Structural dynamics and catalytic mechanism of ATP13A2 (PARK9) from simulations

Teodora Mateeva[1], Marco Klähn[2], Edina Rosta[1,3]

[1]Department of Chemistry, Faculty of Natural & Mathematical Sciences, King's College London
[2]Department of Materials Science and Chemistry, Institute of High Performance Computing, Agency for Science, Technology and Research (A*STAR), Singapore
[3]Department of Physics and Astronomy, Faculty of Maths & Physical Sciences, University College London

Abstract

ATP13A2 is a gene encoding a protein of the P5B subfamily of ATPases and is a PARK gene. Molecular defects of the gene are mainly associated with variations of Parkinson's disease (PD). Despite the established importance of the protein in regulating neuronal integrity, the three-dimensional structure of the protein currently remains unresolved crystallographically. We have modelled the structure and reactivity of the full-length protein in its E1-ATP state. Using Molecular Dynamics (MD), Quantum cluster and Quantum Mechanical/Molecular mechanical (QM/MM) methods, we aimed at describing the main catalytic reaction, leading to the phosphorylation of Asp513. Our MD simulations suggest that two positively charged Mg^{2+} cations are present at the active site during the catalytic reaction, stabilizing a specific triphosphate binding mode. Using QM/MM calculations, we subsequently calculated the reaction profiles for the phosphate transfer step in the presence of one and two Mg^{2+} cations. The calculated barrier heights in both cases are found to be around 10.5 and 13.0 kcal mol^{-1}, respectively. We elucidated details of the catalytically competent ATP conformation and the binding mode of the second Mg^{2+} cofactor. We also examined the role of the conserved Arg686 and Lys859 catalytic residues. We observed that by lowering significantly the barrier height of the ATP hydrolysis reaction, Arg686 had significant effect on the reaction. The removal of Arg686 increased the barrier height for the ATP hydrolysis by ~3.5 kcal mol^{-1} while the removal of key electrostatic interactions created by Lys859 to the γ-phosphate and Arg513 destabilizes the reactant state. When missense mutations occur in close proximity to an active site residue, they can interfere with the barrier height of the reaction, which can halt the normal enzymatic rate of the protein. We also identified the main binding pockets in the full-length structure, including the pocket in the transmembrane region, which is likely where ATP13A2 cargo binds.

Introduction

The ATP13A2 gene has emerged as one of the genes strongly correlated with Parkinson's disease (PD) and is also known as PARK 9 [1,2]. The ATP13A2 gene encodes the P5B ATPase ATP13A2 which has attracted interest as an enzyme implicated in a range of neurodegenerative disorders: Spastic Paraplegia (SPG78), Kufor-Rakeb syndrome, neuronal ceroid lipofuscinosis and various other types of neurodegenerative disorders [2–6]. Currently, a multitude of molecular defects associated with the gene have been identified, [2,4,7–10] including loss-of-function missense mutations of the protein [4,5,15,16,7–14]. The precise role of most of these missense mutations remains unexplored, as well as the overall structural dynamics of the protein.

ATP13A2 belongs to the haloacid dehydrogenase-like (HAD) superfamily of enzymes which all share a hydrolase fold. The HAD superfamily is very diverse, encompassing phosphoesterases, ATPases, dehalogenases, and sugar phosphomutases, which act on a wide range of substrates, typically catalyzing carbon or phosphoryl group transfer reactions [17]. Phosphotransferase enzymes typically require Mg^{2+} cofactor for their catalytic activity [18,19]. ATP13A2, in particular, belongs to the big family of P-type ATPases which is split in five distinct subfamilies: P1, P2, P3, P4, P5 [6]. Most of these proteins are well studied and have resolved crystallographic structures, including ones in different functional states. ATP13A2 is part of the least studied subfamily P5B, which remains the only subfamily without any three-dimensional structures resolved.

The cytoplasmic domains of the protein include: Nucleotide-binding domain N, Phosphorylation domain P and an actuator domain A (Fig. 1A). Additionally, a transmembrane domain (T) connects the catalytic domains located in the cytoplasm to the extracytoplasmic area (Fig. 1B). Interestingly, the various missense mutations of the ATP13A2 protein currently described in the literature [5,8,9,12,15] are not confined to one spatial region (Fig. 1A), but are scattered across the entirety of the protein and encompass all domains. The catalytically active domains, N, P and A are involved in ATP binding, ATP hydrolysis, auto and de-phosphorylation. ATP13A2 has been classified as a membrane transporter protein [6] with the proposed candidates ranging from: heavy metals [2], to Ca^{2+} cations [20] and polyamine spermidine (SPD) [21,22]. Recent studies have revealed the role of ATP13A2 in polyamine export [23]. All enzymes belonging to the HAD superfamily contain a specific form of the Rossmannoid fold. This fold has two characteristic features that distinguish it from other superfamilies with Rossmannoid type-folds: a β-hairpin motif (also called a "flap") located immediately downstream of the first β-strand of the core Rossmanoid fold and a single helical turn ("the squiggle") [17]. This is important for ATP13A2 and other P-type ATPases, as these motifs provide mobility which allows the protein to alternate between the E1 "open" conformation (before the binding of any cargo) and the E2 "closed" conformation. The E1 state is associated with the binding and subsequent hydrolysis of ATP. In this state, the protein

has a high affinity for the cargo that is to be transported from the cytoplasm to the other side of the membrane. In ATP13A2, the ATP hydrolysis reaction results in autophosphorylation of the strictly conserved Asp513. Similarly, the E2 state is associated with the process of dephosphorylation of the aspartate [6]. In this work, we are interested in the change from the E1-ATP to the E1P functional state.

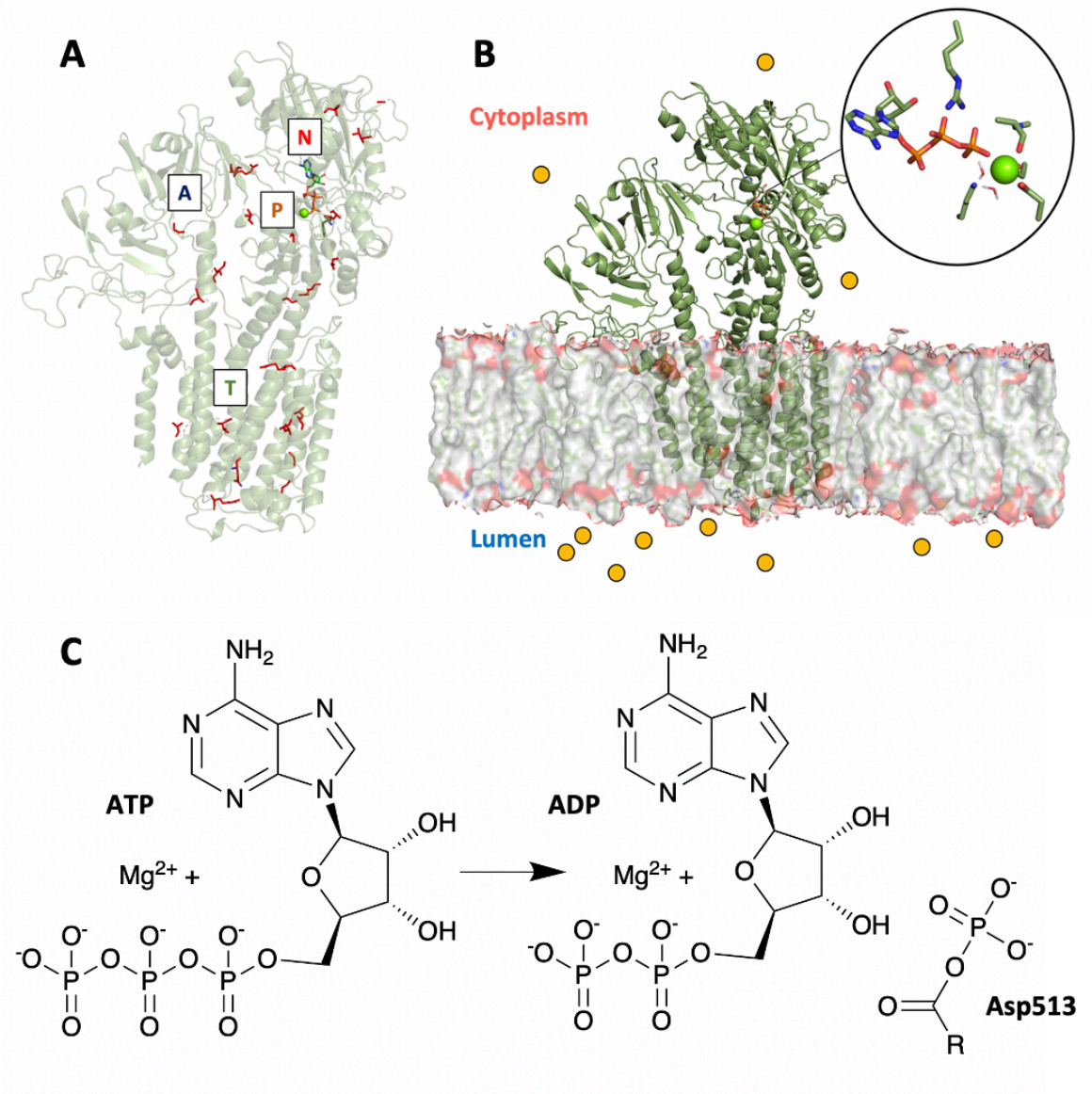

Figure 1. (A) Three-dimensional homology model of the ATP13A2 protein depicting missense mutations identified clinically [10–12,24,25] (red sticks) on the protein (green cartoon) and **(B)** a homology model of ATP13A2 with the transmembrane domain buried in the lipid-rich membrane while the N, P and A domains are located in the cytoplasm of the cell. **(C)** Substrate in the E1 state and the products of the autophosphorylation reaction.

The active site motif DKTGT is strictly conserved among all P-type ATPases [26]. Two other amino acids, which are highly conserved and located immediately in the active site, are

Arg686 and Lys859. Arg686 and Lys859 were structurally conserved in all enzymes whose crystal structures were used further in this study [27–32], Fig. S1.

Currently, none of the proteins within the P5B ATPase family have been resolved crystallographically, including ATP13A2, therefore, no three-dimensional structure is resolved in any of the functional states of the protein. Nevertheless, ATPases of the P2A and P2C subfamilies, which are highly homologous, are available with experimentally determined structures. Most recently, a crystal structure of the P5A ATPase ATP13A1 was resolved, which is currently the most homologous protein to ATP13A2 whose three-dimensional structure has been determined [33]. Many of the three-dimensional P-type ATPase structures contain active-site bound ATP-analogues, typically with synthetic non-hydrolysable derivatives of ATP, such as ACP or AMPPCP [27–29]. Most of these structures feature only one bound active site Mg^{2+} cation [27–29]. However, there are structures obtained with ADP and AlF_3 that feature two Mg^{2+} cations [30,32,34] bound in the active site. This brings up the question whether one or two Mg^{2+} ions are present and/or required for the phosphate transfer to proceed? Importantly, due to the different charge distribution of the synthetic derivative analogues, the second ion coordination may be captured incorrectly or not at all. There are no structures which have been crystallised with the catalytically competent ATP substrate, and, it is currently unknown what the precise ATP conformation during the hydrolysis reaction is, especially in terms of its proposed second Mg^{2+} coordination [34]. This leaves open the question of what the precise ATP-Mg^{2+} coordination is, as well as overall Mg^{2+}-Mg^{2+} distance and position.

Currently, the catalytic mechanism is not available using atomistic details for any P5 ATPase. Previous short Molecular Dynamics (MD) simulations were carried out on the P2A Endoplasmic Reticulum Ca^{2+}-ATPase [35], however, those did not provide any detailed insight on the catalytic mechanism or overall conformational dynamics of the protein. Multiscale reactive molecular dynamics (MS-RMD) and free energy sampling have been used to quantify the free energy profile and timescale of the proton transport in SERCA [36]. Quantum Mechanical/Molecular Mechanical (QM/MM) calculations have also previously been performed for the Phosphoserine Phosphatase, which also belongs to the large HAD-like family of proteins, but it is classified in a distinctly different family [37]. We therefore performed MD, QM cluster and QM/MM calculations with the aim to describe this important catalytic mechanism and quantify the role of active site residues and active site cations in the ATP hydrolysis reaction in the catalytically competent E1-ATP functional state. Our QM/MM calculations found that Arg686 had a significant effect on the barrier height similarly to arginine fingers, however interacting with the β-phosphate of the ATP backbone [38]. Accordingly, experimental data from mutagenesis studies of the Ca^{2+}-ATPase suggests that this conserved arginine is detrimental to the ATPase activity of the homologous enzyme [30]. We further show the precise effect on the barrier height of Lys859, which is similarly very important through interactions with both the γ-phosphate of the ATP and Asp513 in the

reactant state. Additionally, we present strong evidence from MD simulations and QM/MM calculations that the enzyme operates via a two-ion catalytic mechanism. The results presented in this work can suggest how missense mutations disrupting crucial barrier-lowering interactions can have abolishing effect on the enzymatic activity of ATP13A2 and other P5B ATPases with homologous active site, specifically G877R [5].

Results

Homology Modelling

To identify the closest protein sequences to ATP13A2 and any available structural information on these, we performed a BLAST [39] and FASTA [40] searches. The most similar proteins in humans, by sequence similarity, are as follows: ATP13A3, ATP13A4, ATP13A5 (Fig. 2), which are all part of the P5B ATPase protein subfamily. As it was proposed earlier [23], those proteins have high conservation in the substrate binding domain and are likely transporting and/or interacting with the same cargo within the cell. Unfortunately, none of these proteins have three-dimensional structures deposited in the Protein Data Bank (PDB). The closest available proteins with experimentally resolved structures are: ATP13A1 (P5A subfamily), the Na^+/K^+-transporting ATPase (P2C subfamily), the endoplasmic Reticulum Ca^{2+}-ATPase (SERCA) (P2A subfamily) and the H^+/K^+-ATPase (P2C subfamily). We selected the very recently resolved structure of ATP13A1 [33], and the endoplasmic Reticulum Ca^{2+}-ATPase (SERCA) [29] for modelling the active site (Fig. B and 2C). The SERCA ATPase has a complete conservation with ATP13A2 in the active site region of interest (Fig. S1) and was already available with several ATP analogues bound, including water molecules crystallographically resolved [29]. Our final active site models match both experimental structures very accurately (Fig. 2B and 2C for SERCA and ATP13A1, respectively).

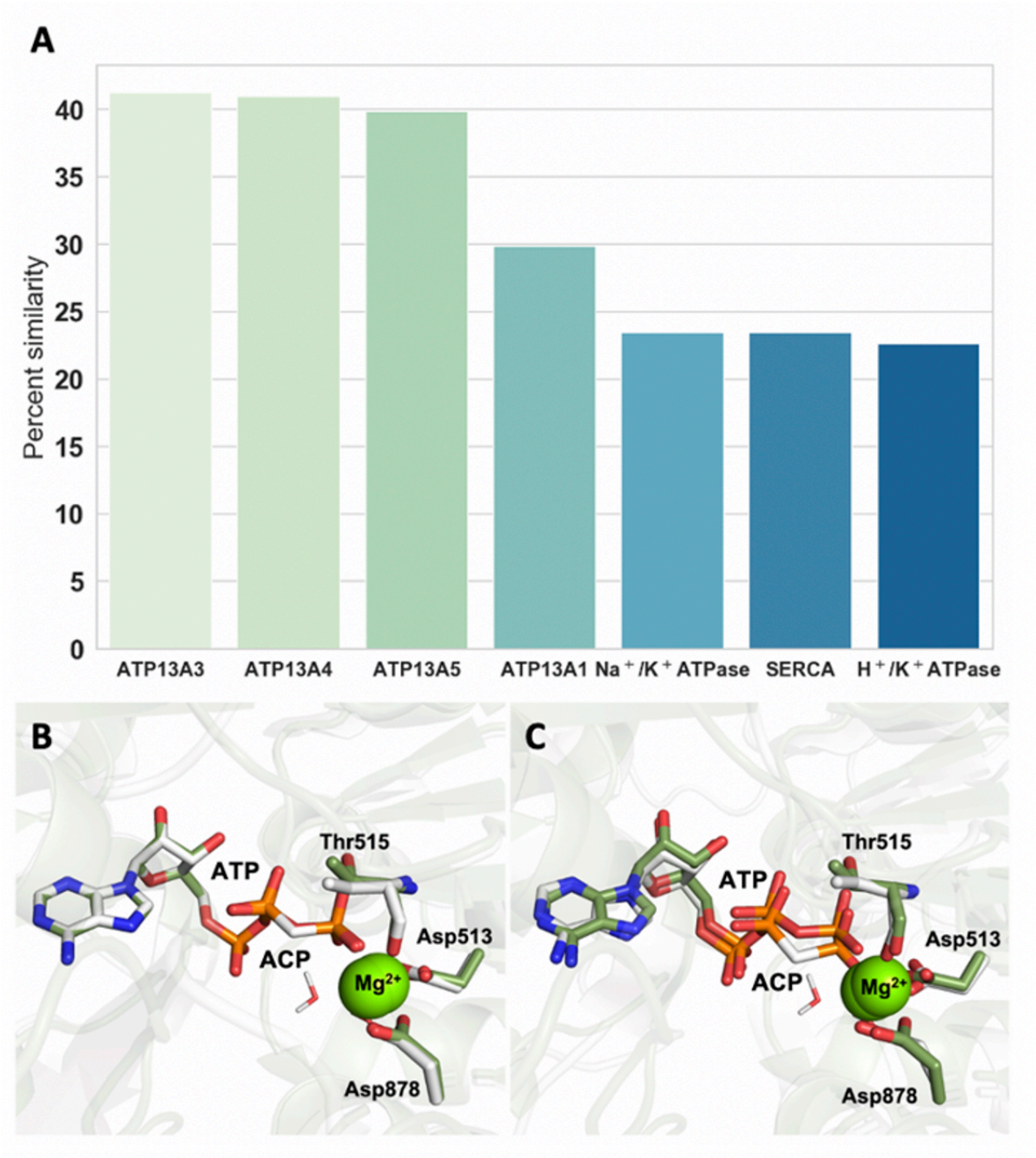

Figure 2. (A) Most similar human proteins to ATP13A2 based on overall fold, percent of sequence similarity and overall query coverage, ordered from most to least similar. **(B)** Active site of the Endoplasmic Reticulum Ca^{2+}-ATPase (SERCA) in E1 state with bound ACP molecule and 1 Mg^{2+} cation, (PDB code: 3tlm [29], grey) and our homology model of ATP13A2, (green sticks). **(C)** Active site of ATP13A1 in the E1 state with bound ACP molecule and 1 Mg^{2+} cation, (PDB code: 6xmq, grey sticks) and our homology model of ATP13A2, (green sticks).

We also note that any of the ATPases listed in Fig. 2A would be a suitable choice for modelling the active site of ATP13A2 as the Mg^{2+}-binding residues in the catalytic P-domain are highly conserved among P-type ATPases, as well as the other amino acids found in the immediate active site.

Molecular Dynamics (MD) Simulations

To probe the conformational dynamics of the protein, first, MD simulations were performed on the membrane-embedded model based on SERCA. The overall structure and in particular, the active site, was well conserved during the simulations (Fig. S2).
We focused on the Mg^{2+} ion coordination, which was initially octahedrally coordinated, coordinating: two water molecules, the side chain of the Asp513 residue (monodentate), the side chain of Asp878 (monodentate), the main chain of Thr515 (monodentate) and the γ-phosphate of the ATP. During the simulations the coordination of the Mg^{2+} ion has remained octahedral, and importantly, the binding mode of the Asp513 residue has remained monodentate. The γ-phosphate of the ATP phosphate chain was also in the correct orientation for the hydrolysis reaction to occur.

Importantly, the ATP molecule no longer preserved its original "zig-zag" conformation but immediately adopted a "straight" conformation, from the first nanosecond of the simulation (Fig. S3A). This has also been observed before in the short MD simulations of SERCA [35]. Komuro Y., et. al suggested that the "straight" ATP conformation is incorrect and reparameterization of the ATP molecule is needed, to ensure that the original conformation from the crystal structure ACP is preserved. We also noticed that there was a charge imbalance in the active site that could possibly cause this conformational change. The initial simulation based on the crystal structure involved a single Mg^{2+} cation only in the active site, however, we observed that a second cation constantly occupied a position very close to the α and β-phosphates (Fig. S3B). We conducted subsequent MD simulations where the ATP molecule was fixed at its original conformation. The rest of the ions, water and the protein were free to move without any constrains. In this simulation, we also observed that K^+ ions approached the active site for the full simulation time as before, and were located in the same position where a second Mg^{2+} cation was found in the crystal structures of the homologous SERCA active site TS-like crystal structure [32] (Fig. 3A). Throughout the duration of the MD simulations, the K^+ ions remained a constant presence at the active site (Fig. 3B and Fig. S3C). As crystallographic structures often lack catalytically essential Mg^{2+} ions, we propose that the second Mg^{2+} ion could be needed to stabilize the catalytically active conformation of the ATP. Two Mg^{2+} ions have been resolved in the structures of homologous enzymes, however, only in TS-like states, most likely because the chain of the synthetic derivative analogue has a different charge than the catalytically active ATP.

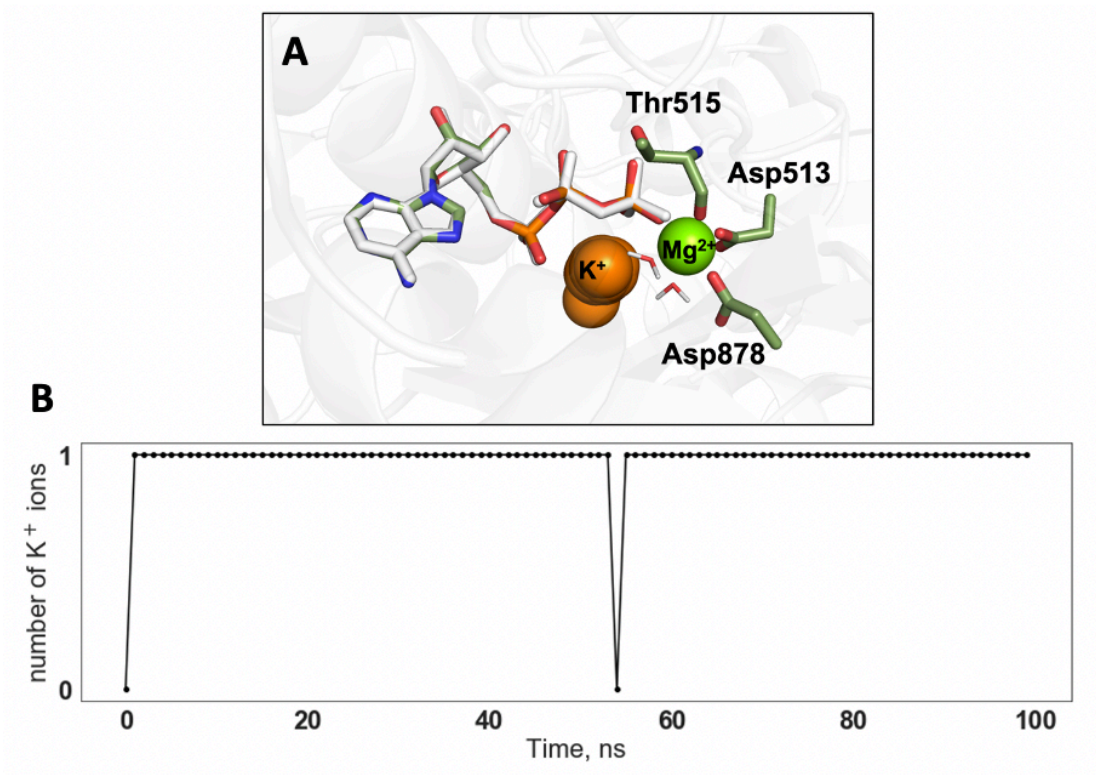

Figure 3. (A) Region of the active site where the K^+ ion clustering is observed during the simulation (orange spheres); **(B)** Number of K^+ ions in the active site during the duration of the 100 ns simulation.

QM cluster Calculations

To probe the catalytically active conformation of the ATP molecule at the active site, and, specifically, the second Mg^{2+} ion coordination, we performed QM cluster calculations. We generated four different ATP-Mg^{2+} starting geometries, G1-G4 (Fig. S4), by using information from the crystal structures containing two active site cations and ADP [30,32,34] (G1) and from preliminary results of the initial MD simulations with one Mg^{2+} in the active site (G3) or by manual initial guess (G2 and G4). All four starting geometries G1-G4 were optimized and upon convergence yielded three distinct conformations. The optimized structures of G1 and G2 (Fig. 4, green sticks) are most energetically favourable and agree particularly well with the conformation and coordination mode observed in the crystal structures containing two Mg^{2+} ions and ADP [32] (Fig. 4, grey sticks), where the second Mg^{2+} ion is coordinating only two oxygen atoms coming from the α and β phosphate of the ATP phosphate chain, and four water molecules. Importantly, the phosphate chain in both G1 and G2 is not in a straight conformation such as in geometries G3 and G4. Conformations G3 and G4 in which the phosphate chain is "straight" (Fig. S4C and D) have considerably higher energy and were therefore not used in any further QM/MM calculations. Additionally, upon aligning the optimized geometries G3 and G4 with "straight" chain to the crystal structure [41], larger deviations are clearly visible (Fig. S5C and S5D). Conformations in which the second Mg^{2+} cation is coordinated by three oxygen atoms are not favourable either (Fig. S4B) and converge

to the coordination mode observed in G1 and G2, Fig. 4. This ATP conformation and coordination mode were subsequently used for QM/MM calculations to determine the corresponding reaction profile. The good agreement between the crystal structure and the most energetically favourable optimized geometries demonstrates that the QM cluster method is indeed useful in the search for active site conformers.

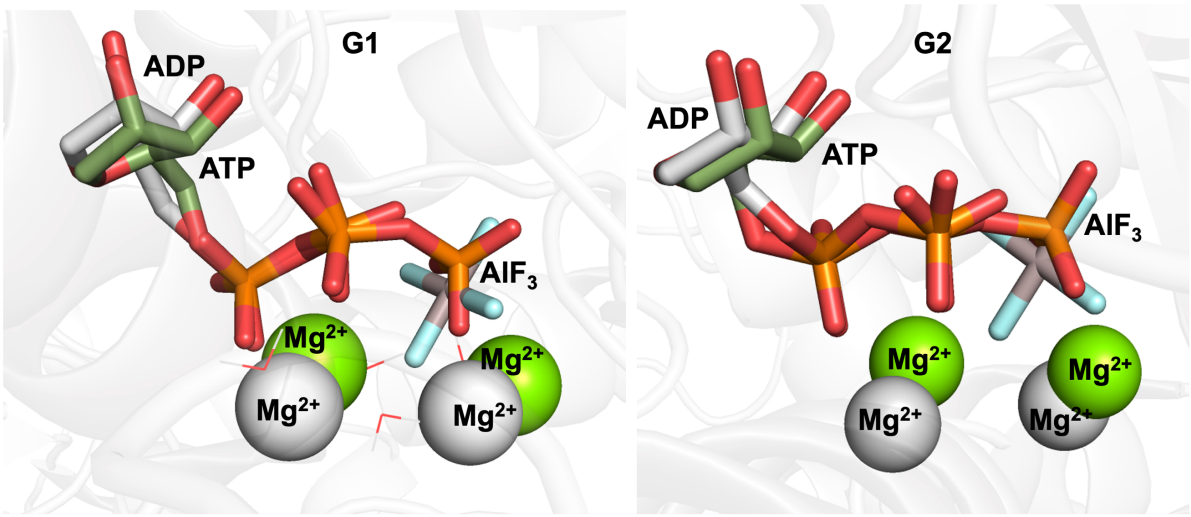

Figure 4. Optimized active site geometries for G1 and G2 (green sticks), aligned to the existing homologous active site crystal structure of the Na^+/K^+ ATPase (grey sticks, pdb: 3wgu) [32] which is resolved with two Mg^{2+} ions (grey spheres) and an ADP molecule (grey sticks). The green spheres represent the two Mg^{2+} ions in the optimized geometries and the grey spheres represent the two Mg^{2+} ions in the crystal structure. Both structures represent the same binding mode to the second Mg^{2+} via the two oxygen atoms of the phosphate chain of the ATP.

QM/MM Calculations

For the QM/MM calculations, six systems (P1-P6) were created using the same number and species of atoms in the MM region. However, we varied the number of atoms in the QM region to explore various effects related to the role of the number of Mg^{2+} ions, and the Lys859 and Arg686 residues for the reaction. System P1 included one Mg^{2+} ion, six water molecules, ATP, Asp513, Asp878, Thr515, Lys859 and Arg686 in the QM region. System P2 contained the same number and atom species but also a second Mg^{2+}. P3 and P4 (Table S2) contain one Mg^{2+} ion in the active site and Lys859 or Arg686, respectively, are removed from the QM region and their MM atomic charges are set to 0. P5 and P6 contain two Mg^{2+} ions in the active site (Table S2) with Lys859 or Arg686, respectively, electrostatically removed from the system as detailed above. Reaction scans were performed on all six systems, P1-P6, to determine the potential energy barriers. We performed forward and backward scans until we observed consistent energy profiles (Fig. 5A) and the energy minimum was obtained. Both scans show a good convergence and very small hysteresis between the forward and backward directions. For P1, in the presence of one Mg^{2+} only, the barrier height is approximately 13.0

kcal mol^{-1}. During the ATP hydrolysis, the ion is octahedrally coordinated via two water molecules, the γ-phosphate of the ATP phosphate chain, the carboxylate side chains of Asp878 and Asp513 (in a monodentate mode) and the main chain carbonyl of Thr515 (Fig. 5B). The barrier height for P2, containing two Mg^{2+} ions was calculated to be 10.5 kcal mol^{-1}. The second Mg^{2+} ion coordinates four water molecules and two oxygen atoms coming from the α and β phosphates of the ATP chain. Importantly, it does not coordinate a third oxygen atom from the phosphate chain.

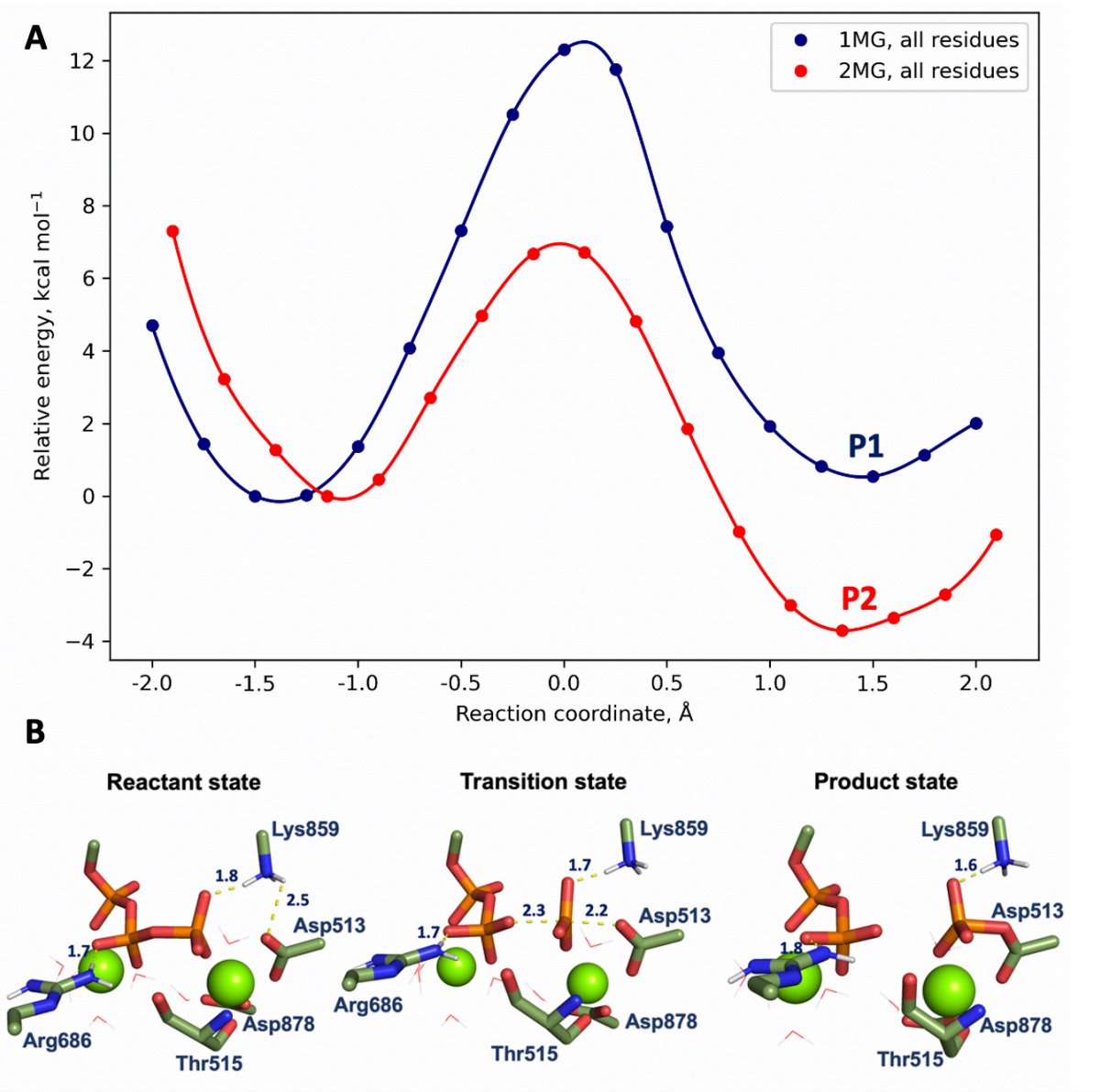

Figure 5. (A) Reaction coordinate scan with 1 (blue) and 2 (red) Mg^{2+} ions in the active site of ATP13A2. The reactant state is set to 0. **(B)** Reactant, transition state and product state during the ATP hydrolysis reaction with 2 Mg^{2+} ions in the active site are shown (green sticks). Distances are shown between the Lys859 and Arg686 and the ATP molecule. O_{2D}–P_G–O_{3B} distance is shown in the TS state. Hydrogens on the rest of the amino acids are not shown for clarity. Throughout the RCS, the Mg^{2+} ion coordinating Asp513 is always bound to two water molecules and the Mg^{2+} ion which is coordinating the α and β-phosphates of the ATP chain, coordinates four water molecules.

We find that the reaction proceeds as a one-step reaction, without any stable pentavalent phosphate intermediate. The reaction pathway in phosphate hydrolysis can be classified as associative, dissociative or concerted [42,43]. In the associative pathway, the attacking nucleophile approaches the phosphorus atom, decreasing the bond length between the attacking nucleophile and the reactive phosphorus while the bond to the leaving group simultaneously increases. This is generally a lower energy pathway for biological phosphate reactions. Alternatively, in the dissociative pathway, the leaving group departure precedes the nucleophilic attack. In the concerted pathway, partial bond formation and bond breaking occur simultaneously. Based on the evolution of the distances between the attacking nucleophile and the reacting phosphate, and the phosphate and the leaving group, the reaction pathway can be classified. In this ATP hydrolysis mechanism of ATP13A2, there is an energetically stable leaving group, a diphosphate anion, and a dissociative pathway could be expected to potentially be more favourable [44]. To confirm, we considered the distances between O_{2D}-P_G and P_G-O_{3B} (Fig. 6A). Our results show that the path from the reactant state (RS) to the product state (PS) follows an associative pathway in the protein environment, for both one and two Mg^{2+} cases. Furthermore, we note that while the reaction can follow an associative pathway, the TS itself does not have to be associative [43,44]. We calculated the bond order in the transition state and obtained Wiberg bond indices for the P_G-O_{2D} and P_G-O_{3B} bonds of 0.162 and 0.125. We therefore found that while the reaction clearly follows an associative pathway (Fig. 6A), the TS itself corresponds to a loose, dissociative structure, where both P-O bonds are already broken (Fig. 6B). This is consistent both with computational work demonstrating an associative path [42] and with experimental findings that can capture the loose TS. [45]

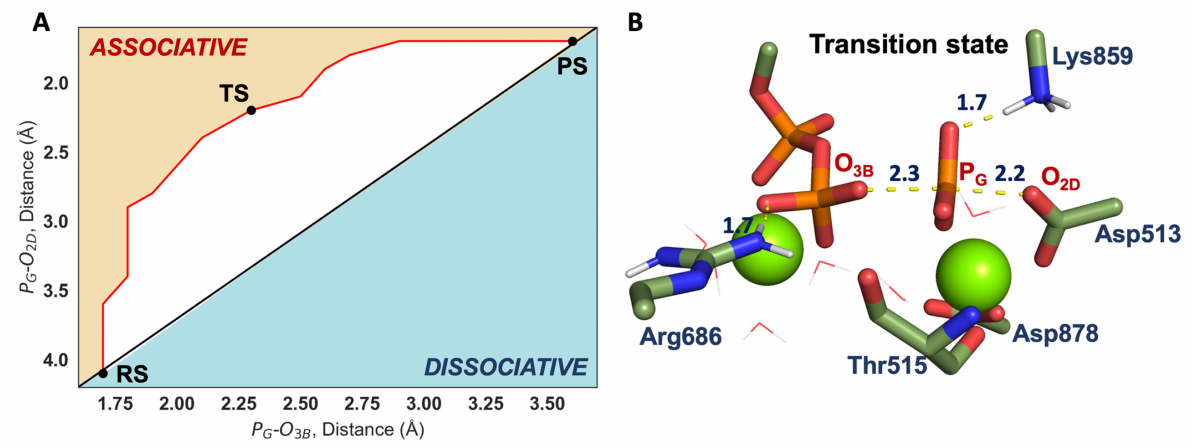

Figure 6. (A) The path adopted in the ATP hydrolysis of ATP13A2 is associative, based on the distance evolution from the reacting nucleophile O_{2D} to P_G and P_G-O_{3B} (leaving group), respectively. **(B)** Details of the transition state structure for the reaction. Key residues at the active site (sticks and spheres for Mg^{2+}) and relevant distances (yellow dashed lines) are shown.

To quantify the effects of the interactions between the ATP and Lys859 or Arg686 (P3-P6), we obtained converged Reaction Coordinate Scans by eliminating these interactions (Fig. 7). This allows us to observe the reaction coordinate without the stabilizing electrostatic interactions formed between Arg686 and the β-phosphate of the ATP chain; and between Lys859 and both the γ-phosphate and the Asp513. Without Arg686, the barrier height increases by ~3.5 kcal mol^{-1} (Fig. 7), which is significant in terms of time scales. Without Lys859, which forms key electrostatic interactions to the side chain of Asp513 and the γ-phosphate of ATP (Fig. 5B), the reactant state is considerably higher in energy than in the cases when Lys859 is present in the system. While this results in a lower apparent barrier, the reactant state is destabilized and the ATP binding is likely impaired, which might have additional consequences for the enzyme function and stability.

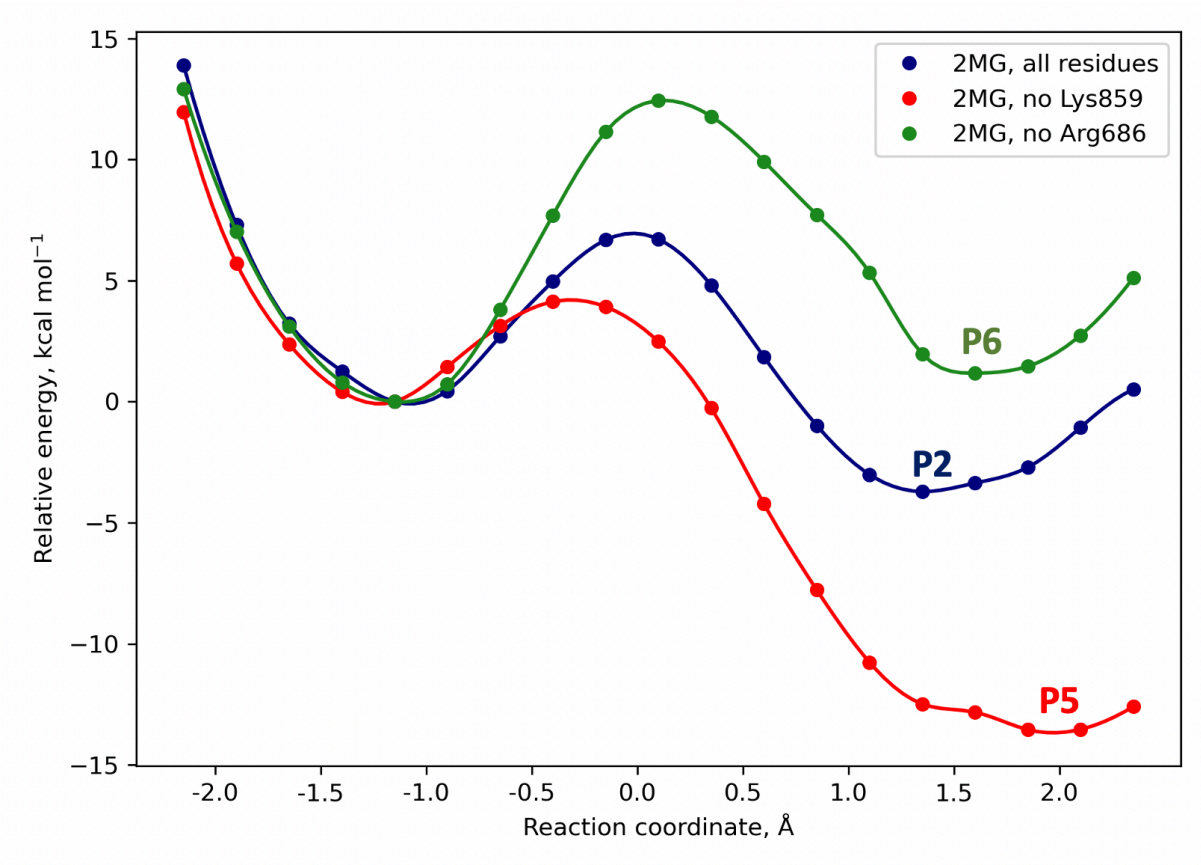

Figure 7. Reaction coordinate scans for the active site of ATP13A2 containing two Mg^{2+} ions. The RCS depicted in red represents the ATP hydrolysis reaction without Lys859 (P5). The RCS in green represents the ATP hydrolysis reaction without Arg686 (P6). The blue RCS shows the reaction in the active site when all amino acids required for the normal enzymatic activity, are present.

Binding Pockets and Transmembrane binding Analysis

While the P5A ATPase ATP13A1 has been shown to be a Mn-transporter [46], it has recently been demonstrated that ATP13A2 is strongly implicated in polyamine export [23]. It has been known that P5B ATPases such as ATP13A2 likely transport different cargo from ATP13A1 due

to major differences in the transmembrane domain sequence conservation [47]. To identify binding regions on the surface of ATP13A2, pocket analysis was performed on twenty equidistantly spaced frames from the MD trajectories of the modelled protein. The four biggest pockets were analysed for each frame. Not surprisingly, for most of the frames the biggest pocket in terms of surface area is the one where ATP binds in the N-domain (Fig. 8A). It is expected that this pocket will be conserved for all homologous P-type ATPases due to the highly conserved ATP binding region and mechanism in the active site.

Interestingly, the second biggest pocket, which appears consistently, was identified in the transmembrane region (Fig. 8A), more specifically, where inorganic ion binding has been observed in the crystal structures of homologous enzymes such as the P2C Na^+/K^+-ATPase [31] and the P2A ATPase SERCA [29]. We call this spatial region Site 1 (S1), Fig. 8D. For the Na^+/K^+-ATPase, the specific amino acid scaffold surrounding the three Na^+ ions which bind in the transmembrane domain consists of: Val322, Ala323, Val325, Pro326, Glu327, Tyr771, Thr774, Ser755, Asn776, Glu 779, Asp804 and Gln924. From this sequence motif we observe conservation in ATP13A2 for the amino acids that bind one of the inorganic ions in S1 (Fig. 8D). In ATP13A2, these correspond to: Val469, Pro470 and Asp967 (Fig. 8D). The remaining amino acids that coordinate the additional two Na^+ ions in the Na^+/K^+-ATPase (Site 2 and 3, Fig. 8D) are not conserved in ATP13A2. Val469 and Pro470 in S1 are also conserved between ATP13A2 and SERCA, which transports Ca^{2+} (Fig. 8D), but the amino acids coordinating the second ion in SERCA and Na^+/K^+-ATPase (S2 and S3) are not conserved in ATP13A2. Considering the amino acid conservation in the ion-binding region of the transmembrane (Fig. 8D), and overall similar shape of the ion-binding scaffold (Fig. 8D) it is possible that ATP13A2 also can bind and transport one inorganic ion, although this is unlikely to be either Na^+ or Ca^{2+}. Additionally, this transmembrane pocket was consistently (90% of the time of the analysed frames) found within 1.5 Å of ion-binding amino acids (Val469, Pro470, Asp967) in the S1 region, which additionally supports the idea that the protein binds cargo in this area [2,23]. It is important to note that the surface area of this pocket is considerably larger than expected from an ion binding site alone – the average size of the transmembrane pocket is 2264.61 $Å^3$. This suggests that this region of the protein additionally could interact with a much bigger substrate, likely in S2 and S3 region which is not conserved between the Na^+/K^+-ATPase or SERCA, Fig. 8D.

Two additional pockets also appeared frequently, however, not as consistently as the ATP pocket and the main transmembrane pocket. These include a considerably smaller pocket in the upper part of the transmembrane (Fig. S7) and an additional pocket around residues 985-995 and 800-807 (Fig. S7).

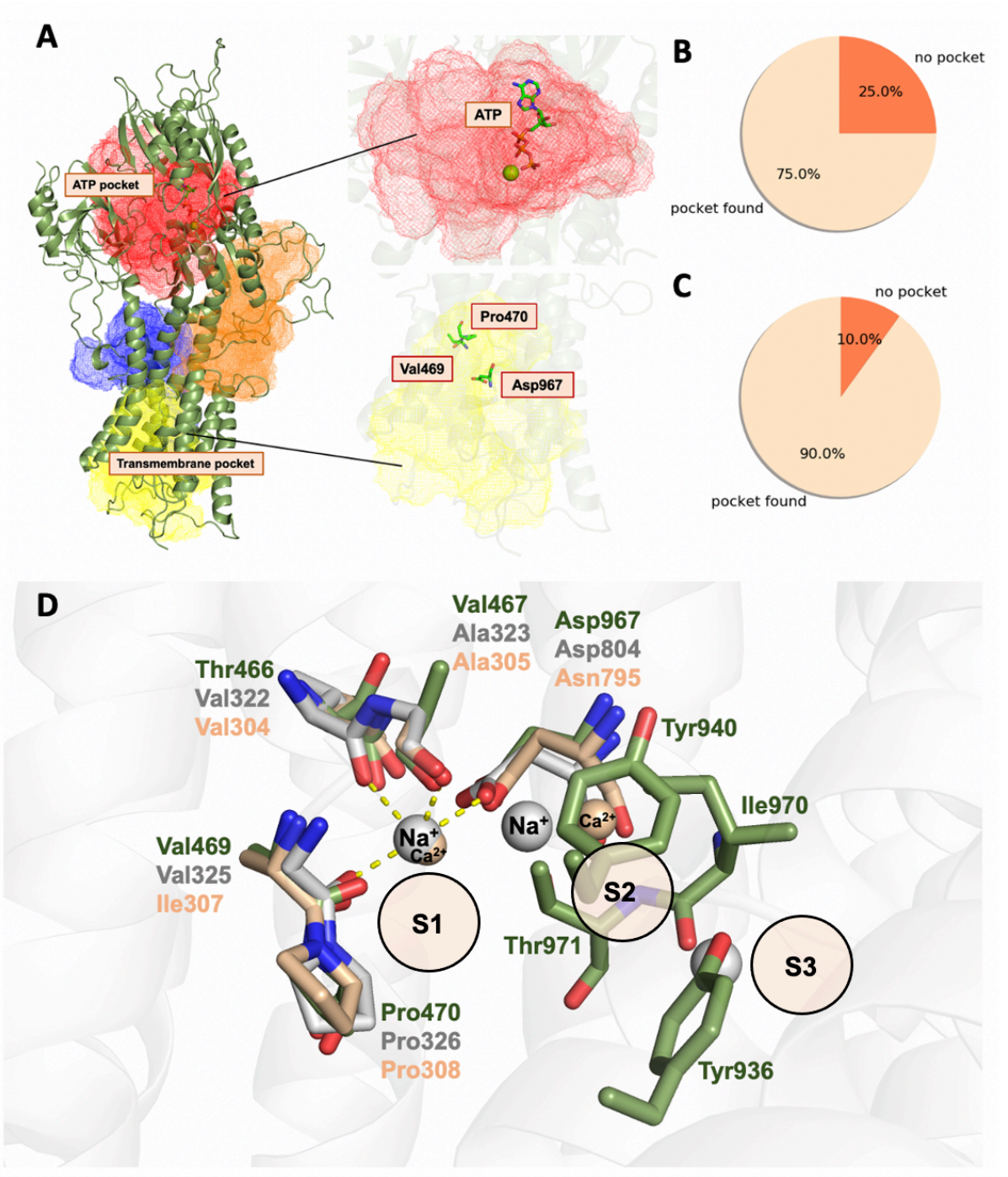

Figure 8. (A) The two main binding pockets (red mesh and yellow mesh) found consistently on the surface of the homology model of ATP13A2 (green cartoon). **(B)** Frequency of ATP-pocket occurrence calculated from the 20 frames of the unconstrained MD simulations. **(C)** Frequency of transmembrane pocket occurrence calculated for the same frames. **(D)** Sequence conservation in the inorganic ion-binding region in the transmembrane between the Na^+/K^+-ATPase (grey sticks), SERCA (wheat sticks) and ATP13A2 (green sticks). S1, S2 and S3 stand for binding sites 1, 2 and 3.

Methods

Homology Modelling

The E1-ATP-bound state was modelled based on the Endoplasmic Reticulum Ca^{2+}-ATPase (SERCA), (PDB code: 3tlm) [29]. The position of the ATP molecule and the Mg^{2+} cation was based on the position of the ACP moiety and the Mg^{2+} cation, respectively, in this crystal structure [29]. The crystal structure has the position of only one Mg^{2+} resolved, hence, the initial model contained only one Mg^{2+}. The modelling server used was SWISS-MODEL [48] and the sequence of ATP13A2 was obtained from Uniprot [48] for Homo Sapiens (Uniprot code: Q9HD20). To account for the most recent crystal structure available of ATP13A1, ATP13A2 was also modelled based on the P5A ATPase ATP13A1 (PDB code: 6xms) [33]. Both templates result in the same three-dimensional structure in the active site region of interest (Fig. 2).

Molecular Dynamics (MD) Simulations

All Molecular Dynamics simulations were performed by using the program NAMD [49]. The force field used in the simulations was CHARMM36 [50] with periodic boundary conditions and to evaluate the non-bonded long-range interactions the particle mesh Ewald method [51] was utilised with a 12 Å cut-off. The NPT ensemble was maintained with a Langevin thermostat (303 K) and an anisotropic Langevin piston barostat (1 atm). The simulation was repeated at 309.15 K. The box type was rectangular, 22.5 Å per side. The water model was TIP3P [52]. To neutralize the system 0.15 M KCl solution was added. The ion placing method was by distance. The energy of the system was minimized via steepest descent algorithm, followed by a standard six-step equilibration for membrane-embedded systems with restrained heavy atoms via standard CHARMM-GUI [53] procedure with a time step of 2 fs. The first step of the equilibration was done with a time step of 1 fs. SHAKE algorithm [54] was deployed to constraint the covalent bonds involving hydrogen atoms. The equilibration was followed by 100 ns production with all the atoms completely unconstrained and free to move. A second 100 ns MD simulation was performed where the coordinates of the ATP molecule were fixed in their original position (following the crystal structure coordinates of the homologous template) in order to preserve the original coordinates of the crystal structure ACP (PDB code: 3tlm) [32]. The protein, solvent and ions were completely unconstrained. The PPM server was used for orientation of the protein in the membrane [55]. The membrane had the following composition: 40% Cholesterol, 30% Phosphatidylcholine lipids (PC) and 30% Phosphatidylethanolamine lipids (PE) in order to mimic a membrane environment in a lysosome-like cell. The same protocol was repeated for the second homology model of ATP13A2, which was based on the P5A ATPase ATP13A1.

QM Cluster Calculations

All of the QM cluster calculations were performed by using the Gaussian 09 program [56]. The QM region was treated with the B3LYP hybrid density functional [57] and the 6-31+G* basis set [58]. The QM region consisted of 2 Mg^{2+} cations, 6 water molecules, Asp513, Asp878 and Thr515, as well as the full ATP molecule (Fig. S4). The geometry optimization followed standard QM cluster procedure where the C atom where the amino acids are truncated, is frozen. Where a single C-C bond is cut, three H atoms are added to satisfy the C atom valency. The exact atoms which are frozen and truncated in the calculation are illustrated in Fig. S6. The second Mg^{2+} cation in the starting geometry of G1 (Fig. S4A) was placed based on alignment with the crystal structure of the Na^{+}/K^{+} -transporting ATPase, which has full sequence conservation within 4.5 Å of the Mg^{2+} ions and is resolved with ADP and two Mg^{2+} cations (PDB code: 3wgu) [32]. The starting structure of G3 was taken from a snapshot of the last nanosecond of the unconstrained MD simulation (Fig. S4C). In G2 and G4 the second Mg^{2+} ion was placed in such way to explore different coordination possibilities.

Quantum Mechanical/Molecular Mechanical Methods (QM/MM)

All QM/MM calculations were performed by Q-Chem [59], coupled with CHARMM [60]. The QM region contained: the ATP moiety, the Mg^{2+} cation/s, six water molecules, Asp513, Asp878, Thr515, Lys859 and Arg686. The QM region during the RCS was treated with the B3LYP hybrid density functional and 6-31+G* basis set [57,58]. The MM region contained all residues and solvent within 25 Å of the QM region. The residues included in the QM region were separated from the rest of the chain by cleaving homonuclear C-C bonds and introducing link atoms, which were treated as hydrogen atoms in the QM calculations. An initial energy minimization was carried out, which constituted 1000 steps via the SCF DIIS algorithm. Each minimized geometry was supplied for the RCS as a reactant state starting point. The active site containing two Mg^{2+} ions was obtained by aligning the ATP and second Mg^{2+} ion from the QM cluster optimization and translating the optimized geometry to the QM/MM model. The active site conformation containing one Mg^{2+} model was taken directly from the one observed in the crystal structure of the E1-ATP state of SERCA [29], however, substituting one carbon atom of the crystal structure ACP molecule to a phosphorus, in order to have the catalytically-active ATP. This structure was further minimized for 1000 steps. We defined the reaction coordinate by the distance from the reacting nucleophile to the reacting phosphate O_{2D}-P_G (R1) and the reacting phosphate and the leaving group P_G-O_{3B} (R2). Starting from the reactant state and moving along this coordinate, we simultaneously decrease the R1 distance and increase the R2 distance, to reach the product state. The distances were changed linearly. This forward-backward scanning is performed until energy convergence is observed. The solvent molecules in the QM region do not undergo reorganization from reactant to product state so the system has not been constrained additionally. 40 minimization steps were completed each time before a datapoint was recorded during the RCS. All presented scans

are converged and show the forward scan direction, going from the reactant to the product state. Six systems were independently minimized. The minimized structure of each was supplied for a starting structure (reactant) of the RCS. For the purpose of this manuscript, "all amino acids" refers to Asp513, Asp878, Thr515, Lys859 and Arg686. Each system studied and its corresponding QM-region overall charge and atoms is summarised in Table S2.

Pocket Analysis

Twenty frames were extracted from the 100 ns MD trajectory of the unconstrained simulation. The frames were spaced equidistantly and covered the duration of the MD simulation (Table S3). The twenty biggest pockets were calculated for each frame, using the Pymol [61] plugin PyVOL [62]. To find pockets on the surface of ATP13A2, PyVOL [62] was provided with the protein chain only without the ATP or any of the Mg^{2+} ions. The four biggest pockets in terms of surface area were chosen for further analysis. Upon inspection, it was observed that the biggest pockets for every frame were observed in the N-binding domain where ATP binding normally occurs, and in the transmembrane domain, respectively. For this work, we define a pocket as an "ATP-binding pocket" if the pocket was found in a location within 1.5 Å of the ATP in the respective frame from the simulation. We define a "transmembrane pocket" if a pocket is located within 1.5 Å of where small inorganic ion binding has been observed in the crystal structures [29,41] of homologous proteins. If a pocket was calculated by PyVOL [62] to be occupying this area of interest, it was recorded as found. If it was not calculated within this area, it was recorded as not found (Table S3). In this way, we were able to calculate the frequency of occurrence of those two pockets (Table S3). Other pockets which were found consistently are illustrated in S7. However, their occurrence was not as consistent as for the ATP pocket and the main transmembrane pocket. The results presented are obtained from the MD simulation of the homology model based on SERCA but the same analysis was performed on a homology model based on ATP13A1 and the main transmembrane pocket was consistently found as well.

Conclusions

In this work we present a structural model of the P5B enzyme ATP13A2 which has been complemented with MD, QM cluster and QM/MM calculations. This has allowed us to find an accurate conformation for the catalytically competent ATP structure and a reliable position for the second Mg^{2+} ion with respect to the ATP and the other Mg^{2+} active site cation. Using this information, we have subsequently calculated the barrier height for the ATP hydrolysis reaction in different active site conformations, including with one and two Mg^{2+} ions. We present strong evidence that the active site of the protein contains two divalent Mg^{2+} ions and operates via a two-ion catalytic mechanism. Additionally, we present the first quantitative analysis of the role of Arg686 and Lys859 on the barrier height of the ATP hydrolysis reaction. This work can suggest how missense mutations close to important active

site interactions in the respective catalytic domains, can have diminishing effect on the catalytic activity of the enzyme.
Additionally, we have analysed the surface of the protein for binding pockets and found two pockets which occur consistently. The pocket which appeared most consistently was found in the transmembrane domain of the protein. From the sequence analysis performed and the binding pocket calculations, we suggest that ATP13A2 also likely binds a substrate in this part of the transmembrane, other than an inorganic ion, which is consistent with the big size of the calculated pocket in this part of the transmembrane.

BIBLIOGRAPHY

[1] Park JS, Blair NF, Sue CM. The role of ATP13A2 in Parkinson's disease: Clinical phenotypes and molecular mechanisms. Mov Disord 2015. https://doi.org/10.1002/mds.26243.

[2] van Veen S, Sørensen DM, Holemans T, Holen HW, Palmgren MG, Vangheluwe P. Cellular function and pathological role of ATP13A2 and related P-type transport ATPases in Parkinson's disease and other neurological disorders. Front Mol Neurosci 2014. https://doi.org/10.3389/fnmol.2014.00048.

[3] Martin S, Holemans T, Vangheluwe P. Unlocking atp13a2/park9 activity. Cell Cycle 2015;14:3341–2. https://doi.org/10.1080/15384101.2015.1093420.

[4] Holemans T, Sørensen DM, Van Veen S, Martin S, Hermans D, Kemmer GC, et al. A lipid switch unlocks Parkinson's disease-associated ATP13A2. Proc Natl Acad Sci U S A 2015. https://doi.org/10.1073/pnas.1508220112.

[5] Podhajska A, Musso A, Trancikova A, Stafa K, Moser R, Sonnay S, et al. Common pathogenic effects of missense mutations in the P-type ATPase ATP13A2 (PARK9) associated with early-onset parkinsonism. PLoS One 2012. https://doi.org/10.1371/journal.pone.0039942.

[6] Bublitz M, Morth JP, Nissen P. P-type ATPases at a glance. J Cell Sci 2011. https://doi.org/10.1242/jcs.102921.

[7] Santoro L, Breedveld GJ, Manganelli F, Iodice R, Pisciotta C, Nolano M, et al. Novel ATP13A2 (PARK9) homozygous mutation in a family with marked phenotype variability. Neurogenetics 2011. https://doi.org/10.1007/s10048-010-0259-0.

[8] C.H. L, E.K. T, M.L. C, L.C. T, H.Q. L, G.S. C. Novel ATP13A2 variant associated with Parkinson disease in Taiwan and Singapore. Neurology 2008.

[9] H. S, S. M, D. D, J. C, F. G, P. B, et al. Kufor-Rakeb syndrome and a rare ATP13A2 missense mutation in a Portuguese patient. Mov Disord 2018.

[10] Yang X, Xu Y. Mutations in the ATP13A2 gene and Parkinsonism: A preliminary review. Biomed Res Int 2014. https://doi.org/10.1155/2014/371256.

[11] Di Fonzo A, Chien HF, Socal M, Giraudo S, Tassorelli C, Iliceto G, et al. ATP13A2 missense mutations in juvenile parkinsonism and young onset Parkinson disease. Neurology 2007. https://doi.org/10.1212/01.wnl.0000260963.08711.08.

[12] Abbas MM, Govindappa ST, Sheerin UM, Bhatia KP, Muthane UB. Exome Sequencing Identifies a Novel Homozygous Missense ATP13A2 Mutation. Mov Disord Clin Pract 2017. https://doi.org/10.1002/mdc3.12353.

[13] Schneider SA, Paisan-Ruiz C, Quinn NP, Lees AJ, Houlden H, Hardy J, et al. ATP13A2 mutations (PARK9) cause neurodegeneration with brain iron accumulation. Mov Disord 2010. https://doi.org/10.1002/mds.22947.
[14] Brüggemann N, Hagenah J, Reetz K, Schmidt A, Kasten M, Buchmann I, et al. Recessively inherited parkinsonism: Effect of ATP13A2 mutations on the clinical and neuroimaging phenotype. Arch Neurol 2010. https://doi.org/10.1001/archneurol.2010.281.
[15] Estrada-Cuzcano A, Martin S, Chamova T, Synofzik M, Timmann D, Holemans T, et al. Loss-of-function mutations in the ATP13A2/PARK9 gene cause complicated hereditary spastic paraplegia (SPG78). Brain 2017. https://doi.org/10.1093/brain/aww307.
[16] Behrens MI, Brüggemann N, Chana P, Venegas P, Kägi M, Parrao T, et al. Clinical spectrum of Kufor-Rakeb syndrome in the Chilean kindred with ATP13A2 mutations. Mov Disord 2010. https://doi.org/10.1002/mds.22996.
[17] Burroughs AM, Allen KN, Dunaway-Mariano D, Aravind L. Evolutionary Genomics of the HAD Superfamily: Understanding the Structural Adaptations and Catalytic Diversity in a Superfamily of Phosphoesterases and Allied Enzymes. J Mol Biol 2006. https://doi.org/10.1016/j.jmb.2006.06.049.
[18] Koonin E V., Tatusove RL. Computer analysis of bacterial haloacid dehalogenases defines a large superfamily of hydrolases with diverse specificity. Application of an iterative approach to database search. J Mol Biol 1994. https://doi.org/10.1006/jmbi.1994.1711.
[19] Aravind L, Galperin MY, Koonin E V. The catalytic domain of the P-type ATPase has the haloacid dehalogenase fold. Trends Biochem Sci 1998. https://doi.org/10.1016/S0968-0004(98)01189-X.
[20] Narayanaswamy N, Chakraborty K, Saminathan A, Zeichner E, Leung KH, Devany J, et al. A pH-correctable, DNA-based fluorescent reporter for organellar calcium. Nat Methods 2019. https://doi.org/10.1038/s41592-018-0232-7.
[21] De La Hera DP, Corradi GR, Adamo HP, De Tezanos Pinto F. Parkinson's disease-associated human P5B-ATPase ATP13A2 increases spermidine uptake. Biochem J 2013. https://doi.org/10.1042/BJ20120739.
[22] Heinick A, Urban K, Roth S, Spies D, Nunes F, Phanstiel O, et al. Caenorhabditis elegans P 5B -type ATPase CATP-5 operates in polyamine transport and is crucial for norspermidine-mediated suppression of RNA interference . FASEB J 2010. https://doi.org/10.1096/fj.09-135889.
[23] van Veen S, Martin S, Van den Haute C, Benoy V, Lyons J, Vanhoutte R, et al. ATP13A2 deficiency disrupts lysosomal polyamine export. Nature 2020. https://doi.org/10.1038/s41586-020-1968-7.
[24] Vilariño-Güell C, Soto AI, Lincoln SJ, Yahmed S Ben, Kefi M, Heckman MG, et al. ATP13A2 variability in Parkinson disease. Hum Mutat 2009. https://doi.org/10.1002/humu.20877.
[25] Odake Y, Koh K, Takiyama Y, Ishiura H, Tsuji S, Yamada M, et al. Identification of a novel mutation in ATP13A2 associated with a complicated form of hereditary spastic paraplegia . Neurol Genet 2020. https://doi.org/10.1212/nxg.0000000000000514.
[26] Thever MD, Saier MH. Bioinformatic characterization of P-Type ATPases encoded within the fully sequenced genomes of 26 eukaryotes. J Membr Biol 2009. https://doi.org/10.1007/s00232-009-9176-2.
[27] Toyoshima C, Nakasako M, Nomura H, Ogawa H. Crystal structure of the calcium

pump of sarcoplasmic reticulum at 2.6 Å resolution. Nature 2000. https://doi.org/10.1038/35015017.
[28] Toyoshima C, Mizutani T. Crystal structure of the calcium pump with a bound ATP analogue. Nature 2004. https://doi.org/10.1038/nature02680.
[29] Sacchetto R, Bertipaglia I, Giannetti S, Cendron L, Mascarello F, Damiani E, et al. Crystal structure of sarcoplasmic reticulum Ca 2+-ATPase (SERCA) from bovine muscle. J Struct Biol 2012. https://doi.org/10.1016/j.jsb.2012.02.008.
[30] Clausen JD, McIntosh DB, Vilsen B, Woolley DG, Andersen JP. Importance of conserved N-domain residues Thr441, Glu442, Lys515, Arg560, and Leu562 of sarcoplasmic reticulum Ca2+-ATPase for MgATP binding and subsequent catalytic steps. Plasticity of the nucleotide-binding site. J Biol Chem 2003. https://doi.org/10.1074/jbc.M301122200.
[31] Ogawa H, Shinoda T, Cornelius F, Toyoshima C. Crystal structure of the sodium-potassium pump (Na+,K +-ATPase) with bound potassium and ouabain. Proc Natl Acad Sci U S A 2009. https://doi.org/10.1073/pnas.0907054106.
[32] Kanai R, Ogawa H, Vilsen B, Cornelius F, Toyoshima C. Crystal structure of a Na + -bound Na +,K + -ATPase preceding the E1P state. Nature 2013. https://doi.org/10.1038/nature12578.
[33] McKenna MJ, Sim SI, Ordureau A, Wei L, Wade Harper J, Shao S, et al. The endoplasmic reticulum P5A-ATPase is a transmembrane helix dislocase. Science (80-) 2020. https://doi.org/10.1126/science.abc5809.
[34] Sørensen TLM, Møller JV, Nissen P. Phosphoryl transfer and calcium ion occlusion in the calcium pump. Science (80-) 2004. https://doi.org/10.1126/science.1099366.
[35] Komuro Y, Re S, Kobayashi C, Muneyuki E, Sugita Y. CHARMM force-fields with modified polyphosphate parameters allow stable simulation of the ATP-bound structure of Ca2+-ATPase. J Chem Theory Comput 2014. https://doi.org/10.1021/ct5004143.
[36] Li C, Yue Z, Espinoza-Fonseca LM, Voth GA. Multiscale Simulation Reveals Passive Proton Transport Through SERCA on the Microsecond Timescale. Biophys J 2020. https://doi.org/10.1016/j.bpj.2020.07.027.
[37] Krachtus D, Smith JC, Imhof P. Quantum mechanical/molecular mechanical analysis of the catalytic mechanism of phosphoserine phosphatase. Molecules 2018. https://doi.org/10.3390/molecules23123342.
[38] Nagy GN, Suardíaz R, Lopata A, Ozohanics O, Vékey K, Brooks BR, et al. Structural Characterization of Arginine Fingers: Identification of an Arginine Finger for the Pyrophosphatase dUTPases. J Am Chem Soc 2016. https://doi.org/10.1021/jacs.6b09012.
[39] Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, et al. Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. Nucleic Acids Res 1997. https://doi.org/10.1093/nar/25.17.3389.
[40] Pearson WR, Lipman DJ. Improved tools for biological sequence comparison. Proc Natl Acad Sci U S A 1988. https://doi.org/10.1073/pnas.85.8.2444.
[41] Kanai R, Ogawa H, Vilsen B, Cornelius F, Toyoshima C. Crystal structure of a Na + -bound Na +,K + -ATPase preceding the E1P state. Nature 2013. https://doi.org/10.1038/nature12578.
[42] Klähn M, Rosta E, Warshel A. On the mechanism of hydrolysis of phosphate monoesters dianions in solutions and proteins. J Am Chem Soc 2006.

https://doi.org/10.1021/ja065470t.
[43] Kamerlin SCL, Wilkie J. The effect of leaving group on mechanistic preference in phosphate monoester hydrolysis. Org Biomol Chem 2011. https://doi.org/10.1039/c0ob01210f.
[44] Lopata A, Jambrina PG, Sharma PK, Brooks BR, Toth J, Vertessy BG, et al. Mutations decouple proton transfer from phosphate cleavage in the dutpase catalytic reaction. ACS Catal 2015. https://doi.org/10.1021/cs502087f.
[45] Zalatan JG, Herschlag D. Alkaline phosphatase mono- and diesterase reactions: Comparative transition state analysis. J Am Chem Soc 2006. https://doi.org/10.1021/ja056528r.
[46] Cohen Y, Megyeri M, Chen OCW, Condomitti G, Riezman I, Loizides-Mangold U, et al. The yeast P5 type ATPase, Spf1, regulates manganese transport into the endoplasmic reticulum. PLoS One 2013. https://doi.org/10.1371/journal.pone.0085519.
[47] Sørensen DM, Buch-Pedersen MJ, Palmgren MG. Structural divergence between the two subgroups of P5 ATPases. Biochim Biophys Acta - Bioenerg 2010. https://doi.org/10.1016/j.bbabio.2010.04.010.
[48] Hancock JM, Zvelebil MJ, Zvelebil MJ. UniProt. Dict. Bioinforma. Comput. Biol., 2004. https://doi.org/10.1002/9780471650126.dob0721.pub2.
[49] Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. J Comput Chem 2005. https://doi.org/10.1002/jcc.20289.
[50] Pastor RW, MacKerell AD. Development of the CHARMM force field for lipids. J Phys Chem Lett 2011. https://doi.org/10.1021/jz200167q.
[51] Darden T, York D, Pedersen L. Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems. J Chem Phys 1993. https://doi.org/10.1063/1.464397.
[52] MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. J Phys Chem B 1998. https://doi.org/10.1021/jp973084f.
[53] Jo S, Kim T, Iyer VG, Im W. CHARMM-GUI: A web-based graphical user interface for CHARMM. J Comput Chem 2008. https://doi.org/10.1002/jcc.20945.
[54] Kräutler V, Van Gunsteren WF, Hünenberger PH. A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. J Comput Chem 2001. https://doi.org/10.1002/1096-987X(20010415)22:5<501::AID-JCC1021>3.0.CO;2-V.
[55] Lomize MA, Pogozheva ID, Joo H, Mosberg HI, Lomize AL. OPM database and PPM web server: Resources for positioning of proteins in membranes. Nucleic Acids Res 2012. https://doi.org/10.1093/nar/gkr703.
[56] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Had and DJF, Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, et al. Gaussian 09, Revision D.01. Gaussian Inc, Wallingford 2013. https://doi.org/10.1017/CBO9781107415324.004.
[57] Becke A. B3LYP. J Chem Phys 1993.
[58] Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, Cheeseman JR, et al. Gaussian 03, revision B. 05; Gaussian. Inc, Pittsburgh, PA 2003.
[59] Shao Y, Gan Z, Epifanovsky E, Gilbert ATB, Wormit M, Kussmann J, et al. Advances in

molecular quantum chemistry contained in the Q-Chem 4 program package. Mol Phys 2015. https://doi.org/10.1080/00268976.2014.952696.
[60] Brooks BR, Brooks CL, Mackerell AD, Nilsson L, Petrella RJ, Roux B, et al. CHARMM: The biomolecular simulation program. J Comput Chem 2009. https://doi.org/10.1002/jcc.21287.
[61] Schrödinger LLC. The PyMOL Molecular Graphics System, Version 2.4. Schrödinger LLC 2020.
[62] Smith RHB, Dar AC, Schlessinger A. PyVOL: A PyMOL plugin for visualization, comparison, and volume calculation of drug-binding sites. BioRxiv 2019. https://doi.org/10.1101/816702.