J Mol Model (2001) 7:184–192
DOI 10.1007/s008940100031

ORIGINAL PAPER

Rüdiger Ettrich · Milan Melichercik · Jan Teisinger
Olga Ettrichova · Rita Krumscheid
Katerina Hofbauerova · Peter Kvasnicka
Wilhelm Schoner · Evzen Amler

Three-dimensional structure of the large cytoplasmic H_4–H_5 loop of Na^+/K^+-ATPase deduced by restraint-based comparative modeling shows only one ATP binding site

Received: 11 December 2000 / Accepted: 28 April 2001

Abstract Homology modeling of the complete structure of the large cytoplasmic loop between the fourth and fifth transmembrane segments (H_4–H_5 loop) of the α subunit of Na^+/K^+-ATPase is reported. The deduced amino acid sequence shows high sequence identity and homology to the Ca^{2+}-ATPase (32.8% identity and 53.3% similarity in our alignment), whose tertiary structure has been solved recently at 2.6-Å resolution by X-ray crystallography. This high homology allowed the construction of a model structure using the MODELLER program. Refinement was achieved through interactive visual and algorithmic analysis and minimization with the TRIPOS force field included in the SYBYL/MAXIMIN2 module. The docking of ATP as a substrate into the active site of the model was explored with the AUTODOCK program followed by molecular mechanics optimization of the most interesting complexes. Thus, the docking of ATP into the resulting model of the H_4–H_5 loop gave evidence for the existence of one ATP binding site only. We were able to specify Cys^{549}, Phe^{548}, Glu^{505}, Lys^{501}, Gln^{482}, Lys^{480}, Ser^{477}, Phe^{475} and Glu^{446} as parts of the ATP binding site with Lys^{501} located in the depth of the positively charged binding pocket.

Keywords Sodium potassium adenosine triphosphatase · Tertiary structure · Adenosine triphosphate binding site · Restraint-based homology modeling

Electronic supplementary material to this paper can be obtained by using the Springer Link server located at http://dx.doi.org/10.1007/s008949010031

R. Ettrich (✉) · K. Hofbauerova
Department of Physical and Macromolecular Chemistry, Charles University, Albertov 2030, 128 40 Prague 2, Czech Republic
e-mail: ettrich@natur.cuni.cz, Tel.: +420-2-4752249

R. Ettrich · J. Teisinger · O. Ettrichova · K. Hofbauerova · E. Amler
Institute of Physiology, Czech Academy of Sciences, Videnska 1083, 142 20 Prague 4, Czech Republic

M. Melichercik · P. Kvasnicka
Faculty of Mathematics and Physics, Comenius University, Mlynska dolina, 842 48 Bratislava, Slovak Republic

R. Krumscheid · W. Schoner
Institute of Biochemistry and Endocrinology, Justus-Liebig-University Giessen, Frankfurter Str. 100, 35392 Giessen, Germany

Introduction

Na^+/K^+-ATPase (EC 3.6.1.37) is an integral membrane protein that transports sodium and potassium ions against an electrochemical gradient. This enzyme consists of two subunits, the catalytic α subunit with a molecular mass of about 112,000, which carries out all transport and catalytic functions, and the associated β subunit, a glycoprotein with the molecular mass of 35,000 (neglecting the oligosaccharides). [1] Transport of Na^+ and K^+ ions occurs by an oscillation of the enzyme between two major conformational states, the E_1Na^+ and the E_2K^+-conformations, which seem to be intimately connected to the E_1ATP and E_2ATP sites with approximate K_d values of 1 μM (E_1ATP site) and 200 μM (E_2ATP site). [2] The high affinity E_1ATP site (from where the Na^+ export starts by phosphorylation of the Asp^{369} residue of the α subunit) and the low affinity E_2ATP site (which is involved in the K^+ import) seem to interchange in their affinities during the transport process. Contrary to the Albers–Post model assuming a consecutive change of a single ATP site [3, 4] on a single (αβ) protomer between E_1ATP and E_2ATP conformations, extensive kinetic analysis with various ATP analogs [5, 6] as well as energy transfer measurements [7] revealed that E_1ATP and E_2ATP conformers coexist during catalysis and that both ATP binding sites cannot be localized on the same α subunit. This favors the postulate that Na^+/K^+-ATPase may work as a functional $(\alpha\beta)_2$ diprotomer. Most of the information available localizes

the ATP binding site to the large cytoplasmic loop between the transmembrane helices H_4 and H_5 (the H_4–H_5 loop) [2]. The existence of two ATP-binding sites per H_4–H_5 loop or α subunit has been postulated. [8] A number of amino acid residues are thought to be involved in ATP recognition and hydrolysis such as Glu^{472}, Lys^{480}, Gly^{502}, Asp^{710}, Asp^{714} [9, 10, 11] as well as Lys^{501}, Cys^{656}, Lys^{719} and Lys^{766}. [12] Site-directed mutagenesis of Glu^{472} and Lys^{480} suggests that interaction of the α and β phosphate groups of ATP arrests the energy substrate in a proper position prior to hydrolysis of the phosphate group. [11] Affinity labeling by ATP analogs revealed that a cystein residue is involved in ATP binding of Na^+/K^+-ATPase and Ca^{2+}-ATPase of sarcoplasmic reticulum. [13, 14, 15, 16] Cys^{549} has been identified to reside in the ATP binding site by labeling with dihydro-4,4-diidothiocyanatostilbene-2,2´-disulfonate (H_2DIDS). [12] Affinity labeling at pH 7.4 of the E_2ATP site by erythrosin 5-isothiocyanate showed that loss of K^+-phosphatase activity is most probably due to modification of Cys^{549}. [17]

P-type ATPases belong to the family of hydrolases that are structurally typified by the L-2-haloacid dehalogenase. The amino acid homology suggests that the phosphorylation domain adopts a Rossman fold. [18] Recently the tertiary structure of Ca^{2+}-ATPase, also a member of the P-type ATPases, has been solved at 2.6-Å resolution by X-ray crystallography and a single ATP binding pocket was identified [19].

To date several different groups have been able to overexpress parts of the H_4–H_5 loop and to characterize properties of its ATP sites. [20, 21] Gatto et al. [20] have expressed the sequence Leu^{354}–Leu^{774} and showed that the construct is able to bind nucleotide triphosphates. Similar results were obtained in our laboratory using a shorter sequence, Met^{386}–Ile^{784}. [22] We now show that there is evidence only for one ATP binding site per H_4–H_5 loop construct and thus anticipate that our model will serve as a tool to understand ATP binding on the H_4–H_5 loop.

Methods

Choice of the templates and alignment

The choice of the templates was restricted to the Ca^{2+}-ATPase, whose tertiary structure has been solved recently at 2.6-Å resolution by X-ray crystallography and for which the PDB coordinates were available [19]. The structure (1EUL) was extracted from the Brookhaven Protein Data Bank [23] and loaded into SYBYL, [24] where we extracted a shortened construct containing only the large cytoplasmic loop from Arg^{334} to Phe^{909} without any membrane parts of the protein.

The primary structure of pig-kidney Na^+/K^+-ATPase [25, 26] from Leu^{354}–Leu^{773} was aligned with the template sequence by CLUSTALX. [27] The slow–accurate mode with a gap opening penalty of 10 and a gap extensions penalty of 0.1 for the local alignment was used as well as the Gonnet 250 protein weight matrix and hydrophobic penalties for the amino acids GPSNDQEKR. For the global alignment a penalty of 9 has been employed for gap opening and 0.2 for extension with the Gonnet series. The regions from Gln^{390} to Ile^{417} and Leu^{450} to Arg^{464} were not aligned satisfactorily with these parameters. Thus, we have adjusted the alignment by hand using the SwissPDBviewer alignment editor. [28] As a basis for this manual adjustment we took a local alignment performed with LALIGN [29] and the results of an alignment together with secondary structures of several ATPases by PHD. [30] An opening gap penalty of 14 and for the extension a penalty of 4 together with the Blossum50 matrix were used for LALIGN.

Modeling

Three-dimensional models comprising all non-hydrogen atoms were generated by the MODELLER4 package. [31] This is based on a distance restraint algorithm, satisfying spatial constraints extracted from the alignment of the known protein, which is the template structure, with the target sequence [31, 32, 33, 34, 35] and from the CHARMM-22 force-field. [36] The constraints for each of the 15 features used to build 3D models of the loop are listed in Table 1. These constraints are expressed as probability density functions for the features to be restrained. The features are structural properties at residue positions and relationships between residues including solvent accessibility, secondary structure and hydrogen bonding. These constraints, which determine stereochemistry in the structures, combined with energy terms are added up into an objective function. The 3D model of a protein is obtained by optimization of this molecular

Table 1 Spatial constraints used to model the H_4–H_5 loop

Type of constraint	n[a]	RMS[b]
1. Bond length (Å)	3237	0.009
2. Bond angles (°)	4384	2.634
3. Stereochemical cosine torsion (°)[c]	1918	51.519
4. Stereochemical improper torsion (°) [c]	1271	1.187
5. Van der Waals contacts (Å)	7417	0.003
6. C_α–C_α distances (Å)	8100	0.476
7. Main chain N–O distances (Å)	8240	0.450
8. Main chain Ω dihedral angles (°)	419	3.872
9. Side-chain χ_1 dihedral angles (°)	352	79.726
10. Side-chain χ_2 dihedral angles (°)	259	75.744
11. Side-chain χ_3 dihedral angles (°)	111	74.159
12. Side-chain χ_4 dihedral angles (°)	41	96.560
13. Side-chain–main-chain CH–CH distances (Å)	3709	0.390
14. Ψ/Φ pair of dihedral angles (°)	418	39.742
15. Side-chain–side-chain CH–CH distances (Å)	1034	1.031

[a] Number of constraints of a given type that were used to model the loop

[b] Root mean square deviation between the actual and the most likely values

[c] These dihedral angles constrain the planarity of peptide bonds and rings as well as the chirality of the chiral carbon atoms

probability density function such that the model violates the input restraints as little as possible. This optimization procedure is done by the program using the variable target function method that applies the conjugate gradients algorithm to positions of all non-hydrogen atoms. A bundle of five models from random generation of the starting structure was calculated. It was shown that no significant improvement of the target function is obtained if more models are calculated. [34] The representative model was that with the lowest MODELLER target function value (3016).

Model refinement

All models obtained were subjected to a short simulated annealing refinement protocol available in MODELLER. For further model refinement of the representative model, the SYBYL package [24] was used. Hydrogen atoms were added and the model of the generated structure was minimized to convergence of the energy gradient of less than 0.01 kcal mol^{-1} $Å^{-1}$ using the TRIPOS force field included in the SYBYL/MAXIMIN2 module. [25] The minimization included electrostatic interactions based on Gasteiger–Hückel partial charge distributions using a dielectric constant with a distance-dependent function $\varepsilon=4r$ and a non-bonded interaction cutoff of 8 Å. [37, 38]

Structure validation

The geometry of the minimized structure was inspected with the program PROSA. [39, 40] The method can be used to identify misfolded structures as well as faulty parts of structural models. PROSA calculates a score for the modeled structure that indicates the quality of the protein structure. A polyprotein was used for the *Z* score calculation that comprises 230 proteins of known structures with a total length of about 50,000 residues. The conformations of these proteins have a good stereochemistry and many features of native protein folds. The set of conformations derived from the polyprotein represents a sample of the conformation space of a given protein. The amino acid sequences of the loop models were combined with all conformations in the polyprotein and the energies were calculated. The *Z* score is derived from the resulting energy distributions.

Moreover, the method constructs an energy graph for the energetic architecture of the protein folds as a function of the amino acid sequence position. These energy graphs represent a detailed view of the energy distributions in protein folds from the residue interaction energies e_{ij}. The interaction energies e_{ij} (i, j=1,..., l) from the energy matrix E of a conformation, where the sequence length l corresponds to the dimension of the matrix. From E the interaction energy $e_i=\Sigma_j e_{ij}$ of a particular residue i is derived. When e_i is plotted as a function of i, an energy graph will result displaying the energy distribution of the sequence structure pair in terms of sequence position. In this energy graph, positive values point to strained sections of the chain and negative values correspond to stable parts of the molecule.

Finally, the tertiary structure models were checked with PROCHECK. [41] It produces a Ramachandran diagram and allows examination of various structural features such as bond lengths and angles, secondary structures and exposure of residues to the solvent.

Docking procedure

The 3D structure of adenosine triphosphate was built using the Builder module in SYBYL and geometrically optimized using the TRIPOS force field. It was prepared as MgATP with Gasteiger–Hückel partial charges. To allow flexibility of the ligand, all bonds, with the exception of those fixed in rings, were assigned as rotatable with the deftors program, included in AutoDock. [42] AutoDock [42] is a program suite for automated docking of flexible ligands to receptors. Initially the MgATP was placed in arbitrary positions close to six amino acids belonging to the probable binding sites published in the literature (Cys^{457}, [43] K^{480}, [44] Glu^{472}, [11] K^{501}, [12] Cys^{549}, [17] Lys^{691}, [45]). The positioning of the ligand in these arbitrary sites was done with the DOCK module included in SYBYL/MAXIMIN2 that calculates energies of interaction based on steric contributions from the TRIPOS force field and electrostatic contributions from any atomic charges present in the ligand. AutoDock requires precalculated grid maps, one of each atom type present in the ligand being docked. A grid map consists of a 3D lattice of regularly spaced points, centered on the region of interest, the arbitrary position of the ligand. AutoGrid, which is included in the AutoDock package, was used with 120 grid points in each direction and a grid point spacing of 0.2 Å. Each point within the grid map stores the potential energy experienced by a probe atom or functional group due to all the atoms in the macromolecule. In addition to the atomic affinity grid maps, AutoDock requires an electrostatic potential grid map, calculated by solvation of the linearized Poisson–Boltzmann equation. For the explicit modeling of hydrogen bonds it was necessary to add polar hydrogens and assign partial atomic charges to the macromolecule. For the docking the Lamarckian Genetic Algorithm (GA-LS), a hybrid search technique that implements an adaptive global optimizer with local search [46] was used. The global search method is a modified genetic algorithm, with 2-point crossover and random mutation. The genome consists of floating point genes each of which encodes one state variable describing the molecular position, orientation and conformation. By setting the rate of genetic crossover to zero, and increasing the rate of genetic mutation, this hybrid genetic algorithm can mimic an evolutionary programming method. The local search method is based on the optimization algorithm of Solis and Wets, [46] which modifies the phenotype. This is allowed to update the genotype and utilizes the Lamarckian notion

that an adaptation of an individual to its environment can be inherited by its offspring. Variances are used for probabilistically determining the change to a particular state variable. The relative magnitudes of translation and rotation are also taken into account in these variances. The exploration of docking positions included for every arbitrary position 50 hybrid GA-LS docking runs using a population size of 50, a maximum number of energy evaluations of 25,000, a maximum number of generations of 27,000 and 300 iterations of Solis and Wets' local search. The resulting positions were clustered according to an rms criterion of 1 Å and the most energetically favorable position of every cluster was analyzed visually. The selected docking positions were subjected to 2000 steps of steepest descent minimization, optimizing both internal and relative geometries of the substrate and the binding site residues.

Calculation of binding energies

The non-bonded interaction energy between the loop-model and the ATP within the optimized complex was calculated using the TRIPOS force field. This estimation of real interaction energy neglects solvation and desolvation effects.

Results and discussion

Sequence similarity and confidence in the alignment

The quality of the alignment can be seen as the most important step in homology modeling. Therefore, the degree of similarity between the target sequence and the template and the reliability of the alignment are the most critical problems. These two problems are of course partially interconnected, since the degree of similarity of two structures decreases with the degree of sequence identity. [47] In our case the pairwise identity with Ca^{2+}-ATPase was 32.8% and the similarity was 53.3%. Similarity in this case included not only identical amino acids, but also indicated that amino acids of the stronger groups were conserved. Stronger groups are: CSTA, NEQK, NHQK, NDEQ, QHRK, MILF, HY, FYW. These amino acids should conserve the structure and are marked in the alignment with two stars. For such a degree of similarity, alignment errors were possible. [48] One basis of homology modeling is the assumption that it is possible to define a unique optimal sequence-based alignment that coincides with a structure-based alignment. This is not true in general because every alignment program tries to maximize the number of alignable residues, although these residues might not be spatially superposable. This limitation and source of error is intrinsic and should always be taken into account when estimating the degree of confidence of a certain model. We tried to minimize this error by adjusting the alignment by hand when necessary, especially taking into account the regions of proposed functionally key residues. The alignment used for further modeling is shown in Fig.1. A number of experiments with ATP analogs have shown the existence of an SH group within the ATP binding site [11, 13, 14, 15, 16] and indicated this cystein as Cys^{549}. [17] Although the presence of a sulfhydryl group was also shown in the ATP binding site of Ca^{2+}-ATPase, [14] both cysteins are in non-conserved regions and cannot be properly aligned to each other.

Modeling and model validation

Energy minimization of a model is necessary in order to relieve short contacts and to rectify bad geometry that may be present in the model. Inconsistencies are particularly possible in regions where a loop is melded with the core of the protein. Despite an rmsd value of 0.74 Å between the unminimized post-MODELLER structure and the final model, there were a number of reasons to justify this approach. The post-MODELLER model had several amino acids not properly modeled, five were even in disallowed regions of the Ramachandran plot, but especially there were shortcomings at side-chains. The correctness of side-chains is necessary for further docking and therefore a higher rmsd value has been accepted. In order to have a more realistic model, it is desirable to take into account the effect of the solvent. We have therefore chosen a distance-dependent dielectric constant equal to 4*r*, which was especially suitable for mimicking the solvent environment in the inner part of the protein. The use of a distant-dependent dielectric constant has become customary in molecular mechanics protein computations to account for the effect of the solvent implicitly. [49, 50, 51, 52] There are a number of reasons that justify this approach. An explicit solvation by soaking the protein with bulk water would be a computationally highly demanding method, which makes it unwieldy for medium-sized proteins. [50]

The Ramachandran plot of the predicted structure calculated with PROCHECK, shown in Fig.2, revealed a good quality stereochemistry, as indicated by the torsion angles Φ and Ψ. The Φ, Ψ torsion angles of 73% of the residues had values within the most favored areas and 24% of the residues had values within additionally allowed regions of the Ramachandran plot. No residues were found in disallowed regions. The overall *g* factor of the structures obtained showed a value of –0.38. The *g* factor should be above –0.5 and values below –1.0 may need investigation.

The energy graph calculated with PROSA from the modeled structure of the H_4–H_5 loop of the α subunit of Na^+/K^+-ATPase is shown in Fig.3. In general, the energy graphs of the observed sequence structure pairs have negative values that correspond to stable parts of these molecules. The *Z* score of this structure was –8.54. *Z* scores indicate the quality of protein structures. The scores of native folds are in a characteristic range and they depend on the sequence length. For a sequence

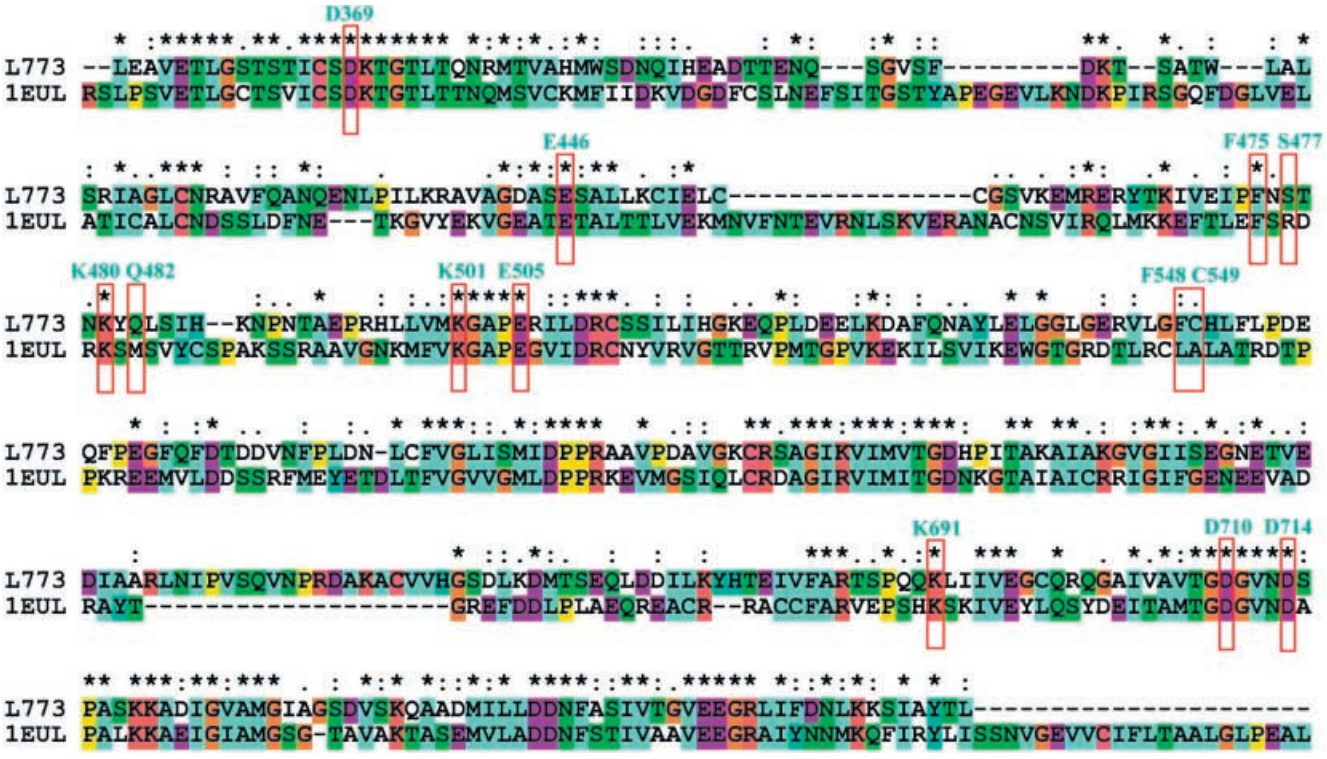

Fig. 1 Sequence alignment of the large cytoplasmic loop of the α subunit of pig-kidney Na^+/K^+-ATPase with Ca^{2+}-ATPase, whose tertiary structure has been solved recently at 2.6-Å resolution by X-ray crystallography and for which the PDB coordinates were available. [19] The structure (1EUL) was extracted from the Brookhaven Protein Data Bank. [22] Sequences are colored on the basis of an automatic calculated alignment consensus

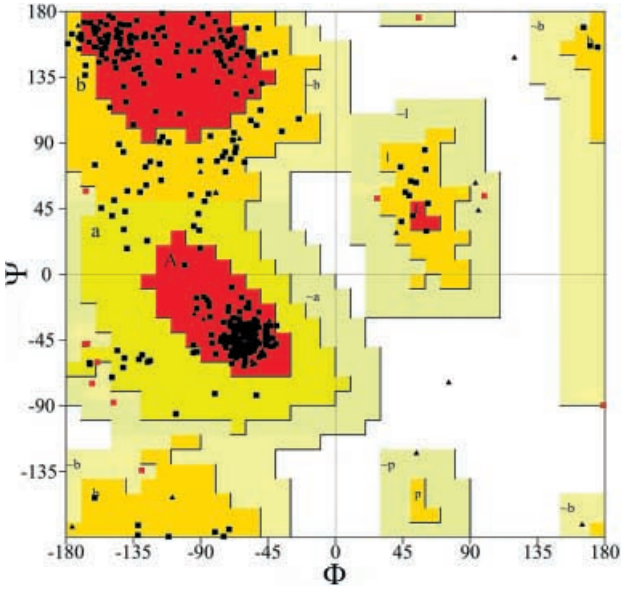

Fig. 2 Ramachandran plot of the predicted structure of the large cytoplasmic loop of the α subunit of Na^+/K^+-ATPase. The good stereochemical quality is shown by the presence of 73% of the residues in the most favored regions. In *red* are shown the most favored regions, *bright yellow* are shown additionally allowed regions and *light yellow* are shown generously allowed regions. Disallowed regions are *white*. Glycin residues are shown as *triangles*

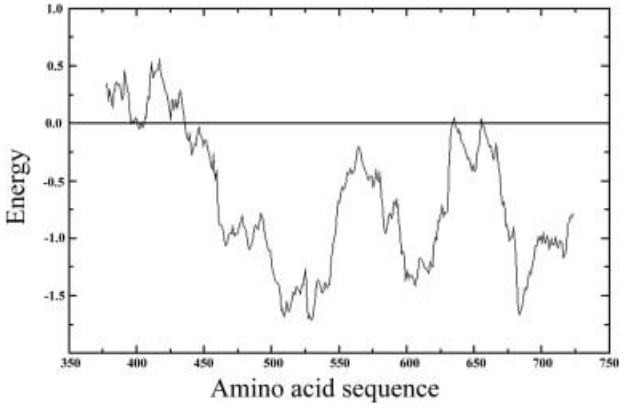

Fig. 3 Energy graph of the tertiary structure model of the large cytoplasmic loop. The graph is smoothened by a window size of 50 residues. In this energy graph negative values correspond to stable parts of the molecule

length of 420 amino acids, the value should be within a range from –8 to –10. The *Z* score of our structure was slightly lower but still in very good agreement compared with the X-ray structure of the large cytoplasmic loop of Ca^{2+}-ATPase, which has a *Z* score of –9.55. The validity of the model is supported by a negative energy graph and the correlation of this energy graph with the template structure. We can establish that the structure corresponds to stable parts of this molecule.

Both the energy graph and the results from the PROCHECK procedure support the principle correctness of our model structure.

Fig. 4 The large cytoplasmic loop of the α subunit of Na^+/K^+-ATPase consists of two well-separated domains. Domain P comprises the N-terminal and C-terminal ending and contains the phosphorylation site. Domain N binds nucleotides, and has only one binding site for ATP

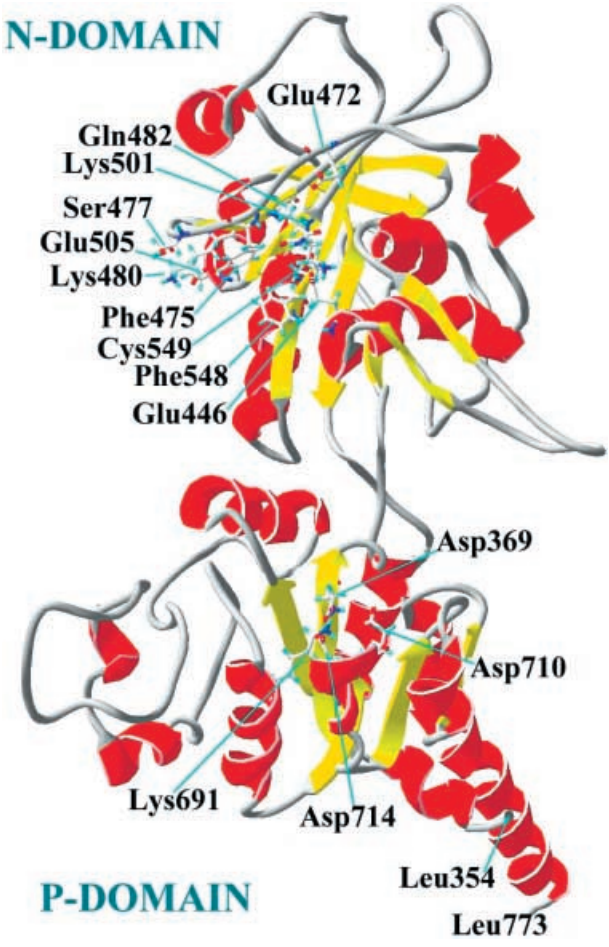

Fig. 5 Three-dimensional representation of the structure of the large cytoplasmic loop of the α subunit of Na^+/K^+-ATPase from Leu^{354} to Leu^{773}. The picture shows the phosphorylation site Asp^{369}, Lys^{691}, Asp^{710} and Asp^{714} on the P-domain and the ATP binding site comprising Cys^{549}, Phe^{548}, Glu^{505}, Lys^{501}, Gln^{482}, Lys^{480}, Ser^{477}, Phe^{475} and Glu^{446} on the N-domain. Glu^{472} is a part of the structurally important β strand directly above the binding pocket

Function and structure of the large cytoplasmic loop

The large cytoplasmic loop of the α subunit of Na^+/K^+-ATPase consists of two well-separated domains (Fig.4). The domain closer to the membrane contains the Asp^{369} residue of phosphorylation and therefore we call it domain P. This domain is composed of the N-terminal part, which is connected to the fourth transmembrane segment H_4 in the complete enzyme, and the C-terminal part, which is connected to the fifth transmembrane segment H_5 in the complete enzyme. These two parts form a typical Rossman fold as predicted by sequence homology, [18] and the secondary structure of this domain can be divided into a seven-stranded parallel β sheet with eight short associated helices (Fig.5). Asp^{369}, the residue of phosphorylation, is situated in the C-terminal end of the central β strand. Thus, Asp^{369} is the residue of phosphorylation as shown previously experimentally. [53] This is a typical position for nucleotide-binding proteins containing a Rossman fold, also observed with the Ca^{2+}-ATPase [19] and conserved in the alignment. Around the residue of phosphorylation, we observe a highly negatively charged surface that was accessible to solvent (Fig. 5). The large haloacid dehalogenase superfamily of hydrolases comprises P-type ATPases, phosphatases, epoxide hydrolases and L-2 haloacid dehalogenases. There are identified highly conserved motifs playing a key role in phosphorylation in a multiple sequence alignment of members of this family: [54] Asp^{369}, Lys^{691}, Asp^{710} and Asp^{714} (Fig.6). A mechanism of ATPase and phosphate activity is proposed: [54] A magnesium ion is coordinated by oxygen atoms from the three Asp residues. One of the carboxylate oxygen atoms of Asp^{369}, mediated by the Mg^{2+} ion, performs a nucleophilic attack on the phosphorus atom of the terminal phosphate of ATP. The ADP is cleaved off and an acyl phosphate intermediate is formed. The other carboxylate oxygen could be held in position by the Mg^{2+} ion. In the next step, the acyl phosphate intermediate is hydrolyzed by nucleophilic attack of a water molecule on the phosphorus atom to yield phosphate and free enzyme. The positive charge of the Lys^{691} (Fig. 6) may be involved in binding the terminal phosphate and supply charge shielding of the negatively charged phosphate group. In addition, it might also be involved in the activation of the attacking nucleophile and the stabilization of the leaving group.

The second domain binds nucleotides and therefore we call it domain N. It is larger than domain P and has $M_r \approx 24$ K. This domain is formed by the middle part of the loop sequence (roughly Arg^{378}–Arg^{589}). Its secondary structure showed a seven-stranded antiparallel β sheet with two helix bundles sandwiching it. On this domain Phe^{548}, Glu^{505}, Lys^{501}, Gln^{482}, Lys^{480}, Ser^{477}, Phe^{475} and Glu^{446} formed a positively charged binding pocket to be able to dock MgATP (Figs. 7 and 8). The attempts to

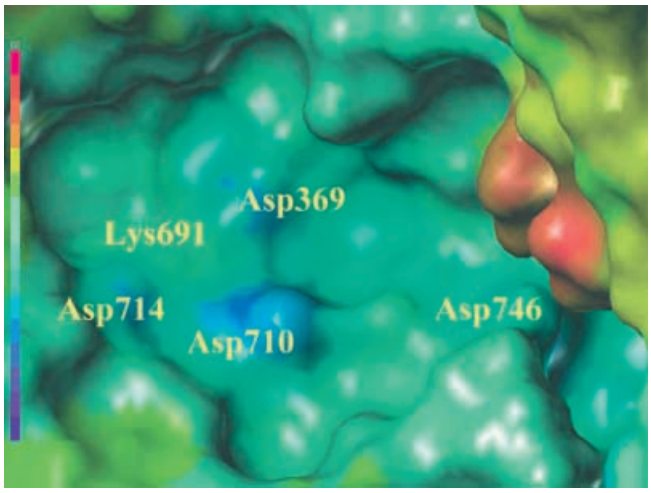

Fig. 6 The negative charged cavity around the residue of phosphorylation Asp ^{369} on a Conolly-type surface, representing the surface accessibility, with the electrostatic potential as a property. *Blue* color stands for negative charged areas and *red* color for positive charged areas on the solvent accessible surface. The three residues Lys^{691}, Asp^{710} and Asp^{714} playing probably a major role in phosphorylation are highly conserved

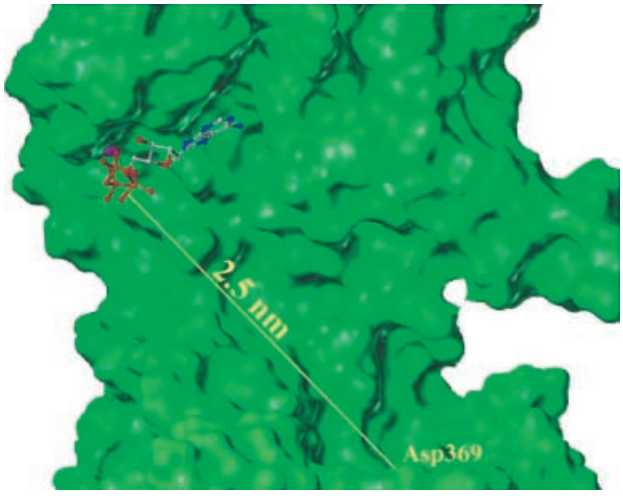

Fig. 7 ATP docked into the positively charged binding pocket on the N-domain. The ATP bound on the N-domain is located in a distance about 25 Å far from the phosphorylation site

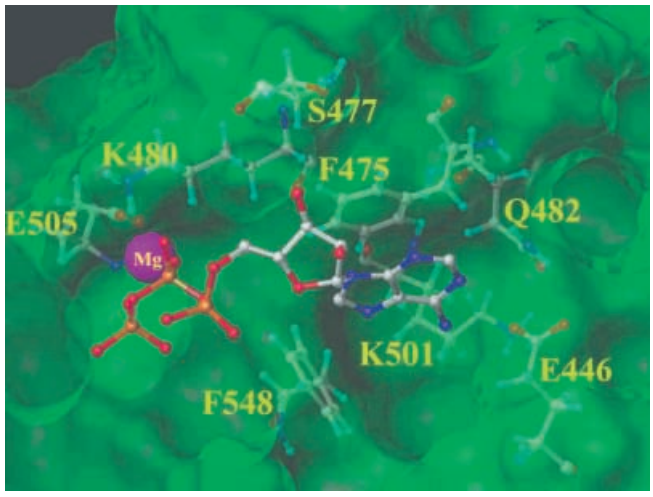

Fig. 8 Phe^{548}, Glu^{505}, Lys^{501}, Gln^{482}, Lys^{480}, Ser^{477}, Phe^{475} and Glu^{446} on the N-domain of the large cytoplasmic loop of the α subunit of Na^+/K^+-ATPase are taking part in MgATP binding. The magnesium ion, Mg^{2+}, is coordinated to two phosphate oxygen atoms, one each from the γ and β phosphates. Oxygen atoms from the side-chain of Glu^{505} anchor the magnesium ion to the protein. Glu^{446} forms a hydrogen bond to the NH_2 hydrogen donor of the adenine moiety of ATP. The side-chain of Glu^{446} is stabilized by hydrogen bonding with the Lys^{501} and Gln^{482} side-chains

dock MgATP into other binding sites proposed in the literature were not successful, although we had around every arbitrary position a grid map with 24 Å in every direction. Our results thus support the hypothesis of only one ATP binding site on each H_4–H_5 loop of the α subunit of Na^+/K^+-ATPase. However, the presence of another binding site in a different conformation of the protein cannot be excluded. Since entropy contributions are not considered in the docking procedure, the calculated interaction energies represented only an estimation of the interaction enthalpy. Nevertheless more negative interaction energy should correlate with a higher affinity of the ligand to the substrate. The calculated non-bonded interaction energy ΔE of ATP with the H_4–H_5 loop of the α subunit of Na^+/K^+-ATPase was –156.9 kcal mol^{-1}. The K_d measured for MgATP on the native enzyme was for the E_1 conformation 0.3 µM and for the E_2 conformation 120 µM. [55] Despite the fact that the calculated interaction energy could hardly be compared directly with the K_d values, we have obtained a strong binding of MgATP and therefore our structure was probably in an E_1-like conformation. Almost all nucleotide triphosphate hydrolyzing enzymes require magnesium ions. Magnesium is presumably required in these enzymes both for proper positioning of the phosphate groups and for weakening the P–O bond that is split during catalysis. A magnesium ion, Mg^{2+}, is coordinated to two phosphate oxygen atoms, one each from the γ and β phosphates. Oxygen atoms from the side-chain of Glu^{505} anchored the magnesium ion to the protein. The side-chain of Lys^{480} that was located in a short loop connecting the β strands, formed an ion-pair interaction with the β-phosphate group. The aromatic ring of ATP was located near Phe^{475}, which seemed to be a critical residue for ATP binding. Residue Glu^{446} formed a hydrogen bond to the NH_2 hydrogen donor of the adenine moiety of ATP. The side-chain of Glu^{446} was stabilized by hydrogen bonding with the Lys^{501} and Gln^{482} side-chains. Thus, Lys^{501} was located in the depth of the binding pocket at a distance of around 5 Å to the adenosine moiety, which also explains the inhibition of ATP binding by chemical modification of this residue with fluorescein 5-isothiocyanate [56] (Fig.8). Glu^{472}, a structurally conserved amino acid in several P-type ATPases which has been shown to be essential for the enzyme activity of the sodium pump, [11] seems not

to interact during binding of ATP. However, as a part of the structurally important β strand directly above the binding pocket, it probably plays an important role in phosphorylation, which has to be an extraordinarily dynamic process to overcome the distance of over 20 Å between the binding pocket and the residue of phosphorylation. Cys^{549}, which has been shown by the use of erythrosin 5-isothiocyanate to be part of the E_2ATP site, [17] lies in the binding pocket just beneath the solvent-accessible surface behind Phe^{548}, and the sulfhydryl group is accessible by induced fit.

Conclusion

On the basis of the recently solved tertiary structure of Ca^{2+}-ATPase, the complete 3D structure of the large cytoplasmic loop between the fourth and fifth transmembrane segment (H_4–H_5 loop) of the α subunit of Na^+/K^+-ATPase, beginning with Lys^{354} and ending with Lys^{773}, was modeled using the method of restraint-based comparative modeling. Due to the relatively high degree of homology of more than 50% with respect to the cytoplasmic loop of Ca^{2+}-ATPase and the sufficient amount of background information, contributed by different labeling techniques of certain residues and kinetic studies, it was possible to create a model in which all amino acids were modeled convincingly. Molecular modeling of protein structures is becoming of increasing importance in predicting the folds of proteins for which homology with known templates can be reliably identified. We have shown that the ATP binding site and the phosphorylation site are located on two different, well-separated domains, which together form the large cytoplasmic loop (Fig. 4). For these domains we propose to use the following names: domain P comprises the N-terminal and C-terminal ending of the H_4–H_5 loop and contains Asp^{369}, the residue of phosphorylation surrounded by a highly negatively charged surface; and domain N, which binds nucleotides. These names were recently proposed by Toyoshima et al. [19] for the analogical loop of Ca^{2+}-ATPase and we consider it useful to have a common nomenclature within the family of P-type ATPases. We show that there is evidence for only one ATP binding site on the N domain of the H_4–H_5 loop, and we were able to specify Cys^{549}, Phe^{548}, Glu^{505}, Lys^{501}, Gln^{482}, Lys^{480}, Ser^{477}, Phe^{475} and Glu^{446} as parts of the ATP binding site with Lys^{501} located in the depth of the positively charged binding pocket. According to Linnertz et al. [57] the microenvironment of the high affinity ATP binding site is slightly acidic, which correlates to the positive charge in parts of the pocket of our model. Even though we might still be far from replacing experimental structure determinations with calculated model structures, we believe that the model presented here is of general validity and especially helpful for the expression of shortened constructs and site-directed mutagenesis. At this time we know neither the crystal structure of the sodium pump, nor the exact pumping mechanism. We thus believe that this calculated model of the large cytoplasmic loop is a valuable step towards the understanding of the sodium pump and will be a source of inspiration for further studies.

Supplementary material Model of the H_4–H_5 loop of the α subunit of Na^+/K^+-ATPase in PDB format (H4H5loop.pdb)

Acknowledgements This work was supported by the German and Czech Governments by IWTZ TSR-088-97, by the Ministry of Education of Czech Republic (VS 961410), by the Volkswagen Foundation (I/74 679) and by grants No. 204/98/0468 of the GACR, No. A7011801 of GACAS, No. 204/98/0416 of the GACR.

References

1. Jørgensen PL (1986) Kidney Int 29:10
2. Lingrel JB, Kuntzweiler TJ (1994) Biol Chem 269:19659
3. Albers RW (1967) Annu Rev Biochem 3:727
4. Post RL, Kume L, Tobin T, Orcutt HB, Sen AK (1969) J Gen Physiol 54:306
5. Thoenges D, Schoner W (1997) J Biol Chem 272:16315
6. Schoner W, Thönges D, Hamer E, Antolovic R, Buxbaum E, Willeke M, Serpersu EH, Scheiner-Bobis G (1994) In: Bamberg E, Schoner W (eds) The Sodium Pump. Springer, New York, pp 332–341
7. Linnertz H, Urbanova P, Obsil T, Herman P, Amler E, Schoner W (1998) J Biol Chem 273:28813
8. Ward DG, Carvieres JD (1996) J Biol Chem 271:12317
9. Ovchinnikov YA, Dzandzugazyan KN, Lutsenko SV, Mustayev AA, Mondyanov NN (1987) FEBS Lett 217:111
10. Tran CM, Houston EE, Farley RA (1994) J Biol Chem 269:10260
11. Scheiner-Bobis G, Schreiber S (1999) Biochemistry 38:9198
12. Pedemonte CH, Kaplan JH (1990) Am J Physiol 258:C1-C23
13. Patzelt-Wenczler R, Schoner W (1975) Biochim Biophys Acta 403:538
14. Patzelt-Wenczler R, Kreickmann H, Schoner W (1980) Eur J Biochem 109:167
15. Patzelt-Wenczler R, Schoner W (1981) Eur J Biochem 114:79
16. Scheiner-Bobis G, Mertens W, Willeke M, Schoner W (1992) Biochemistry 31:2107
17. Linnertz H, Kost H, Obsil T, Kotyk A, Amler E, Schoner W (1998) FEBS Lett 441:103
18. Aravind L, Galperin MY, Koonin EV (1998) Trends Biol Sci 23:127
19. Toyoshima C, Nakasako M, Nomura H, Ogawa H (2000) Nature 405:647
20. Gatto C, Wang AX, Kaplan JH (1998) J Biol Chem 273:10578
21. Tran CM, Farley RA (1999) Biophys J 77:258
22. Obsil T, Merola F, Lewit-Bentley A, Amler E (1998) FEBS Lett 426:297
23. http://www.rcsb.org/pdb/
24. TRIPOS Associates Inc, 1699 S Hanley Road, Suite 303, St. Louis, MO 63144
25. Ovchinnikov YA, Modyanov NN, Broude NE, Petrukhin KE, Grishin AV, Arzamazova NM, Aldanova NA, Monastyrskaya GS, Sverdlov ED (1986) FEBS Lett 201:237
26. Ovchinnikov YA, Arsenyan SG, Broude NE, Petrukhin KE, Grishin AV, Aldanova NA, Arzamazova NM, Aristarkhova EA, Melkov AM, Smirnov YV, Gur'Ev SO, Monastyrskaya GS, Modyanov NN (1986) Dokl Biochem 285:403
27. Thompson JD, Gibson TJ, Plewniak F, Jeanmougin F, Higgins DG (1997) Nucl Acids Res 25:4876
28. Guex N, Peitsch MC (1997) Electrophoresis 18:2714
29. Huang X, Miller W (1991) Adv Appl Math 12:373
30. Rost B, Sander CJ (1993) Mol Biol 232:584
31. Sali A, Blundell TL (1993) J Mol Biol 234:779

32. Sali A, Overington, JP (1994) Protein Sci 3:1582
33. Sali A (1995) Mol Med Today 1:270
34. Sali A, Potterton L, Yuan F, van Vlijmen H, Karplus M (1995) Proteins: Struct Funct Genet 23:318
35. Sanchez R, Sali A (1997) Proteins: Struct Funct Genet Suppl 1:50
36. Brooks BR, Bruccoleri RE, Olafson BD, States DJ, Swaminathan S, Karplus M (1983) J Comput Chem 4:187
37. Clark M, Cramer III RD, Van Opdenbosch NJ (1989) J Comput Chem 10:982
38. Gasteiger J, Marsili M (1980) Tetrahedron 36:3219
39. Sippl MJ (1993) Proteins 17:355
40. Sippl MJ (1993) J Comput-Aided Mol Design 7:473
41. Laskowski RA, McArthur MW, Moss DS, Thornton JM (1993) J Appl Crystallogr 26:283
42. Morris GM, Goodsell D, Huey R, Olson AJ (1996) J Comput-Aided Mol Design 10:293
43. McInstosh D (1998) Adv Mol Cell Biol 23A:33
44. Ward DG, Carvieres JD (1998) J Biol Chem 273:19277
45. Stokes DL, Green NM (2000) Biophys J 78:1765
46. Solis FJ, Wets RJ (1981) Math Oper Res 6:19
47. Chothia C, Lesk AM (1986) EMBO J 5:823
48. Venclovas C, Zemla A, Fidelis K, Mould J (1997) Proteins: Struct Funct Genet Suppl 1:7
49. Koymans LMH, Vermeulen NPE, Baarslag A, Donne-Op den Kelder GM (1993) J Comput-Aided Mol Design 7:281
50. Arnold GE, Ornstein RL (1994) Proteins: Struct Funct Genet 18:19
51. Guenot JM, Kollman PA (1992) Protein Sci 1:1185
52. Guenot JM, Kollman PA (1993) J Comput Chem 14:295
53. Ellies-Davies GC, Kaplan JH (1990) J Biol Chem 265:20570
54. Ridder IS, Bauke WD (1999) Biochem J 339:223
55. Thoenges D, Amler E, Eckert T, Schoner W (1999) J Biol Chem 274:1971
56. Abott AJ, Amler E, Ball jr. WJ (1991) Biochemistry 30:1692
57. Linnertz H, Lanz E, Gregor M, Antolovic R, Krumscheid R, Obsil T, Slavik J, Kovarik Z, Schoner W, Amler E (1999) Biochem Biophys Res Commun 254:215