Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

The structure of Mg-ATPase nucleotide-binding domain at 1.6 Å resolution reveals a unique ATP-binding motif

Kjell O. Håkansson

Department of Biology, August Krogh Building, University of Copenhagen, Universitetsparken 13, DK-2100 Copenhagen O, Denmark

Correspondence e-mail: kohakansson@bio.ku.dk

Received 9 July 2009
Accepted 19 August 2009

PDB Reference: Mg-ATPase nucleotide-binding domain, 3gwi, r3gwisf.

The structure of the nucleotide-binding domain of the Mg-ATPase MgtA from *Escherichia coli* has been solved and refined to 1.6 Å resolution. The structure is made up of a six-stranded β-sheet and a bundle of three α-helices, with the nucleotide-binding site sandwiched in between. The MgtA nucleotide-binding domain is shorter and more compact compared with that of the related Ca-ATPase and lacks one of the β-strands at the edge of the β-sheet. The ATP-binding pocket is surrounded by three sequence and structural motifs known from other P-type ATPases and a fourth unique motif that is found only in Mg-ATPases. This motif consists of a short polypeptide stretch running very close to the ATP-binding site, while in Ca-ATPase the binding site is more open, with the corresponding polypeptide segment folded away from the active site.

1. Introduction

P-type ATPases are found in most living organisms and fulfil a number of important functions involving cation transport in eukaryotes: *e.g.* maintaining electrochemical membrane potentials, transcellular transport processes, calcium-mediated signal transduction, muscle contraction and gastric juice production. In prokaryotes, P-type ATPases are mainly involved in homeostasis or detoxification of cells.

To date, three-dimensional structures at atomic resolution are known for only three P-type ATPases, Ca-ATPase (Toyoshima *et al.*, 2000), Na,K-ATPase (Morth *et al.*, 2007; Shinoda *et al.*, 2009) and H-ATPase (Pedersen *et al.*, 2007), although structures have been obtained for several conformations and ligand complexes (Olesen *et al.*, 2007; Toyoshima, 2008). In addition, crystallographic structures are available for the N-domain of Na,K-ATPase (Håkansson, 2003), the N-domain of Cu-ATPase (Sazinsky *et al.*, 2006) and a fragment of Cu-ATPase containing both the N-domain and the P-domain (Lubben *et al.*, 2007).

P-type ATPases contain a number of transmembrane helices and three cytoplasmic domains: the actuator or A-domain, the nucleotide binding or N-domain and the phosphorylation or P-domain. The N-domain is made up of an antiparallel β-sheet with up-and-down topology and is responsible for binding and delivering the substrate ATP towards the P-domain. The P-domain belongs to the HAD family of folds and contains a conserved aspartic acid residue which functions as a transient covalent carrier of the ATP γ-phosphate during ATP hydrolysis (Degani & Boyer, 1973). The A-domain has a jelly-roll fold and plays an important role in dephosphorylation of this residue. The three domains thus interact during ATP hydrolysis and undergo large movements. These movements are

transmitted to the transmembrane region, where the positions of some of the transmembrane helices are shifted. Since four of these helices collectively make up the binding sites for the transported cations, these events result in sequentially ordered changes in binding affinity and occlusion (Jorgensen, 1975; Jorgensen & Andersen, 1988; Toyoshima, 2008).

The Mg-ATPase MgtA is found in a wide variety of prokaryotes, including *Escherichia coli*. Some bacteria, *e.g. Salmonella typhimurium*, have an additional Mg-ATPase called MgtB (Park *et al.*, 1976; Snavely *et al.*, 1989). These two Mg-ATPases mediate the uptake of Mg ions. They are more closely related to eukaryotic Ca-ATPases and Na,K-ATPases than to their prokaryotic paralogues both in terms of sequence similarity and the arrangement of the ten transmembrane helices (Snavely *et al.*, 1991; Tao *et al.*, 1995). Although MgtB was the first P-type ATPase for which the membrane topology was experimentally established (Smith *et al.*, 1993), the Mg-ATPases have not been characterized to the same extent as the Ca-ATPases, Na,K-ATPases, H-ATPases or Cu-ATPases. In this paper, the crystallographic structure of the ATP-binding 19.5 kDa nucleotide-binding domain of Mg-ATPase is reported at 1.6 Å resolution, which is the highest resolution achieved to date for any P-type ATPase structure or domain thereof. This domain is the most divergent part of the P-type ATPases and the normal algorithms used to align sequences fail to align these sequences properly. The structure presented in this paper indicates that the Mg-ATPase and Ca-ATPase N-domains have 31 identities among 133 structurally homologous residues, representing 19% of the Mg-ATPase N-domain sequence. The structure also suggests some interesting differences in the way these proteins interact with ATP.

2. Material and methods

The *E. coli* Mg-ATPase N-domain was cloned and crystallized as described previously (Håkansson & Ćurović, 2009). Synchrotron data were collected using a CCD detector at MAX-lab beam station I911-2 (Lund) and were processed with *XDS* (Kabsch, 1993). The structure was solved by molecular replacement using the *CCP*4 version of *AMoRe* (Collaborative Computational Project, Number 4, 1994; Navaza, 1994) in the two alternative space groups $P4_12_12$ and $P4_32_11$. The search model was a truncated version of the 2.4 Å resolution structure of the Na,K-ATPase N-domain from which those parts of the structure in which the Na,K-ATPase and Ca-ATPase (PDB codes 1q3i and 1eul; Håkansson, 2003; Toyoshima *et al.*, 2000) backbones deviate had first been removed, leaving a core consisting of approximately 75% of the Na,K-ATPase domain. Although no solution was significantly better than any other in terms of correlation coefficient or *R* factor, the solution with the highest correlation coefficient (13.9%, refined to 31.1% with *AMoRe FITFUN*) and lowest *R* value (54.0%, refined to 50.5% with *AMoRe FITFUN*) also had a realistic minimum distance between molecule centres (33 Å). This solution in space group $P4_12_12$ was rebuilt into the MgtA N-domain with the familiar iterative improvements of structure statistics and electron-density maps. The graphics programs *Coot* and *O* (Jones *et al.*, 1991; Emsley & Cowtan, 2004) were used to build and analyze the structure, while electron-density map calculations and refinement were performed with *CNS* (Brünger *et al.*, 1998). Solvent building and final refinement were performed with *ARP*/*wARP* and *REFMAC* (Lamzin & Wilson, 1997; Murshudov *et al.*, 1997). Data-collection statistics have been published elsewhere (Håkansson & Ćurović, 2009) and the structure-refinement statistics are shown in Table 1. The coordinates and X-ray data have been submitted to the Protein Data Bank with accession code 3gwi. ATP inhibits crystallization of *E. coli* MgtA and attempts to soak ATP into the crystals failed. Inspection of the crystallographic symmetry revealed that the expected phosphate-chain location overlaps with a symmetry-related protein molecule.

Table 1
Refinement statistics.

Data statistics have been published elsewhere (Håkansson & Ćurović, 2009).

Resolution (Å)	20–1.6 (1.64–1.60)
R (%)	19.1 (20.0)
R_{free} (%)	22.0 (24.7)
R.m.s.d. bonds (Å)	0.009
R.m.s.d. angles (°)	1.2
No. of protein atoms	1314
No. of solvent atoms	282
B_{ave}, protein (Å^2)	15.1
B_{ave}, solvent (Å^2)	33.0
Ramachandran plot (%)	
Most favoured	94.0
Allowed	6.0
Generously allowed	0.0
Disallowed	0.0

3. Results and discussion

3.1. Overall fold

As predicted from the sequence homology, the Mg-ATPase N-domain has the same fold as the N-domains of the known structures of Na,K-ATPase, which was used as a search model in the molecular replacement, and sarcoplasmic reticulum Ca-ATPase. The final Mg-ATPase N-domain structure comprised 1314 protein atoms, 272 water molecules and two sulfate ions. This included a methionine added to the N-terminus and 163 Mg-ATPase residues from Ile383 to Asp545. The structure was in general well defined and the electron-density maps were easy to interpret, except for the C-terminal hexahistidine affinity tag, for which the density was weak and difficult to interpret. Consequently, these disordered residues were not defined and may be the cause of the relatively high *R* factors. In the following discussion, Ca-ATPase will be used as a reference, since this protein is better characterized at the structural level, and similarities and differences between the two proteins will be highlighted. Furthermore, the Ca-ATPase structure with bound AMPPCP can be used to analyze the residues involved in ATP binding (Toyoshima & Mizutani, 2004).

Comparison of Mg-ATPase and Ca-ATPase revealed that the backbone trace appears to be similar for about 133 resi-

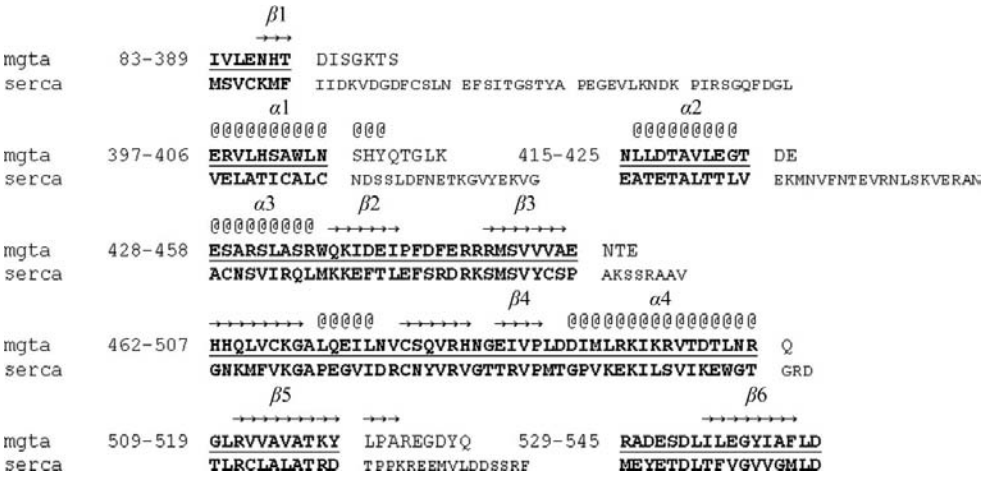

Figure 1
Structure-based sequence alignment of Mg-ATPase and Ca-ATPase. Mg-ATPase residues that have structurally homologous residues in Ca-ATPase are underlined and shown in bold. Ca-ATPase residues that have no structurally homologous residues in Mg-ATPase are shown in a smaller font. Arrows indicate β-strands and '@' indicates α-helices or 3_{10}-helices. Each segment that can be structurally aligned is preceded by Mg-ATPase residue numbers.

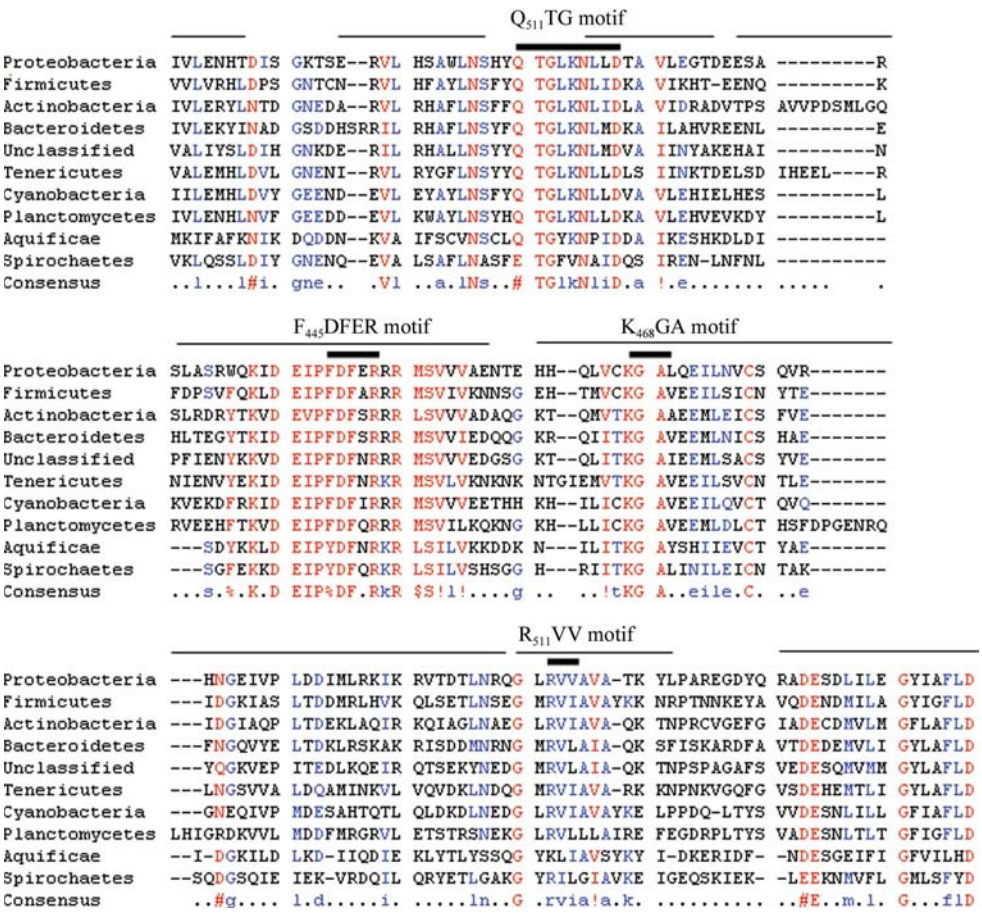

Figure 2
Sequence alignment of the closest homologous sequences found by *BLAST* in a series of searches limited to the following groups: proteobacteria (taxid 1224), actinobacteria (taxid 201174), aquificae (taxid 200783), bacteroidetes (taxid 68336), cyanobacteria (taxid 1117), firmicutes (taxid 1239), planctomycetes (taxid 112), spirochaetes (taxid 203691), tenericutes (taxid 31969) and unclassified sequences (taxid 1223). The location of the four conserved ATP-binding motifs are indicated with thick bars and the regions that are superimposable on Ca-ATPase with thin bars. The consensus symbols are ! for (I, V), $ for (L, M), % for (F, Y) and # for (N, D, Q, E).

dues. They were selected merely by human judgment to be locally superimposable. Although some hairpin loops might deviate when the two domains are globally superimposed, they have been defined as homologous if they have similar backbone profiles. These 133 residues correspond to 81% of the Mg-ATPase N-domain and make up seven blocks separated by six non-homologous segments with insertions and deletions. Superimposition of the C^{α} atoms of these residues results in a root-mean-square deviation of 1.6 Å for these 133 atoms. All eight β-strands, all four α-helices and the single 3_{10}-helix are within the superimposable regions. There are only 31 identical residues among these superimposable residues, amounting to 19% of the Mg-ATPase N-domain or 13% of the larger Ca-ATPase N-domain. Fig. 1 shows the structure-based alignment together with the secondary-structure elements. Fig. 2 shows a multiple sequence alignment of Mg-ATPases from nine different prokaryotic groups and the most closely related 'unclassified' prokaryotic sequence. The regions with the highest sequence conservation are generally found in regions which are superimposable on Ca-ATPase. However, there is a notable exception: the QTG motif between helix 2 and 3, which approaches the ATP-binding site, is unique to Mg-ATPase (see below). The three-dimensional structure is shown in Fig. 3, where it is superimposed on the homologous Ca-ATPase structure. Fig. 4 shows the Mg-ATPase structure from the same orientation coloured according to sequence conservation. Since no nucleotide-ligand structure is available for MgtA, the ATP molecule is modelled from the Ca-ATPase structure.

The first β-strand is similar in its backbone conformation to the

homologous strand in Ca-ATPase but is displaced by 2–3 Å. There are two possible reasons for this. Firstly, this β-strand is on the edge of the β-sheet in Mg-ATPase, having only one adjacent β-strand, while in Ca-ATPase this strand is found within the β-sheet and has neighbouring strands on both sides. Secondly, it is much shorter in Mg-ATPase than in Ca-ATPase and is abruptly terminated after Thr389. In Ca-ATPase the β-strand continues for an additional six residues before making a hairpin loop followed by an antiparallel β-strand on the edge of the sheet; this strand has no counterpart in MgtA.

The first α-helix, α1, is followed by two other α-helices forming a three-helix bundle on the concave side (the right side in Figs. 3 and 4) of the β-sheet. These are followed by β-strand β2, which makes up the second, opposite edge of the β-sheet. The rest of the β-sheet is filled out by strands β3, β4, β7 and β8 with a simple antiparallel up-and-down topology. β8 is then hydrogen bonded to the first strand, β1. β-Strands β5–β6 make a little hairpin insertion which can also be regarded as a bifurcation of the β-sheet as they approach β8. This hairpin is flanked by a 3_{10}-helix and an α-helix α4 on the convex side of the β-sheet.

Comparison of the MgtA N-domain with the N-domain of Ca-ATPase and with the N-domain of Na,K-ATPase gives a similar picture; after all, the two eukaryotic calcium and sodium potassium pumps have more in common with each other than they have with MgtA. They are both larger, with larger loop regions than found in MgtA, although Na,K-ATPase is less exuberant than Ca-ATPase in the segment between the α2 and α3 helices. The first β-strand in the Na,K-ATPase N-domain is also shorter, like in MgtA, but is followed by a hairpin and a β-strand at the edge of the sheet, like in Ca-ATPase (see above). The presence of this β-strand in the molecular-replacement model and its absence from the final MgtA structure represents the major flaw in the model, which only consisted of elements that are structurally similar in Ca-ATPase and Na,K-ATPase.

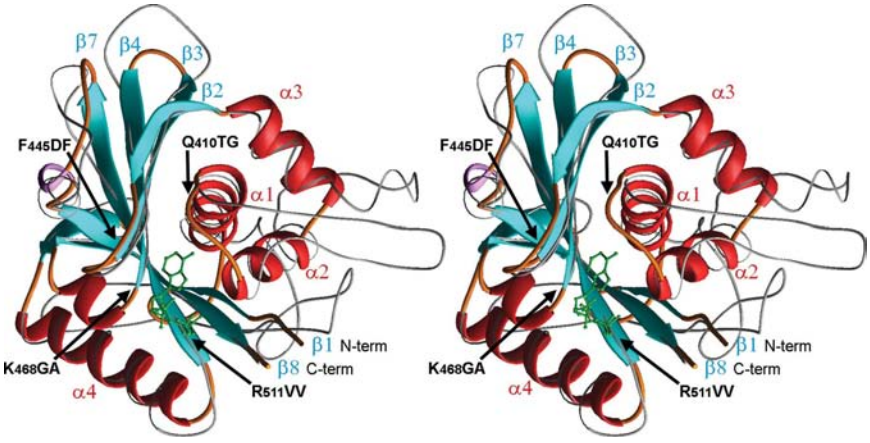

Figure 3
Ribbon representation of the Mg-ATPase structure. β-Strands are shown in cyan, α-helices in red and the single 3_{10}-helix in magenta. For comparison, Ca-ATPase (PDB code 2dqs) is shown in grey and its AMPPCP ligand in green. The Ca-ATPase structure was superimposed using the C^{α} atoms from the homologous residues displayed in Fig. 1. The four motifs discussed in the text are indicated with arrows; the first of these, Q_{410}TG, is unique to Mg-ATPase and is found between the first two α-helices; the other three motifs, on the edge of the β-sheet, are shared with Ca-ATPase and Na,K-ATPase. The strand order from front to back is 2, 3, 4, 7, 8 and 1 (labelled) or 5, 6 (on the left side, not labelled). Labelling is the same as in Fig. 1. Note the large difference between Mg-ATPase and Ca-ATPase in the N-terminal part and between the first two helices.

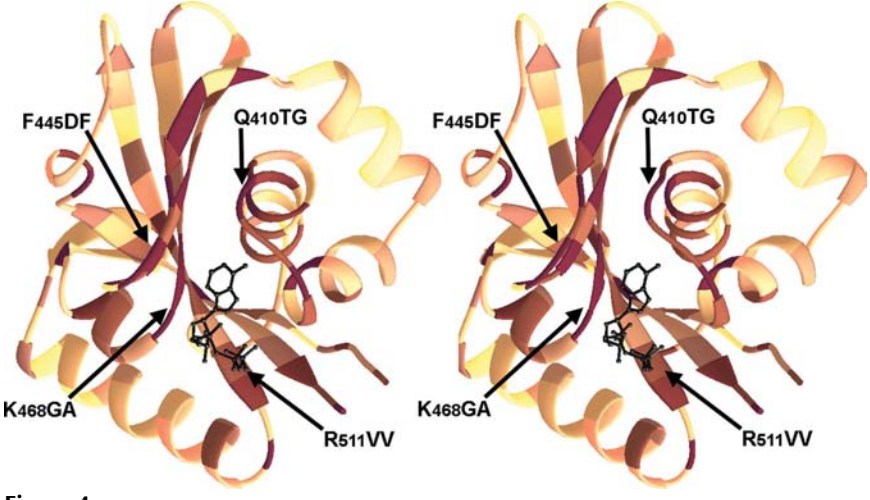

Figure 4
Ribbon representation from the same orientation as in Fig. 3. The polypeptide chain is coloured according to the number of sequence identities at each position according to the alignment in Fig. 2. The higher the sequence conservation, the darker the colour.

3.2. Conserved motifs and the ATP-binding site

There are four conserved polypeptide segments around the ATP-binding site: Gln410–Asp418, Phe445–Arg450, Lys468–Ala470 and Arg/Lys511–Leu/Ile/Val513. The last three of these motifs are shared with Ca-ATPase, while the first motif is unique to Mg-ATPase. This segment between helices α1 and α2 forms an extra flap over the binding site. In Ca-ATPase the structure of the segment between helices α1 and α2 is much longer and is folded away from the ATP-binding site. The conserved motifs are indicated with arrows in the structural representations of Figs. 3 and 4.

As shown in Fig. 2, the Q_{410}TG motif is preserved in the closest relative in all nine represented groups of bacteria, except in the case of *Spirochaetes*, in which Gln410 is substituted by a glutamic acid. This stretch of amino acids is defined as a signature motif for Mg-ATPases according to *Print-S* (Attwood *et al.*, 2003). Interestingly, this has started to diverge somewhat between MgtA and MgtB; in *S. typhimurium* MgtB this motif is represented by the amino acids QSG and is flanked by a Ser and an Ala,

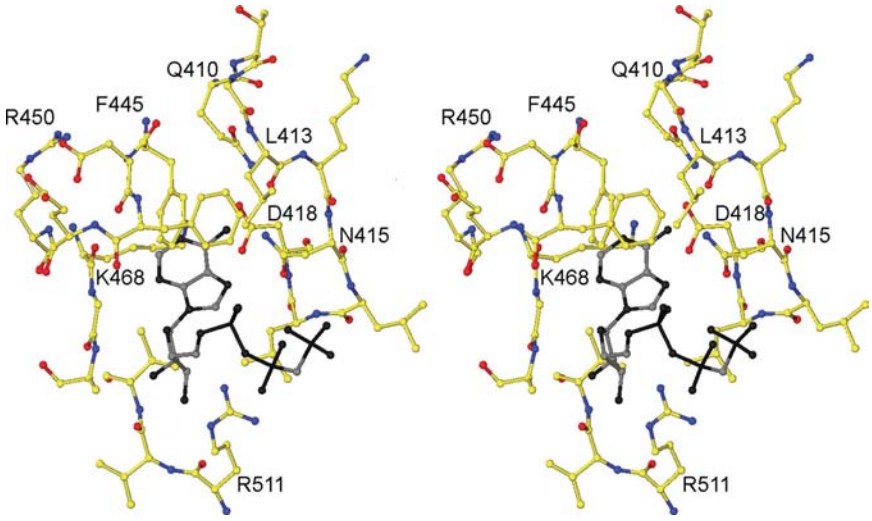

Figure 5
The ATP-binding site in MgtA, viewed from approximately the same orientation as in Figs. 3 and 4. The polypeptide segments of the four conserved motifs are shown, together with some of the side chains discussed in the text. The MgtA unique motif can be seen on the right side of the nucleotide and the other three motifs are on the left side. The AMPPCP ligand is shown in black and white to emphasize that its position has not been determined experimentally but has been superimposed from the Ca-ATPase structure 2dqs.

rather than the bulkier tyrosine or hydrophobic residues that flank this motif in the sequences in Fig. 2. The ATP analogue AMPPCP has been docked into the active site as shown in Fig. 5 by superimposing the known structure of the Ca-ATPase–AMPPCP complex. MgtA residue Asn415 will with all probability hydrogen bond to the adenine N6 group, together with Asp418. The backbone structure of these residues appears to be stabilized by the hydrogen bonds between Gln410 of the Q_{410}TG motif and the backbone N atom and carbonyl group of Leu413. There is also a direct hydrogen bond from Gln410 to the side chain of Asp418. While the hydrogen bond between Asn415 and adenine N6 is unique to MgtA, the hydrogen bond between Asp418 and adenine N6 does have a counterpart in Ca-ATPase in the hydrogen bond between adenine and Glu442 of Ca-ATPase. In Ca-ATPase, which lacks the Q_{410}TQ motif, Glu442 is fixed to the backbone N atom of Asp422 by a hydrogen bond. The Glu/Asp carboxylate group of both enzymes binds to a conserved lysine: Lys468 in MgtA and Lys515 in Ca-ATPase (see below). Overall, there are more interactions possible between this part of the adenine group and MgtA than there are between adenine and Ca-ATPase, perhaps indicating that MgtA has a higher selectivity for the nucleotide base. While no MgtA–nucleotide structure has been determined, there are several other observations that suggest this hypothesis. (i) The long loop found in Ca-ATPase that extends some 30 Å away from the ATP-binding site has the same structure in the presence and absence of bound nucleotide. (ii) In the three available Na,K-ATPase structures this loop is as long as in Ca-ATPase and has the same structure in all three cases, except that the remotest tip of this loop is disordered in the α2 porcine Na,K-ATPase N-domain. (iii) The much shorter MgtA segment runs almost the shortest way between the stable and conserved α-helices α1 and α2, leaving little room for conformational changes. It is hard to see how interactions between the Q_{410}TQ motif and bound ATP could be avoided in view of the limited space in the active site. In Ca-ATPase and Na,K-ATPase the lid is open and folded away. In MgtA the lid is closed over a part of the nucleotide-binding site and appears too small to be opened and removed.

The second conserved active-site motif F_{445}DFER contains, in addition to Phe445, a second Phe residue at position 447. These two residues are homologous to Phe487 and Arg489 in Ca-ATPase, in which the former is stacked on the ATP adenosine ring. In MgtA these two phenylalanines help to anchor the MgtA-specific conserved Gln410–Asp418 loop through hydrophobic interactions with Leu413, but this interaction does not preclude a possible interaction between the phenyl group and ATP in Mg-ATPase also.

The third of these motifs, K_{468}GA, is also found in Ca-ATPase and Na,K-ATPase. In the Ca-ATPase structure 2dqs the side chain of Lys515 (equivalent to Lys468 in MgtA) makes a long hydrogen bond to the N6 atom of ATP. The presence of the conserved Gly469 can be explained by the potential sterical hindrance a side chain would face. The Ramachandran values are also slightly outside those allowed for amino acids with side chains and the bulky Arg450 and Arg451 (not shown in the figure) will come close to any side chain. While the two residues following the K_{468}GA motif are a proline and a glutamate in both Ca-ATPase and Na,K-ATPase, MgtA does not seem to need the assistance of a proline residue in order to terminate β-strand β4.

The last of these motifs, R_{511}VV, corresponds to the R560XL motif in Ca-ATPase and the importance of this motif has been established by site-directed mutagenesis in both Ca-ATPase and Na,K-ATPase (Jacobsen *et al.*, 2002; Hua *et al.*, 2002). The arginine interacts with the β-phosphate of ATP, thus aiding the release of ADP as well as directing the γ-phosphate to the phosphorylation site in analogy with the phosphate-binding site of HAD enzymes such as phosphoserine phosphatase (Wang *et al.*, 2002). In Ca-ATPase, Arg560 also hydrogen bonds to Thr441. The homologue of Thr441 in MgtA is Leu417 and this hydrogen bond is therefore lost.

Aida Ćurović is thanked for technical assistance, Dr Johan Unge is thanked for his help during data collection and Professors Sine Larsen and Leila Lo Leggio are thanked for sharing synchrotron beam time. This work was supported by the Carlsberg foundation and the Danish Research Council FNU.

References

Attwood, T. K., Bradley, P., Flower, D. R., Gaulton, A., Maudling, N., Mitchell, A. L., Moulton, G., Nordle, A., Paine, K., Taylor, P., Uddin, A. & Zygouri, C. (2003). *Nucleic Acids Res.* **31**, 400–402.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* D**54**, 905–921.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Degani, C. & Boyer, P. D. (1973). *J. Biol. Chem.* **248**, 8222–8226.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Håkansson, K. O. (2003). *J. Mol. Biol.* **332**, 1175–1182.
Håkansson, K. O. & Ćurović, A. (2009). *Acta Cryst.* F**65**, 223–225.
Hua, S., Ma, H., Lewis, D., Inesi, G. & Toyoshima, C. (2002). *Biochemistry*, **41**, 2264–2272.
Jacobsen, M. D., Pedersen, P. A. & Jorgensen, P. L. (2002). *Biochemistry*, **41**, 1451–1456.
Jones, T. A., Zou, J.-Y., Cowan, S. W. & Kjeldgaard, M. (1991). *Acta Cryst.* A**47**, 110–119.
Jorgensen, P. L. (1975). *Biochim. Biophys. Acta*, **401**, 399–415.
Jorgensen, P. L. & Andersen, J. P. (1988). *J. Membr. Biol.* **103**, 95–120.
Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
Lamzin, V. S. & Wilson, K. S. (1997). *Methods Enzymol.* **277**, 269–305.
Lubben, M., Guldenhaupt, J., Zoltner, M., Deigweiher, K., Haebel, P., Urbanke, C. & Scheidig, A. J. (2007). *J. Mol. Biol.* **369**, 368–385.
Morth, J. P., Pedersen, B. P., Toustrup-Jensen, M. S., Sorensen, T. L., Petersen, J., Andersen, J. P., Vilsen, B. & Nissen, P. (2007). *Nature (London)*, **450**, 1043–1049.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Navaza, J. (1994). *Acta Cryst.* A**50**, 157–163.
Olesen, C., Picard, M., Winther, A. M., Gyrup, C., Morth, J. P., Oxvig, C., Moller, J. V. & Nissen, P. (2007). *Nature (London)*, **450**, 1036–1042.
Park, M. H., Wong, B. B. & Lusk, J. E. (1976). *J. Bacteriol.* **126**, 1096–1103.
Pedersen, B. P., Buch-Pedersen, M. J., Morth, J. P., Palmgren, M. G. & Nissen, P. (2007). *Nature (London)*, **450**, 1111–1114.
Sazinsky, M. H., Mandal, A. K., Arguello, J. M. & Rosenzweig, A. C. (2006). *J. Biol. Chem.* **281**, 11161–11166.
Shinoda, T., Ogawa, H., Cornelius, F. & Toyoshima, C. (2009). *Nature (London)*, **459**, 446–450.
Smith, D. L., Tao, T. & Maguire, M. E. (1993). *J. Biol. Chem.* **268**, 22469–22479.
Snavely, M. D., Florer, J. B., Miller, C. G. & Maguire, M. E. (1989). *J. Bacteriol.* **171**, 4752–4760.
Snavely, M. D., Miller, C. G. & Maguire, M. E. (1991). *J. Biol. Chem.* **266**, 815–823.
Tao, T., Snavely, M. D., Farr, S. G. & Maguire, M. E. (1995). *J. Bacteriol.* **177**, 2654–2662.
Toyoshima, C. (2008). *Arch. Biochem. Biophys.* **476**, 3–11.
Toyoshima, C. & Mizutani, T. (2004). *Nature (London)*, **430**, 529–535.
Toyoshima, C., Nakasako, M., Nomura, H. & Ogawa, H. (2000). *Nature (London)*, **405**, 647–655.
Wang, W., Cho, H. S., Kim, R., Jancarik, J., Yokota, H., Nguyen, H. H., Grigoriev, I. V., Wemmer, D. E. & Kim, S.-H. (2002). *J. Mol. Biol.* **319**, 421–431.