. guttata*, *G. gallus*, *M. gallopavo*) and ATP7B (from *T. guttata*, *E. caballus*, *G. gallus*, *M. domestica*, *C. familiaris*, *M. gallopavo*, *A. melanoleuca*, *N. leucogenys*, *M. musculus*, *C. porcellus*, *A. carolinensis*, *R. norvegicus*, *H. sapiens*, *B. taurus*, *C. jacchus*, *O. cuniculus*, *O. aries*) ATPases were aligned using MUSCLE and their consensus extracted from JALVIEW. Insertions in the consensus sequences have been removed.
Figure 3.

Figure legend Figure 3: Homology models of ATP7A and ATP7B in a transition state of dephosphorylation stabilized by the phosphate mimic aluminum fluoride. HMBD5, HMBD6 and the A-, P-, N-domains are shown in green, cyan, yellow, blue and red, respectively. The PIB-specific TM helices MA and MB, connected to HMBD6, are colored in cyan, TM helices M1-M6 in wheat, while black color and grey indicate less reliable insertions. Black portions display major differences between ATP7A and ATP7B. A) Overview of the human ATP7A model. Green spheres mark positions of missense mutations that cause Menkes’ disease (numbers as in Supplementary Table 2 in (Gourdon et al., 2011)) while brown spheres represent glycosylation sites (N686 and N897). B) Overview of the human ATP7B model in the identical orientation as in A. Magenta spheres represent Wilson’s disease mutations (numbers as in Table 1) and the orange sphere marks the non-catalytic phosphorylation site on S1121.
Figure legend Figure 4: Details of ATP7A. The domains and mutations are colored as in figure 3. Annotation of known disease-causing mutations in ATP7A and ATP7B also follows that of Figure 3 and stars (*) that the residue is targeted in both proteins. A) Close-view of the phosphorylation site of ATP7A in the copper-free E2-conformation. The phosphate mimic aluminum-fluoride at the catalytic phosphorylation site and the adjacent, functionally relevant, Mg$^{2+}$ are shown as pink and green spheres, respectively. Residues known to be involved in ATP binding are shown as sticks and labeled in cyan. B) Close-view of the linker connecting HMBD6 and the core, the platform with its positively charged and surface exposed residues as well as the putative copper pathway of ATP7A (suggested copper binding residues as sticks and labeled in orange). A copper ion at the putative exit site is indicated.
Table 1.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>20. Only selected references are linked to the original studies</td>
<td></td>
</tr>
</tbody>
</table>

Table legend Table 1. Positions of Wilson's disease-causing missense mutations in ATP7B.

Only positions identified in two independent studies were selected in order to filter against polymorphisms, although neither false positives nor false negatives can be excluded. For each position, two selected references to the original studies linking the specific residue to disease are indicated. Data acquired from http://www.wilsondisease.med.ualberta.ca/database.asp (Kenney and Cox, 2007).
Table 2.

<table>
<thead>
<tr>
<th># CXXCs</th>
<th># Sequences</th>
<th>N_term</th>
<th>L1/2</th>
<th>L2/3</th>
<th>L3/4</th>
<th>L4/5</th>
<th>L5/6</th>
<th>core L</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>14</td>
</tr>
<tr>
<td>1</td>
<td>697</td>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>518</td>
<td>16</td>
<td></td>
<td></td>
<td></td>
<td>1</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>205</td>
<td>14</td>
<td></td>
<td></td>
<td>-5</td>
<td>6</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>42</td>
<td>24</td>
<td>18</td>
<td>45</td>
<td>17</td>
<td>18</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>34</td>
<td>40</td>
<td>37</td>
<td>16</td>
<td>17</td>
<td>18</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>28</td>
<td>32</td>
<td>18</td>
<td>43</td>
<td>33</td>
<td>60</td>
<td>9</td>
<td>15</td>
</tr>
</tbody>
</table>

Table legend: Table 2. Sequence analysis for strict CXXC motifs from 1599 putative class IB P-type ATPases with the subclass specific sequence motifs CPC, YN and MXXSS (see Materials and Methods for details). The number of entries (# Sequences) with a certain amount of CXXCs (# CXXCs) is indicated as well as the distance of the first motif from the amino-terminus (N_term). The linkers between HMBD1/HMBD2 (L1/2), HMBD2/HMBD3 (L2/3) etcetera have been using the formula: Linker length="CXXC position X" – “CXXC position X-1” – 67 (67 being the length of one classical ferredoxin fold). For the core linker between the core and the last HMBD the same formula has been employed, except that 67 was replaced for 53 (the distance from the last CXXC motif to the start of the core). The negative linker length suggests that the HMBD may be absent or smaller than typical.
forms of the first metal binding sequence of ATP7A associated with Menkes disease. Proteins 61, 1038-1049.

References


Banci, L., Bertini, I., Del Conte, R., D’Onofrio, M., and Rosato, A. (2004). Solution structure and backbone dynamics of the Cu(I) and apo forms of the second metal-binding domain of the Menkes protein ATP7A. Biochemistry 43, 3396-3403.


Gonzalez, Gu


Calcium Signalling and Disease. In, E. Carafoli, and M. Brini, eds. (Springer Netherlands), pp. 385-404.
