Molecular modelling of the nucleotide-binding domain of Wilson's disease protein: location of the ATP-binding site, domain dynamics and potential effects of the major disease mutations

Roman G. EFREMOV^{*1}, Yuri A. KOSINSKY^{*}, Dmitry E. NOLDE^{*}, Ruslan TSIVKOVSKI[†], Alexander S. ARSENIYEV^{*} and Svetlana LUTSENKO[†]

^{*}M. M. Shemyakin & Yu. A. Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Ul. Miklukho-Maklaya, 16/10, Moscow V-437, 117997 GSP, Russia, and [†]Department of Biochemistry and Molecular Biology, Oregon Health and Science University, Portland, OR 97239, U.S.A.

W NDP (Wilson's disease protein) is a copper-transporting ATPase that plays an essential role in human physiology. Mutations in W NDP result in copper accumulation in tissues and cause a severe hepato-neurological disorder known as Wilson's disease. Several mutations were surmised to affect the nucleotide binding and hydrolysis by W NDP; however, how the nucleotides bind to normal and mutated W NDP remains unknown. To aid such studies, we performed the molecular modelling of the spatial structure and dynamics of the ATP-binding domain of W NDP and its interactions with ATP. The three-dimensional models of this domain in two conformations were built using the X-ray structures of the Ca²⁺-ATPase in the E1 and E2 states. To study the functional aspects of the models, they were subjected to long-term molecular dynamics simulations in an explicit solvent; similar calculations were performed for the ATP-binding domain of Ca²⁺-ATPase. In

both cases, we found large-scale motions that lead to significant changes of distances between several functionally important residues. The ATP docking revealed two possible modes of ATP binding: via adenosine buried in the cleft near residues H1069, R1151 and D1164, and via phosphate moiety 'anchored' by H-bonds with residues in the vicinity of catalytic D1027. Furthermore, interaction of ATP with both sites occurs if they are spatially close to each other. This may be achieved after relative domain motions of the 'closure' type observed in molecular dynamics simulations. The results provide a framework for analysis of disease mutations and for future mutagenesis studies.

Key words: ATP7B, ATP-binding site, molecular dynamics, molecular modelling, P-type ATPase, Wilson's disease protein.

INTRODUCTION

W NDP (Wilson's disease protein) is a human P-type ATPase that uses the energy of ATP hydrolysis to deliver copper into the secretory pathway and to export excess copper out of the cell [1]. Mutations in Wilson's disease gene, *ATP7B*, result in massive accumulation of copper in a number of tissues, particularly in the liver and brain. Copper overload induces severe pathological changes, including liver disease, neurological problems and psychiatric abnormalities [2]. In recent years, numerous mutations in *ATP7B* have been identified [3–7], but for a vast majority of the disease-causing mutations, their effects on structure, function and regulation of W NDP remain unknown.

W NDP belongs to the large family of the P-type ATPases. During their catalytic cycle, the P-type ATPases become transiently phosphorylated at the invariant aspartic acid residue in the sequence motif DKTG located in the ATP-BD (ATP-binding domain) of the protein. The structural and functional properties of several members of this family, such as Ca²⁺-ATPase of SR (sarcoplasmic reticulum) and plasma membrane Na⁺,K⁺-ATPase have been studied fairly extensively. Recently, the high-resolution structure of SR Ca²⁺-ATPase has been determined in two conformational states (the calcium-bound form, E1 [8], and the calcium-empty form in a complex with thapsigargin, E2 [9], providing the structural basis for understanding the mechanism of the ATP-driven ion transport.

Ca²⁺-ATPase is composed of three major domains. The TM (transmembrane) domain consists of 10 TM segments and forms sites for binding and translocation of ions. It is linked to a cytosolic ATP-BD, which contains the catalytic site and to an actuator domain, which plays an important role in conformational transitions of Ca²⁺-ATPase. In turn, ATP-BD was shown to consist of two structural units connected by a linker: the P-domain, which houses the DKTG motif, a site of catalytic phosphorylation and the N-domain, which contributes to the binding of nucleotides [8]. The structures also suggested that domain motion played a central role in the catalytic activity of P-type ATPases [8,9]. At present, it remains unclear whether binding of ATP to the N-domain induces movement of the N-domain towards the P-domain, resulting in the formation of the high-affinity site for ATP, and initiation of catalysis or such a movement is the result of large-scale thermal motions independent of ATP binding. Our work addresses this important issue.

In contrast with Ca²⁺-ATPase, very little is known about the structure and mechanistic properties of W NDP. Our previous studies [10] demonstrated that W NDP undergoes catalytic phosphorylation at D1027 in the DKTG motif, as expected for the P-type ATPases. We also found that the mutation H1069Q (the most frequent mutation in patients suffering from Wilson's disease) disrupts catalytic phosphorylation from ATP, suggesting that H1069 plays an important role in ATP co-ordination [11]. However, it remains unclear whether H1069 is the residue

Abbreviations used: ATP-BD, ATP-binding domain; 3D, three-dimensional; MD, molecular dynamics; RMSD, root-mean-square deviation; RMSF, root-mean-square fluctuation; SR, sarcoplasmic reticulum; TM, transmembrane; W NDP, Wilson's disease protein; for brevity single-letter code has been used for amino acids, e.g. H1069 stands for His-1069.

¹ To whom correspondence should be addressed (email efremov@nmr.ru).

co-ordinating the nucleotide or whether its effect on the ATP binding is indirect. Similarly, several disease-causing mutations were identified in the ATP-BD of WNDP. Whereas for some of them the effect on WNDP function can be predicted (e.g. the mutation T103I adjacent to the DKTG motif will probably disrupt catalytic phosphorylation), the specific consequences of other mutations are unclear. Therefore one of the important goals of this study was to generate a useful structural tool to aid analysis of WNDP mutants.

Another important goal was to identify the candidate residues involved in nucleotide binding in WNDP. Binding of ATP and catalytic phosphorylation are the key steps in WNDP-mediated transport of copper. However, currently, very little information is available on how these steps occur. Furthermore, the residues known to be involved in ATP binding in Ca^{2+} -ATPase and Na^+ , K^+ -ATPase are not conserved in the structure of WNDP, suggesting that the organization of the ATP-binding site in WNDP is distinct. In recent years, computational approaches have been increasingly useful in providing important information about the structure and functional activity of a variety of proteins. Therefore, to understand better the structure and function of the ATP-BD of WNDP, we employed a computational approach, which combines homology modelling, MD (molecular dynamics) and docking simulations.

MATERIALS AND METHODS

Model building and verification

The secondary structure of WNDP ATP-BD (residues M996–R1322) was predicted using the PHD method [12]. (PHD is a suite of programs predicting protein's secondary structure and solvent accessibility from multiple sequence alignments.) These predictions and threading algorithms were utilized to search the database of high-resolution structures and identify proteins with similar fold. The mGenTHREADER program [13] running on the PSIPRED server (<http://bioinf.cs.ucl.ac.uk/psipred/>) detected the best fit to the 3D (three-dimensional) structure of Ca^{2+} -ATPase in the E1 state, residues A320–K758 (Protein Data Bank entry 1EUL [8]). The fold 1EUL was recognized with the E value of 8×10^{-5} (the model is considered 'certain' if $E < 1 \times 10^{-3}$ [13]). The generated alignment (Figure 1) served as a basis for building the model of WNDP ATP-BD. Ten models with slightly different conformations of the loop regions were generated using the Modeller software [14]. Their quality was verified with the Profiles_3D program [15], and the best-scoring model was then used in MD simulations. A similar approach was employed to build a model of WNDP ATP-BD based on the 3D structure of Ca^{2+} -ATPase in the E2 state (Protein Data Bank entry 1IWO [9]).

The quality of 3D models of WNDP ATP-BD was evaluated by calculating their total (for the entire molecule) and local (for residue i) 3D–1D compatibility scores, i.e. S and S_i respectively. This was done using the Profiles_3D program [15]. The S_i value characterizes how well a given model satisfies general principles of folding observed in high-resolution structures of globular proteins [15]. The overall quality of the models was assessed by comparison of their S values with S_{cor} , the so-called 'correct score' for a protein of a given size. (S_{cor} is obtained from statistical analysis of the high-resolution 3D structures.) If $S/S_{\text{cor}} < 0.45$, the model was treated as incorrect [15].

The 3D-profile algorithm was also employed to examine whether the constructed spatial models of WNDP ATP-BD satisfy general principles of folding for globular proteins. The generated atomic co-ordinates of WNDP ATP-BD in the E1 and E2 states were used to calculate the so-called 3D-profiles for the entire WNDP ATP-BD and its two subdomains. The profiles were generated using the 3D–1D table of Bowie et al. [16]. The resulting profiles were employed in the compatibility search against the Swiss-Prot sequence database (release 41, 2003) containing approx. 1.22×10^5 entries. The alignments of a given 3D profile with the sequences found by this profile in the database were considered significant if their Z scores were > 10 and the fitted length was more than 100 residues.

Spatial distribution of the electrostatic potential on the solvent-accessible surface of WNDP was obtained from finite-difference solutions to the Poisson–Boltzmann equation. This was done using the DelPhi program [17]. The dielectric constants of 2 and 80 were assigned to the protein interior and the surrounding solvent respectively.

MD simulations

MD of WNDP ATP-BD models and ATP-BD of Ca^{2+} -ATPase (residues 320–758) in the E1 and E2 states was simulated using the GROMACS program [18] and the GROMOS96 force field [19]. For Ca^{2+} -ATPase, the starting structures were taken from the Protein Data Bank (entries 1EUL and 1IWO for the E1 and E2 states respectively). Molecules with uncharged N- and C-termini were placed in rectangular boxes of a simple point charge (SPC) model of water molecules with edges of 10 Å (1 Å = 0.1 nm). The 3D periodic boundary conditions were imposed. The energy of the system was minimized using the 300 steepest descent iterations with fixed protein atoms followed by 300 conjugate-gradient steps with fixed backbone, and final relaxation (300 conjugate gradient steps) without constraints. Then the system was subjected to a 10 ps MD run in an NPT (constant pressure and temperature) ensemble with fixed protein atoms. Finally, it was heated from 5 to 300 K for 60 ps in an NVT (constant volume and temperature) ensemble followed by 5–6 ns of unrestrained NPT MD collection run at 300 K. Non-bonded interactions were truncated with the double cut-off of 10/18 Å.

The processing of MD trajectories (including essential dynamics analysis) was performed using the GROMACS software. The Dyndom program [20] was used to delineate dynamic domains and to define the parameters of their movements. Changes in mutual disposition of the domains were characterized by an angle Θ (Figure 2) between the vectors connecting the centre-of-mass of the hinge region (residues 358–362, 599–602 for Ca^{2+} -ATPase or 1033–1039, 1193–1198 for WNDP) with the centres-of-mass of the N-domain (residues 368–595 for Ca^{2+} -ATPase and 1045–1190 for WNDP) and the P-domain (residues 358–362, 599–602 for Ca^{2+} -ATPase and 1007–1030, 1200–1305 for WNDP). The highly flexible N- and C-terminal parts of the P-domain were excluded from these calculations.

Nucleotide docking

Docking simulations of the ATP molecule with the ATP-BD of Ca^{2+} -ATPase and with WNDP ATP-BD models were performed

calculated from the X-ray and NMR structures, (Protein Data Bank codes 1EUL and 1M08 respectively). Secondary structure of WNDP, as predicted by the 3D model (using DSSP software [32]) or with the mGenTHREADER [13] and PHD [12] programs, are also shown. Identical and iso-functional amino acids in the alignment are marked with symbols '=' and ':' respectively. Residues important for ATP binding are hatched in grey.

Figure 2 (A) Molecular model of the ATP-BD of WNDP in the open (light grey) and closed (dark grey) states

Models are built on the X-ray structures of Ca^{2+} -ATPase in E1 and E2 states respectively (Protein Data Bank entries 1EUL and 1IWO respectively). The molecules are given in a ribbon presentation and superimposed over their P-domains. ATP moiety at the top is shown to illustrate relative distances between the N- and P-domains in both states. Inset: definition of the angle Θ between the vectors connecting the centre of mass of the hinge region with centres of mass of the N- and P-domains. (B) The structures of the N-domains of the P-type ATPases. 1, WNDP (homology-based model; this study); 2, Na^+ , K^+ -ATPase (NMR model [23]; Protein Data Bank code 1M08); 3, Ca^{2+} -ATPase (X-ray structure [9]; Protein Data Bank code 1IWO). The molecules are given in a ribbon presentation. Several common structural elements in all three N-domains are in dark grey. They correspond to the following segments in WNDP 1054–1062 (segment $\alpha 1$ in Figure 1), 1096–1099 ($\beta 2$), 1105–1108 ($\beta 3$), 1126–1130 ($\alpha 3$), 1141–1154 ($\alpha 4$). Positions of the ATP molecule found in NMR experiments (2) and predicted by docking (3, this study) are indicated.

using the GOLD program [21], version 2.0 for two sets of 26 and 20 conformers respectively extracted from the equilibrium parts (1–3 ns) of MD trajectories with the interval 0.1 ns. To estimate the quality of the protein–ligand complexes, the scoring function GOLDScore [21] was employed. The centre of the docking sphere with radius 21 Å was placed on the atoms CG:F487 and CB:H1069 in Ca^{2+} -ATPase and WNDP respectively. This sphere covers approx. 30 and 40 % of the protein-accessible surface area in the E1 and E2 forms, including the putative nucleotide-binding and phosphorylation sites. The resulting complexes were analysed using the following parameters: (i) the GOLDScore value; (ii) the distance between the γ -phosphate of ATP and carboxy

group of D1027; (iii) the residues in close contact with ATP; and (iv) the protein–ligand H-bonds.

CD

The expression and purification of WNDP ATP-BD containing both the N- and the P-subdomains were described previously [11]. The individual N-subdomain was expressed in *Escherichia coli* and purified using the intein-mediated system (New England Biolabs, Beverly, MA, U.S.A.) [33]. CD spectra for purified proteins were recorded on an AVIV CD model 215 spectrometer in 50 mM sodium phosphate (pH 7.0) at a protein concentration of

Table 1 Comparison of the experimental and predicted secondary-structure content

For the results in columns 3, 4, 6 and 7, the secondary-structure content was calculated using the DSSP program [32].

	W NDP					
	CD data*	Initial 3D-model†	3D model, MD data‡	PHD prediction§	Ca ²⁺ -ATPase (X-ray)	Na ⁺ ,K ⁺ -ATPase (NMR)¶
N-domain						
α -helix	26	30	26 $\pm$ 1	35	30	25
β -sheet	20	28	17 $\pm$ 3	28	37	27
Entire ATP-BD						
α -helix	38	33	36 $\pm$ 1	40	35	NA
β -sheet	16	21	17 $\pm$ 2	22	27	
P-domain						
α -helix	49	45	46 $\pm$ 2	40	45	NA
β -sheet	20	22	15 $\pm$ 1	16	16	

* Results averaged over four CD-processing algorithms [22]. Details of CD spectra processing are given in the Materials and methods section.

† Calculated for the homology-built 3D model of W NDP-NB.

‡ Calculated for the set of equilibrated MD conformers (means $\pm$ S.D.).

§ Secondary-structure prediction using the PHD [12] algorithm.

|| PDB entry 1IWO.

¶ PDB entry 1M08.

0.4 mg/ml. The protein concentration was verified by the amino acid analysis of the corresponding samples and further used to calculate the molecular ellipticity. The secondary structure of W NDP ATP-BD and the N-domain were calculated using the sets of 37 and 43 reference protein spectra and the following algorithms of CD spectra deconvolution: CONTIN, SELCON3, CDSSTR and CDNN [22].

RESULTS AND DISCUSSION

The initial and challenging task in building the spatial model of W NDP ATP-BD (M996-R1322) was obtaining the alignment with the sequence of the corresponding region of SR Ca²⁺-ATPase (A320-K758). As shown in Figure 1, three W NDP segments, M996–A1065, D1185–I1236 and F1240–I1311, show homology to sequences in the ATP-BD of Ca²⁺-ATPase with 21, 33 and 47 % identity respectively. In the spatial structure of SR Ca²⁺-ATPase, these regions belong mainly to the P-domain, with the exception of two short segments, M1356–A1385 and D1585–L1595, which form a three-strand β -sheet layer connecting the N- and P-domains. The strong homology in the P-domain region is consistent with our earlier result showing that W NDP is phosphorylated at D1027, similar to Ca²⁺-ATPase and other P-type ATPases [10].

In contrast, the rest of W NDP ATP-BD sequence (residues S1066–I1184) does not show even distant sequence similarity to either the corresponding segment of Ca²⁺-ATPase (P391–T589) or to other proteins with known 3D structure. Consequently, to determine a putative spatial fold of this region of W NDP, we used the secondary-structure predictions and a number of threading algorithms. These experiments revealed that the predicted fold of the entire W NDP ATP-BD matches unequivocally to the fold of the Ca²⁺-ATPase ATP-BD in both conserved and non-conserved regions. The resulting alignment shown in Figure 1 was used to build the 3D models of W NDP ATP-BD.

Evaluation of the molecular models

The generated models reveal reasonably high *S* values, for all of them the ratio $S/S_{cor} > 0.68$. This considerably exceeds the 'confidence level' 0.45 [15], thus confirming the overall good quality of the models. For future considerations, two top-scoring

models (one in the open and another in the closed states) were selected. As expected, the quality of the models for the P-domain is better than for the N-domain, e.g. for the open model the ratios S/S_{cor} are 0.82 and 0.53 respectively. Analysis of the local 3D–1D scores for amino acid residues in the resulting models shows that the regions with relatively low *S_i* values are the following: 1090–1093, 1131–1132, 1139–1141, 1178–1179, 1244–1249. All these segments, except the last one, belong to the N-domain.

Another criterion for the validity of the models is similarity of their secondary structure to that observed in experiments. To make this comparison, we measured the CD spectra for the entire W NDP ATP-BD and for the isolated N-domain. The derived secondary-structure content was then compared with the results of PHD predictions and with the modelling data (Table 1). Assuming that the folding of the N-domain within W NDP ATP-BD is the same as folding of the isolated N-domain, the secondary-structure content of the P-domain could be derived using the CD data. As shown in Table 1, there is a fairly good agreement between the CD results, PHD predictions and modelling data for either the entire ATP-BD or the N-domain. Interestingly, the best fit to CD data is observed for W NDP models subjected to MD simulations (see below). Close values of the α -helical and β -sheet contents for the highly homologous P-domains of W NDP (CD-based data) and Ca²⁺-ATPase (the X-ray structure) strongly suggest that the predicted folding of W NDP ATP-BD and the N-domain closely resembles folding of these domains in native protein.

Finally, we utilized the 3D-profiles algorithm to verify whether the residue environments in the models are consistent with those found in proteins with known high-resolution structures. If the residue environment in a model resembles that in experimental structures and if a given profile efficiently recognizes its own amino acid sequence, then the model is seen as 'correct', and vice versa. Consequently, the profiles for the entire W NDP ATP-BD (E1 state) and the individual N- and P-domains, calculated based on their atomic co-ordinates, were used to screen the Swiss-Prot sequence database (see the Materials and methods section). In all cases, the profiles not only 'pulled out' the sequence of W NDP or its subdomains, but did this at the highest confidence level expressed in the highest Z-score values (Table 2). It is important to emphasize that the profiles do not contain sequence information, and any sequence compatible with the environmental classes in profiles may be a candidate for the best score. It is

Table 2 Results of screening the Swiss-Prot database with environmental profiles

Protein part used to build the profile*	Best-scoring sequence†	Z score
Ca-ATPase		
(N + P)-domain	ata1_ranes	85.1
P-domain	ata1_ranes	39.3
Ca-ATPase, N-domain	ata1_ranes	49.9
Na,K-ATPase, N-domain	ata1_oremo	30.6
WNDP		
(N + P)-domain	at7b_human	48.4
P-domain	at7b_human	27.6
N-domain	at7b_human	13.3

* Ca²⁺-ATPase and WNDP were analysed in the open state.

† Swiss-Prot names of the recognized sequences: ata1_ranes, sarcoplasmic reticulum Ca²⁺-ATPase 1; ata1_oremo, Na⁺,K⁺-ATPase α -1 chain; at7b_human, human copper-transporting ATPase (WNDP).

also important that the resulting sequence–profile alignments are 100 % compatible (i.e. the environmental class of *i* residue in the profile corresponds to *i* residue in the sequence).

It should be noted that, among the best Z scores for the profile–sequence alignment, the highest value was observed for the P-domain, whereas the values for the N-domain were lower (Table 2). This seems reasonable because the best sequence homology between WNDP and Ca²⁺-ATPase is observed for the P-domain and, therefore, its modelling is more certain. Although due to low sequence homology of the N-domains the modelling of this region is less certain, good results obtained in the profile–sequence compatibility search indicate that the 3D model of this part of WNDP ATP-BD is acceptable. Taken together, the results demonstrate the overall correctness of the constructed models of WNDP ATP-BD.

WNDP ATP-BD model

The generated model is shown in Figure 2(A). Similar to the template, WNDP ATP-BD is formed by two subdomains, the P-domain with the $\alpha\beta\alpha\beta\beta\alpha\beta\alpha\beta\alpha\beta\alpha\beta$ fold and the N-domain with the $\beta\alpha\alpha\beta\beta\alpha\alpha\beta\alpha\beta$ fold (Figure 1). The domains are connected by a short linker formed by residues 1033–1038 and 1194–1198. There is a visible cleft in the N-domain, which is adjacent to the linker region. The cleft is formed by α -helices 1056–1061 and 1071–1081, connected by a loop that contains H1069 and β -strands 1190–1194, 1159–1164, 1121–1125. As we show below, this cleft contains a putative ATP-binding site. Non-polar side chains, except that of D1164, form the accessible surface of the nucleotide-binding site.

Recently, the solution structure became available for the N-domain of Na⁺,K⁺-ATPase, another member of the family of P-type ATPases [23]. It is interesting to compare it with the model of the N-domain of WNDP. Despite the difference in length and poor sequence homology (approx. 12 %), the 3D structures of the N-domains are quite similar, particularly with respect to the spatial position of helical segments α 1, α 3 and α 4, along with the β -strands β 2 and β 3 (Figure 2). The most significant differences between the two structures are observed in flexible loops. Therefore the overall spatial fold of the N-domains in Na⁺,K⁺- and copper-transporting ATPases is well preserved. Obviously, this is also true for Ca²⁺-ATPase, which was used as a scaffold to build the model of the N-domain of WNDP. As discussed below, the position of the ATP-binding site within the N-domains of these ATPases is also similar.

MD simulations of the ATP-BDs

Knowing the 3D structure of a protein is often insufficient for understanding the functioning of a molecule. In many cases, deeper insight into the intimate mechanism of protein functioning can be gained by studying its dynamic behaviour. We describe the results of MD simulations of the constructed models in explicit water surroundings below. These simulations pursued the following goals: (i) to evaluate the quality of the models and to inspect their global and local conformational lability; (ii) to analyse the large-scale motions in protein and delineate the dynamical domains (if any) and their relative movements; and (iii) to determine how conformational dynamics affects the ability of models to bind ATP.

To test the validity of the proposed MD procedure for WNDP models, calculations were also performed for the ATP-BD of Ca²⁺-ATPase in the ‘open’ and ‘closed’ forms. As described below, the stability of the experimentally derived structures and consistency of the results obtained for both ATPases strongly support the correctness of the model building and MD protocols.

ATP-BD of Ca²⁺-ATPase

Analysis of the MD trajectories obtained for the X-ray structures of Ca²⁺-ATPase ATP-BD in the E1 and E2 forms reveals that both structures equilibrate after approx. 1–2 ns of the dynamics run. (Hereafter we will use the terms ‘MD_{E1}’ and ‘MD_{E2}’ to indicate the MD trajectories of ATP-BD in Ca²⁺-ATPase or WNDP obtained with the open and closed starting models respectively.) For the C α atoms of Ca²⁺-ATPase ATP-BD, the RMSDs (root-mean-square deviations) from the starting conformations are approx. 2–3 Å (Figure 3A), indicating that the overall structures of the N- and P-domains remain stable. However, RMSFs (root-mean-square fluctuations), which characterize the mobility of each residue during MD (results not shown), suggest that several regions in the N-domain (455–462, 500–510 and 575–585) may undergo conformational rearrangements. In addition, the N- and C-terminal α -helices in the P-domain can significantly deviate from their initial positions during MD.

Whereas the individual domains of ATP-BD are stable, their mutual positions change considerably during MD. Superposition of the P-domains shows that, at equilibrium, the overall backbone RMSDs may reach 13 Å in MD_{E1} (Figure 3A). This signifies possible large-scale conformational transitions in water. Such rearrangements also occur in MD_{E2}, although the corresponding RMSDs are less than 6 Å (Figure 3A). The essential dynamic analysis of the MD results permits delineation of the low-frequency motions related to global conformational transitions in the protein structure. They represent movements of two dynamic domains with respect to each other. As determined by the Dyndom software, these protein parts correspond well to the N- and P-domains. The ‘closure movements’ provide the most important contribution to the domain motions, although relative rotation of two subdomains also occurs. These observations are consistent with the crystallographic data [8,9] and with the proteolysis results showing that, in Ca²⁺-ATPase, binding of nucleotide is associated with the closure of the nucleotide-binding domain [24].

The scale of the domain motions is different in MD_{E1} and MD_{E2}. In the first case, two subdomains move towards each other, and their relative position becomes quite close to that observed in the X-ray structure of Ca²⁺-ATPase in the E2 state. In the second case, the molecule undergoes slow conformational transitions near the initial ‘closed’ state. Such changes in the mutual position of the domains are illustrated by the angle Θ (see the Materials and methods section). As shown in Figure 3(B), at equilibrium,

Figure 3 MD of the ATP-BDs of Ca^{2+} -ATPase (A, B) and WNDP (C, D)

(A) Backbone RMSD from the X-ray structure (E1 and E2 forms) as a function of time. Curves 1, 2, RMSD values calculated for individual N- and P-domains with starting structures in E1 form; curves 3, 4, RMSD between the backbone atoms with starting E1 and E2 forms respectively; in this case, the P-domain structures in both forms are superimposed. (B) Dependence of the angle Θ on the simulation time. Solid and broken lines correspond to starting structures in open and closed states respectively. (C) Backbone RMSDs in the models of WNDP ATP-BD (open and closed starting states) as a function of time. (D) Dependence of the angle Θ on the simulation time. Other details are the same as in (A, B).

the values of Θ are approx. 135° for both the MD trajectories, whereas in the starting crystal structures they are 157° and 126° for the E1 and E2 forms respectively.

The domain motions lead to significant changes in the position of the ATP-binding region in the N-domain with respect to the phosphorylation site in the P-domain (D351) (see the section Nucleotide docking). Thus, in MD_{E1} and MD_{E2} , the characteristic distance between the N- and P-domains (measured for C_α atoms of residues E439 and D351) fluctuates in the ranges 18–23 and 16–19 Å respectively (results not shown). This change in the relative positions of two subdomains is probably the first step necessary for orientating ATP within the binding site and for aligning the γ -phosphate of ATP with the catalytic aspartate. Although for some MD_{E1} conformers, it was possible to obtain a simultaneous tight fit of ATP to both sites, for most of the structures these distances are still too large for ATP to reach D351. Therefore one may suggest that other domains of the ATPase play an important role in the fine-tuning of the ATP-BD movements. Indeed, in the native enzyme, conformational rearrangements are not limited to the nucleotide binding part; they also involve the TM region and the so-called actuator-domain [8,9]. Modelling the entire protein in the heterogeneous membrane–water environment would account for the effects of other domains in simulations. However, such an approach is even more computationally demanding to reach the time scales discussed below.

The simulation results agree reasonably well with the experimental results. For example, in Ca^{2+} -ATPase, binding of Ca^{2+} ions induces marked conformational transitions in the enzyme, with

the P- and N-domains moving relative to each other [8]. These displacements correspond mainly to a pure ‘closure motion’ of the P-domain with respect to the N-domain. Both, the X-ray and simulation results delineate similar dynamic domains, which may form the open and closed states. In summary, the high degree of stability of the individual N- and P-domains of Ca^{2+} -ATPase and the character of their relative movements in MD simulations give us confidence that the employed computational methods are correct and can be used for the analysis of WNDP.

ATP-BD of WNDP

The MD simulations performed using WNDP ATP-BD models revealed that the conclusions made for Ca^{2+} -ATPase were largely true for WNDP. However, there are some differences in the quantitative characteristics of the effects. Thus, the N- and P-domains in WNDP retain their spatial structure well, although the RMSDs between the backbone atoms of the MD structures and the starting open or closed structures are slightly higher than those in Ca^{2+} -ATPase (3.5/3.0 Å compared with 2.0/3.0 Å; Figure 3C). Rigid-body movements of the N- and C-terminal α -helices cause the relatively high value of RMSD for the P-domain. Conformational rearrangements in the N-domain are more pronounced. Note that RMSDs were similar in MD_{E1} and MD_{E2} simulations. It is also important that the individual N- and P-domains do not unfold during MD in the explicit solvent. These results also prove that the constructed models are physically reasonable.

To inspect the local conformational mobility of the models, we analysed the RMSF (root-mean-square fluctuation) values for each residue in MD runs. Equilibrium RMSF values (results not shown) demonstrate that the P-domain is more stable for both models (for each domain fitting was done separately). Interestingly, the residues involved in the binding of ATP in both domains have low RMSFs and, hence, are quite stable during MD (see also the discussion below). The most flexible parts correspond to the regions 1065–1070, 1084–1091, 1099–1104, 1109–1117 and 1170–1183 in the N-domain. This difference in flexibility between the domains may reflect the division of labour between these two portions of WNDP ATP-BD. Whereas the major function of the P-domain is to house the catalytic aspartate and transmit the changes that occur during catalysis to the membrane portion of the protein, the N-domain seems to participate in a wider scope of the reactions: it binds the nucleotide, interacts with the N-terminal copper-binding domain in a ligand-dependent manner [25] and may play a role in other WNDP functions such as intracellular trafficking. Therefore the flexibility of this subdomain could be a structural feature that allows the N-domain to perform all these functions through regulated conformational transitions. At the same time, we cannot exclude that the higher flexibility of the N-domain may be partly explained by inaccuracies of its 3D model compared with the P-domain.

As seen in Figure 3(C), the total equilibrium RMSD between the backbone atoms of the structures reached 12 Å when models were superimposed over their P-domains. This result points to large conformational rearrangements, which involve relative motions of two dynamic domains. Analysis of the equilibrium angles Θ obtained in both trajectories (Figure 3D) shows that the protein models built on E1 and E2 forms converge to different conformations. Similar to Ca^{2+} -ATPase, in MD_{E1}, Θ reaches approx. $135 \pm 10^\circ$ from the starting value of 160° . In MD_{E2}, an additional closure of the domains occurs: Θ becomes equal to $105 \pm 5^\circ$ (starting value 122°).

The dynamic domains of WNDP ATP-BD were identified by Dyndom as the entire N-domain and the larger portion of the P-domain. The residues 1193–1196 of the hinge region represent the linker between them. Overall, the character of domain motions resembles very much that described above for Ca^{2+} -ATPase. The most prominent movements are observed in MD_{E1} of WNDP ATP-BD (Figure 3C). It takes less than 1 ns for the model to approach significantly the closed form; the movement is associated with the decrease of RMSD from 15 to ~ 4 Å. In contrast, the amplitude of relative domain movements is smaller in MD_{E2}, where the structure becomes even more closed (Figure 3C). This was not observed for Ca^{2+} -ATPase. Probably, the last effect might be explained by some inaccuracies of the model of WNDP ATP-BD, especially in the loop regions of the N-domain involved in interactions with the P-domain.

According to MD_{E1} data, juxtaposition of the domains is accompanied by the increase in the inter-domain surface area from 30 to 100–140 Å². Under these conditions, the domains interact via the loop 1152–1156 in the N-domain and α -helix 1225–1232 in the P-domain. Interestingly, both regions are spatially close to the putative ATP-binding site. It seems that such a contact may stabilize a certain orientation of functionally important residues in WNDP. In addition, enlargement of the inter-domain contact area increases the overall quality of the model, as indicated by the 3D score. Specifically, the *S* value averaged over the equilibrium part of the MD trajectory reaches 120–140 instead of 105 for the starting model. For MD_{E2}, the contact area is 80–160 Å², and the *S* values vary between 130 and 135. Therefore domain closure considerably improves the environments of residues in WNDP compared with the open state and stabilizes the spatial structure.

Also, as seen in Table 1, the overall secondary-structure content observed in MD fits better the CD results compared with the starting models of WNDP ATP-BD. Qualitatively, the same picture is observed also for MD_{E1} and MD_{E2} simulations of Ca^{2+} -ATPase; the domain motions are accompanied by significant increase (approx. 60–70 Å²) in the contact area between them as well as by simultaneous improvement in the 3D score. Moreover, similar loop regions in both ATPases are involved in interactions between N- and P-domains, which occur during MD runs. Therefore, despite some minor differences, the dynamic behaviour of the two proteins is principally the same.

It is also interesting to see how the described domain motions affect the distance between the phosphorylation site in the P-domain (C_α atom of D1027) and the nucleotide-binding site in the N-domain. To define the latter, we have chosen the C_α atom of H1069, since recent experimental results pointed to the role of this residue in the binding of ATP [11,26]. In MD_{E1}, the equilibrium value of the distance $R_{1069-1027}$ varies in the range 16–20 Å (results not shown). In this case, the length of the ATP molecule bound in the N-domain is insufficient to reach the phosphorylated site. In contrast, in MD_{E2}, the value of $R_{1069-1027}$ may decrease to 11 Å, thus making the formation of the two-domain complex with ATP possible. This observation may also indicate that domain closure does not require binding of ATP.

ATP-binding site in Ca^{2+} -ATPase and WNDP: results of docking simulations

Exploration of the conformational space via MD simulations provides an opportunity to examine binding of ATP by protein present in different equilibrium states. Below we describe the results of ATP docking to the set of WNDP ATP-BD conformers extracted from the MD trajectories. Initially, validity of the computational procedure was tested in docking calculations with the X-ray structures of Ca^{2+} -ATPase, and the results were compared with the available experimental data. The ATP-binding site in the N-domain of Ca^{2+} -ATPase is well characterized (see e.g. [24,27]). In this domain, ATP appears to have contacts with the residues T441, E442, F487, K515, R560 and V562, although the spatial structure of the complex is yet to be resolved.

MD conformers obtained for Ca^{2+} -ATPase ATP-BD have different mutual orientations of the N- and P-domains; the angle Θ varies between 145 and 120° . Docking of ATP demonstrated the presence of only one putative binding site in the N-domain, and also revealed that the ligand could adopt different orientations with respect to the protein. For further studies, we selected only those complexes where the γ -phosphate is directed towards the phosphorylation site (D351). The resulting complexes are in good agreement with the experimental results. Specifically, the nucleic acid base and the ribose moiety of ATP are buried within the binding cleft of the N-domain and have contacts with known functionally important residues. Thus the carboxy groups of E439 and E442 form H-bonds with the amino group of adenine. The side chains of F487, K515, T441 and V562 are in close vicinity of the adenine ring, whereas the ribose has contacts with the side chains of K492 and V562. The observed position of ATP may explain the important role of R560, which forms up to three H-bonds with the α -, β - or γ -PO₃ group of ATP, thus determining the orientation of its phosphate tail [28].

The docking results show that the domains' closure observed in MD_{E1} leads to juxtaposition of the binding cleft in the N-domain to the phosphorylated D351 in the P-domain. Thus, in the open MD conformers ($\Theta \sim 140^\circ$), ATP anchored in the N-domain has no contacts with the residues in the P-domain, the distance

Figure 4 Structural model of the ATP-binding site in Ca^{2+} -ATPase obtained via MD and docking simulations

(A) Optimal orientation of the ATP moiety (coloured in black) bound to the equilibrium structure of the ATP-BD part of Ca^{2+} -ATPase (with angle $\Theta = 120^\circ$). The protein molecule is given in a ribbon presentation. Key residues in contact with ATP are marked with their names and numbers in the sequence. Their side chains are shown with black sticks. Protein–ligand H-bonds are indicated with broken lines. (B) ATP-binding sites in the N-domains of Na^+ , K^+ -ATPase (1) and Ca^{2+} -ATPase (2) as revealed by NMR spectroscopy (1; [23]) and molecular modelling (2; this study).

between the γ - PO_3 of ATP and the carboxy group of D351 being approx. 8 Å. In conformers with $\Theta \sim 130^\circ$, this distance decreases up to 6 Å. In this case, the γ - PO_3 group of ATP is H-bonded to the side chain of conserved K352 located in the P-domain. Another conserved residue in this domain, D627, creates a salt bridge with K352 and stabilizes its orientation. D627 may also form a

H-bond with the γ -phosphate of ATP. Finally, at $\Theta \sim 120^\circ$ (this corresponds to the extreme states of the protein obtained in MD simulations), the γ - PO_3 group of the bound ATP is engaged in a direct contact with the carboxy group of D351 and forms a H-bond with D627 (Figure 4A). In such complexes, R560 is H-bonded to α -phosphate, whereas K352 is bonded either to β - or

γ -phosphate. The γ -OH group of the conserved T353 may also participate in H-bond formation with the γ -PO₃ group of ATP. Therefore we propose that binding of ATP stabilizes the closed conformation of ATP-BD.

The good agreement of docking results with the known experimental data (see above) indicates that such simulations can provide a realistic model of the ligand binding without *a priori* information about the active site. They are also consistent with the mutagenesis results reported by Andersen and co-workers [27]. Specifically, it was shown that the mutation R560V affects not only nucleotide binding, but also inhibits phosphoryl transfer. On the basis of the available experimental results and our modelling results, we propose the following sequence of events that precede catalytic phosphorylation. ATP binds initially in the N-domain and its γ - and β -PO₃ groups form H-bonds with R560. Because of conformational transitions, such as domain motions and/or local protein mobility, R560 approaches D627 and K352 in the P-domain, thus providing close contact between γ -PO₃ and D351. As a result, the side chain of R560 moves towards the α -PO₃ group. We propose that the interaction between R560 and D627 results in close contacts of K352 and D351 with phosphate groups of ATP anchored in the binding cleft of the N-domain.

It is interesting to compare our docking results with the NMR-derived structure of the complex of the N-domain of Na⁺,K⁺-ATPase with ATP [23]. As seen in Figure 4(B), in both cases, location of the active site and orientation of the ligand in the N-domains are similar. The main difference is related to conformation of the loop containing R560 and R551 in Ca²⁺- and Na⁺,K⁺-ATPases respectively. In the latter case, this arginine residue has no contacts with bound ATP. This may explain the very low affinity of the isolated N-domain of Na⁺,K⁺-ATPase for ATP [23]. One of the possible reasons for the different orientations of R551 in the N-domain of Na⁺,K⁺-ATPase is the absence of the P-domain, which stabilizes the conformation of the R560-containing loop in Ca²⁺-ATPase.

The docking procedure developed for Ca²⁺-ATPase was further used to study ATP binding in WNDP ATP-BD. A representative set of protein conformations was extracted from trajectories MD_{E1} and MD_{E2}. As indicated before, the Θ values for MD_{E1} and MD_{E2} lie in the ranges 125–140° and 100–110° respectively. Thus, in the first case, the conformers are more open. In all generated complexes, the nucleotide binds near the interface between the N- and P-domains, whereas the phosphate groups can be coordinated by residues from both domains. Similar to Ca²⁺-ATPase, for further studies, we selected only complexes where the γ -phosphate is directed towards the P-domain. Docking results show that, for a large majority of complexes, the adenosine moiety occupies a wide cavity in the N-domain and may adopt different orientations in the active site. In our opinion, such structural heterogeneity is caused by the large size of the cavity, which is considerably wider than that in Ca²⁺-ATPase. The 3D model of the N-domain of WNDP is approximate and further biochemical and mutagenesis studies are necessary to verify the above conclusions. At the same time, several structural features inherent in most of the complexes can be delineated. Thus the environment of adenosine in the N-domain is formed mainly by hydrophobic residues (V1059, V1060, A1063, A1065, L1071, V1075 and A1124). At the binding site, few polar or charged residues, H1069, R1151 and D1164, exist, which interact with ATP. Moreover, all these residues, except H1069, are stable in MD runs and their backbone RMSFs do not exceed 1.0 Å (results not shown). The corresponding value for H1069 located on the top of a flexible loop is 1.5 Å.

We should note that the equilibrium MD structures of the N-domain obtained for the open and closed starting models are somewhat different. Owing to differences in the conformation of

the loop 1151–1161, the side chain of R1151 stabilizes bound ATP only in the MD_{E1} conformers, but does not interact with the ligand in the MD_{E2} conformers. R1151 probably plays an important role in WNDP function, because mutation of this residue leads to Wilson's disease. Such variations in docking results may be caused by a slightly different mobility of the N-domain in MD.

Analysis of the complexes shows that the γ -phosphate of ATP bound in the N-domain via the adenosine moiety may reach the carboxy group of D1027, if $\Theta \leq 105^\circ$. Such conformations are observed only in MD_{E2} (Figure 3C). In the resulting complexes (Figure 5B), the phosphate tail of the ligand is stabilized by H-bonds with conserved residues D1027, K1028 and T1029, whereas orientation of adenosine may vary significantly. Nevertheless, frequently, the amino group of ATP forms H-bonds with the carboxy group of D1164. In addition, in some models, the ligand reveals contacts with the side chain of H1069.

Docking ATP to the MD_{E1} conformers with $\Theta > 125^\circ$ results in complexes, where the distance between the binding site in the N-domain and the residue D1027 is too large to permit simultaneous accommodation of the ligand by both domains. Nevertheless, analysis of such structures also leads to interesting conclusions. These complexes may be divided into the following two groups: (i) Similar to WNDP-MD_{E2}, the adenosine part binds in the cleft of the N-domain (Figure 5A, site 1). The majority of the complexes possess a H-bond between the amino group of ATP and the carboxy group of D1164. The phosphate tail of ATP forms one or two H-bonds with the side chain of R1151; this residue approaches the binding site as a result of domain motions observed in MD_{E1}. Together with residues K1028 and D1222 from the P-domain, it 'anchors' the phosphate moiety of ATP and orients it towards functionally important D1027. On the other hand, γ -phosphate of the ligand is still far (> 8 Å) from the phosphorylation site. The γ -PO₃ group of ATP is in direct contact with D1027, whereas the adenosine moiety binds on the interface between the N- and P-domains (Figure 5A, site 2). In this case, the protein–ligand interaction is determined by the network of H-bonds between the phosphate groups of ATP and residues D1027, K1028, T1029 and T1220 of the P-domain. H1069 may be H-bonded to the OH2' or OH3' groups of ribose.

In summary, WNDP ATP-BD contains two distinct motifs responsible for ATP binding. These include the hydrophobic cleft in the N-domain that shields the adenosine moiety from the solvent, and the cluster of residues in the P-domain, which anchor the phosphate tail of ATP via the network of H-bonds. The high-affinity binding of ATP probably occurs after simultaneous contacts with both motifs. Spatial juxtaposition of the binding sites takes place as a result of domain closure observed in MD.

Correlation of docking results with mutagenesis data

Finally, we utilized the generated model of WNDP ATP-BD to understand better the potential effect of various mutations identified in patients with Wilson's disease. The disease mutations are found in both the P- and the N-domains, as well as in the linker region. A large number of mutated residues are located either in close proximity to catalytic D1027 or clustered around the purine-binding pocket in the N-domain. It seems probable that mutations found in the P-domain would alter precise protein folding and disrupt catalytic phosphorylation, whereas mutation in the N-domain may have a profound effect on affinity for ATP and orientation of ATP with respect to the catalytic aspartate. It seems also quite probable that mutations in the linker region would affect the dynamics of the domains and may alter conformational transitions during catalysis. Below we consider in more detail the role of several such residues delineated via docking simulations.

Figure 5 Structural model of the ATP-binding site in WNDP obtained via MD and docking simulations

(A) Two optimal orientations of ATP bound to the equilibrium structure of the ATP-BD part of WNDP (with angle $\Theta = 135^\circ$). Results of MD with starting open state. Site 1, ATP (coloured in black) binds in the hydrophobic cleft in the N-domain; site 2, ATP (grey) binds in the P-domain via the phosphate tail. (B) Results of MD with starting closed state. ATP-binding site in the equilibrium structure of WNDP ATP-BD ($\Theta = 100^\circ$). ATP is shown in CPK (Corey–Pauling–Koltun) representation. Other details are the same as in Figure 4.

Thus the model confirms the functional importance of the highly conserved residues D1027, K1028, T1029 and D1222 in the P-domain. It is probable that mutations of these residues would eliminate catalytic phosphorylation of WNDP by disrupting the network of H-bonds which anchors the phosphate tail of bound ATP. D1222 probably stabilizes the position of the side chain of K1028 via electrostatic interactions.

Despite the lack of sequence homology in the N-domains of WNDP and Ca^{2+} -ATPase, there are analogies to the locations of several key residues in the binding site. We propose that the role of the functionally important R1151 in WNDP is similar to R560 in Ca^{2+} -ATPase despite their different positions in the alignment (Figure 1). Being located at the interface of two domains, R1151 orients the phosphate tail of ATP towards the phosphorylation site. WNDP residues K1028 and D1222 are analogous to K352 and D627 in the P-domain of Ca^{2+} -ATPase (they are aligned at the same positions in Figure 1). In Ca^{2+} -ATPase, the purine moiety of ATP may be stabilized by an H-bond between its amino and carboxy group of E439 or E442. In WNDP, the amino group may

form an H-bond with D1164, which is not aligned against E439 or E442 (Figure 1).

The specific role of another important residue of WNDP, H1069, is still unclear. This residue was shown to be important for catalytic phosphorylation of D1027 by ATP [11] probably by positioning ATP before transfer of the γ -phosphate. Our docking results show that, in most of the complexes, H1069 does not form specific contacts (e.g. H-bonds) with ATP, although it is not far from the ribose moiety and α -phosphate. On the other hand, we cannot exclude that the conformation of the flexible loop, which carries H1069, is inaccurately predicted by the model (correct generation of protein loops is a common problem for all homology modelling algorithms [29]). In principle, this region of WNDP ATP-BD can be built in such a way that H1069 would be closer to the ligand, whereas the structural core of the N-domain would be preserved. According to the alignment in Figure 1, H1069 may correspond to E439 of Ca^{2+} -ATPase, which forms an H-bond with the adenine ring of ATP. If this is the case, the residue H1069 could be indispensable for the stabilization of ATP in the vicinity of the

P-domain. Obviously, future experiments are required to shed light on the precise role of H1069 in the functioning of WNDP.

As discussed above, orientation of ATP in WNDP ATP-BD demonstrates a higher degree of conformational heterogeneity compared with that in Ca^{2+} -ATPase. This could be due to limitations of WNDP N-domain model, caused by low sequence homology between two proteins. As a consequence, there are several contradictions between the docking and mutagenesis results. Thus mutation of the highly conserved residue E1064 decreases considerably the affinity of WNDP for ATP [33]. However, in our model, the side chain of E1064 is exposed to the solvent and not towards the binding site. At the same time, in possible alternative models, in which the loop 1062–1070 has different orientations with respect to the protein core, the side chain of E1064 may be positioned in close proximity to the amino group of ATP and can form the H-bond with it. In most of the complexes obtained in this study, the purine moiety creates a H-bond with the carboxy group of D1164, but in the literature there are no data regarding the functionality of this residue.

Another moot point is related to the region 1098–1105 in the N-domain: it contains the glycine-rich motif GxG conserved in the subfamily of heavy-metal transporting ATPases. In Zn^{2+} -ATPase, which has approx. 33% of sequence identity with WNDP in the ATP-BD region, this motif is important for ATP binding [30]. A similar motif has also been found in cAMP-dependent protein kinases, where the backbone amino groups form several H-bonds with bound ATP [31]. In our model, this region (G¹⁰⁹⁹CGIGC) is some distance from the binding site, although in the generated complexes there are no residues shielding it from the ligand. As seen in Figure 3(D), this fragment is flexible and, in principle, may approach the binding site during dynamics and induce additional stabilization of ATP in the N-domain. Further studies are required to test these hypotheses and to adjust the model. Such experiments have been undertaken by our group, and are currently in progress.

Conclusion

To summarize, two spatial models of the nucleotide-binding part of WNDP were generated. They are based on the high-resolution X-ray structures of Ca^{2+} -ATPase in its open and closed forms respectively. Inspection of the models using independent techniques (CD and MD data, analysis with environmental profiles) revealed their overall correctness. To study functional aspects of the models, they were subjected to MD simulations, whereas docking with ATP was employed to find the active site and to understand the role of particular residues in ligand binding. These molecular modelling approaches were tested in similar calculations performed for the ATP-BD part of Ca^{2+} -ATPase. Overall, in both cases, the results were in agreement, thus ensuring the validity of the conclusions made for WNDP. As a result of these experiments, large-scale motions (mainly of type closure) were found in both ATPases. They induce significant changes in the distances between several key residues in the N- and P-domains. The ATP docking revealed two possible regions involved in ATP binding via adenosine buried in the cleft near residues H1069, R1115 and D1164 and via the phosphate moiety anchored by H-bonds with residues in the vicinity of catalytic D1027. Furthermore, high-affinity interaction of ATP with these sites occurs if they are spatially close to each other. This can be achieved after relative domain motions of the closure type observed in MD simulations. A number of residues potentially involved in the coordination of ATP were delineated. Their predicted role will be tested in future site-directed mutagenesis studies.

We should emphasize that this MD study was not designed to reproduce the full atomic details of conformational rearrangements in the nucleotide-binding domains of Ca^{2+} -ATPase and WNDP. The aforementioned limitations and approximations of the theoretical treatment would make such studies difficult. Rather, the goal of our modelling was to find the principal factors, such as structural stability, putative large-scale motions and location of the active site(s), which may be important for the functioning of WNDP. Overall agreement of the simulation results with the experimental results makes the modelling a powerful tool that may be applied to explore spatial structure, dynamics and functioning of the ATP-BD part of WNDP, its mutants and homologous enzymes. Recently, the proposed 3D model was successfully applied to explain the role of several mutations in the N-domain of WNDP ATP-BD [33].

We thank Dr M. Hilge for providing co-ordinates of the nucleotide-binding domain of Na^+, K^+ -ATPase. Access to computational facilities of the Joint Supercomputer Center (Moscow) is gratefully acknowledged. This work was supported by the Russian Foundation for Basic Research (grant 04-04-48875-a), by the Ministry of Science and Technology of the Russian Federation (the State contract 43.073.1.1.1508/31.01.2002) and by the National Institute of Health grant PPG 1-P01-GM067166-01 to S. L. R. G. E. and D. E. N. are grateful to the Science Support Foundation (Russia) for the grants awarded.

REFERENCES

- 1 Lutsenko, S., Efremov, R. G., Tsvikovskii, R. and Walker, J. M. (2002) Human copper-transporting ATPase ATP7B (the Wilson's disease protein): biochemical properties and regulation. *J. Bioenerg. Biomembr.* **34**, 351–362.
- 2 Gittlin, J. D. (2003) Wilson disease. *Gastroenterology* **125**, 1868–1877.
- 3 Ferenci, P., Caca, K., Loudianos, G., Mieli-Vergani, G., Tanner, S., Sternlieb, I., Schilsky, M., Cox, D. and Berr, F. (2003) Diagnosis and phenotypic classification of Wilson disease. *Liver Int.* **23**, 139–142.
- 4 Kim, E. K., Yoo, O. J., Song, K. Y., Yoo, H. W., Choi, S. Y., Cho, S. W. and Hahn, S. H. (1998) Identification of three novel mutations and a high frequency of the Arg778Leu mutation in Korean patients with Wilson disease. *Hum. Mutat.* **11**, 275–278.
- 5 Loudianos, G., Dessi, V., Lovicu, M., Angius, A., Altuntas, B., Giacchino, R., Marazzi, M., Marcellini, M., Sartorelli, M. R., Sturniolo, G. C. et al. (1999) Mutation analysis in patients of Mediterranean descent with Wilson disease: identification of 19 novel mutations. *J. Med. Genet.* **36**, 833–836.
- 6 Nanji, M. S., Nguyen, V. T., Kawasoe, J. H., Inui, K., Endo, F., Nakajima, T., Anezaki, T. and Cox, D. W. (1997) Haplotype and mutation analysis in Japanese patients with Wilson disease. *Am. J. Hum. Genet.* **60**, 1423–1429.
- 7 Yamaguchi, A., Matsuura, A., Arashima, S., Kikuchi, Y. and Kikuchi, K. (1998) Mutations of ATP7B gene in Wilson disease in Japan: identification of nine mutations and lack of clear founder effect in a Japanese population. *Hum. Mutat.* **1** (Suppl.), S320–S322.
- 8 Toyoshima, C., Nakasako, M., Nomura, H. and Ogawa, H. (2000) Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature (London)* **405**, 647–655.
- 9 Toyoshima, C. and Nomura, H. (2002) Structural changes in the calcium pump accompanying the dissociation of calcium. *Nature (London)* **418**, 605–611.
- 10 Tsvikovskii, R., Eisses, J. F., Kaplan, J. H. and Lutsenko, S. (2002) Functional properties of the copper-transporting ATPase ATP7B (the Wilson's disease protein) expressed in insect cells. *J. Biol. Chem.* **277**, 976–983.
- 11 Tsvikovskii, R., Efremov, R. G. and Lutsenko, S. (2003) The role of the invariant His-1069 in folding and function of the Wilson's disease protein, the human copper-transporting ATPase ATP7B. *J. Biol. Chem.* **278**, 13302–13308.
- 12 Rost, B. and Sander, C. (1993) Prediction of protein secondary structure at better than 70% accuracy. *J. Mol. Biol.* **232**, 584–599.
- 13 Jones, D. T. (1999) GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J. Mol. Biol.* **287**, 797–815.
- 14 Sali, A. and Overington, J. P. (1994) Derivation of rules for comparative protein modelling from a database of protein structure alignments. *Protein Sci.* **3**, 1582–1596.
- 15 Lüthy, R., Bowie, J. U. and Eisenberg, D. (1992) Assessment of protein models with three-dimensional profiles. *Nature (London)* **356**, 83–85.
- 16 Bowie, J. U., Lüthy, R. and Eisenberg, D. (1991) A method to identify protein sequences that fold into a known three-dimensional structure. *Science* **253**, 164–170.
- 17 Gilson, M. K., Sharp, K. and Honig, B. (1987) Calculating the electrostatic potential of molecules in solution: method and error assessment. *J. Comp. Chem.* **9**, 327–335.

- 18 Berendsen, H. J. C., van der Spoel, D. and van Drunen, R. (1995) GROMACS. *Comp. Phys. Comm.* **91**, 43–56
- 19 van Gunsteren, W. F. and Berendsen, H. J. C. (1987) Gromos-87 Manual. Biomos BV Nijenborgh 4, 9747 AG Groningen, The Netherlands
- 20 Hayward, S. and Berendsen, H. J. C. (1998) Systematic analysis of domain motions in proteins from conformational change; new results on citrate synthase and T4 lysozyme. *Proteins Struct. Funct. Genet.* **30**, 144–154
- 21 Jones, G., Willett, P., Glen, R. C., Leach, A. R. and Taylor, R. (1997) Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* **267**, 727–748
- 22 Sreerama, N. and Woody, R. W. (2000) Estimation of protein secondary structure from CD spectra: comparison of CONTIN, SELCON and CDSSTR methods with an expanded reference set. *Anal. Biochem.* **282**, 252–260
- 23 Hilge, M., Siegal, G., Vuister, G. W., Guntert, P., Gloor, S. M. and Abrahams, J. P. (2003) ATP-induced conformational changes of the nucleotide-binding domain of Na,K-ATPase. *Nat. Struct. Biol.* **10**, 468–474
- 24 Ma, H., Inesi, G. and Toyoshima, C. (2003) Substrate-induced conformational fit and headpiece closure in the Ca^{2+} -ATPase (SERCA). *J. Biol. Chem.* **278**, 28938–28943
- 25 Tsivkovskii, R., MacArthur, B. C. and Lutsenko, S. (2001) The Lys1010-Lys1325 fragment of the Wilson's disease protein binds nucleotides and interacts with the N-terminal domain of this protein in a copper-dependent manner. *J. Biol. Chem.* **276**, 2234–2242
- 26 Voskoboinik, I., Mar, J. and Camakaris, J. (2003) Mutational analysis of the Menkes copper P-type ATPase (ATP7A). *Biochem. Biophys. Res. Commun.* **301**, 488–494
- 27 Clausen, J. D., McIntosh, D. B., Vilsen, B., Woolley, D. G. and Andersen, J. P. (2003) Importance of conserved N-domain residues Thr441, Glu442, Lys515, Arg560 and Leu562 of sarcoplasmic reticulum Ca^{2+} -ATPase for MgATP binding and subsequent catalytic steps. Plasticity of the nucleotide-binding site. *J. Biol. Chem.* **278**, 20245–20258
- 28 Toyoshima, C., Nomura, H. and Sugita, Y. (2003) Structural basis of ion pumping by Ca^{2+} -ATPase of sarcoplasmic reticulum. *FEBS Lett.* **555**, 106–110
- 29 Schonbrun, J., Wedemeyer, W. J. and Baker, D. (2002) Protein structure prediction in (2002). *Curr. Opin. Struct. Biol.* **12**, 348–354
- 30 Okkeri, J., Laakkonen, L. and Haltia, T. (2004) The nucleotide-binding domain of the Zn^{2+} -transporting P-type ATPase from *Escherichia coli* carries a glycine motif that may be involved in binding of ATP. *Biochem. J.* **377**, 95–105
- 31 Bossemeyer, D. (1994) The glycine-rich sequence of protein kinases: a multifunctional element. *Trends Biochem. Sci.* **19**, 201–205
- 32 Kabsch, W. and Sander, C. (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. *Biopolymers* **22**, 2577–2637
- 33 Morgan, C. T., Tsivkovskii, R., Kosinsky, Yu. A., Efremov, R. G. and Lutsenko, S. (2004) The distinct functional properties of the nucleotide-binding domain of ATP7B, the human copper-transporting ATPase. Analysis of the Wilson disease mutations E1064A, H1069Q, R1151H, and C1104F. *J. Biol. Chem.*, in the press

Received 1 March 2004/6 May 2004; accepted 17 May 2004

Published as BJ Immediate Publication 17 May 2004, DOI 10.1042/BJ20040326