Conformational transitions and alternating-access mechanism in the sarcoplasmic reticulum calcium pump

Avishek Das¹, Huan Rui¹, Robert Nakamoto² & Benoît Roux^{1*}

¹Department of Biochemistry and Molecular Biology, The University of Chicago, 929 East 57th Street, Chicago Illinois 60637, USA

²Department of Molecular Physiology and Biological Physics, University of Virginia School of Medicine, PO Box 800886, 480 Ray C. Hunt Drive, Charlottesville, Virginia 22908, USA

*Email: roux@uchicago.edu Phone: 773-834-3557

Keywords: SERCA, P-type ATPase, Ion Pumps, Conformational Transition, Molecular Dynamics Simulation, String Method

Abstract

Ion pumps are integral membrane proteins responsible for transporting ions against concentration gradients across biological membranes. Sarco/endoplasmic reticulum Ca^{2+} -ATPase (SERCA), a member of the P-type ATPases family, transports two calcium ions per hydrolyzed ATP molecule via an “alternating-access” mechanism. High-resolution crystallographic structures provide invaluable insight on the structural mechanism of the ion pumping process. However, to understand the molecular details of how ATP hydrolysis is coupled to calcium transport, it is necessary to gain knowledge about the conformational transition pathways connecting the crystallographically resolved conformations. Large-scale transitions in SERCA occur at time-scales beyond the current reach of unbiased molecular dynamics (MD) simulations. Here, we overcome this challenge by employing the string method, which represents a transition pathway as a chain-of-states linking two conformational end-points. Using a highly scalable multiscale methodology, we have determined all-atom transition pathways for three main conformational transitions responsible for the alternating-access. The present pathways provide a clear chronology of the key events underlying the active transport of calcium ions by SERCA. Important conclusions are that the conformational transition that leads to occlusion with bound ATP and calcium is highly concerted and cooperative, the phosphorylation of Asp351 causes a reorganization of the cytoplasmic domains that subsequently drives the opening of the luminal gate, and re-closing of luminal gate induces a shift in the cytoplasmic domains that subsequently enables dephosphorylation of Asp351-P. Transient residue-residue contacts during the conformational transitions predicted by the computations provide an experimental route to test the general validity of the computational pathways.

Introduction

P-type ATPases [1, 2] form a large class of transmembrane ion pumps named upon their ability to catalyze the phosphorylation of a key conserved aspartate residue (Asp351) within the pump by hydrolyzing adenosine triphosphate (ATP) used as a source of chemical energy to actively transport ions against their concentration gradient across a biological membrane. An important member of the P-type ATPase family is the SERCA calcium pump [2-5], which pumps two Ca^{2+} ions per ATP molecule from the calcium-poor cytoplasm to the calcium-rich lumen of the sarcoplasmic reticulum (SR). SERCA has been extensively characterized by enzymatic, thermodynamic as well as structural analyses, which led to the resolution of multiple high-resolution X-ray crystal structures [2-5]. The overall topology of the protein is depicted in Fig. 1. It consists of a single chain of 994 amino acids comprising three cytoplasmic domains and one transmembrane (TM) domain (Fig. 1A). The cytoplasmic headpiece comprises three regions denoted as the nucleotide binding (N), phosphorylation (P) and actuator (A) domains. The TM domain is made of ten TM helices (labeled M1 to M10). During the transport cycle, the γ -phosphate of ATP bound to the N domain is transferred to Asp351 located in the P domain, which is almost 50 Å away from the Ca^{2+} binding sites located within the TM region (Fig. 1A).

Like other P-type ATPase, SERCA functions according to an “alternating-access” mechanism, whereby the ion binding sites are accessible from only one side of the membrane at a time while the protein undergoes a transport cycle. The alternating-access mechanism [6, 7] posits that the protein toggles between two main conformational states, denoted by E1 and E2. In the context of Ca^{2+} transport by SERCA, the E1 and E2 states correspond to states with high and low affinity for Ca^{2+} ions, respectively.

In the light of the available crystal structures [8-18], the transport cycle of the pump may be briefly described by the following elementary steps (Fig. S1B). Starting in a state facing the cytoplasmic side (E1), the pump binds 2 cytoplasmic Ca^{2+} ions in a positive cooperative manner and one ATP molecule with high affinity, leading to occlusion (E1-2 Ca^{2+} -ATP). Auto-phosphorylation of Asp351 in the P domain (E1P-2 Ca^{2+} -ADP) is followed by a large conformational change that renders the ion binding sites accessible to the luminal environment (E2P) [16, 17]. Two Ca^{2+} ions are then released and 2-3 luminal protons enter the binding site, causing occlusion of the protons and dephosphorylation (E2-Pi). Release of the bound inorganic phosphate (Pi) yields the dephosphorylated proton-occluded conformation E2, which then returns to the cytoplasm-facing conformation E1 to initiate a new cycle of the pump.

While X-ray crystallography provides high-resolution snapshots about the stable experimentally resolved states of the transport cycle (Fig. S2), it is very difficult to detect the short-lived intermediate conformations occurring transiently during the transport cycle. As a result, there is generally a paucity of information about many of the key microscopic events underlying the mechanism of the pump [19-25]. Computational methods can be used to help supplement the missing information by providing atomic models for the transient intermediates along the transport cycle. Our goal with the present effort is to elucidate the details of the alternating access mechanism in SERCA by simulating the large-scale conformational transitions between the experimentally resolved stable states. The main question we seek to answer in the paper is the following. In the light of the available crystal structures, how one can explain the relationship between the active transport of Ca^{2+} ions and the hydrolysis of ATP catalyzed by the same molecule. This broad question can be further subdivided into several more pointed questions:

- 1) Why is the hydrolysis of ATP strongly dependent on the binding of both the Ca^{2+} ions?

- 2) What is the relation between the hydrolysis of ATP and the occlusion of the Ca^{2+} ions?
- 3) How is phosphorylation of Asp351 driving the luminal opening?
- 4) What is the nature of coupling between the closing of luminal gate and the dephosphorylation of the phosphorylated Asp351 (Asp351-P)?

By answering these questions, we hope to elucidate the long-range information transfer between the catalytic site and the TM binding sites that are separated by more than ~ 50 Å. The functionally relevant conformational transitions take place on time-scales that are far beyond what can be attained with unbiased MD simulations of large systems [20]. To circumvent this challenge, we use a specialized computational approach called the “string method” [26-28] to determine the transition pathway between pairs of stable endpoint states as a chain of evenly distributed “image” conformations.

Results

Alternating access of binding sites in the stable states

Key structural features underlying the coupling between the N, A, and P cytoplasmic domains and the Ca^{2+} binding sites in the transmembrane region for four functionally important states, E1 [8], E1-2 Ca^{2+} -ATP [9], E2P [16], and E2 [8], are displayed in Fig. 2. The four states are depicted schematically in Fig. 2A, the alternating access pathway highlighted by water molecules from MD simulations is shown in Fig. 2B, and a view of the three cytoplasmic domains is provided in Fig. 2C. The cytoplasm-facing E1 state presents a large opening between the M1 and M3 helices, providing a wide aqueous pathway for Ca^{2+} diffusion and binding from the cytoplasmic side (Fig. 2B, extreme left). The E1-2 Ca^{2+} -ATP and the E1P-2 Ca^{2+} -ADP states are occluded states with bound Ca^{2+} ions in the TM binding site, as evident from the lack of water density from either side of the membrane (Fig.

2B, second from the left). In the E2P state, a large water-filled lumen-facing access channel is created between the M1, M2, M4 and M6 helices, but no water density is found near the cytoplasmic pathway i.e. between the M1 and M3 helices (Fig. 2B, second from the right). Finally, the E2 state with the protonated binding sites may be considered as a proton-occluded state, since there is no continuous water density leading to the binding sites from either side of the membrane, indicating that both gates are closed (Fig. 2B, extreme right).

At a coarse level, the alternating access pathways towards the ion binding sites controlled by the arrangement of the TM helices directly correlates with the large variations in the arrangement of cytoplasmic domains for the four functionally important states (Fig. 2B and C). However, considerations based only on static structures associated with the main steps of the transport cycle are ultimately insufficient to understand the molecular details that are responsible for the long-range coupling between the cytoplasmic domains and the TM region underlying the alternating access mechanism. To address this issue, we have determined computational transition pathways between the stable states along the pumping cycle using the string method. The string method represents a pathway between two endpoints as a chain of evenly distributed discrete snapshots called images [26, 27]. All paths comprise 32 images. Starting from an initial pathway, we use the mean drifts of each image, determined by unbiased MD simulations from swarms-of-trajectories, to iteratively refine the string in the subspace of collective variables until convergence is reached [28]. By virtue of the thermal noise added from the swarms-of-trajectories, the algorithm samples the ensemble of plausible (high probability) paths in the interior of an effective multi-dimensional least free energy “reaction tube” between the two endpoints [29, 30]. For this reason, all the results presented here are averages from the last 20 iterations of the string with swarms-of-trajectories protocol (including error bars). In the following sections, we will focus on the molecular details of conformational transition pathways underlying specific step along the transport cycle (Fig. S1B): (1) cytoplasmic

calcium binding and occlusion (E1 to E1-2Ca²⁺-ATP), (2) luminal opening and calcium release (E1P-2Ca²⁺-ADP to E2P) and (3) closing of the luminal gate and dephosphorylation (E2P to E2-Pi). These three transitions involve the main steps in the alternating-access mechanism and are most relevant for understanding the relationship between the ATP hydrolysis and active transport. The transition between E2-Pi and E2 states does not involve major changes in the TM region and not important for alternating-access. The final leg of the transport cycle between E2 and E1 involves proton counter-transport and will be the subject of a separate study.

Convergence of the string pathways

For proper convergence to a physically meaningful pathway, it is essential to start from a reasonable initial guess. Our multiscale approach with the coarse-grained (CG) pathway ensures that our initial path has a reasonable description of the large-scale domain motions [31]. The subsequent refinement provided by the all-atom string method can then be expected to capture the atomic details at the residue level and correct the domain arrangement beyond the simple CG initial path. The combined procedure systematically builds complexity on top of a simpler model and provides the optimal description of the real system with the best use of computational resources and within the limited accuracy of the classical non-polarizable force fields. The convergence of the all-atom pathway by the swarms-of-trajectory string method was critically assessed by monitoring the evolution of the collective variables and the root mean squared deviations (RMSDs) of the backbone C^α of each image calculated from the initial CG pathway during the iterative refinement. The image C^α RMSDs are shown in Fig. 3 for the three transition pathways. One may note that the images at the ends of the pathways display the smallest changes. These images display a RMSD on the order of 1-2 Å relative to the initial conformation (the collective variables for the end points were not restrained during the swarms-of-trajectory optimization procedure), which is reasonable since these images correspond to high-resolution X-ray structures that are expected to be very stable. Interestingly, the evolution of

the all-atom strings appears to be clearly non-monotonic, with different parts of the pathways displaying much larger changes than other parts relative to the initial CG pathways. For example, for the string between E1 and E1-2Ca²⁺-ATP states, the RMSD of image 14 end up deviating by more than 3 Å relative to the initial guess. Similarly, image 22 ends up deviating by more than 5 Å RMSD for the pathway between the E1P-2Ca²⁺ and E2P states, and image 26 ends up deviating by almost 3.5 Å RMSD for the pathway between the E2P and E2-Pi states. It is apparent from Fig. 3 that the images of each string pathway evolved vigorously during the optimization cycle, with several images changing by ~3 Å RMSD. The magnitude of the RMSD is strongly indicative that the optimization procedure based on the swarms-of-trajectories did not simply stall at the nearest minimum. As an additional diagnostic of convergence, the RMSD was calculated for the last five iterations using the final status of the string pathway as the reference. For all three transitions, the RMSD values are in within the 0.3 to 0.5 Å, providing further evidence of that the iterative optimization of the string is converged. Further tests could be carried out to rigorously assess the robustness of the convergence, although most would be computationally prohibitive due to the system size and complexity. Our working assumption is that the swarms-of-trajectories algorithm effectively samples the set of probable paths within the interior of the dominant “reaction tube” (~3 Å RMSD) underlying the three conformational transitions studied here [30]. The assumption that the existence of widely different “transition tubes” for these functional transitions is unlikely, is generally supported by a number of previous studies suggesting that conformational transitions in proteins typically proceed via a well-defined and ordered series of microscopic events [30-36].

Cytoplasmic calcium binding and occlusion (E1 to E1-2Ca²⁺-ATP)

Under normal physiological conditions, the pump is activated by the binding of Ca²⁺ ions and ATP, leading to the occlusion of the pump and the hydrolysis of ATP. A complete chronology of the

binding of Ca^{2+} ions and the concomitant large-scale conformational changes leading to the occluded state are provided by the string pathway between the E1 and the E1-2 Ca^{2+} -ATP states. The first state is referred as E1 because it was constructed from the crystal structure of the E1 state without calcium (PDB ID: 3W5A). The main events along the pathway are analyzed by monitoring the evolution of a few key order parameters as shown in Fig. 4. Illustrative snapshots taken at critical stages of the transition pathway are presented in Fig. 5. The complete transition pathway may also be visualized in supplementary movies S1 and S2.

The results from Fig. 4 suggest that there are several stages to the conformational transition, comprising N-A and N-P rearrangements (images 1 to 20 in Fig. 4A and B), motions of hydrophobic residues in the TM region (images 12 to 20 in Fig. 4C), and dehydration of the ion binding sites (images 15 to 20 in Fig. 4D). At an early stage corresponding roughly to images 1 to 11 in Fig. 4A, the ATP molecule bound to the N domain is not yet interacting with the P domain. As the N domain approaches the P domain through a hinge-like motion (movie S1), the ATP progressively forms new interactions with several residues in the P domain, including Asp351 (images 7 to 12 in Figs. 4A). As shown in Fig. 4B, the movement of the A domain is accompanied by a reorganization of the interface between the N and A domains, with some initial contacts being broken and new ones being formed (the change in N-A interactions are detailed in Fig. S4B). Furthermore, the side-chain of Asn706 flips from facing the outside of the protein to the interior of the ATP binding pocket in the P domain (Fig. S5) and interact directly with the ATP molecule. The cumulative effects of these interactions help position the ATP near Asp351. At image 13 in Fig. 4A, the Mg^{2+} ion bound to ATP comes to close proximity to Asp351, and by image 16, the interaction of ATP and Mg^{2+} ion with the P domain is almost complete (Figs. 5A). This triggers some local structural rearrangements in the P domain near Asp351: The hydroxyl side-chain of Thr355 breaks its contacts with Asp351 backbone

oxygen, a hydrogen bond between Thr357 and Lys352 is also broken, and Lys352 forms a new interaction with Asp627 (image 16 in Fig. S5).

Immediately after the changes in the P domain, the side-chain of Glu309 flips and brings Ca(II) inside the membrane. The ion progressively forms interactions with its binding partners from the M4 helix (Val304, Ile307, Ala305) and M6 helix (Asn706) (Fig. S8). The complete formation of the second binding site induces subtle changes in the cytoplasmic part of the M4 helix, which affect the P domain through P-M4 junction (Figs. 5A, S6). As a result, the interaction between backbone atoms of Thr355 and Phe740 is broken, as shown in image 26 of Fig. 5A, and the loop containing Asp351 moves away from the P-M5 junction. Because of this, Thr353 backbone carbonyl can coordinate with the Mg^{2+} ion completing the proper coordination at the catalytic site. By image 26, the Ca^{2+} ions are fully coordinates in their binding sites, and the catalytic site with bound ATP, equally shared by the N and P domains, has been fully formed (Fig. S8).

After completion of new N-P interface mediated by ATP, the A domain completes its rigid body rotation. This results in significant movements of the M1 and M2 helices toward the cytoplasmic side, at the same time M2 helix moves sideways and changes its internal structure (image 26 in Fig. S7 and movie S2). During the motion of the helices, the side-chains of several large hydrophobic residues (Leu65, Ile94, Ile97) on the M1 and M2 helices move in close proximity to the side-chain of Glu309 from the second ion binding site [17]. As shown by Figs. 4C and 4D, these events, which take place rapidly from images 13 to 20, coincide with the dehydration of the Ca^{2+} binding sites (see also movie S2). Beyond this point, the system is essentially in the Ca^{2+} -bound occluded state shown in Fig. 5B.

The results indicate that the binding of Ca^{2+} might be required to reach the final occluded state. To further buttress this assertion, an unbiased 50 ns MD simulation of the occluded end-state

was carried out without the bound Ca^{2+} ions. This brief simulation shows that the final occluded state is unstable in the absence of bound Ca^{2+} : water molecules enter the binding site to hydrate the crucial Ca^{2+} -binding residue Glu309 and there is a downward shift of the M1 helix. In comparison with the images along the transition string pathway, image 19 most closely resembles this Ca^{2+} -free state with respect to the conformation of M1 helix. This corresponds to an intermediate stage along the transition pathway at which point the catalytic site between the N and P domains is still not fully formed. This strongly suggests that—without the bound Ca^{2+} ions—the transition pathway would get stuck around image 19 with the downward shifted helix M1, unable to reach the occluded state with proper N-P catalytic site for the phosphoryl transfer reaction.

The access pathway of the ions from the cytoplasm to the TM binding sites at the early stage of the transition is of interest. It should be noted that the coordinates of the Ca^{2+} ions were included in the set of CVs used in the string method, so they do not diffuse freely but evolve along the optimal pathway connecting the two end-states. Here, the initial end-state E1, generated by steered molecular dynamics (SMD), has Ca(II) in the cytoplasmic solution, interaction with Glu309 and Asp59 on the kinked part of the M1 helix (Fig. 5B). The Ca(I) is located between the two binding sites. Following the motions of the ions along the optimized computational pathway, the following picture emerges. The in-coming Ca^{2+} ions are first attracted by negatively charged residues on the M1 and M3 helices, including Asp59, and bind to the Glu309 when it is projecting its side-chain away from the binding site and toward the cytoplasmic solution. Subsequent conformational change of the same side-chain brings the bound ion inside the membrane where it starts to interact with other binding site residues. This picture is supported by the biochemical evidence in favor of sequential Ca^{2+} binding [37].

Luminal opening and calcium release (E1P-2 Ca^{2+} -ADP to E2P)

Once the protein is stably locked in the occluded E1-2Ca²⁺-ATP state for an extended period of time, the phosphoryl transfer chemical reaction between ATP and Asp351 is expected to occur spontaneously, bringing the pump to the E1P-2Ca²⁺-ADP state (this reaction is not simulated here). This leads us to the next step in the transport cycle, which is the isomerization and conversion of the phosphorylated pump into the E2P state, where the luminal access channel is open (Fig. 2B, second from the right). The character of the E1P-2Ca²⁺-ADP state was first investigated by carrying out an unbiased MD simulation. The purpose of this initial simulation is to stabilize and equilibrate the end-state before the calculation of the transition pathway with the string method. During the unbiased MD simulation, the N domain moves away from the P domain and the ADP molecule dissociates from the binding site, while maintaining some interactions with Arg489 and Arg560 (Fig. S10). The repulsion between the phosphorylated Asp351 and the phosphate group of ADP contributes to the breaking up of the N-P complex and initiates the subsequent large-scale conformational transition. Again, the main events along the pathway are characterized by several order parameters. The results are shown in Fig. 6 and snapshots taken at key stages of the transition pathway are presented in Fig. 7. The transition pathway can also be viewed in supplementary movies S3 and S4.

At the early stage of the computational pathway (images 1-3), some concerted motion of the entire cytoplasmic headpiece and the cytoplasmic extension of the TM helices are observed. After the initial stage, the A domain begins to rotate independently, breaking several contacts with the N domain to form new contacts as shown in the profiles of the order parameters characterizing the N-A interactions in Fig. 6A (also see images 17, 29 in Figs. S11, S12). Interestingly, the dynamically evolving N-A interface also involves some transient interaction, e.g., between Glu439 and Gly217 (see Fig. 6A, grey curve). The large rotational motion of the A domain is correlated with structural rearrangement in the A-M3 junction loop which progressively folds and forms a hydrophobic cluster with the P7 and P8 helices in the P domain (images 17, 20 in Fig. S14 and movie S3). There are

significant changes in the junction region of A-M2 and cytoplasmic part of M1 happening at the same time as the large movement of the A domain (images 17, 20 Fig. S14). The A-M2 interface becomes more helical and a hydrophobic interface is formed involving several residues in this region (118 to 122) with the P domain and P-M4 junction (Fig. S14). The top part of M1, which is amphiphilic and partially structured in the occluded state, progressively refolds into a helix (image 17, 20, 29 in Figs. S14 and S16).

The rotation of the A domain mentioned above is accompanied by a reorganization of the P-A complex via creation of new interactions between the TGES loop (Thr181-Gly182-Glu183-Ser184) in the A domain and the region surrounding the phosphorylated Asp351 residue in the P domain (movie S3). This takes place roughly between images 10 and 22 in the order parameter profiles depicted in Fig. 6B (also see images 17, 29 in Fig. 7A, images 17, 20, 29 in Fig. S13). The broken contacts involve Gly156, Asn213 in A domain and Thr724, Ala725, Val726 and Thr729 in the P domain (Fig. 6B). The new P-A contacts near the catalytic site involve Ser178, Ile179, Leu180, Thr181, Gly182, Glu183, and Ser184 in the A domain, and Asp351, Thr353, Gly626, Asp627, Arg678, and Asn706 in the P domain (Fig. 6B, images 17, 20, 29 in Fig. S13). Other P-A interactions form between Asp176, Ile235, Met239, and Thr242 in the A domain, and Gly704, Val705, and Lys712 in the P domain. These contacts result from hydrogen bonding, electrostatic and van der Waals interactions between both side-chain and backbone atoms. Fig. 6A shows the interactions between the TGES loop and the catalytic site region of the P domain. The detailed analysis of these interactions reveal that interacting partners from the P domain are the same residues that were involved in the binding of ATP and formation of the catalytic center for phosphorylation in the previous step (Fig. S13). In this process, the TGES loop displaces the ADP and shields the phosphorylated Asp351 from the cytoplasmic solution preventing the back reaction with free or loosely bound ADP as seen by a comparison of images 5 and 19 in Figs. 6A and S16. By image 22

along the computational pathway, the ADP is quite far away from the catalytic site and Asp351-P is completely shielded from the cytoplasmic solution. It is important to note that there is a clear sequence of motions that result in changes in N-A and P-A interfaces, which cannot be inferred from the static structures of the end-states.

This large-scale reorganization of the cytoplasmic N, P and A domains essentially takes place over the first half of the computational pathway (images 1 to 15). It is followed, in the second half (images 16 to 32), by significant structural rearrangements within the TM region. Those structural rearrangements are mediated by the M1, M2 and M3 helices, which are covalently connected to the A domain (images 17 and 20 in Fig. S14). The cytoplasmic part of the M1 helix, which is partially unfolded in the E1P-2Ca²⁺ state, first unfolds further and then progressively folds after being pulled by the A domain through the long A-M1 linker to adopt an α -helical conformation parallel to the membrane surface in the head-group region (compare images 17, 20, 29 in Fig. S14, images 17, 29 in S16 and movie S3). The M2 helix partially straightens and goes up in the membrane (Figs. S14, S16 and movie S3). The M3 helix moves sideways and up in the membrane, dragging the luminal end of M4 to which it is covalently connected (images 17, 29 Fig. S16), while the cytoplasmic end of M4 remains in a stable position. The motion of M4 disrupts the Ca²⁺ binding sites, first it breaks interactions between the Ca(II) and the carbonyl oxygen of Ile307 (image 17 in Fig. S15A, Fig. S15C), this is followed by rearrangement in both site, e.g. Asp800 no longer interacts with Ca(II) and solely interacts with Ca(I) (images 17, 20 in Fig. S15A, Figs. S15 B, C, D). These changes lead to new coordination environments for Ca²⁺ ions (image 29 Fig. S15A), though still without direct exposure to the luminal solution. Immediately after this, further movements of luminal part of M4 away the M6 and M5 helices breaks several hydrophobic contacts involving these three helices (Figs. 6C and 6D, image 29 in Fig. 7B and movie S4). Along with these important changes, slight upward movements of M1 and M2 create a wide aqueous opening towards the luminal side between

the M1, M2, M4 and M6 helices (Fig. 7B and movie S4). The motion of M4 with respect to M5 and M6 are quantified in Fig. 6C, which depicts the opening of the luminal gate. For a pictorial representation of this important event compare images 17 and 29 in Fig. 7B. The opening created by the motion of M4 is perfectly correlated to the hydration of the binding sites as captured in Fig. 6D. Inspection of the images 17 and 29 in Fig. 7B shows that these water molecules come from the luminal side, providing the final evidence of luminal opening.

Ion dissociation and protonation of the E2P state

We examined the character of the E2P-2Ca²⁺ state by carrying out a 50 ns unbiased MD simulation. During the simulation, a slight rearrangement of the lumenally-exposed Ca²⁺ binding sites is observed, although both Ca²⁺ ions remain bound (Fig. S17). In the new binding sites, the Ca(I) ion is coordinated by both side-chain oxygens of Asp800, side-chains of Asn796, Glu771 and three water molecules (Fig. S17). The Ca(II) ion interacts with side-chains of Glu309, Asn768, Glu771 backbone carbonyl of Ala305 and two to three water molecules (Fig. S17). Recent biochemical experiments have suggested that exchange of H⁺ and Ca²⁺ ions does take place in the luminal release of Ca²⁺ [38]. To further investigate the role of H⁺ countercharge at this stage, we carried out a number of additional unbiased MD simulations of the calcium-bound E2P state considering different protonation states for the acidic residues Glu309, Glu771 and Asp800 that are part of the ion binding sites. Protonation of Glu309 or Asp800 leads to the significant movement one Ca²⁺, which comes off the binding site and interacts with Glu90 on the luminal end of the M2 helix (Fig. S18). Protonation of Glu771 does not induce any appreciable ion movement (Fig. S18). Double protonation of Glu309 and Asp771 results in leaving of Ca(II) but not Ca(I). Triple protonation of Gu309, Glu771 and Asp800 leads to significant movement of both ions toward the lumen. It appears that the protonation of Asp800 is essential for departure of both ions, even though it is well established that in the

unphosphorylated E2 states, this residue is deprotonated [21]. It is possible that Asp800 releases the proton to the luminal side upon initial protonation and Ca^{2+} release, as it is constantly interacting with the luminal aqueous medium. The simulations highlight the importance of shifts in pKa values of the binding site residues Glu309, Glu771 and Asp800 in the luminal release of Ca^{2+} ions.

Closing of luminal gate and dephosphorylation (E2P to E2-Pi)

After the release of Ca^{2+} ions, the next step in the transport cycle corresponds to the transition from the E2P state to the E2-Pi state, comprising two important molecular events: the closing of the luminal gate, and the dephosphorylation of Asp351 in the P domain. Of importance for the dehydration of the binding sites and the successful closing of the luminal gate and, residues Glu309, Glu771 and Glu908 were protonated at this stage. The profiles of the order parameters characterizing the key events in the cytoplasmic and the TM regions are shown in Fig. 8 and relevant snapshots are presented in Fig. 9. The complete transition pathway can be visualized in supplementary movies S5 and S6. During the initial stage of the transition (images 1 to 8), the entire cytoplasmic headpiece along with the cytoplasmic part of M5 moves slightly (Fig. S19 and movie S5). Then, there is an unfolding-to-folding transition at the junction of the helix M2 and the A domain and folding-to-unfolding transition at the top of the M2 helix between images 10 and 22. These events are key triggers for large-scale changes in the protein conformation along this transition. The structural changes at the A-M2 junction initiates a rotation of the A domain (compare images 18, 26 in Figs. S22, S24). The M2 helix shifts towards the luminal side, inducing a similar motion in the covalently connected M1 helix (compare images 16, 18, 26 in Figs. S22, S24). Simultaneously, the luminal part of M4 starts to move to make contact with the luminal part of M1. Two important hydrogen bonds in the TM ion binding sites region are formed involving residues Glu771 along M5 with Asn796 along M6, and Glu309 along M4 with Asn796 along M6 (compare images 16, 18 in Fig. S23). The

movements of the M4 with respect to the M5 and M6 helices are captured in Fig. 8C. The latter, together with Fig. 8D monitoring the dehydration of the Ca^{2+} binding sites (devoid of ions), shows that the motions of M4 are correlated with the dehydration of the ion binding sites (movie S6). The two correlated processes may be visualized by comparing images 18 and 26 in Fig. 9B. The same images reveal that luminal end of M4 restores a more compact packing with the M1, M2, M5 and M6 helices, a process that ultimately results in the closing of the luminal gate (image 26 in Figs. 9B, S24 and movie S6).

Once the conformational changes in the TM region resulting in the closing of the luminal gate have taken place (around image 20), the A domain moves further toward N domain through a rigid-body rotation (image 26 in Fig. S19). While there are no important changes in the N-P interface, the N-A interface is significantly reorganized. The change in the N-A interface is captured by the order parameter profiles in Fig. 8A. These events occurring in the TM region drive major changes in the P-A interface, ultimately creating the proper orientation of crucial residues to favor the dephosphorylation reaction. The changes in the P-A interface is presented in Fig. 8B. As depicted in Fig. 9A, the TGES loop moves away from the phosphorylated Asp351 (also see Fig. S19 and movie S5). As illustrated in Figs. 8B and S21, several important interactions between the P domain (phosphorylated Asp351, Thr353, Asp601, Asp627, Arg678, Ser682, Asp703, and Gly704) and the A domain (Ile179, Gly182, Glu183, Ser184, Gln238, and Thr242) are lost while some new contacts are formed (Arg656, Glu680, Asn706, Lys712, Thr724 in the P domain, and Tyr122, Ile179, Arg198, Asn201, Ala241 in the A domain). In this new position of the A domain, the phosphorylated Asp351 is no longer shielded from the cytoplasmic aqueous medium (image 26 Figs. 9A, S21 and movie S5); there is a narrow access toward the phosphate group, which interacts directly with a few cytoplasmic water molecules. The N-A and P-A interactions place the TGES loop such that Glu183 can hydrogen bond to a water molecule, which can then serve as a nucleophile

for the dephosphorylation reaction [18, 39]. These two events are coupled because the movements of the M4, M2 and M1 helices close the gate at the luminal end, and induce a rotation in the A domain to poise the chemical surrounding of Asp351-P for the subsequent dephosphorylation reaction, leading ultimately to the E2-Pi state. It should be noted that the actual chemical reaction step is not simulated here, but would be expected to occur spontaneously once the favorable conditions for the dephosphorylation reaction appear after the closing of the luminal gate.

Discussion

Because many of the transitions underlying the transport cycle of SERCA take place on time-scales that are far beyond what can be attained with unbiased MD simulations, we use a specialized computational approach called the string method, to determine the transition pathway connecting stable endpoint conformations. The string method represents a pathway between two endpoints as a chain of evenly distributed discrete snapshots (the “images”) within a subspace of collective variables [27, 28]. Starting from an initial pathway, we use the mean drifts of each image, determined by unbiased MD simulations from 10ps swarms-of-trajectories, to iteratively refine the string in the subspace of collective variables until convergence is reached [28]. Formally, the string method aims at determining the least free energy pathway between these two endpoints by refining the images [27, 28]. In practice, by virtue of the thermal noise added from the swarms-of-trajectories, the present algorithm is not expected to yield a single “optimal” path, but rather to sample the ensemble of plausible (high probability) paths within the interior of a broad “reaction tube” connecting the two endpoints [30]. For this reason, all the results presented here are averages from the last 20 iterations of the string with swarms-of-trajectories protocol (including error bars).

The most important information extracted from the present computational pathways is the spatial and sequential reorganization underlying the conformational transitions along each of the

main steps forming the transport cycle. While there are computational approaches to determine free energy barriers and transition rates from the string method, this is extremely difficult for systems as large as SERCA. For this reason, we limit the present analysis to the “chronology” of microscopic events during the transitions. This nontrivial information generated here from the string method is in sharp contrast with the naïve picture generated from simple “morphing” procedures, which are often based on linear coordinate interpolation of the endpoint states [40]. The latter always produce artificial pathways where, by default, all the events occur in a concerted progressive fashion. Transition pathways deduced from steered molecular dynamics (SMD) non-equilibrium trajectories, while typically more advanced than plain morphing, also display a similarly artificial concerted character [29]. In Fig. 2, we highlighted how the variations in the organization of the N, P, and A cytoplasmic domains in response to the phosphorylation and dephosphorylation events directly correlate with the Ca^{2+} binding sites and the arrangement of the TM helices controlling the alternating access to the ion binding sites. For a P-type ATPase pump like SERCA to function properly, “conformational information” must be transmitted within the protein structure via domain-domain contacts that are formed and broken as well as direct pulls exerted by the cytoplasmic domains on the TM region. By virtue of the structural topology of SERCA (Fig. 1B), two main modes of coupling exist between the catalytic site in the P cytoplasmic domain and the Ca^{2+} binding sites in the membrane. The polypeptide connectivity underlying this coupling is depicted in Fig. 1B. The first one involves direct interactions through the M4 and M5 helices, which are both connected to the P domain. The second one involves indirect interactions mediated by the A domain and the M1-M3 helices, as M3 is connected to M4. Both modes of communication contribute to the cytoplasmic calcium binding and occlusion, while the second mode plays a dominant role for the remaining transitions (E1P-2Ca^{2+} to E2P and E2P to E2-Pi). In the following, we discuss the main

results from the computational transition pathways and correlate our findings with experimental results.

Cytoplasmic calcium binding and occlusion (E1 to E1-2Ca²⁺-ATP)

According to the computational results shown in Figs. 4-5, this entire transition takes place in a very concerted and cooperative manner. The bound Ca²⁺ ions are needed in their TM sites to lock the N-P interface, and configure the ATP near the Asp351. The ATP bound in the N domain is needed to strongly bind the two Ca²⁺ ions. All these concerted events are completed by image 20 in Fig. 4, which is essentially the occluded state. The results from the computational pathway show, in particular, and that the existence of a stable occluded state does not require hydrolysis of ATP and phosphorylation of Asp351. In addition, the presence of two bound Ca²⁺ ions appears to be necessary for reaching the final occluded catalytically competent state.

These observations provide a clear explanation of the strong coupling between Ca²⁺ binding and ATP hydrolysis. It is known that despite significant ATP binding, the phosphoryl transfer does not happen in the absence of Ca²⁺ binding and the disruption of binding of Ca(II) severely interferes with phosphorylation [41, 42]. Analysis of the computational pathway shows that the motion of the N domain can bring the ATP close to the phosphorylation site in the P domain but the catalytic site is not fully formed. Only after binding of both Ca²⁺ ions in the TM binding sites, Thr353 can make the final contact with Mg²⁺ to form the correct geometry for the phosphoryl transfer reaction, thus explaining the tight coupling between Ca²⁺ binding and phosphorylation. Cooperative and sequential binding of two Ca²⁺ ions to SERCA have been studied in detail by mutagenesis experiments over several decades [37, 42-46]. While the functional impact of many mutations can be rationalized from the X-ray structures, some appear to involve more subtle effects and are not readily explained. For instance, several mutants of Val304 exhibit a slower rate and partial inhibition of calcium transport

activity [47], a puzzling observation given that Val304 contributes to the binding site through its backbone carbonyl oxygen. The side-chain of Val304 is facing the M1 and M2 helices, which exhibit large movements during the occlusion transition. These movements involve some direct interactions between the side-chains of Leu65, the residue involved in blocking the cytoplasmic pathway, and Val304. The space near Val304 seems to be optimal, such that a sliding motion of the TM helices can occur. Substitution of Val304 by a charged residue (Glu or Lys) can destabilize the TM helix arrangement. On the other hand, substitutions by an amino acid with either a smaller (alanine) or a larger (isoleucine) nonpolar side-chain can interfere with the packing of helices via Leu65, thus hindering the occlusion. These effects take place in the middle of the conformational transition and perhaps affects mainly the transition rate between the end states, thus explaining the observed inhibition of Ca^{2+} transport activity [47].

In the calcium binding experiments in the absence of ATP, it has been shown that D59R mutant has a 10-fold increase in rate of dissociation of Ca^{2+} ions towards the cytoplasm [48], which can be readily explained by our simulations, as in-coming Ca^{2+} ions interact directly with Asp59 while diffusing toward the binding sites (Fig. 5B). Multiple investigators have demonstrated impaired ATP binding and ATPase activity of SERCA mutants with mutations at the following residues: Phe487 [49, 50], Arg489 [49-51], Lys492 [49, 52, 53], Arg560 [50-53], Cys561 [52], Leu562 [52]. All these residues interact with the ATP molecule throughout the transition pathway, thus explaining the mutagenesis results. Analyses of the pathway as well as the crystal structures also show how peripheral interactions around the ATP binding site play an important role in ATP binding. For example Ser488 forms a hydrogen bond with Asp490, which is important for maintaining the hairpin structure near the ATP binding site. Mutations of these residues lead to appreciable functional consequence [49].

According to the computational pathway, several residues in the P domain play important roles in directing and positioning of the ATP near the phosphorylation site. Identities of these residues can be deduced from the crystal structure, but their roles in the stepwise formation of contacts with the ATP molecule only become clear in the analysis of the discretized transition pathway. The analysis provides a clearer picture and explains the sensitivity of these residues to mutagenesis: these residues are Lys352 [50, 51, 54], Thr625 [50], Gly626 [50, 54, 55], Lys684 [50, 51, 55, 56] and Asn706 [50, 57]. As mentioned previously, upon binding of the ATP to the P domain, there are important intra-domain changes with important functional consequences. For example, during this transition Asp627 forms new interactions with Lys352 (P) and Arg560 (N) and mediates transmission of structural changes. These observations provide an explanation for the reduced rate of phosphorylation in D627A and D627N mutants [50, 54, 55]. Similar roles are played by Thr355, Thr357 and Asp707. The side-chain of Thr355 interacts with Asp351 backbone and contributes to its position in the catalytic site, this can explain the very low calcium transport activity of T355A mutant [54]. The side-chain of Thr357 makes a hydrogen bond with Lys352 which is involved in ATP binding as well as mediating structural changes, and the T357A mutation almost completely destroys transport activity [54]. It is observed in the computational pathway that Asp707 interacts with Mg^{2+} binding residue Asp703 and ATP binding residue Lys684 through its side-chain oxygen atoms (Fig. S5). The mutations removing the charge or the hydrogen bond capacity will destroy these interactions and interfere with the utilization of ATP. This is precisely the case, as several Asp707 mutants show negligible ATPase activity [50, 51, 55].

There are significant changes at the N-A interface over the course of this transition. For example, Asp426 and Arg134 form strong interaction after initial weak encounters. It is clear that the Asp426-Arg134 interaction is important for the proper positioning of the A domain in the first two legs of the transport cycle and disruption of these interaction will affect the function. Indeed,

mutagenesis of both these residues has been shown to affect ATPase activity and phosphoenzyme production [50]. A similar reasoning can also explain the importance of the Asp426-Arg139 interaction, with the R139A mutant, giving rise to reduced phosphoenzyme production [50].

Luminal opening and calcium release (E1P-2Ca²⁺-ADP to E2P)

The most striking observation from the results displayed in Figs. 6-7 is that this conformational transition takes place in two highly distinct steps. In the first part of the path (images 1-15), there is considerable rearrangement of the N, P and A cytoplasmic domains with minimal movements within the TM region, until the P-A contact is completed. Only then, do the conformational changes within the luminal TM region and the leaking of water into the binding sites start to take place, both of which happen in the second part of the path (images 16-32). This description is somewhat an oversimplification because the formation of the N-A interactions actually extends over an intermediate stage covering images 10 to 20 in Fig. 6B, but the main implication is that, once Asp351 is phosphorylated, the P-A interface must first reorganize, and it is this process that then drives the TM region to open the luminal gate. In other words, phosphorylation of Asp351 is needed to open the gate on the luminal side. While the initial and final pathway display some differences (Fig. 3B), analysis indicates that the two-step character of the conformational transition was already present in the initial CG pathway calculated by the *ANMP* pathway method [31]. The fact that this behavior emerges even at the CG level suggests that it is a robust feature of the macromolecular structure.

Alanine and glutamate mutants of Arg324 (M4) and Arg334 (P) severely slow down E1P to E2P transition [58]. The computational pathway indicates that structural changes in the cytoplasmic end of M2 helix and at the A-M2 junction are correlated with the motion of the A domain. During these events, the M2 helix remains in contact with the cytoplasmic part of the M4 helix. It appears

that the specific rotation of the A domain is probably dependent on the proper positioning of the M2 helix, which, in turn, is controlled by the M2-M4 interaction. Mutation at Arg324 on M4 severely disrupts its interactions with Gln108 and Asn111, which are both on M2, and this will affect the E1P to E2P transition. The pathway clearly shows important salt-bridges formed by Arg334. At the initial stage, Arg334 interacts with Glu44 and this salt-bridge is broken and new interactions are formed with Glu109 and Glu113 on M2 during the major changes in the cytoplasmic end of M2. These interactions are destroyed in the mutants, which should affect the E1P to E2P transition.

The E1P to E2P transition involves large changes in the N-A and P-A interfaces. The most detailed study of inter-domain interaction was conducted on the Glu439-Ser186 contact; mutations of one or both residues were shown to affect the E1P to E2P transition [59, 60]. But the swap mutation (E439S/S186E) was found to retain normal behavior, clearly indicating the importance of this interaction. The pathway shows that the formation of this interaction between the N and the A domains is correlated with the formation of several interactions at the P-A interface, therefore, is important for positioning the A domain with respect to the N and P domain and destruction of this contact can lead to disruption of sequential structural changes needed for function. During the same transition, Thr171 forms an interaction with Glu486 which is important for motion of the A domain, mutagenesis of these residues to alanine severely impairs ATPase activity [50]. Another important feature of this transition is the interactions of the TGES loop in the A domain with the residues near the catalytic site in the P domain. The substitution of Gly182 by alanine affects the E1P to E2P transition [61], which could be a result of steric hindrance created by the alanine side-chain or alteration of the loop structure. The slowing down of the same transition observed with the E183D mutant [61] can be explained by the fact that replacing Glu183 by a shorter Asp side-chain can no longer make all the required contacts with the P domain.

To provide experimentally testable predictions from the computational pathways, we have searched for the formation of non-native residue-residue contacts during the transition from E1P-2Ca²⁺-ADP to E2P. Here, a non-native contact is defined by two residues that are farther apart by more than 10 Å at the end-states but come within 5.5 Å at some point along the transition pathway (distances are estimated based on the positions of C^β or C^α atoms). A number of residues involved in non-native contacts could be identified in the cytoplasmic domains along the transition pathway (Table S2). It is important to note that the existence of such non-native contacts represent a non-trivial prediction from the computational pathways—these could not be deduced solely on the basis of the X-ray structures of the known stable endpoint states. It is reasonable to expect that double mutants, designed to either stabilize or destabilize the contact-forming transient states, could have a considerable impact on the function of the pump, thus providing a powerful route to test the validity of the computational pathways.

Closing of luminal gate and dephosphorylation (E2P to E2-Pi)

This conformational transition also takes place in two steps, although their character is relatively subtle. In the first part of the path (image 1 to image 17), there are movements within the TM region leading to the closing of the luminal gate (Figs. 8 and 9). Then, the cytoplasmic domains start to gently shift, creating an environment that allows for the dephosphorylation of Asp351 in the second part of the path (image 18 to image 32). The implication is that, once the luminal gate has closed, the A domain respond to this event and present Asp351-P with a chemical environment favoring dephosphorylation. In other words, it is the reorganization in the TM region that drives the cytoplasmic domains. The transition involves considerable movement of the A domain, with relatively small movements of the N and the P domains, causing the TGES loop to move away from the catalytic site in the P domain. It is noteworthy that the TGES loop has been implicated in

phosphatase function of the pump [39]. The shift in the N and P domains creates a narrow access pathway for the cytoplasmic water molecules toward Asp351-P and rearrangements of the key residues in the TGES loop with respect to the catalytic residue. This helps to create a favorable situation for the dephosphorylation reaction.

The pathway shows that Thr171 forms a transient interaction with Arg489, which correlates with the motion of the A domain that is ultimately responsible for dephosphorylation. Substitutions of these residues by alanine have been shown to significantly affect the function of the pump [50, 51]. More generally, Suzuki and co-workers have investigated the importance of several residues at the interface of A, P, M2 and M4 domains [58, 62-64] with respect to the E2P to E2-Pi transition. Careful inspection of the computational pathway reveals that structural changes at the A-M2 junction (residues 118 to 122) and the upper part of the M2 helix (residues 112 to 116) correlate with the crucial rotation of the A domain. Another consequence of these events is the formation of several interactions involving hydrophobic groups from the P domain (Val705, Val726), the A domain near the TGES loop (Ile179), the A-M2 junction (Leu119, Tyr122) and the A-M3 (Ile232, Ile235) junction. Referred to as the Tyr122-hydrophobic cluster by Suzuki and co-workers [58, 62-64], these contacts are likely to provide the driving force for the structural changes ultimately responsible for preparing the catalytic site to participate in the dephosphorylation reaction. Substitution of large nonpolar residues (Val, Ile, Tyr) by alanine will significantly weaken the side-chain-side-chain hydrophobic interaction and will affect the conformational transition. Similar correlated structural events captured by the pathway help explain the inhibition of dephosphorylation by the alanine and lysine mutants of Glu123 [58]. During the transition, Glu123 forms a dynamic salt-bridge interaction with Lys120, which is retained at the end of the transition and could provide stability to newly formed helical region at the A-M2 junction. Mutations will disrupt this interaction and consequently interfere with the above-mentioned conformational changes.

Conclusion

The nature of the long distance coupling between the phosphorylation and the TM binding sites is clarified only in the light of atomistic information about the intermediate structures, highlighting the need to go beyond static crystal structures. Analyses of specific sequence of motions in the complex structural transitions reveals the ordered cascade of changes propagating through the protein that helps explain the correlation between important events at the cytoplasmic and TM parts of the protein. To summarize, the main conclusions regarding three critical steps of the transport cycle $E1 \rightarrow E1-2Ca-ATP \rightarrow E2P \rightarrow E2-Pi$ are:

- ($E1 \rightarrow E1-2Ca-ATP$) Starting from the cytoplasmic open-access $E1$ state, the binding of ATP and the formation of the phosphoryl transfer catalytic site at the interface of the N and P domain together with the binding and sequestration of the two Ca^{2+} ions in the TM binding sites leading to occlusion are highly concerted and cooperative. The binding of Ca^{2+} ions is critical to enable the transition to the occluded state. Careful analysis of the structural events presented in the paper (highlighted text in the revised version), rationalizes this conclusion. The text was revised to reflect this discussion.
- The highly concerted transition explains the tight coupling between Ca^{2+} binding and ATP hydrolysis that is observed experimentally.
- ($E1-2Ca-ATP \rightarrow E2P$) Phosphorylation of Asp351 when the pump is in the occluded $E1-2Ca-ATP$ state drives the opening of the luminal gate via a two-step mechanism. In a first step, phosphorylation of Asp351 triggers a large-scale rotational movement of the cytoplasmic A domain to form a strong interaction with Asp351-P at the surface of the P domain (blocking of the catalytic site and preventing the back reaction by the ADP). In a second step, the large

movement of the A domain is subsequently transmitted to the M4 helix via M3 helix leading to the disruption of the Ca^{2+} binding sites and the opening of the TM region toward the luminal side.

- (E2P→E2-Pi) The closing of the luminal gate following the dissociation of the Ca^{2+} ions to the luminal solution and protonation of binding sites residues is correlated with the dephosphorylation of Asp351-P via a two-step mechanism. In a first step, dehydration of the (protonated) binding sites together with the hydrophobic re-packing between the TM helices M4, M5 and M6 spontaneously drives the closing of the luminal gate. In a second step, the closed luminal gate mediated by the junction with the M2 helix induces a rotation in the position of the A domain, positioning the $^{181}\text{TGES}$ loop in the A domain together with increased water access near Asp351-P needed for the dephosphorylation reaction.

While the present results are consistent with experiments, further efforts will necessarily be required to validate and confirm the detailed information provided by the computational pathways. In this endeavor, the occurrence of transient non-native residue-residue contacts predicted by the computational pathways offers a powerful route to experimentally probe the spatial and sequential organization underlying the conformational transitions in SERCA. The insight provided by the detailed atomistic pathways should be invaluable to elucidate the mechanisms of other ion pumps and ATP-driven molecular machines. The present computational approach can be used to understand any large-scale transitions in any biomolecular system, paving the way for deeper understanding of molecular processes fundamental to biology.

Materials and Methods

The present computational study is based on the crystal structures of six states (PDB IDs are given in parentheses): E1 (3W5A) [8], E1-2Ca²⁺-ATP (3AR2) [65], E1P-2Ca²⁺-ADP (3BA6) [16], E2P (3B9B) [16], E2-Pi (1WPG) [11] and E2 (3W5C) [8]. The E1 state was generated by extracting the protein without sarcolipin in the asymmetric unit [8], followed by removal of the Mg²⁺ ion from the TM region. The E1-2Ca²⁺-ATP state was generated by converting the ATP-analogue to ATP and the Ca²⁺ ion near the catalytic site to Mg²⁺ [65]. The Asp351 in the phosphorylated states was modelled as doubly negative phosphorylated aspartate. In the E1P-2Ca²⁺-ADP state, ADP-analogue was converted to ADP [17]. The protonation states of various systems were adjusted by PropKa [66] calculation on the crystal structures, MD simulations and published data [19, 21]. The following protonation states of the ionizable binding site residues (Glu309, Glu771, Asp800, Glu908) were used (H next to a residue name indicates that the residue is protonated): E1, E1-2Ca²⁺-ATP and E1P-2Ca²⁺-ADP states: Glu309, Glu771, Asp800, Glu908H; E2P, E2-Pi and E2 states: Glu309H, Glu771H, Asp800, Glu908H. Specifically, all the binding site residues except Glu908 were in their deprotonated states during the transition path between E1 and E1-2Ca²⁺-ATP states, in accord with experiments, while the residues Glu309, Glu771 and Glu908 were protonated during the transition path between E2P and E2-Pi. The orientation of the protein in the membrane was determined by the PPM webserver [67] and solvated systems were generated by using the CHARMM-GUI server [68]. A solvated protein system consisted of protein with necessary co-factors, 480 POPC (1-Palmitoyl-2-oleoylphosphatidylcholine) lipids, 70328 TIP3P [69] water molecules, potassium and chloride ions such that the total KCl concentration of the system was 0.15M. The sizes of different systems were typically around ~290,000 atoms. Molecular dynamics simulations were performed with the NAMD 2.9 [70] package and the CHARMM 36 force field for protein [71-73], lipids [74], ATP or ADP [75], and ions [76]. The force field parameters for aspartyl phosphate were obtained by quantum

chemical calculations using the General Automated Atomic Model Parameterization (GAAMP) webserver [77]. The parameters are given in Supplemental Information. Each solvated system was equilibrated for 8 ns with a combination of restrained and free MD runs. For the production run, equilibrated systems were simulated for 50ns in the NPT ensemble. For constant temperature and pressure, the Langevin thermostat and the Langevin piston method [78, 79] were used, respectively. The temperature and pressure were fixed at 300K and 1 atmosphere, respectively. The time-step for integration was 2 femtosecond and constraints were handled with the RATTLE algorithm [80]. The cut-off distance for non-bonded interaction was set at 12 Å with a switching function at 10 Å. Long range electrostatic interactions were calculated by the PME method with a grid-spacing of 1 Å [81]. For the calculation of water densities 50 ns MD simulations were performed after equilibration.

The string method [26, 27] represents the pathway as a chain of M evenly distributed “images” (copies of the full system) in a multidimensional subspace of collective variables $\mathbf{Z}$, e.g., the pathway is given as the sequence $[\mathbf{Z}^{(1)}, \mathbf{Z}^{(2)}, \dots, \mathbf{Z}^{(M)}]$. A multi-step and multiscale procedure was used to calculate the all-atom transition pathways. In the first step both end-states were equilibrated for 50 ns. In the second step, a coarse-grained (CG) transition pathway was calculated by the *ANMP* pathway method [31], which is based on the anisotropic network model (ANM) [82]. In the CG pathway, the protein was represented as a C^α trace. In the third step, an initial all-atom pathway (with explicit solvent and membrane) was generated from the CG pathway by successive image-to-image steered MD (SMD) runs starting from one or both the end-states. In the fourth step, the initial all-atom pathway was iteratively refined by the string method with swarms-of-trajectories [28, 29]. The steered MD (SMD) procedure for reconstructing an all-atom pathway from a CG pathway is briefly outlined here. Starting from one end-state, steering forces were applied on the C^α atoms with the C^α trace of the next image along the path as the target. Typical duration of an SMD run between two consecutive images was 1ns. At the end of the SMD run, the target image had a solvated

structure, where the C^α positions were fixed at the values given by the CG pathway. The same procedure was continued for the next two consecutive images, and the image-to-image SMD procedure was repeated until both the end-states were connected. To maintain proper orientation of co-factors (i.e. ATP) and key side-chains, successive SMD simulations were sometimes performed in a bidirectional manner, where the all-atom reconstructions were started from both endpoints and the final pathway was a composition of two segments of pulling in opposite directions. Each transition pathway between two stable states comprised 32 images. Data from last 20 iterations were used for analysis and error estimation. Critical details regarding the system setup for all the pathways calculated in the present study are given in Supplemental Table S1. The initial pathway for the E1 to E1-2Ca²⁺-ATP transition was generated by SMD starting from the occluded state with the bound ions. The E1 state for the initial string had the Ca(II) in the cytoplasmic side and Ca(I) in between two binding sites. The initial pathway for the E1P to E2P transition was generated by SMD from the E1P state. For E2P to E2-Pi transition, the initial SMD was performed from the E2P state, which resulted in the final state with the overall structure of the E2-Pi state with a phosphorylated Asp351. The following protocol was used for each pathway calculation. For each iteration, the length of the trajectories in the swarm was 10ps, and an additional 10ps was used for restrained equilibration after reparametrization of the collective variables (the restraining force constant was 50 kcal/mol/Å²). The mean drift was calculated from the swarms by averaging over 32 trajectories. A converged string was obtained after roughly 325 to 375 iterations. The string calculations were executed in a fully parallelized fashion, with a total of 1024 replicas of the system, with 32 images and 32 trajectories per swarm, using the multiple copy algorithm framework recently introduced in the NAMD 2.9 program [83].

Acknowledgements

This work was supported by the Membrane Protein Structural Dynamics Consortium funded by NIH/NIGMS through grant U54-GM087519. An award of computer time was provided by the Innovative and Novel Computational Impact on Theory and Experiment (INCITE) program from the Department of Energy (DOE) of the United States of America. This research used resources of the Argonne Leadership Computing Facility (ALCF), which is a DOE Office of Science User Facility supported under Contract DE-AC02-06CH11357. Additional computational resources were provided by the Great Lake Consortium for Petascale Computing and the XSEDE program of the National Science Foundation of the United States of America. We thank Dr. Lei Huang for his help in the parametrization of the force field for the aspartyl phosphate.

References

- [1] Moller JV, Juul B, le Maire M. Structural organization, ion transport, and energy transduction of P-type ATPases. *Biochim Biophys Acta*. 1996;1286:1-51.
- [2] Palmgren MG, Nissen P. P-type ATPases. *Annu Rev Biophys*. 2011;40:243-66.
- [3] de Meis L, Vianna AL. Energy interconversion by the Ca^{2+} -dependent ATPase of the sarcoplasmic reticulum. *Annu Rev Biochem*. 1979;48:275-92.
- [4] MacLennan DH, Brandl CJ, Korczak B, Green NM. Amino-acid sequence of a Ca^{2+} + Mg^{2+} -dependent ATPase from rabbit muscle sarcoplasmic reticulum, deduced from its complementary DNA sequence. *Nature*. 1985;316:696-700.
- [5] Toyoshima C, Inesi G. Structural basis of ion pumping by Ca^{2+} -ATPase of the sarcoplasmic reticulum. *Annu Rev Biochem*. 2004;73:269-92.
- [6] Post RL, Hegyvary C, Kume S. Activation by Adenosine Triphosphate in the Phosphorylation Kinetics of Sodium and Potassium Ion Transport Adenosine Triphosphatase. *Journal of Biological Chemistry*. 1972;247:6530-40.
- [7] Albers RW. Biochemical aspects of active transport. *Annu Rev Biochem*. 1967;36:727-56.
- [8] Toyoshima C, Iwasawa S, Ogawa H, Hirata A, Tsueda J, Inesi G. Crystal structures of the calcium pump and sarcolipin in the Mg^{2+} -bound E1 state. *Nature*. 2013;495:260-4.
- [9] Toyoshima C, Norimatsu Y, Iwasawa S, Tsuda T, Ogawa H. How processing of aspartylphosphate is coupled to lumenal gating of the ion pathway in the calcium pump. *Proc Natl Acad Sci U S A*. 2007;104:19831-6.
- [10] Obara K, Miyashita N, Xu C, Toyoshima I, Sugita Y, Inesi G, et al. Structural role of countertransport revealed in Ca^{2+} pump crystal structure in the absence of Ca^{2+} . *Proc Natl Acad Sci U S A*. 2005;102:14489-96.
- [11] Toyoshima C, Nomura H, Tsuda T. Lumenal gating mechanism revealed in calcium pump crystal structures with phosphate analogues. *Nature*. 2004;432:361-8.
- [12] Toyoshima C, Mizutani T. Crystal structure of the calcium pump with a bound ATP analogue. *Nature*. 2004;430:529-35.
- [13] Toyoshima C, Nomura H. Structural changes in the calcium pump accompanying the dissociation of calcium. *Nature*. 2002;418:605-11.
- [14] Toyoshima C, Nakasako M, Nomura H, Ogawa H. Crystal structure of the calcium pump of sarcoplasmic reticulum at 2.6 Å resolution. *Nature*. 2000;405:647-55.
- [15] Winther AM, Bubltz M, Karlsen JL, Moller JV, Hansen JB, Nissen P, et al. The sarcolipin-bound calcium pump stabilizes calcium sites exposed to the cytoplasm. *Nature*. 2013;495:265-9.
- [16] Olesen C, Picard M, Winther AM, Gyru C, Morth JP, Oxvig C, et al. The structural basis of calcium transport by the calcium pump. *Nature*. 2007;450:1036-42.
- [17] Sorensen TL, Moller JV, Nissen P. Phosphoryl transfer and calcium ion occlusion in the calcium pump. *Science*. 2004;304:1672-5.
- [18] Olesen C, Sorensen TL, Nielsen RC, Moller JV, Nissen P. Dephosphorylation of the calcium pump coupled to counterion occlusion. *Science*. 2004;306:2251-5.
- [19] Sugita Y, Miyashita N, Ikeguchi M, Kidera A, Toyoshima C. Protonation of the acidic residues in the transmembrane cation-binding sites of the Ca^{2+} pump. *J Am Chem Soc*. 2005;127:6150-1.
- [20] Espinoza-Fonseca LM, Thomas DD. Atomic-Level Characterization of the Activation Mechanism of SERCA by Calcium. *PLoS ONE*. 2011;6:e26936.
- [21] Musgaard M, Thøgersen L, Schiøtt B. Protonation States of Important Acidic Residues in the Central Ca^{2+} Ion Binding Sites of the Ca^{2+} -ATPase: A Molecular Modeling Study. *Biochemistry*. 2011;50:11109-20.
- [22] Sugita Y, Ikeguchi M, Toyoshima C. Relationship between Ca^{2+} -affinity and shielding of bulk water in the Ca^{2+} -pump from molecular dynamics simulations. *Proc Natl Acad Sci U S A*. 2010;107:21465-9.

- [23] Musgaard M, Thogersen L, Schiott B, Tajkhorshid E. Tracing cytoplasmic Ca(2+) ion and water access points in the Ca(2+)-ATPase. *Biophysical journal*. 2012;102:268-77.
- [24] Nagarajan A, Andersen JP, Woolf TB. Coarse- Grained Simulations of Transitions in the E2- to- E1 Conformations for Ca ATPase (SERCA) Show Entropy–Enthalpy Compensation. *J Mol Biol*. 2012;422:575-93.
- [25] Espinoza-Fonseca LM, Autry JM, Thomas DD. Microsecond molecular dynamics simulations of Mg(2)(+)- and K(+)-bound E1 intermediate states of the calcium pump. *PLoS One*. 2014;9:e95979.
- [26] Weinan E, Ren W, Eijnden E. String method for the study of rare events. *Physical Review B*. 2002;66:052301.
- [27] Maragliano L, Fischer A, Vanden-Eijnden E, Ciccotti G. String method in collective variables: minimum free energy paths and isocommittor surfaces. *The Journal of chemical physics*. 2006;125:24106.
- [28] Pan AC, Sezer D, Roux B. Finding transition pathways using the string method with swarms of trajectories. *J Phys Chem B*. 2008;112:3432-40.
- [29] Gan W, Yang S, Roux B. Atomistic view of the conformational activation of Src kinase using the string method with swarms-of-trajectories. *Biophysical journal*. 2009;97:L8-L10.
- [30] Meng Y, Shukla D, Pande VS, Roux B. Transition path theory analysis of c-Src kinase activation. *Proc Natl Acad Sci U S A*. 2016;113:9193-8.
- [31] Das A, Gur M, Cheng MH, Jo S, Bahar I, Roux B. Exploring the conformational transitions of biomolecular systems using a simple two-state anisotropic network model. *PLoS computational biology*. 2014;10:e1003521.
- [32] Henzler-Wildman KA, Lei M, Thai V, Kerns SJ, Karplus M, Kern D. A hierarchy of timescales in protein dynamics is linked to enzyme catalysis. *Nature*. 2007;450:913-6.
- [33] Henzler-Wildman KA, Thai V, Lei M, Ott M, Wolf-Watz M, Fenn T, et al. Intrinsic motions along an enzymatic reaction trajectory. *Nature*. 2007;450:838-44.
- [34] Lei M, Velos J, Gardino A, Kivenson A, Karplus M, Kern D. Segmented transition pathway of the signaling protein nitrogen regulatory protein C. *J Mol Biol*. 2009;392:823-36.
- [35] Fischer S, Olsen KW, Nam K, Karplus M. Unsuspected pathway of the allosteric transition in hemoglobin. *Proc Natl Acad Sci U S A*. 2011;108:5608-13.
- [36] Shukla D, Meng Y, Roux B, Pande VS. Activation pathway of Src kinase reveals intermediate states as targets for drug design. *Nat Commun*. 2014;5:3397.
- [37] Inesi G. Sequential mechanism of calcium binding and translocation in sarcoplasmic reticulum adenosine triphosphatase. *J Biol Chem*. 1987;262:16338-42.
- [38] Inesi G, Tadini-Buoninsegni F. Ca(2+)/H (+) exchange, lumenal Ca(2+) release and Ca (2+)/ATP coupling ratios in the sarcoplasmic reticulum ATPase. *J Cell Commun Signal*. 2014;8:5-11.
- [39] Clausen JD, Vilsen B, McIntosh DB, Einholm AP, Andersen JP. Glutamate-183 in the conserved TGES motif of domain A of sarcoplasmic reticulum Ca2+-ATPase assists in catalysis of E2/E2P partial reactions. *Proc Natl Acad Sci U S A*. 2004;101:2776-81.
- [40] Flores S, Echols N, Milburn D, Hespeneheide B, Keating K, Lu J, et al. The Database of Macromolecular Motions: new features added at the decade mark. *Nucleic Acids Res*. 2006;34:D296-301.
- [41] Bartolommei G, Tadini-Buoninsegni F, Hua S, Moncelli MR, Inesi G, Guidelli R. Clotrimazole inhibits the Ca2+-ATPase (SERCA) by interfering with Ca2+ binding and favoring the E2 conformation. *J Biol Chem*. 2006;281:9547-51.
- [42] Inesi G, Zhang Z, Lewis D. Cooperative setting for long-range linkage of Ca(2+) binding and ATP synthesis in the Ca(2+) ATPase. *Biophysical journal*. 2002;83:2327-32.
- [43] Andersen JP. Mutational analysis of Glu771 of the Ca(2+)-ATPase of sarcoplasmic reticulum. Effect of positive charge on dephosphorylation. *FEBS Lett*. 1994;354:93-6.
- [44] Strock C, Cavagna M, Peiffer WE, Sumbilla C, Lewis D, Inesi G. Direct demonstration of Ca2+ binding defects in sarco-endoplasmic reticulum Ca2+ ATPase mutants overexpressed in COS-1 cells transfected with adenovirus vectors. *J Biol Chem*. 1998;273:15104-9.
- [45] Zhang Z, Lewis D, Strock C, Inesi G, Nakasako M, Nomura H, et al. Detailed characterization of the cooperative mechanism of Ca(2+) binding and catalytic activation in the Ca(2+) transport (SERCA) ATPase. *Biochemistry*. 2000;39:8758-67.

- [46] Inesi G, Ma H, Lewis D, Xu C. Ca²⁺ occlusion and gating function of Glu309 in the ADP-fluoroaluminate analog of the Ca²⁺-ATPase phosphoenzyme intermediate. *J Biol Chem.* 2004;279:31629-37.
- [47] Xu C, Prasad AM, Inesi G, Toyoshima C. Critical role of Val-304 in conformational transitions that allow Ca²⁺ occlusion and phosphoenzyme turnover in the Ca²⁺ transport ATPase. *J Biol Chem.* 2008;283:3297-304.
- [48] Andersen JP, Clausen JD, Einholm AP, Vilsen B. Mutagenesis of residues involved in control of the Ca²⁺ entry pathway and conformational changes associated with Ca²⁺ binding in the SR Ca²⁺-ATPase. *Ann N Y Acad Sci.* 2003;986:72-81.
- [49] McIntosh DB, Woolley DG, Vilsen B, Andersen JP. Mutagenesis of segment 487Phe-Ser-Arg-Asp-Arg-Lys492 of sarcoplasmic reticulum Ca²⁺-ATPase produces pumps defective in ATP binding. *J Biol Chem.* 1996;271:25778-89.
- [50] Ma H, Lewis D, Xu C, Inesi G, Toyoshima C. Functional and structural roles of critical amino acids within the "N", "P", and "A" domains of the Ca²⁺ ATPase (SERCA) headpiece. *Biochemistry.* 2005;44:8090-100.
- [51] Ma H, Inesi G, Toyoshima C. Substrate-induced conformational fit and headpiece closure in the Ca²⁺ATPase (SERCA). *J Biol Chem.* 2003;278:28938-43.
- [52] 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;278:20245-58.
- [53] Hua S, Ma H, Lewis D, Inesi G, Toyoshima C. Functional role of "N" (nucleotide) and "P" (phosphorylation) domain interactions in the sarcoplasmic reticulum (SERCA) ATPase. *Biochemistry.* 2002;41:2264-72.
- [54] Maruyama K, Clarke DM, Fujii J, Inesi G, Loo TW, MacLennan DH. Functional consequences of alterations to amino acids located in the catalytic center (isoleucine 348 to threonine 357) and nucleotide-binding domain of the Ca²⁺-ATPase of sarcoplasmic reticulum. *J Biol Chem.* 1989;264:13038-42.
- [55] McIntosh DB, Clausen JD, Woolley DG, MacLennan DH, Vilsen B, Andersen JP. Roles of conserved P domain residues and Mg²⁺ in ATP binding in the ground and Ca²⁺-activated states of sarcoplasmic reticulum Ca²⁺-ATPase. *J Biol Chem.* 2004;279:32515-23.
- [56] Vilsen B, Andersen JP, MacLennan DH. Functional consequences of alterations to amino acids located in the hinge domain of the Ca(2+)-ATPase of sarcoplasmic reticulum. *J Biol Chem.* 1991;266:16157-64.
- [57] Clausen JD, McIntosh DB, Woolley DG, Anthonisen AN, Vilsen B, Andersen JP. Asparagine 706 and glutamate 183 at the catalytic site of sarcoplasmic reticulum Ca²⁺-ATPase play critical but distinct roles in E2 states. *J Biol Chem.* 2006;281:9471-81.
- [58] Yamasaki K, Daiho T, Danko S, Suzuki H. Multiple and Distinct Effects of Mutations of Tyr122, Glu123, Arg324, and Arg334 Involved in Interactions between the Top Part of Second and Fourth Transmembrane Helices in Sarcoplasmic Reticulum Ca²⁺-ATPase: CHANGES IN CYTOPLASMIC DOMAIN ORGANIZATION DURING ISOMETRIC TRANSITION OF PHOSPHOENZYME INTERMEDIATE AND SUBSEQUENT Ca²⁺ RELEASE. *Journal of Biological Chemistry.* 2004;279:2202-10.
- [59] Liu X, Daiho T, Yamasaki K, Wang G, Danko S, Suzuki H. Roles of Interaction between Actuator and Nucleotide Binding Domains of Sarco(endo)plasmic Reticulum Ca²⁺-ATPase as Revealed by Single and Swap Mutational Analyses of Serine 186 and Glutamate 439. *Journal of Biological Chemistry.* 2009;284:25190-8.
- [60] Clausen JD, McIntosh DB, Anthonisen AN, Woolley DG, Vilsen B, Andersen JP. ATP-binding modes and functionally important interdomain bonds of sarcoplasmic reticulum Ca²⁺-ATPase revealed by mutation of glycine 438, glutamate 439, and arginine 678. *J Biol Chem.* 2007;282:20686-97.
- [61] Anthonisen AN, Clausen JD, Andersen JP. Mutational analysis of the conserved TGES loop of sarcoplasmic reticulum Ca²⁺-ATPase. *J Biol Chem.* 2006;281:31572-82.

- [62] Wang G, Yamasaki K, Daiho T, Suzuki H. Critical Hydrophobic Interactions between Phosphorylation and Actuator Domains of Ca^{2+} -ATPase for Hydrolysis of Phosphorylated Intermediate. *Journal of Biological Chemistry*. 2005;280:26508-16.
- [63] Yamasaki K, Wang G, Daiho T, Danko S, Suzuki H. Roles of Tyr122-hydrophobic Cluster and K^{+} Binding in Ca^{2+} -releasing Process of ADP-insensitive Phosphoenzyme of Sarcoplasmic Reticulum Ca^{2+} -ATPase. *Journal of Biological Chemistry*. 2008;283:29144-55.
- [64] Yamasaki K, Daiho T, Danko S, Suzuki H. Assembly of a Tyr122-hydrophobic Cluster in Sarcoplasmic Reticulum Ca^{2+} -ATPase Synchronizes Ca^{2+} Affinity-reduction and Release with Phosphoenzyme Isomerization. *Journal of Biological Chemistry*. 2015.
- [65] Toyoshima C, Yonekura S, Tsueda J, Iwasawa S. Trinitrophenyl derivatives bind differently from parent adenine nucleotides to Ca^{2+} -ATPase in the absence of Ca^{2+} . *Proc Natl Acad Sci U S A*. 2011;108:1833-8.
- [66] Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK_a Predictions. *Journal of Chemical Theory and Computation*. 2011;7:525-37.
- [67] 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;40:D370-D6.
- [68] Jo S, Kim T, Im W. Automated Builder and Database of Protein/Membrane Complexes for Molecular Dynamics Simulations. *PLoS ONE*. 2007;2:e880.
- [69] Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of simple potential functions for simulating liquid water. *The Journal of chemical physics*. 1983;79:926-35.
- [70] Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, et al. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry*. 2005;26:1781-802.
- [71] Mackerell AD, Jr., Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, et al. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *Journal of Physical Chemistry B*. 1998;102:3586-616.
- [72] MacKerell JA, Feig M, 3rd CB. Improved treatment of the protein backbone in empirical force fields. *J Am Chem Soc*. 2004;126:698-9.
- [73] Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, et al. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ , ψ and Side-Chain χ_1 and χ_2 Dihedral Angles. *Journal of Chemical Theory and Computation*. 2012;8:3257-73.
- [74] Klauda JB, Venable RM, Freites JA, O'Connor JW, Tobias DJ, Mondragon-Ramirez C, et al. Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *The Journal of Physical Chemistry B*. 2010;114:7830-43.
- [75] Foloppe N, MacKerell AD. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J Comput Chem*. 2000;21:86-104.
- [76] Beglov D, Roux B. Finite Representation of an Infinite Bulk System - Solvent Boundary Potential for Computer-Simulations. *Journal of Chemical Physics*. 1994;100:9050-63.
- [77] Huang L, Roux B. Automated Force Field Parameterization for Nonpolarizable and Polarizable Atomic Models Based on Ab Initio Target Data. *Journal of Chemical Theory and Computation*. 2013;9:3543-56.
- [78] Martyna GJ, Tobias DJ, Klein ML. Constant pressure molecular dynamics algorithms. *The Journal of Chemical Physics*. 1994;101:4177-89.
- [79] Feller SE, Zhang Y, Pastor RW, Brooks BR. Constant pressure molecular dynamics simulation: The Langevin piston method. *The Journal of Chemical Physics*. 1995;103:4613-21.
- [80] Andersen HC. Rattle - a Velocity Version of the Shake Algorithm for Molecular-Dynamics Calculations. *J Comput Phys*. 1983;52:24-34.
- [81] Essmann U, Perera L, Berkowitz M, Darden T, Lee H, Pedersen L. A smooth particle mesh Ewald method. *J Chem Phys*. 1995;103.
- [82] Bahar I, Rader AJ. Coarse-grained normal mode analysis in structural biology. *Curr Opin Struct Biol*. 2005;15:586-92.

- [83] Jiang W, Phillips JC, Huang L, Fajer M, Meng Y, Gumbart JC, et al. Generalized scalable multiple copy algorithms for molecular dynamics simulations in NAMD. *Computer Physics Communications*. 2014;185:908-16.

Figure Legends

Figure 1. Overall topology of the P-type ATPase calcium pump SERCA. (A) The cytoplasmic headpiece comprises the nucleotide binding (N), phosphorylation (P) and actuator (A) domains. The ATP binds to the N domain. The conserved aspartic acid that gets phosphorylated is in the P domain. The calcium binding sites are located in the transmembrane (TM) domain. The TM domain is made of ten TM helices (labeled M1 to M10). (B) The A domain is covalently linked to the helices M1, M2 and M3, and the P domain is covalently linked to the helices M4 and M5.

Figure 2: Water access of the binding sites and arrangements of three cytoplasmic domains in four important states: **A.** Cartoon representation of the alternating access mechanism. **B.** TM helices M1 to M6, the binding site residues and the three dimensional water densities, averaged over 50ns trajectories, are presented to highlight the alternating access of binding sites along the transport cycle. Water densities are shown in sky blue surface representations (a threshold value of 0.6 was used to display the iso-density surface). **C.** Interactions of cytoplasmic domains after 50ns of MD (viewed from the top with fixed orientation of the P domain). The P domain (blue) is shown in surface representation. The N (green) and A (red) domains are shown as C $^{\alpha}$ traces. The green and red patches on the blue surface are the contact forming residues with N and A domains, respectively. Two residues are considered to be in contact, if pairwise distance between any two non-hydrogen atoms is less than 4 Å.

Figure 3: Convergence of the computational pathways. The C $^{\alpha}$ RMSDs of all the images, from the initial string, are plotted for several iterations. **A.** Between E1 and E1-2Ca $^{2+}$ -ATP states. **B.** Between E1P-2Ca $^{2+}$ and E2P states. **C.** Between E2P and E2-Pi states.

Figure 4: Profiles of key order parameters along the transition between E1 and E1-2Ca $^{2+}$ -ATP states leading to the occlusion of Ca $^{2+}$ ions. Distance between a pair of residues is estimated by the corresponding C $^{\beta}$ or C $^{\alpha}$ distance unless otherwise noted. **A.** Interactions between ATP/Mg and phosphorylation site in the P domain. **B.** Interactions between N and A domains. **C.** Hydrophobic side residues responsible for occlusion of Ca $^{2+}$ ions by locking the side-chain conformation of Glu309. **D.** Water access of several coordinating residues giving a quantitative understanding of the occlusion process.

Figure 5: Snapshots of the protein near the cytoplasmic catalytic site and TM Ca $^{2+}$ binding sites along the transition between E1 and E1-2Ca $^{2+}$ -ATP states leading to the occlusion of Ca $^{2+}$ ions. Numbers shown on the panels are indices of selected snapshots of the discretized transition pathway. Snapshots 1 and 32 correspond to the cytoplasmic open and occluded states, respectively. **A.** Region near the catalytic site in the P domain. Interactions of ATP and Mg $^{2+}$ ion with key residues are shown. Crucial movement at the interface of P and M5 TM helix (pink) is also presented. **B.** The region around the Ca $^{2+}$ binding sites and binding of second Ca $^{2+}$ ion. Side-chains of Leu65, Leu94 and Ile97 are shown in black space-filling representation. Ca $^{2+}$ ions are depicted as purple spheres and coordinating residues are shown in space-filling representation with an elemental color code.

Figure 6: Profiles of key order parameters along the transition between E1P-2Ca $^{2+}$ and E2P states, leading to the opening of the luminal exit channel for the release of Ca $^{2+}$ ions. **A.** Interactions between N and A domains. **B.** Interactions between P and A domains. **C.** Motion of TM helices responsible for opening at the luminal side. **D.** Water access of several coordinating residues from the luminal side indicating an open pathway from the lumen to the TM binding sites.

Figure 7: Snapshots of the protein near the cytoplasmic catalytic site and TM Ca^{2+} binding sites along the transition between E1P-2 Ca^{2+} and E2P states, leading to the opening of the luminal exit channel for release of Ca^{2+} ions. Numbers shown on the panels are indices of selected snapshots of the discretized transition pathway. Snapshots 1 and 32 correspond to the occluded and a luminal-open state with loosely bound Ca^{2+} ions, respectively. **A.** Catalytic site in the P domain with phosphorylated Asp351 and ADP. The important interaction between P and the TGES region of the A domain is shown. **B.** Movements of key TM helices, disruption of the Ca^{2+} binding sites and creation of a water filled exit pathway towards the lumen. Side-chains of Ile293, Tyr294, Lys297, Ile298, Val300, Ala301, Leu302, Val772, Ile775, Pro789, Leu792, Leu793 and Val795 are shown in black space-filling representations. These residues are important for luminal gating. Ca^{2+} ions are depicted as purple spheres and coordinating residues are shown in space-filling representation with an elemental color code.

Figure 8: Profiles of key order parameters along the transition between E2P and E2-Pi states, leading to the closing of the luminal exit pathway. **A.** Interactions between N and A domains. **B.** Interactions between P and A domains. **C.** Motion of TM helices responsible for opening at the luminal side. **D.** Water access of several coordinating residues from the luminal side indicating the closing of the luminal pathway.

Figure 9: Snapshots of the protein near the cytoplasmic catalytic site and TM Ca^{2+} binding sites along the transition between E2P and E2-Pi states, leading to the closing of the luminal exit pathway. Numbers shown on the panels are indices of selected snapshots of the discretized transition pathway. Snapshots 1 and 32 correspond to the luminal-open state and a state with luminal gate closed, respectively. **A.** Water access at the catalytic site. Movement of the TGES region of the A domain away from Asp351-P allows water to interact with it, setting the stage for dephosphorylation. **B.** Movements of helices and dehydration of the binding site residues. Side-chains of Ile293, Tyr294, Lys297, Ile298, Val300, Ala301, Leu302, Val772, Ile775, Pro789, Leu792, Leu793 and Val795 are shown in black space-filling representation. Coordinating residues are shown in space-filling representation with an elemental color code.

Figure 1. Overall topology of the P-type ATPase calcium pump SERCA. (A) The cytoplasmic headpiece comprises the nucleotide binding (N), phosphorylation (P) and actuator (A) domains. The ATP binds to the N domain. The conserved aspartic acid that gets phosphorylated is in the P domain. The calcium binding sites are located in the transmembrane (TM) domain. The TM domain is made of ten TM helices (labeled M1 to M10). (B) The A domain is covalently linked to the helices M1, M2 and M3, and the P domain is covalently linked to the helices M4 and M5.

Figure 2: Water access of the binding sites and arrangements of three cytoplasmic domains in four important states: **A.** Cartoon representation of the alternating access mechanism. **B.** TM helices M1 to M6, the binding site residues and the three dimensional water densities, averaged over 50ns trajectories, are presented to highlight the alternating access of binding sites along the transport cycle. Water densities are shown in sky blue surface representations. **C.** Interactions of cytoplasmic domains after 50ns of MD (viewed from the top with fixed orientation of the P domain). The P domain (blue) is shown in surface representation. The N (green) and A (red) domains are shown as C $^{\alpha}$ traces. The green and red patches on the blue surface are the contact forming residues with N and A domains, respectively. Two residues are considered to be in contact, if pairwise distance between any two non-hydrogen atoms is less than 4 Å.

Figure 3: Convergence of the computational pathways. The C α RMSDs of all the images, from the initial string, are plotted for several iterations. **A.** Between E1 and E1-2Ca $^{2+}$ -ATP states. **B.** Between E1P-2Ca $^{2+}$ and E2P states. **C.** Between E2P and E2-Pi states.

Figure 4: Profiles of key order parameters along the transition between E1 and E1-2Ca²⁺-ATP states leading to the occlusion of Ca²⁺ ions. Distance between a pair of residues is estimated by the corresponding C ^{β} or C ^{α} distance unless otherwise noted. **A.** Interactions between ATP/Mg and phosphorylation site in the P domain. **B.** Interactions between N and A domains. **C.** Hydrophobic side residues responsible for occlusion of Ca²⁺ ions by locking the side-chain conformation of Glu309. **D.** Water access of several coordinating residues giving a quantitative understanding of the occlusion process.

Figure 5: Snapshots of the protein near the cytoplasmic catalytic site and TM Ca²⁺ binding sites along the transition between E1 and E1-2Ca²⁺-ATP states leading to the occlusion of Ca²⁺ ions. Numbers shown on the panels are indices of selected snapshots of the discretized transition pathway. Snapshots 1 and 32 correspond to the cytoplasmic open and occluded states, respectively. **A.** Region near the catalytic site in the P domain. Interactions of ATP and Mg²⁺ ion with key residues are shown. Crucial movement at the interface of P and M5 TM helix (pink) is also presented. **B.** The region around the Ca²⁺ binding sites and binding of second Ca²⁺ ion. Side-chains of Leu65, Leu94 and Ile97 are shown in black space-filling representation. Ca²⁺ ions are depicted as purple spheres and coordinating residues are shown in space-filling representation with an elemental color code.

Figure 6: Profiles of key order parameters along the transition between E1P-2Ca²⁺ and E2P states, leading to the opening of the luminal exit channel for the release of Ca²⁺ ions. **A. Interactions between N and A domains. **B.** Interactions between P and A domains. **C.** Motion of TM helices responsible for opening at the luminal side. **D.** Water access of several coordinating residues from the liminal side indicating an open pathway from the lumen to the TM binding sites.**

Figure 7: Snapshots of the protein near the cytoplasmic catalytic site and TM Ca^{2+} binding sites along the transition between $E1P-2Ca^{2+}$ and $E2P$ states, leading to the opening of the luminal exit channel for release of Ca^{2+} ions. Numbers shown on the panels are indices of selected snapshots of the discretized transition pathway. Snapshots 1 and 32 correspond to the occluded and a luminal-open state with loosely bound Ca^{2+} ions, respectively. **A.** Catalytic site in the P domain with phosphorylated Asp351 and ADP. The important interaction between P and the TGES region of the A domain is shown. **B.** Movements of key TM helices, disruption of the Ca^{2+} binding sites and creation of a water filled exit pathway towards the lumen. Side-chains of Ile293, Tyr294, Lys297, Ile298, Val300, Ala301, Leu302, Val772, Ile775, Pro789, Leu792, Leu793 and Val795 are shown in black space-filling representations. These residues are important for luminal gating. Ca^{2+} ions are depicted as purple spheres and coordinating residues are shown in space-filling representation with an elemental color code.

Figure 8: Profiles of key order parameters along the transition between E2P and E2-Pi states, leading to the closing of the luminal exit pathway. A. Interactions between N and A domains. **B.** Interactions between P and A domains. **C.** Motion of TM helices responsible for opening at the luminal side. **D.** Water access of several coordinating residues from the liminal side indicating the closing of the luminal pathway.

Figure 9: Snapshots of the protein near the cytoplasmic catalytic site and TM Ca^{2+} binding sites along the transition between E2P and E2-Pi states, leading to the closing of the luminal exit pathway. Numbers shown on the panels are indices of selected snapshots of the discretized transition pathway. Snapshots 1 and 32 correspond to the luminal-open state and a state with luminal gate closed, respectively. **A.** Water access at the catalytic site. Movement of the TGES region of the A domain away from Asp351-P allows water to interact with it, setting the stage for dephosphorylation. **B.** Movements of helices and dehydration of the binding site residues. Side-chains of Ile293, Tyr294, Lys297, Ile298, Val300, Ala301, Leu302, Val772, Ile775, Pro789, Leu792, Leu793 and Val795 are shown in black space-filling representation. Coordinating residues are shown in space-filling representation with an elemental color code.