Energy landscape steering in SecYEG mediates dynamic coupling in ATP driven protein translocation

Joel Crossley[1,3,4], Matthew A. Watson[1], Tomas Fessl[3], Daniel Watkins[2], Robin A. Corey[5], Tara Sabir[4], Sheena E. Radford[1], Ian Collinson[2†], Roman Tuma[1,3†].

[1]Astbury Centre for Structural Molecular Biology, School of Molecular and Cellular Biology, Faculty of Biological Sciences, University of Leeds, Leeds, United Kingdom; [2]School of Biochemistry, University of Bristol, Bristol, United Kingdom; [3]University of South Bohemia in Ceske Budejovice, Czech Republic; [4]School of Clinical and Applied Sciences, Leeds Beckett University, Leeds, United Kingdom; [5]Department of Biochemistry, University of Oxford, Oxford, United Kingdom.

[†]Corresponding authors. Emails: rtuma@prf.jcu.cz, ian.collinson@bristol.ac.uk

Abstract

The Sec translocon is a transmembrane assembly highly conserved among all forms of life as the principal route for transport of polypeptides across or into lipid bilayers. In bacteria translocation is driven by allosteric communication between the membrane pore SecYEG and the associated SecA ATPase. Using time-resolved single molecule fluorescence we reveal that slow conformational changes associated with SecA ATPase (~ 6 s^{-1}) modulate fast opening and closure of the SecY pore (~ 175 s^{-1}). Such mismatch of timescales is not compatible with direct coupling between SecA and SecYEG and the power stroke mechanism. A dynamic allosteric model in which SecA ATPase cycle controls energy landscape for SecY pore opening is proposed and consistent with the Brownian-ratchet mechanism. Analysis of structures and molecular dynamics trajectories identified key molecular interactions involved in the mechanism. This dynamic allostery may be common among motor ATPases that drive conformational changes in molecular machines.

Introduction

A fascinating class of biological molecular machines are those operating upon biopolymer substrates and converting chemical energy derived from ATP binding and hydrolysis into cycles of conformational changes and mechanical work (*e.g.* helicases, translocases, unfoldases, chromatin remodelling complexes and other molecular motors) (Sun and Wang, 2016). The bacterial protein secretion system composed of the integral membrane assembly SecYEG and peripherally associated cytosolic motor ATPase SecA (designated SecYEG:A

here) belongs to this class, purposed to translocate unfolded polypeptides across the inner membrane (Brundage et al., 1990). Secretory proteins (pre-proteins) with N-terminal cleavable signal sequences (SS) are targeted to SecYEG:A for subsequent post-translational transport across the membrane. Alternatively, membrane proteins are targeted by the signal recognition particle (SRP) and its receptor to SecYEG for co-translational insertion, independent of SecA (Beck et al., 2000).

In the SecYEG complex, SecY forms the pore for protein translocation (Fig. 1A) (Akimaru et al., 1991; Brundage et al., 1990; Ito et al., 1983). At the side of the channel is a lateral gate (LG, Fig. 1B) (Van den Berg et al., 2004), which undergoes opening and closure during protein translocation: opening of the LG is required to facilitate the lateral passage of trans-membrane helices (TMHs) into the bilayer during membrane protein insertion (du Plessis et al., 2009; Kater et al., 2019). SecA is a RecA-like ATPase with two nucleotide binding domains forming the ATP binding site (Hunt et al., 2002; Schmidt et al., 1988). The enzyme associates with SecYEG during translocation and exhibits a two-helix finger (2HF) which penetrates into the SecY channel (Hartl et al., 1990; Lill et al., 1989; Zimmer et al., 2008). The 2HF has been proposed to act as a sensor regulating nucleotide exchange (Allen et al., 2016; Whitehouse et al., 2012), or alternatively, to directly push the polypeptide across the membrane (Bauer et al., 2014; Catipovic et al., 2019).

The precise nature of the translocation mechanism has divided opinion owing to the inherent complexity of the system (Allen et al., 2016; Bauer et al., 2014; Catipovic et al., 2019; Collinson, 2019). There are two limiting cases of transducing the chemical potential of NTP hydrolysis into directional motion: power stroke and Brownian ratchet (Wagoner and Dill, 2016). The power-stroke mechanism involves deterministic or direct coupling in which each conformation imposed by the stage of the nucleotide hydrolysis cycle is linked with a well-defined conformation of the effector part (*e.g.* a mechanical lever or substrate binding site) as within a purely mechanical motor. On the other hand, a Brownian ratchet mechanism exhibits loose coupling between the conformations of the nucleotide-binding site of the motor ATPase and the effector part (Ait-Haddou and Herzog, 2003; Astumian, 2007). In the latter case, the nucleotide-hydrolysis cycle biases a shallow energy landscape towards certain conformations of the effector part, which in turn undergoes rapid interconversion between the available states due to thermal fluctuations. While a power-stroke mechanism has been demonstrated for many cytoskeletal motors (Lin et al., 2014), a Brownian ratchet scheme has been implicated for other assemblies such as the ClpX polypeptide unfoldase (Bell et al., 2018). Both mechanisms have been considered for the Sec translocon: the power stroke invokes large, piston-like motion of 2HF imparting force directly to the substrate while SecYEG is considered a passive pore (Bauer et al., 2014; Catipovic et al., 2019).

Alternatively, SecY has been proposed to act as Brownian ratchet while allosterically communicating with SecA (Allen et al., 2016).

A recent single molecule study detected conformational changes of the 2HF taking place on time scales (measured as dwell times in different states) of 100 - 400 ms (Catipovic et al., 2019), *i.e.* similar to the timescale of the ATP hydrolysis cycle (~100 ms) (Robson et al., 2009), suggesting that SecA acts via a directly coupled power stroke. However, in all available structures of the complex (Li et al., 2016; Ma et al., 2019; Park et al., 2014; Zimmer et al., 2008) there is little space left for 2HF movement, and covalent crosslinking the 2HF to SecY does not prevent translocation activity (Whitehouse et al., 2012), suggesting the 2HF does not move much during protein transport. Thus, the conformational changes between SecA and SecYEG maybe more subtle and remain unresolved.

Many consequences of coupling between the motor and channel components have been observed: (1) the SecA ATP hydrolysis rate increases ten-fold when complexed to SecYEG and 100-fold during translocation (Robson et al., 2009); (2) Opening and closing of the SecY LG is linked to the nucleotide state of SecA (Allen et al., 2016); and (3) the ATPase cycle is affected by the LG conformation (Allen et al., 2016). Further evidence for dynamic coupling comes from recent hydrogen-deuterium exchange experiments which report on internal protein dynamics and allostery (Ye et al., 2019; Zander et al., 2016). In these experiments an increased deprotection of SecA against deuteration is associated with protection of SecY and *vice versa*; *i.e* conformational dynamics are coordinated in an anti-correlated fashion (Ahdash et al., 2019). Understanding this dynamic coupling between SecA and the SecYEG protein channel is key to reconciliation of the apparently distinct mechanisms of its action proposed. Several scenarios need to be considered: do the different conformations of SecYEG (*e.g.* open and closed states of the channel) deterministically and rigidly follow the nucleotide state of SecA? Or, alternatively, is SecYEG coupling to SecA more stochastic; manifested by rapid interchange between states controlled by modulations of the underlying energy landscape by the nucleotide status of SecA?

To address these questions, experiments sampling time scales faster than that of ATP hydrolysis by SecA (~100 ms) are necessary. Here, we extend our recent single-molecule FRET (Förster resonance energy transfer) approaches (Allen et al., 2016; Fessl et al., 2018) to access time scales from sub-millisecond to hundreds of milliseconds. We deployed our previously developed system in which the fluorescently labelled LG (Figure 1A and Figure 1 – Supplement 1) acts as a proxy for SecYEG channel opening and closure and is monitored by FRET (Allen et al., 2016). Surprisingly, we found that the LG moves on a time scale an order of magnitude faster than the ATPase cycle and the recently observed conformational

changes in SecA (Catipovic et al., 2019; Chada et al., 2018; Vandenberk et al., 2019). However, the interchange and occupancy of LG conformational states still depends on the nucleotide state and the presence of substrate, suggesting dynamic allostery in which SecA biases the energy landscape for SecY LG opening. Using an analysis of molecular dynamics (MD) trajectories suggest a molecular mechanism underpinning this dynamic allostery, and propose a model involving coupled dynamics of the SecA and SecY.

Results

Identification of conformational states of SecYEG

Previous structural studies have shown that the LG of SecYEG exhibits three conformational states (Zimmer et al., 2008); (Li et al., 2016). This was further confirmed by single molecule fluorescence (Allen et al., 2016) in which open, part open (here called intermediate) and closed LG configurations were identified. Here, we first recapitulated FRET efficiency distributions from the earlier TIRFM study (200 ms temporal resolution) using a faster confocal approach (temporal resolution ~ 0.1 ms) that has the potential to reveal transiently populated states and their interchange. As previously described, a double-cysteine variants of SecYEG labelled with a FRET dye pair (**Figure** 1A) was reconstituted into *E. coli* polar lipid to give singly-occupied proteoliposomes (Allen et al., 2016). During the SecA ATPase-coupled translocation of substrate through the translocon, the FRET signal between the dyes acts as a proxy for the distance across the LG (**Figure 1 – figure supplement 1**).

The conditions most relevant to translocation – SecYEG:A alone and SecYEG:A + ADP (both at equilibrium), and SecYEG:A + ATP and SecYEG:A + ATP + substrate (proSpy, (Pereira et al., 2019)) (both away from equilibrium), were probed using pulsed excitation in order to simultaneously access donor fluorescence lifetime (τ_D) and intensity information. The cumulative FRET efficiency distributions (example shown for SecYEG:A in **Figure 1D)** are well-explained by fitting to three single Gaussian components corresponding to the open (green), intermediate (orange) and closed (red) states. This assignment is further confirmed by modelling fluorophore accessible volumes yielded three distinct FRET distributions for these states (**Figure 1 – figure supplement 1**), closely reproducing the observed FRET efficiencies.

In order to quantitatively analyse the population of molecules in each state, it was first necessary to correct the distributions for the contribution from the Apo SecYEG complexes (**Figure 1 – figure supplement 2**) that arise from SecYEG that are oriented with their cytoplasmic side facing the vesicle lumen. Due to random orientation of reconstituted

complexes, these constitute 50% of all complexes and do not participate in the reactions involving SecA (Allen et al., 2016; Fessl et al., 2018). Following this correction, the relative populations of the different states could be evaluated for samples in the presence of the various nucleotides and soluble periplasmic pre-protein proSpy, (Pereira et al., 2019) (**Figure 1D, Figure 1 – figure supplement 3**); all of which were in good qualitative agreement with the previously published TIRFM conducted with an alternative pre-protein (proOmpA) (Allen et al., 2016). In fact, better time-resolution led to reduced averaging due to rapid state interconversion and helped to discern a shift in the FRET efficiency distribution for the closed state during translocation (compare panels E and F in **Figure 1 – figure supplement 3**). Presumably, this was a result of signal sequence induced unlocking of SecYEG as previously documented by various methods (Fessl et al., 2018; Gouridis et al., 2009; Hizlan et al., 2012). Thus, the improved temporal resolution also helps to detect subtle changes in LG conformation and allows new states to be discerned.

Dynamic behaviour of the SecYEG lateral gate

In order to examine the dynamic equilibria of the LG conformational states, and its response to SecA and different nucleotides, and the pre-protein, we compiled 2D histograms of FRET efficiency (E) and the corresponding donor lifetime obtained for each burst (**Figure 2**). Such two-dimensional analysis allows for the determination of dynamic behaviour (**Figure 2 - figure supplement 1**) on the millisecond timescale, defined by the average burst duration (*i.e.* the time it takes for the lipid vesicle to diffuse through the confocal volume; **Figure 2 - figure supplement 2**). Bursts for which E and τ_D are approximately anticorrelated and fall into the vicinity of the diagonal, ‘static line’ exhibit no transitions, while those deviating from this line to the right indicate dynamic transitions (for further explanations see **Figure 2 - figure supplement 1**).

At equilibrium, the LG in both SecYEG alone and SecYEG:A exhibits intrinsic dynamic interconversion between all three states, although the closed state dominates (**Figure 2**). This interconversion was unexpectedly fast, on the diffusion timescale (represented by the mean burst width 5.75 ms, **Figure 2 -figure supplement 2**), corresponding to a rate constant of ~175 s^{-1}. Similar dynamic behaviour was observed in SecYEG:A when saturated by AMP-PNP (a non-hydrolysable analogue of ATP). However, addition of excess ADP prevents opening of the gate almost entirely, effectively shutting down all dynamics on this timescale (**Figure 2**).

Interestingly, during conditions of steady state ATP turnover the dynamic behaviour is further accentuated relative to the equilibrium condition, including an increased population of the open state. As noted above, in the presence of translocating substrate the closed state shifts to lower E_{FRET}, presumably due to the insertion of the signal sequence (SS) into the lateral gate. However, the intermediate and open states, and their transition dynamics persists in those conditions.

The presence of FRET events along the "dynamic line" in the E_{FRET} / τ_D plots suggests that the dynamic transitions occur on a time scale similar to that of diffusion of the proteoliposomes through the confocal volume *i.e.* a few milliseconds (see situation illustrated in **Figure 2 -figure supplement 1D**), and thus it should be possible to detect transitions between individual states that occur more rapidly than the diffusion time. This in turn allowed us to perform a Hidden Markov Model (HMM) analysis in order to detect transitions between states and recover the associated dwell times (**Figure 2 -figure supplement 3**). As expected for a stochastic process, the dwell times were widely distributed from 23 - 27 ms for the open state, 24 - 50 ms for the intermediate state and 25 - 37 ms for the closed state (**Figure 2**). There seems to be no significant difference between dwell times in states observed in the presence of different nucleotide conditions (except for ADP where paucity of "dynamic" bursts precluded statistically representative analysis, see Figure 2 -figure supplement 3 legend for more discussion). Hence, the steady state populations of these states (apart from ADP conditions) are determined by the frequency of transitions between the states, while the residence in these states is, on average, similarly short (tens of ms).

Further analysis of the transition patterns shows that opening and closure occurs most often *via* the intermediate state (**Figure 2 – figure supplement 4**). Thus, while the intermediate closed state is only transiently populated, it seems to be an obligatory intermediate on the LG reaction coordinate. The low frequency of 'direct' open to closed transitions also suggests that most transitions can be resolved by our detection system and thus LG dynamics mostly happens on the ~ 20 ms time scale. As expected, from bursts clustering along the dynamic line (**Figure 2**), many single molecule FRET traces exhibit discernible transitions with finite and measurable duration on the millisecond time scale (**Figure 2 – figure supplement 5**).

In summary, the results presented here show that the LG undergoes dynamic transitions on an ~ 20 ms time scale, which is an order of magnitude faster that the ATPase cycle, as well as the recently observed conformational changes of SecA (Catipovic et al., 2019; Ernst et al., 2018). However, the average population of the states is clearly linked to the SecA ATPase cycle (**Figure 1E**) and thus the 'fast' LG opening reaction coordinate is

stochastically coupled to the 'slow' SecA chemical reaction coordinate. This observation adds further evidence in support of the two way communication model presented previously (Allen et al., 2016), but due to the large separation of the observed time scales the two reaction coordinates as shown here, cannot be directly coupled. Thus, even if SecA undergoes a power-stroke like cycle, SecYEG LG does not follow it directly state-by-state, but is instead dynamically and indirectly coupled to SecA conformations.

Weak hydrogen bonding networks couple nucleotide occupancy to lateral gate motion

There are now eleven high resolution structures available of the SecY complex, in a range of different states. Therefore, although each one only represents a single discrete conformation, we reasoned that they might together contain enough information to give structural insights into the dynamic coupling between SecYEG and SecA. To investigate this, we first used invariant core analysis (Gerstein and Altman, 1995) to identify parts of the SecYEG structure that remain structurally conserved among species and conditions (*e.g.* different crystal forms, presence of SecA, nucleotides and a translocation substrate). This analysis identified two principal regions of SecY: the N-terminal half (with the exception of the plug and TMH 5) which remains mostly invariant, and the C-terminal half (with the exception of TMH 10) which is variable (**Figure 3**). Hence, the region encompassing TMH 6-9 and the intervening cytoplasmic loops (C4 and C5) may act as mediators of the allosteric changes in response to the nucleotide state of SecA.

Principal component analysis (PCA) enables identification and ranking of structural motions that explain variability across the whole data set (*i.e.* PDB structures of SecY from various prokaryotic organisms, see Methods and **Figure 4 - table supplement 1**) (Yang et al., 2009). PCA ranking yielded two dominant components which account for over 70% variability among the SecY structures. Cluster classification grouped most structures into two main clusters representing the closed and open state, respectively (**Figure 4 – figure supplement 2)**. In fact, variability between these clusters can be explained by a single principal component (PC1 in **Figure 4 – figure supplement 3**) which well represents the transition between the open and closed LG states (**Figure 4** and **Figure 4 – movie supplement 4**). However, the intermediate, part-open state was not significantly represented in the pool of static high-resolution structures, presumably due to its intermediary nature as demonstrated by our smFRET results (**Figure 2**).

The two endpoint structures (**Figure 4**) can be used to identify any distal changes that correlate with LG opening. As expected, the opening of LG is accompanied by changes

localized mostly within the C-terminal region of the protein. The main component is tilting of the TMH 6-9 helical bundle with respect to the mostly static N-terminal part. This is mediated by rotation and displacement of TMH 6 and 7, resulting in a piston-like motion of the helical bundle at the SecA interface (**Figure 4 – movie supplement 4**).

Previous study showed that residue K268 within TMH 7, when substituted by Cys (K268C mutant), can be crosslinked to the two helix finger (2HF) of SecA without loss of activity (Whitehouse et al., 2012), implying close contact and coupled motion of these structural elements. This suggests that the tilting motion of the cytoplasmic ends of SecY TMH 6-9 may be linked to the slow conformational changes observed for the 2HF (Catipovic et al., 2019). Because of the difference in time scale between SecY and SecA movement the interactions mediating this are likely to be weak and dynamic, possibly encompassing changes in hydrogen bonding pattern at the SecYEG:A interface. Hence, we compared hydrogen bonding networks within SecYEG:A in the ATP- and ADP-bound states, focusing on the interface between SecY and SecA.

Given the limited number and resolution of available of SecYEG:A crystal structures (**Figure 4 - table supplement 1**), we used all-atom MD simulations (Allen et al., 2016) to derive possible such networks. Hydrogen bonds (H-bonds) that stabilise each conformation of the translocon can be reduced to a two-dimensional matrix that only reflects changes between conditions (**Figure 5A**). In this visualization, any H-bond that is present between a given pair of residues in the ADP form, but is absent in the ATP bound ensemble, is represented by a red dot. Conversely, H-bond interactions that appears in the ATP-bound ensemble are represented by green dots. Several SecY and SecA regions exhibit nucleotide-related mutual H-bonding changes e.g. those highlighted by boxes in Figure 5A: SecA pre-protein crosslinking domain (PPXD) and 2HF switch H-bonding between the cytoplasmic loops C4 and C5 of SecY in an anti-correlated fashion (the affected regions are highlighted in **Figure 5B** and animated in the context of PCA in **Figure 5 – movie supplement 2**). These interaction patterns are independent of the pre-protein substrate (**Figure 5 – figure supplement 1** c.f. top and middle panels) and thus are likely to reflect intrinsic LG conformational steering by the SecA ATPase cycle, as observed by FRET (**Figure 2**). In contrast, H-bonding patterns between SecY and SecA nucleotide binding domains 1 and 2 (NBD1/2) and helical scaffold domain (HSD), respectively, do depend on the pre-protein substrate (**Figure 5 – figure supplement 1)**, and thus may be involved in the stimulation of the ATPase cycle by the pre-protein binding.

In summary, the above molecular simulation data link the observed LG dynamics to changes in the pattern of H-bond interactions at the SecYEG:A interface which may be responsible

for biasing the LG conformation in response to the SecA ATPase state, but are too weak to lock the LG in one particular state, hence resulting in a dynamic coupling between the two.

Discussion

We have used smFRET combined with analysis of existing crystal structures and MD simulations to directly observe and describe dynamic interchange between different conformational states of the SecYEG LG. Previous smFRET experiments on the immobilized samples (Allen et al., 2016) lacked the temporal resolution needed to capture dynamic transitions on/or faster than the timescale of ATP hydrolysis. Revisiting the system with improved temporal resolution and the detection of fluorescence lifetimes has allowed us to show that the rates of dynamic exchange between states of the LG are on the order ≈ 20 ms, *i.e.* a time scale an order of magnitude faster than the ATPase cycle (≈ 8 s^{-1} in the presence of substrate and ≈ 0.3 s^{-1} without (Robson et al., 2009).

The faster LG dynamics is reminiscent of that exhibited by the PPXD of isolated SecA (Vandenberk et al.), while the ATPase cycle timescale has been associated with conformational changes of the SecA 2HF during translocation based on smFRET experiments (Catipovic et al., 2019). The stark difference in timescales between SecYEG LG dynamics shown here and those of ATP hydrolysis at SecA suggest that thermal fluctuations play an important role in the pre-protein translocation mechanism. Indeed, lateral gating by SecY has been shown previously to be strongly dependent on temperature (Ge et al., 2014).

Rapid dynamic interchange is an intrinsic property of the SecY lateral gate, even in the absence of SecA; although the closed state is the most highly populated state (**Figure 2**). The presence of SecA allosterically controls this rapid equilibrium by selectively promoting transitions between the open and closed LG conformations, dependent on the nucleotide bound in the ATPase active site (**Figure 1 -figure supplement 3** and **Figure 2**). We propose that this is achieved via modulation of the energy landscape associated with LG opening and closing (**Figure 6**).

Invariant core and PCA analyses of the available crystal structures combined with patterns of hydrogen bonding in MD simulations revealed potential movements and changes in interaction patterns in SecYEG:A associated with the LG reaction coordinate (**Figures 3-5**). These structural changes localize to the C-terminal TMH 6-9 bundle and their connecting

loops (**Figures 3-4**) and correlates with the changes in hydrogen bonding at the SecYEG:A interface (**Figure 5**). This regions of SecYEG undergoes significant deprotection against deuteration in the presence of AMP-PNP suggesting that it is more dynamic than in the ADP-bound state (Ahdash et al., 2019). In contrast, SecA becomes more protected against hydrogen exchange in the ATP-bound state, suggesting that it assumes a more stable, relaxed, conformation (Ahdash et al., 2019). The AMP-PNP SecA conformation also has high affinity for SecYEG suggesting extensive contacts at the cytoplasmic interface (Robson et al., 2009). Indeed, MD analysis of hydrogen bonding indicate that the C4 and C5 loops switch contacts between the SecA PPXD and 2HF, respectively (**Figure 5 – panel B** and **Movie supplement 2**). In the ATP state, the C4 loop interacts preferentially with the 2HF, while preferring the PPXD in the ADP state. The SecY C5 loop, on the other hand, exhibits more H-bonds to the PPXD in the ATP state, while switching to 2HF in the ADP state. The SecY C6 loop is engaged in an ATP/ ADP-dependent manner with HSD and 2HF, but these interactions also depend on the presence of the pre-protein substrate **Figure 5 – figure supplement 1**), suggesting that the latter may also act as an allosteric modulator and further bias the energy landscape.

Our smFRET data indicate that in the nucleotide free or ATP-bound state of SecA these contacts are not strong enough to fix the C-terminal bundle involving TMHs 6-9 in SecY in the fully open configuration and thus allow rapid fluctuations between states. In the ADP bound state SecYEG reverts to its stable, closed ground state; while SecA is in a strained, less stable configuration, as shown by the increase in deuterium exchange in this state (Ahdash et al., 2019). SecA:ADP has lower affinity for SecYEG, suggesting less extensive cytoplasmic interface within the complex, which could then lead to retraction of the TMH 6-9 to the resting position. This is likely due to an extensive structural rearrangement within SecA, such as repositioning of the 2HF finger and the PPXD, that were recently observed by smFRET (Catipovic et al., 2019). Evidence for this repositioning also comes from a previous ensemble fluorescence experiment in which attachment of the environment-sensitive dye fluorescein to *E. coli* SecY residue K268 on loop C4 reports on an interaction with SecA (Robson et al., 2007). In the presence of AMP-PNP the excitation fluorescence of this dye is massively quenched, which turned out to to be attributable to a contact with the 2HF: visualised in the same state (ATP) in SecYEG:A (Zimmer et al., 2008). Interestingly, this effect was lost in the presence of ADP, indicative of a conformational rearrangement at the C4-2HF interaction site (**Figure 5B**). Of course, this conformational change need not necessarily involve movements of large distances. This 2HF repositioning may create a steric barrier for the TMH 6-9 motion and further lock it in position and effectively stop LG dynamics (**Figure 6**). This is supported by analysis of MD trajectories that demonstrate the

cytoplasmic cavity in the vicinity of 2HF and TMH 6-9 is significantly smaller in the ADP state (Corey et al., 2019).

So why are SecA and SecYEG not directly coupled in a power stroke-like mechanism? Our data and the recent SecA smFRET studies (Catipovic et al., 2019; Chada et al., 2018; Vandenberk et al., 2019) indicate that SecYEG and SecA operate on separate, but stochastically coupled timescales. While SecA, during its ~ 500 ms cycle, can 'grab' the pre-protein substrate and translocate about 20 amino acids (based on ~ 40 aa/ s processive rate (Fessl et al., 2018)), SecYEG operates on a much shorter time scale of ~ 20 ms, which is roughly equivalent to the passage time of a single amino acid. Thus, the two machines 'process' the translocating polypeptide on different length scales which need to be coupled during translocation. In addition, SecYEG is a versatile machinery, serving in the SecA ATP hydrolysis-driven translocation across the membrane, as well as ribosome GTP hydrolysis-driven co-translational folding into the membrane, with both processes requiring LG dynamics. Since the latter process relies on intimate probing of the lipid environment by the nascent polypeptide chain during initial recognition of the signal anchor and subsequent insertion of TMHs, it requires multiple LG configurations and relies on LG dynamics (Bischoff et al., 2014; Kater et al., 2019; Wickles et al., 2014). Thus, SecYEG may be 'programmed' to exhibit fast LG dynamics which then is controlled by either SecA in nucleotide-dependent manner, or by the ribosome in a nascent chain sequence-dependent manner. We propose that these mechanisms each occur *via* a general, conserved energy landscape biasing, mediated by interactions of SecY cytoplasmic loops with SecA or the ribosome.

Such a dynamic allosteric mechanism is not unique to SecYEG:A, since similar nucleotide-dependent steering of rapid dynamics has also been shown recently for adenylate kinase (Aviram et al., 2018), also enabled by advances in single molecule technology (Pirchi et al., 2016). Dynamic allostery involving motions occurring across multiple time scales coupled *via* energy landscape bias could be ubiquitous among complex molecular machines. Further experiments exploiting the wide range of timescales and motions now available via smFRET measurements will enable discovery of the extent to which such dynamic allostery is exploited in biology.

Methods

SecYEG, SecA$^{A795C\text{-}FL}$ and proSpy were produced as described previously (Collinson et al., 2001; Gold et al., 2007; Pereira et al., 2019). Labelling of SecYEG was also

conducted as described previously (Allen et al., 2016). The reconstitution and sample preparation for the confocal single molecule detection was performed according to established protocols (Allen et al., 2016; Fessl et al., 2018). The ~100 nM diameter proteoliposomes containing SecYEG were mixed with the appropriate proteins or nucleotides added at 100x concentration before being diluted 1:100 with measurement buffer to a concentration of 30 pM proteoliposomes for single-molecule measurements with a SecA concentration of 1 μM, proSpy concentration of 700 nM and nucleotide concentration of 1 mM.

Single-Molecule Lifetime Analysis

Time-resolved smFRET experiments were performed on a custom-built confocal epi-illuminated microscope in a standard inverted-stage configuration. Donor fluorophore molecules were excited with a 480 nm pulsed diode laser with fiber coupling optics (Advanced Laser Diode Systems, Germany) at a frequency of 40 MHz and a power of ~100 μW. Laser light was reflected from a dichroic beam splitter (Chroma Technology, USA) into a 100x 1.45 NA microscope objective (Nikon Instruments, USA) where it is focused onto the sample which is placed on a #1.5 thickness coverslip. Light emitted from the sample is then collected by the same objective lens before being transmitted through a dichroic beam splitter. The emitted photons are then passed through a 100 μM pinhole before being sorted by wavelength into the green and red channels by a long-pass dichroic mirror. Band pass filters remove the excitation wavelength light before it is focused onto single-photon avalanche diodes (MPD-100-CTB, MPD-050-CTB, Picoquant, Germany) with high timing resolution and low dark counts. Signals from the detectors are sent to a PCIe TCSPC card (TimeHarp 260 PICO Dual, Picoquant, Germany). Data was analysed using the FRETBursts python package with custom built scripts for the burst-wise lifetime fitting. A burst search with a minimum threshold of 4x the background signal of ~2 kHz and ~1.3 kHz in the donor and acceptor channels respectively and a minimum burst size of 100 photons was used to distinguish single-molecules. Donor only molecules were removed using a minimum acceptor burst size of 15 photons to remove the zero peak for clarity. Single-molecule fluorescence decays were analysed using a maximum likelihood estimator method (Maus et al., 2001).

Hidden Markov Analysis

To further characterise dynamics of lateral gate of SecY on a time-scale ranging from 0.5 to ~100 ms, we extracted FRET states (and respective dwell times) from time-sliced bursts

using Hidden Markov Models (HMM) (Rabiner, 1989) with Gaussian emissions as implemented in python packages hmmlearn and scikit-learn (Pedregosa et al., 2011).

Hydrogen Bond Analysis

Hydrogen bonds were localised from molecular dynamic simulation trajectories (Allen et al., 2016) using VMD HBonds Plugin (Humphrey et al., 1996) and further visualised in a Python environment.

Principal Component Analysis

Principal Component Analysis (PCA) is a technique commonly used to extract the most significant conformational changes, here it was performed in Bio3D (Grant et al., 2006).

Constructing Experimental Structure Ensembles for a Protein Family. Comparing multiple structures of homologous proteins and carefully analysing large multiple sequence alignments can help identify patterns of sequence and structural conservation and highlight conserved interactions that are crucial for protein stability and function.

Finding Available Sets of Similar Structures. Here we first used BLAST to query the PDB database to find similar sequences (and hence structures) to our chosen representative (PDB Id: 5AWW). Examining the alignment scores and their associated E-values indicated a sensible normalized score (-log(E-Value)) cutoff of 240 bits resulting in a set of structures listed in **Figure 4 - table supplement 1**.

Multiple Sequence Alignment. Next we split the structures into separate chains and store them for subsequent access. Finally, we extracted the sequence of each structure and perform a multiple sequence alignment to determine residue-residue correspondences.

Comparative Structure Analysis. The detailed comparison of homologous protein structures was used to infer mechanisms for conformational change using refined structural superposition and principal component analysis (PCA) to facilitate comparative structure analysis.

Principal Component Analysis. PCA can be employed to examine the relationship between different structures based on their equivalent residues. PCA provides considerable insight into the nature of conformational differences. Briefly, the resulting principal components (orthogonal eigenvectors) describe the axes of maximal variance of the distribution of structures. Projection of the distribution onto the space defined by the largest principal components results in a lower dimensional representation of the structural dataset. The percentage of the total mean square displacement (or variance) of atom positional fluctuations captured in each dimension is characterized by their corresponding eigenvalue (Grant et al., 2006). Applying PCA to all 14 structures revealed that 85.93% of the total

coordinate variance can be captured in three dimensions (39.98% in the first PC, 30.36% in the second, and 15.59% in the third.

Conformer Clustering in PC Space (see figure S4.2 and S4.3). Clustering structures in PC space enabled us to focus on the relationships between individual structures in terms of their major structural displacements, with a controllable level of dynamic details, i.e. helped us to specify the number of PCs used in the clustering. Two principal components were sufficient to cover > 70% of all variance.

Accessible Volumes

Implementation of the accessible volume (AV) algorithm (Sindbert et al., 2011) from the C++ library HandyFRET was used (Walczewska-Szewc and Corry, 2014). The algorithm allows for the determination of the maximum volume which can be sampled by a fluorophore attached by a linker to a specific point on a macromolecule. This method replaces the dye with a spheroid that has dimensions equal to that of the fluorophore which is attached to the biomolecule by a segmented linker. The position of the dye is recorded as the central point of the headgroup and the AV is only calculated for positions which do not overlap with the van der Waals radius of any atoms in the biomolecule. The distance between each point for the donor and acceptor dye is then measured and converted into FRET efficiency values using forster radius of 55 Å.

FRET Histograms

Due to the nature of random insertion of SecYEG into liposomes 50% of molecules will present the cytosolic interface internally into the liposome meaning SecA cannot bind. This random insertion has been shown for wild-type SecYEG (Mao et al., 2013; Schulze et al., 2014) and the SecYEG lateral gate mutant used in this study (Allen et al., 2016). To accurately determine state populations from FRET efficiency distributions for each experimental conditional this SecYEG Apo population must corrected for. A kernel density estimate (KDE) was applied to the SecYEG Apo FRET efficiency histogram (Figure 1S2). This KDE was used to estimate the probability density function (PDF) of state populations in the SecYEG protein alone. This SecYEG Apo PDF was then subtracted from the FRET efficiency histogram PDF for each condition. FRET efficiency histograms for each condition were then fit with three Gaussian distributions with the following fixed parameters for the centres of the fits: 0.25, 0.50, 0.72 and varied for the data in Figure 1 – figure supplement 3 panel F. These centres were derived through primary fitting of all conditions to find what best fits the majority of the data. The remaining parameters were left to vary.

Acknowledgements

We thank Dr. William Allen for invaluable discussions and critical reading of the manuscript. This work was funded by the BBSRC: MAW, RT and SER (BB/N017307/1), DW and IC (BBSRC: BB/N015126/1 and BB/I008675/1), RAC (BBSRC South West Bioscience Doctoral Training Partnership and BB/M003604/I). Additional support was provided by the Wellcome Trust (104632) to IC and ERC ((FP7/2007-2013) / ERC grant agreement 32240) to SER. JC, TF and RT are supported from European Regional Development Fund-Project "Mechanisms and dynamics of macromolecular complexes: from single molecules to cells" (No. CZ.02.1.01/0.0/0.0/15_003/0000441). JC and TS were funded by Leeds Beckett University.

Competing interests

Authors declare no competing interests.

References

Ahdash, Z., Pyle, E., Allen, W.J., Corey, R.A., Collinson, I., and Politis, A. (2019). HDX-MS reveals nucleotide-regulated, anti-correlated opening and closure of SecA and SecY channels of the bacterial translocon. eLife *8*, e47402.

Ait-Haddou, R., and Herzog, W. (2003). Brownian ratchet models of molecular motors. Cell Biochemistry and Biophysics *38*, 191-213.

Akimaru, J., Matsuyama, S., Tokuda, H., and Mizushima, S. (1991). Reconstitution of a protein translocation system containing purified SecY, SecE, and SecA from Escherichia coli. Proc Natl Acad Sci U S A *88*, 6545-6549.

Allen, W.J., Corey, R.A., Oatley, P., Sessions, R.B., Baldwin, S.A., Radford, S.E., Tuma, R., and Collinson, I. (2016). Two-way communication between SecY and SecA suggests a Brownian ratchet mechanism for protein translocation. Elife *5*, e15598.

Astumian, R.D. (2007). Design principles for Brownian molecular machines: how to swim in molasses and walk in a hurricane. Physical Chemistry Chemical Physics *9*, 5067-5083.

Aviram, H.Y., Pirchi, M., Mazal, H., Barak, Y., Riven, I., and Haran, G. (2018). Direct observation of ultrafast large-scale dynamics of an enzyme under turnover conditions. Proceedings of the National Academy of Sciences *115*, 3243.

Bauer, B.W., Shemesh, T., Chen, Y., and Rapoport, T.A. (2014). A "push and slide" mechanism allows sequence-insensitive translocation of secretory proteins by the SecA ATPase. Cell *157*, 1416-1429.

Beck, K., Wu, L.F., Brunner, J., and Muller, M. (2000). Discrimination between SRP- and SecA/SecB-dependent substrates involves selective recognition of nascent chains by SRP and trigger factor. EMBO J *19*, 134-143.

Bell, T.A., Baker, T.A., and Sauer, R.T. (2018). Hinge-Linker Elements in the AAA plus Protein Unfoldase ClpX Mediate Intersubunit Communication, Assembly, and Mechanical Activity. Biochemistry *57*, 6787-6796.

Bischoff, L., Wickles, S., Berninghausen, O., van der Sluis, E.O., and Beckmann, R. (2014). Visualization of a polytopic membrane protein during SecY-mediated membrane insertion. Nat Commun *5*, 4103.

Brundage, L., Hendrick, J.P., Schiebel, E., Driessen, A.J., and Wickner, W. (1990). The purified E. coli integral membrane protein SecY/E is sufficient for reconstitution of SecA-dependent precursor protein translocation. Cell *62*, 649-657.

Catipovic, M.A., Bauer, B.W., Loparo, J.J., and Rapoport, T.A. (2019). Protein translocation by the SecA ATPase occurs by a power-stroke mechanism. EMBO J.

Chada, N., Chattrakun, K., Marsh, B.P., Mao, C.F., Bariya, P., and King, G.M. (2018). Single-molecule observation of nucleotide induced conformational changes in basal SecA-ATP hydrolysis. Sci Adv *4*, eaat8797.

Collinson, I. (2019). The Dynamic ATP-Driven Mechanism of Bacterial Protein Translocation and the Critical Role of Phospholipids. Frontiers in Microbiology *10*, e1217.

Collinson, I., Breyton, C., Duong, F., Tziatzios, C., Schubert, D., Or, E., Rapoport, T., and Kuhlbrandt, W. (2001). Projection structure and oligomeric properties of a bacterial core protein translocase. EMBO J *20*, 2462-2471.

Corey, R.A., Ahdash, Z., Shah, A., Pyle, E., Allen, W.J., Fessl, T., Lovett, J.E., Politis, A., and Collinson, I. (2019). ATP-induced asymmetric pre-protein folding as a driver of protein translocation through the Sec machinery. Elife *8*.

du Plessis, D.J., Berrelkamp, G., Nouwen, N., and Driessen, A.J. (2009). The lateral gate of SecYEG opens during protein translocation. J Biol Chem *284*, 15805-15814.

Ernst, I., Haase, M., Ernst, S., Yuan, S., Kuhn, A., and Leptihn, S. (2018). Large conformational changes of a highly dynamic pre-protein binding domain in SecA. Commun Biol *1*, 130.

Fessl, T., Watkins, D., Oatley, P., Allen, W.J., Corey, R.A., Horne, J., Baldwin, S.A., Radford, S.E., Collinson, I., and Tuma, R. (2018). Dynamic action of the Sec machinery during initiation, protein translocation and termination. Elife *7*, e35112.

Ge, Y., Draycheva, A., Bornemann, T., Rodnina, M.V., and Wintermeyer, W. (2014). Lateral opening of the bacterial translocon on ribosome binding and signal peptide insertion. Nat Commun *5*, 5263.

Gerstein, M., and Altman, R.B. (1995). Average core structures and variability measures for protein families: application to the immunoglobulins. J Mol Biol *251*, 161-175.

Gold, V.A., Robson, A., Clarke, A.R., and Collinson, I. (2007). Allosteric regulation of SecA: magnesium-mediated control of conformation and activity. J Biol Chem *282*, 17424-17432.

Gouridis, G., Karamanou, S., Gelis, I., Kalodimos, C.G., and Economou, A. (2009). Signal peptides are allosteric activators of the protein translocase. Nature *462*, 363-367.

Grant, B.J., Rodrigues, A.P., ElSawy, K.M., McCammon, J.A., and Caves, L.S. (2006). Bio3d: an R package for the comparative analysis of protein structures. Bioinformatics *22*, 2695-2696.

Hartl, F.U., Lecker, S., Schiebel, E., Hendrick, J.P., and Wickner, W. (1990). The binding cascade of SecB to SecA to SecY/E mediates preprotein targeting to the E. coli plasma membrane. Cell *63*, 269-279.

Hizlan, D., Robson, A., Whitehouse, S., Gold, V.A., Vonck, J., Mills, D., Kuhlbrandt, W., and Collinson, I. (2012). Structure of the SecY complex unlocked by a preprotein mimic. Cell Rep *1*, 21-28.

Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. J Mol Graph *14*, 33-38, 27-38.

Hunt, J.F., Weinkauf, S., Henry, L., Fak, J.J., McNicholas, P., Oliver, D.B., and Deisenhofer, J. (2002). Nucleotide control of interdomain interactions in the conformational reaction cycle of SecA. Science *297*, 2018-2026.

Ito, K., Wittekind, M., Nomura, M., Shiba, K., Yura, T., Miura, A., and Nashimoto, H. (1983). A temperature-sensitive mutant of E. coli exhibiting slow processing of exported proteins. Cell *32*, 789-797.

Kater, L., Frieg, B., Berninghausen, O., Gohlke, H., Beckmann, R., and Kedrov, A. (2019). Partially inserted nascent chain unzips the lateral gate of the Sec translocon. EMBO Rep, e48191.

Li, L., Park, E., Ling, J.J., Ingram, J., Ploegh, H., and Rapoport, T.A. (2016). Crystal structure of a substrate-engaged SecY protein-translocation channel. Nature *531*, 395-+.

Lill, R., Cunningham, K., Brundage, L.A., Ito, K., Oliver, D., and Wickner, W. (1989). SecA protein hydrolyzes ATP and is an essential component of the protein translocation ATPase of Escherichia coli. EMBO J *8*, 961-966.

Lin, J.F., Okada, K., Raytchev, M., Smith, M.C., and Nicastro, D. (2014). Structural mechanism of the dynein power stroke. Nature Cell Biology *16*, 479-U208.

Ma, C., Wu, X., Sun, D., Park, E., Catipovic, M.A., Rapoport, T.A., Gao, N., and Li, L. (2019). Structure of the substrate-engaged SecA-SecY protein translocation machine. Nat Commun *10*, 2872.

Mao, C., Cheadle, C.E., Hardy, S.J., Lilly, A.A., Suo, Y., Sanganna Gari, R.R., King, G.M., and Randall, L.L. (2013). Stoichiometry of SecYEG in the active translocase of Escherichia coli varies with precursor species. Proc Natl Acad Sci U S A *110*, 11815-11820.

Maus, M., Cotlet, M., Hofkens, J., Gensch, T., De Schryver, F.C., Schaffer, J., and Seidel, C.A. (2001). An experimental comparison of the maximum likelihood estimation and nonlinear least-squares fluorescence lifetime analysis of single molecules. Anal Chem *73*, 2078-2086.

Park, E., Menetret, J.F., Gumbart, J.C., Ludtke, S.J., Li, W., Whynot, A., Rapoport, T.A., and Akey, C.W. (2014). Structure of the SecY channel during initiation of protein translocation. Nature *506*, 102-106.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V.*, et al.* (2011). Scikit-learn: Machine Learning in Python. J. Mach. Learn. Res *12*, 2825−2830.

Pereira, G.C., Allen, W.J., Watkins, D.W., Buddrus, L., Noone, D., Liu, X., Richardson, A.P., Chacinska, A., and Collinson, I. (2019). A High-Resolution Luminescent Assay for Rapid and Continuous Monitoring of Protein Translocation across Biological Membranes. J Mol Biol *431*, 1689-1699.

Pirchi, M., Tsukanov, R., Khamis, R., Tomov, T.E., Berger, Y., Khara, D.C., Volkov, H., Haran, G., and Nir, E. (2016). Photon-by-Photon Hidden Markov Model Analysis for Microsecond Single-Molecule FRET Kinetics. J Phys Chem B *120*, 13065-13075.

Rabiner, R.L. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. Proc. IEEE *77*, 257-286.

Robson, A., Booth, A.E., Gold, V.A., Clarke, A.R., and Collinson, I. (2007). A large conformational change couples the ATP binding site of SecA to the SecY protein channel. J Mol Biol *374*, 965-976.

Robson, A., Gold, V.A., Hodson, S., Clarke, A.R., and Collinson, I. (2009). Energy transduction in protein transport and the ATP hydrolytic cycle of SecA. Proc Natl Acad Sci U S A *106*, 5111-5116.

Schmidt, M.G., Rollo, E.E., Grodberg, J., and Oliver, D.B. (1988). Nucleotide sequence of the secA gene and secA(Ts) mutations preventing protein export in Escherichia coli. J Bacteriol *170*, 3404-3414.

Schulze, R.J., Komar, J., Botte, M., Allen, W.J., Whitehouse, S., Gold, V.A., Lycklama, A.N.J.A., Huard, K., Berger, I., Schaffitzel, C.*, et al.* (2014). Membrane protein insertion and proton-motive-force-dependent secretion through the bacterial holo-translocon SecYEG-SecDF-YajC-YidC. Proc Natl Acad Sci U S A *111*, 4844-4849.

Sindbert, S., Kalinin, S., Nguyen, H., Kienzler, A., Clima, L., Bannwarth, W., Appel, B., Muller, S., and Seidel, C.A. (2011). Accurate distance determination of nucleic acids via Forster resonance energy transfer: implications of dye linker length and rigidity. J Am Chem Soc *133*, 2463-2480.

Sun, B., and Wang, M.D. (2016). Single-molecule perspectives on helicase mechanisms and functions. Crit Rev Biochem Mol Biol *51*, 15-25.

Tanaka, Y., Sugano, Y., Takemoto, M., Mori, T., Furukawa, A., Kusakizako, T., Kumazaki, K., Kashima, A., Ishitani, R., Sugita, Y.*, et al.* (2015). Crystal Structures of SecYEG in Lipidic Cubic Phase Elucidate a Precise Resting and a Peptide-Bound State. Cell Rep *13*, 1561-1568.

Van den Berg, B., Clemons, W.M., Jr., Collinson, I., Modis, Y., Hartmann, E., Harrison, S.C., and Rapoport, T.A. (2004). X-ray structure of a protein-conducting channel. Nature *427*, 36-44.

Vandenberk, N., Karamanou, S., Portaliou, A.G., Zorzini, V., Hofkens, J., Hendrix, J., and Economou, A. (2019). The Preprotein Binding Domain of SecA Displays Intrinsic Rotational Dynamics. Structure *27*, 90-101.

Wagoner, J.A., and Dill, K.A. (2016). Molecular Motors: Power Strokes Outperform Brownian Ratchets. Journal of Physical Chemistry B *120*, 6327-6336.

Walczewska-Szewc, K., and Corry, B. (2014). Accounting for dye diffusion and orientation when relating FRET measurements to distances: three simple computational methods. Phys Chem Chem Phys *16*, 12317-12326.

Ward, J.H. (1963). Hierarchical Grouping to Optimize an Objective Function. Journal of the American Statistical Association *58*, 236-&.

Whitehouse, S., Gold, V.A., Robson, A., Allen, W.J., Sessions, R.B., and Collinson, I. (2012). Mobility of the SecA 2-helix-finger is not essential for polypeptide translocation via the SecYEG complex. J Cell Biol *199*, 919-929.

Wickles, S., Singharoy, A., Andreani, J., Seemayer, S., Bischoff, L., Berninghausen, O., Soeding, J., Schulten, K., van der Sluis, E.O., and Beckmann, R. (2014). A structural model of the active ribosome-bound membrane protein insertase YidC. Elife *3*, e03035.

Yang, L.W., Eyal, E., Bahar, I., and Kitao, A. (2009). Principal component analysis of native ensembles of biomolecular structures (PCA_NEST): insights into functional dynamics. Bioinformatics *25*, 606-614.

Ye, X., Lin, J.B., Mayne, L., Shorter, J., and Englander, S.W. (2019). Hydrogen exchange reveals Hsp104 architecture, structural dynamics, and energetics in physiological solution. Proceedings of the National Academy of Sciences of the United States of America *116*, 7333-7342.

Zander, C.B., Mayne, L., Englander, S.W., and Zheng, X.L. (2016). Hydrogen Exchange Mass Spectrometry Reveals Allosteric Interactions Between the Distal and the Proximal Domains of Human ADAMTS13. Blood *128*, 2535.

Zimmer, J., Nam, Y., and Rapoport, T.A. (2008). Structure of a complex of the ATPase SecA and the protein-translocation channel. Nature *455*, 936-943.

Figure legends

Figure 1. Conformational states of the LG of SecYEG. (A) Left: The structure of SecYEG:A (PDB 5EUL[(Li et al., 2016), SecY – red, SecE – yellow, SecA – blue) modelled with translocating substrate (green) (Corey et al., 2019). Right: Cartoon representation of the translocating SecYEG:A complex showing fluorophore placement on the lateral gate (green and orange stars, see **Figure 1 –figure supplement 1** for structure based modelling of the dyes). (B) Top down view of SecYEG as seen from the cytoplasm. C-terminus region is shown in orange, N-terminus in blue, SecE in yellow and plug in magenta (PDB: 3DIN (Zimmer et al., 2008)). (C) Side view of helices lining the lateral gate. The open (green) and closed (red) states were directly discernible in the SecY structures while Principal Component Analysis (PCA) of high-resolution structures was used to obtain a model of the intermediate (orange) state (see Figure 4 – supplements for details) in which the separation between LG helices is half-way between open and closed states. (D) Left: Grey bins represent the experimentally observed distribution of E values for SecYEG:SecA while the green, orange and red overlaid Gaussians correspond to the three conformational states (corrected for the population of Apo SecYEG - **Figure 1 – supplement figure 2**). (E) Relative population of closed, intermediate and open conformational states of LG in SecYEG:A in each experimental condition (data in **Figure 1 –figure supplement 3**).

Figure 1 –figure supplement 1. (A) Accessible volumes for the donor and acceptor fluorophore simulated on the MD simulation of SecYEG:A in the ADP bound state. to account for their linker length and dynamics using representative MD structures. (B) FRET efficiency distributions derived from the accessible volume simulations for the three principal lateral gate states.

Figure 1 –figure supplement 2. Kernel density estimate of probability density function of Apo SecYEG FRET efficiency values used to correct the FRET efficiency histograms for each experimental condition.

Figure 1 –figure supplement 3. FRET efficiency histograms for each experimental condition (corrected for the 50% contribution from Apo SecYEG in the opposite orientation) and globally fitted to three Gaussians representing closed, intermediate and open state (panels A-E). (A) SecYEG:A; (B) SecYEG:A + ATP; (C) SecYEG:A + AMP-PMP; (D) SecYEG:A + ADP. ; (E) SecYEG:A + ATP + proSpy exhibiting large discrepancy between

the fit and the experimental data (F) SecYEG:A + ATP + proSpy fitted independently to three Gaussians exhibiting discernible shift in the position of the closed state FRET component.

Figure 2. Dynamic behaviour of SecYEG:A LG in each experimental condition. Top: 2D FRET E and lifetime (τ_D) analysis (see **Figure 2 – figure supplement 1** for details). The solid diagonal and dashed lines represent the static and dynamic fits respectively. Horizontal dashed coloured lines indicate the centres of the E_{FRET} values of the closed (red), intermediate (orange) and open (green) states. Bottom: Dwell times recovered from HMM fitting of individual bursts with 0.5 ms binning (**Figure 2 – figure supplement 3**).

Figure 2 –figure supplement 1. Confocal FRET detection and fluorescence lifetime analysis. (A) Schematic of vesicle passage through the confocal volume by diffusion which determines the timescale for observation (average burst duration 5.75 ms, see **Figure 2 - figure supplement 2**). (B) Donor (green) and acceptor (red) signal bursts originating from single molecules as they pass the confocal volume. Each burst is composed of many single photons with assigned time of arrival relative to the excitation pulse and thus fluorescence lifetime can be estimated (τ_D). Likewise, the total photon count in each channel across the burst yields (after necessary corrections) an estimate of FRET efficiency E. (C) E and donor τ_D are related since FRET represents non-radiative de-excitation of the donor, thus shortening the lifetime proportionally with increasing the transfer efficiency (right). However, this is strictly true only if the single molecule remains in the same state over the burst duration (left). (D) If the molecule changes state during the burst duration (left) then the estimated E and τ_D values represent different averages over the states. In particular, τ_D is weighted towards higher values due to more photons being contributed by the lower E state into the donor channel. Hence, such dynamic averaging produces an arc-like trace connecting the “static” spots of the two states (right). (E) An extreme case of averaging occurs under conditions when dynamic exchange happens on a much shorter time scale than the burst duration (left). This yields a single averaged peak which is displaced from the “static” line and positioned on the “dynamic” line (right). Note that presence of such averaging would go undetected if only burst averaged intensities were used.

Figure 2 –figure supplement 2. Burst duration distribution from SecYEG:A showing a mean burst width of 5.75 ms.

Figure 2 –figure supplement 3. Dwell time distributions derived from HMM for each experimental condition. (A) Apo SecYEG; (B) SecYEG:A; (C) SecYEG:A + ATP; (D) SecYEG:A + ATP + proSpy; (E) SecYEG:A + ADP; (F) SecYEG:A + AMP-PNP. Colours denote the different FRET states: closed – red, intermediate – orange and open – green. Note that it is impossible to discern whether individual bursts come from the SecA-engaged SecYEG (correctly oriented) or Apo SecYEG (oriented with cytoplasmic interface facing the vesicle lumen), hence HMM analyses contain 50% contribution from the Apo SecYEG. Also note that dwell time analysis requires traces with at least 2 transitions identified by HMM, hence it is biased towards dynamic trajectories. For most conditions such trajectories are abundant (bursts mapping to the "dynamic line" in **Figure 2**) hence dwell time analyses are representative, while for ADP this is not the case since most trajectories do not change.

Figure 2 –figure supplement 4. Frequency of transitions between states under different conditions. C – closed, I – Intermediate, O – open.

Figure 2 –figure supplement 5. Finite duration of transitions. (Top) While HMM analysis identifies transitions as essentially instantaneous (blue line), in many cases it is possible to extract a finite transition time (delineated by dotted vertical red lines). (Bottom) Probability density of transition times extracted for SecYEG:A complex (pooled results from different conditions since there were no significant differences, partly due to lower number of events with a well-defined transition time). Solid red line represents a single exponential fit with characteristic time constant of 20±1 ms.

Figure 3. Identification structurally invariant regions in SecYEG. The invariant core was identified using 14 available structures (listed in Methods for PCA analysis) with high resolution and high sequence similarity (Gerstein and Altman, 1995). Here we project the resulting invariant core on the *T. thermophilus* resting complex (5AWW) (Tanaka et al., 2015). (A) Schematic of SecYEG topology. Trans membrane helices (TMHs) are numbered and the primary cytoplasmic loops (C2-C5), periplasmic loops (P1-P5) and the plug (p) are marked. (B) SecYEG viewed from the side, in position in the lipid bilayer (black lines). Invariant core (shown in black) covers TMHs 1-5 of N-terminal domain (light pink) and TMH 10 of C-terminal domain (red) of SecY. The rest of the C-terminal domain (TMHs 6-8) is free to undergo conformational changes leading to opening and closing of the lateral gate (LG). Plug region (magenta) of SecY, SecE (yellow) and SecG (transparent orange) are structurally highly variable. (C) As in panel (b) but viewed from the cytoplasm.

Figure 4. Visualization of dominant motion identified by principal component analysis (PCA) corresponding to conversion between the open (A) and closed (B) LG. In each panel three orthogonal views are shown together with superposed SecA (single conformation from PDB 5EUL (Li et al., 2016)). SecY ribbon is colored according to the invariant core analysis in **Figure 3**. SecA is colored: 2HF - dark blue, PPXD – magenta. The SecYEG reference secondary structure (open state) is shown in the background including SecE (yellow) and SecG (orange). Structures used for PCA are listed in **Figure 4 – table supplement 1, the** clustering dendrogram is shown in **Figure 4 – figure supplement** 2 and PCA eigenvalues and structure clustering are displayed in **Figure 4 – figure supplement 3**.

Figure 4 - table supplement 1. List of high resolution input structures used for PCA.

Figure 4 –figure supplement 2. PCA Cluster Dendrogram. Clustering using the principal component (PC) 1 and 2 (PC1–2) using the Ward's D2 method (Ward, 1963) yields a cluster dendrogram that can be partitioned into 4 cluster groups. Structures belonging to the cluster representing the open LG are shown in black while closed configurations are in red.

Figure 4 –figure supplement 3. (A) Conformer plot: projection of available structures onto the principal coordinates defined by the user-selected principal components (PCs), coloured as in **Figure 4 –figure supplement 2**. (B) Eigenvalue spectrum: results obtained from diagonalization of the covariance matrix of superposed coordinates. The magnitude of each eigenvalue is given alongside, expressed as the percentage of the total variance (mean-square fluctuation) captured by the corresponding eigenvector. Labels beside each point indicate the cumulative sum of the proportion of the total variance accounted for in all preceding eigenvectors.

Figure 4 –movie supplement 4. SecY structure morphing along PC1 principal coordinate in side view showing the LG (top) and cytoplasmic view (bottom). The invariant core is shown in black, plug is purple. The SecYEG reference secondary structure (open state) is shown in the background as solid cylinders including SecE (yellow) and SecG (orange).

Figure 5. Hydrogen bonding at the interface between SecYEG and SecA. (A) Matrix representation of pairwise changes in hydrogen bonding network delineated from all atom MD simulations in the presence of translocating substrate (**Figure 5 –figure supplement 1** further shows comparison with/without polypeptide substrate). Red dots indicate hydrogen bonds gained between two residues in the ADP (closed) state while green dots represent hydrogen bonds formed in the ATP (open) state. Coloured boxes highlight regions with most changes with the colour and intensity indicating hydrogen bond formation or loss according to the scale on the right.

(B) Visualization of the C4 and C5 loop interactions in the context of the two principal states from PCA analysis. The structure is color coded as in **Figure 4**. Loops interacting with the 2HF are colored dark blue while loops interacting with PPXD are shown in magenta.

An animated version of **Figure 5** is shown in **Figure 5 –Movie supplement 2**.

Figure 5 –figure supplement 1. Hydrogen bond analysis of molecular dynamics simulations of the interface between SecY and SecA. Map of hydrogen bond position (dots) and difference (boxes) with respect to secondary structure of SecY (TMHs 1-10, C – cytoplasmic loops) and domains of SecA (PPXD NBD, HSD, HWD – helical wing domain, and 2HF, see bottom panel for reference structure) in the presence (top) and absence (middle) of pre-protein. H-bonds appearing exclusively in the ADP bound state are shown in red while those present only in the ATP ensemble are green.

Figure 5 –Movie supplement 2. Translocon closing and opening motion extracted by PCA. The Invariant Core is shown in black, 2HF is in dark blue and PPXD is in magenta. In top right panel loops C4 and C5 toggle between dark blue and magenta, indicating interactions with 2HF and PPXD, respectively. Note that SecA is held static since no high-resolution models of this structure directly corresponding to the open and closed LG have been identified thus far.

Figure 6. Dynamic coupling of the SecYEG LG and SecA during the SecA ATPase cycle. (i) The SecA dimer delivers pre-protein to the SecYEG translocon (initiation, see (Fessl et al., 2018)). (ii) In the SecA:ATP state, SecYEG:A interfacial hydrogen bonding between the C4 and C5 loops of SecY and the 2HF and PPXD of SecA, steers the energy landscape associated with LG movement (as measured by the E_{FRET} reaction coordinate) to favour the SecY open state, permitting substrate diffusion. (iii) ATP hydrolysis results in a conformational change in SecA that modifies the SecA:SecY interfacial interactions and

steers the energy landscape to a minimum in which the closed state predominates, abrogating substrate diffusion. (iv) ADP dissociation from SecA permits dynamic LG opening while still favouring the closed state, making substrate diffusion possible. (v) After repeated cycles of ATP hydrolysis, protein translocation is terminated by substrate dissociation and cleavage by signal peptidase.

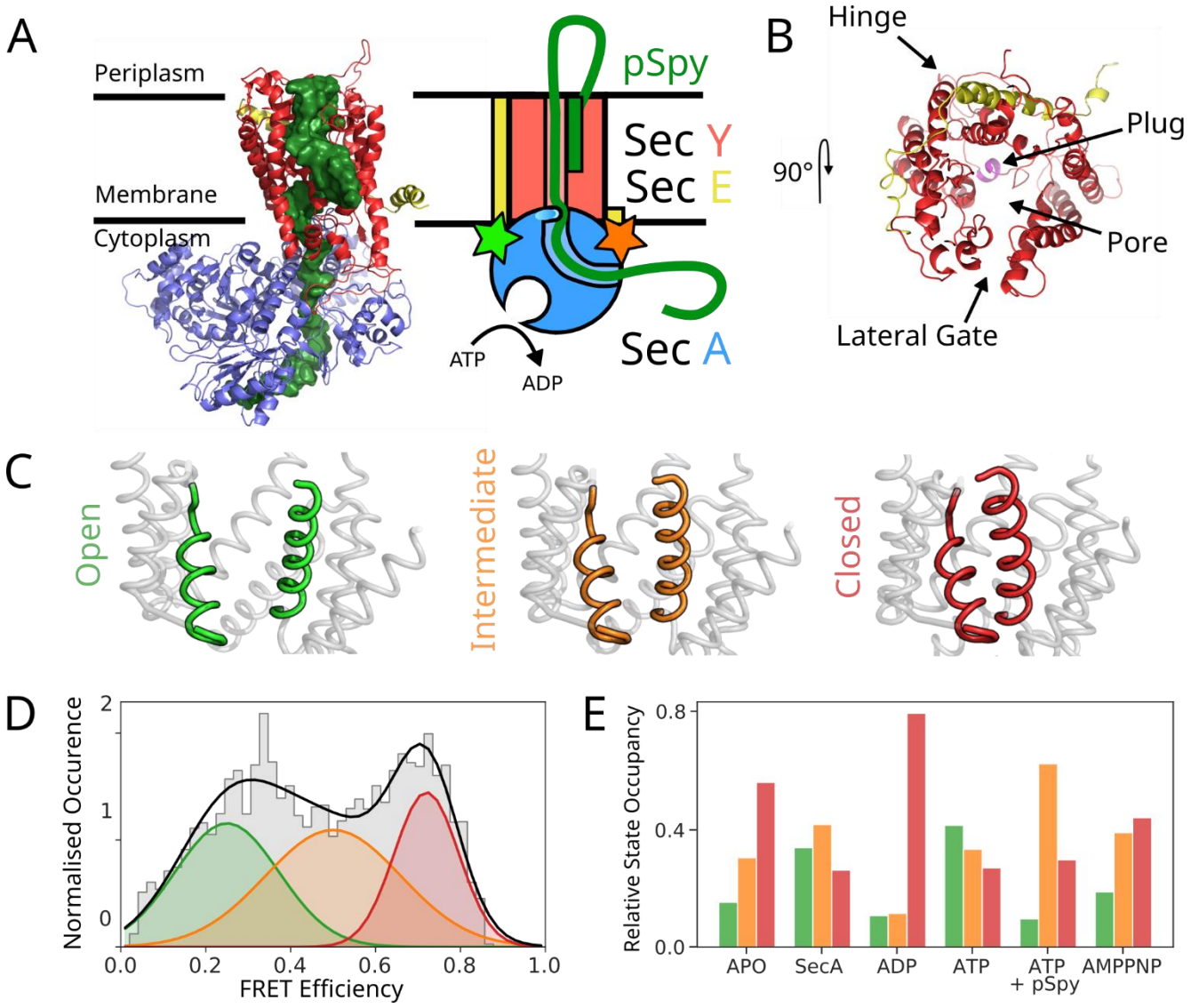

A
Periplasm
Membrane
Cytoplasm
pSpy
Sec Y
Sec E
Sec A
ATP
ADP
B
Hinge
Plug
90°
Pore
Lateral Gate
C
Open
Intermediate
Closed
D
Normalised Occurence
FRET Efficiency
E
Relative State Occupancy
APO
SecA
ADP
ATP
ATP + pSpy
AMPPNP

Figure 1

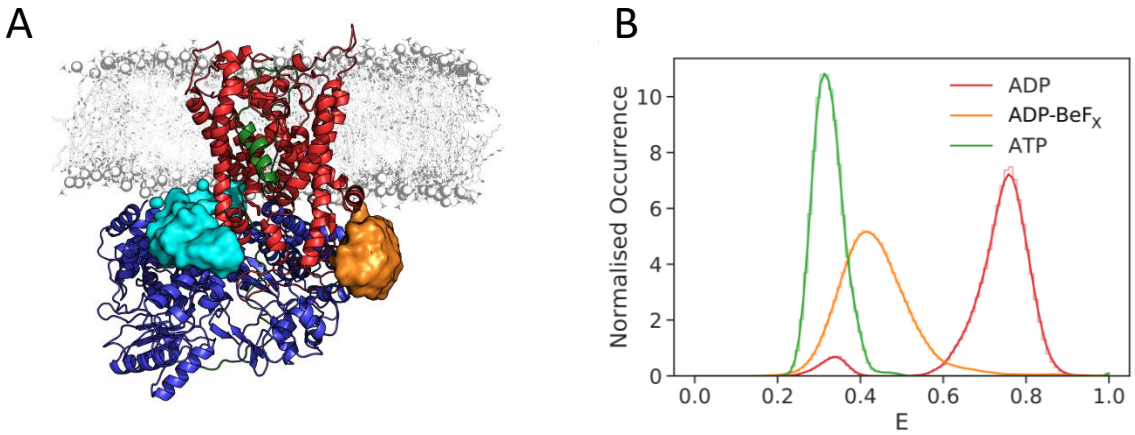


Figure 1- figure supplement 1

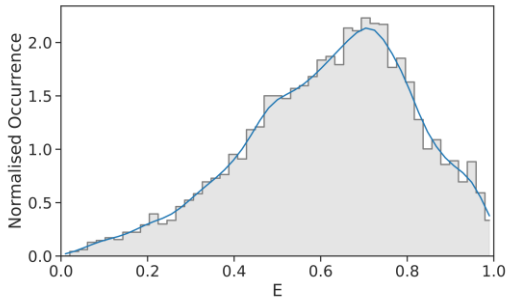


Figure 1 – figure supplement 2

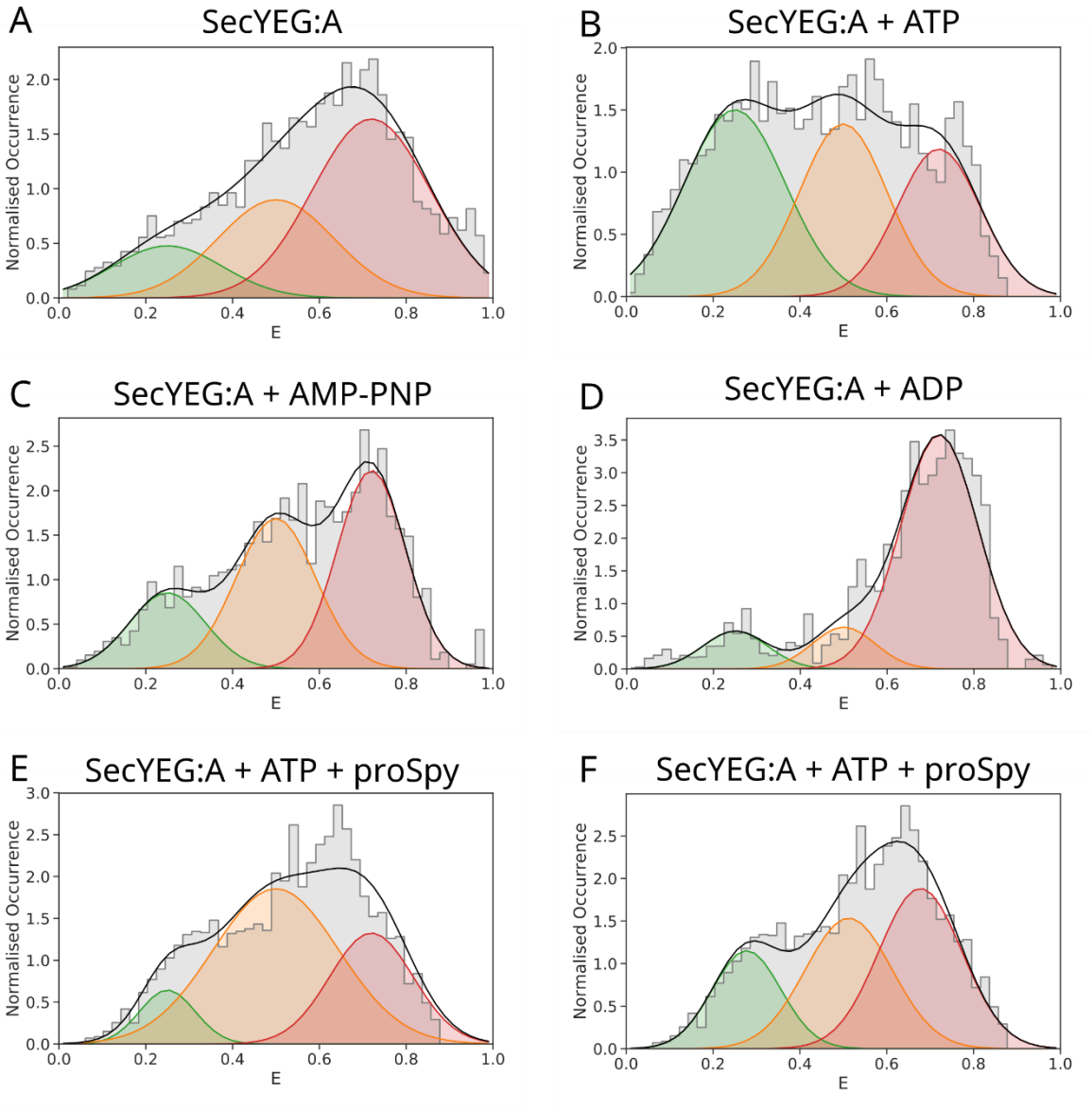


Figure 1 – figure supplement 3

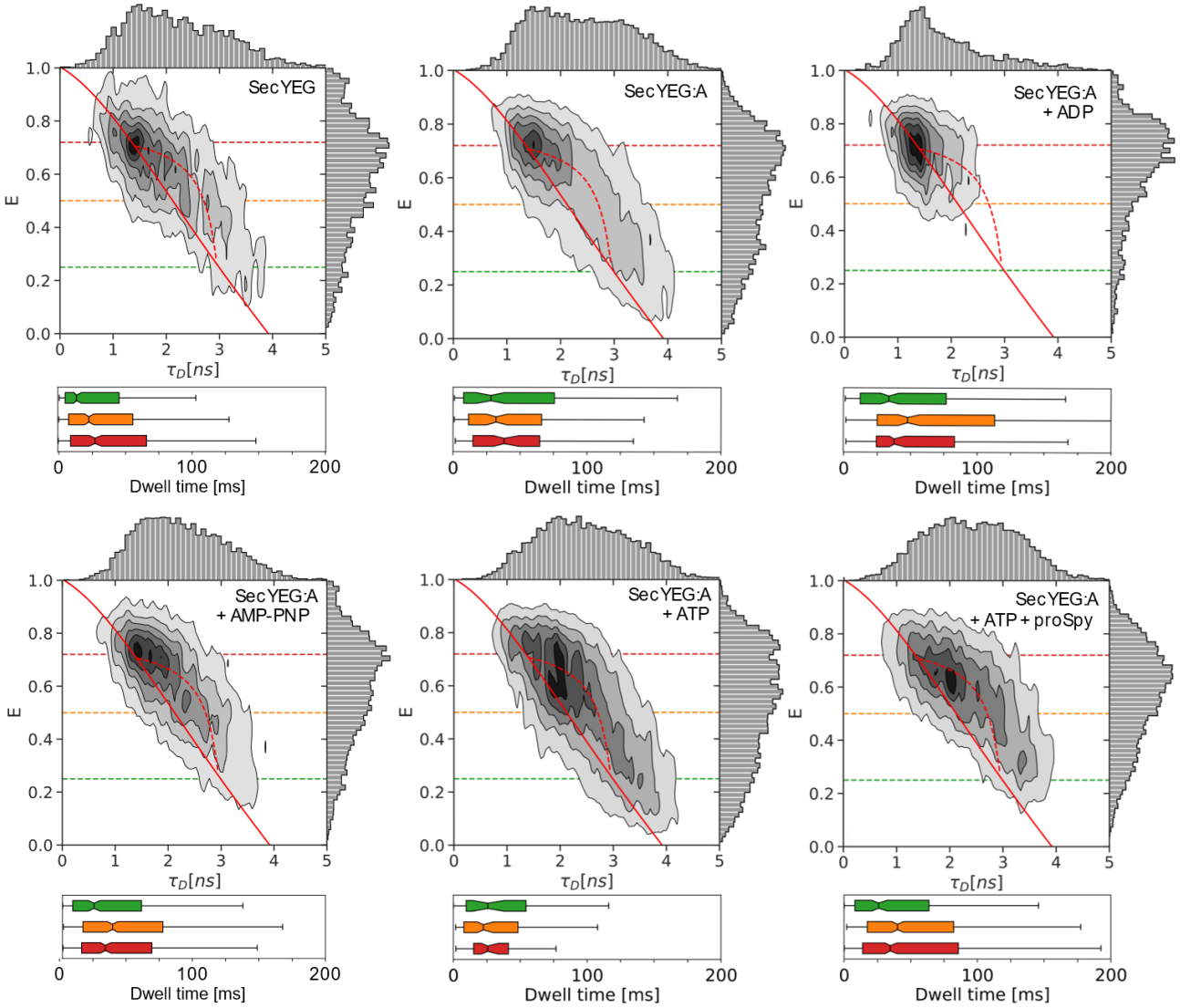

SecYEG
SecYEG:A
SecYEG:A + ADP
SecYEG:A + AMP-PNP
SecYEG:A + ATP
SecYEG:A + ATP + proSpy
E
$\tau_D[ns]$
Dwell time [ms]

Figure 2

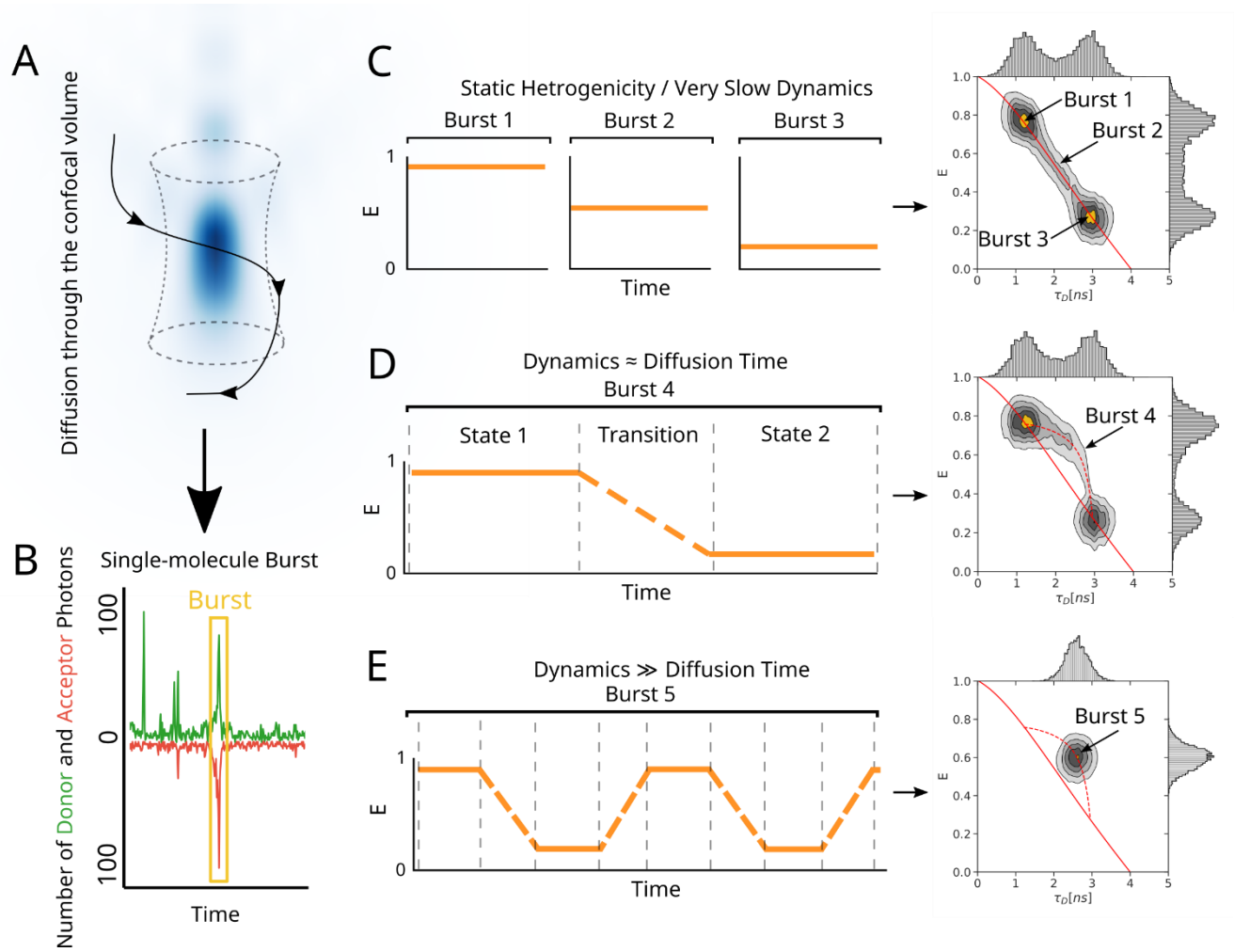


Figure 2 – figure supplement 1

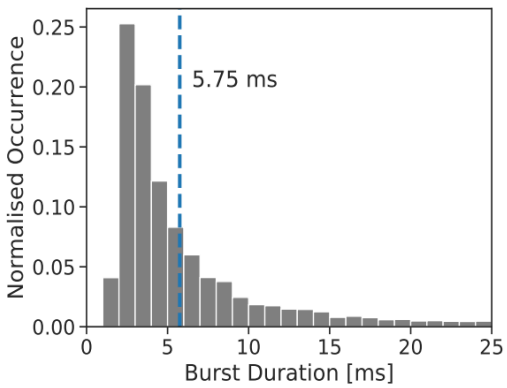


Figure 2 - figure supplement 2

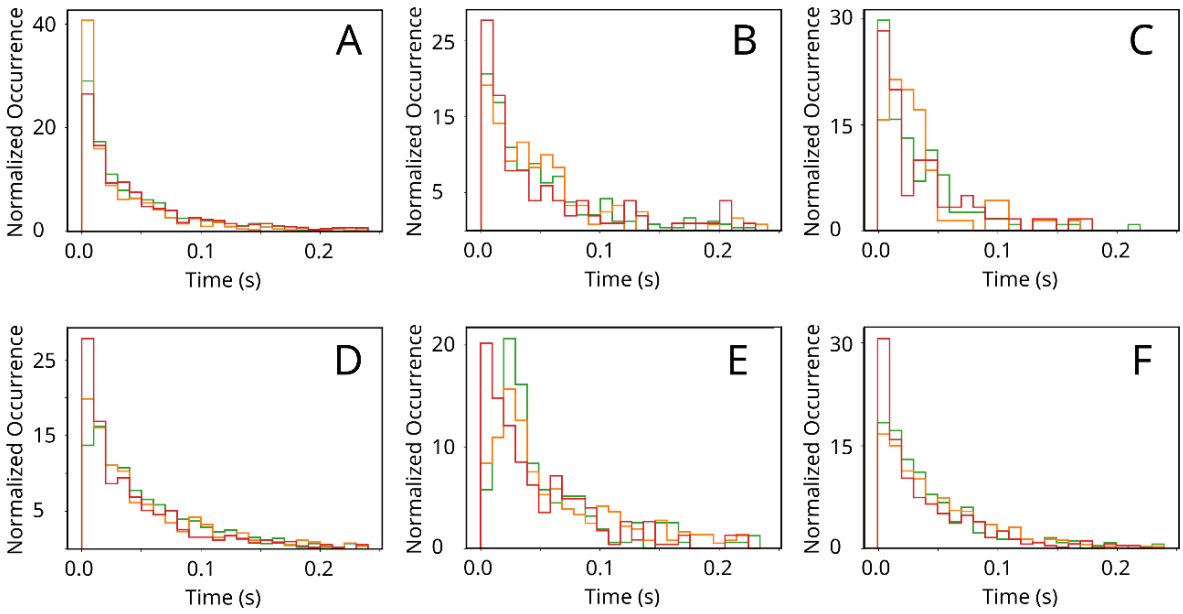


Figure 2 - figure supplement 3

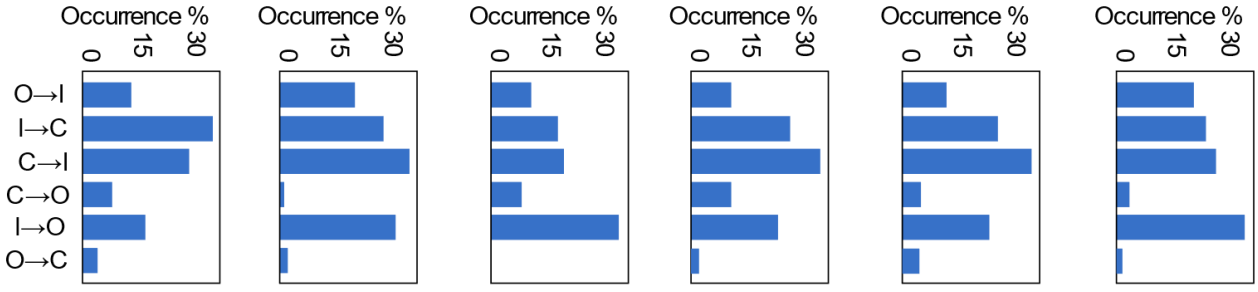
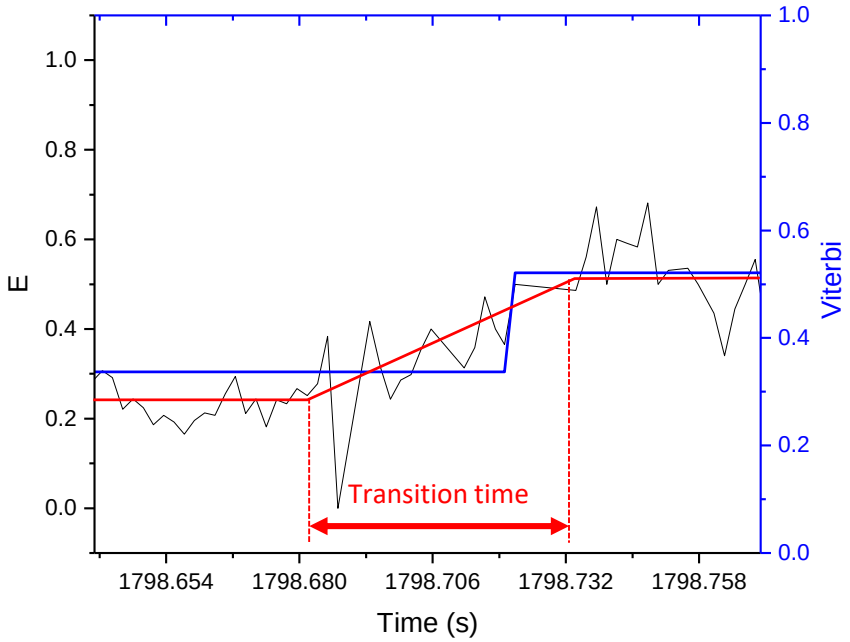


Figure 2 - figure supplement 4



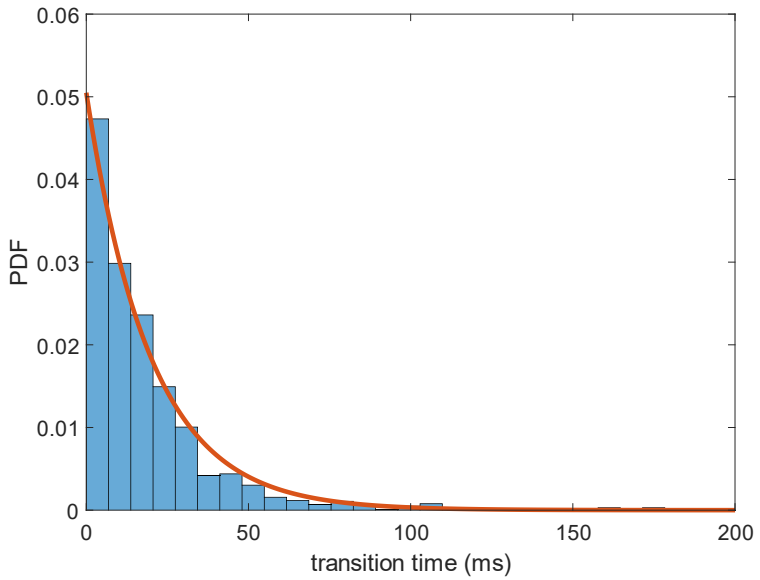


Figure 2 – figure supplement 5

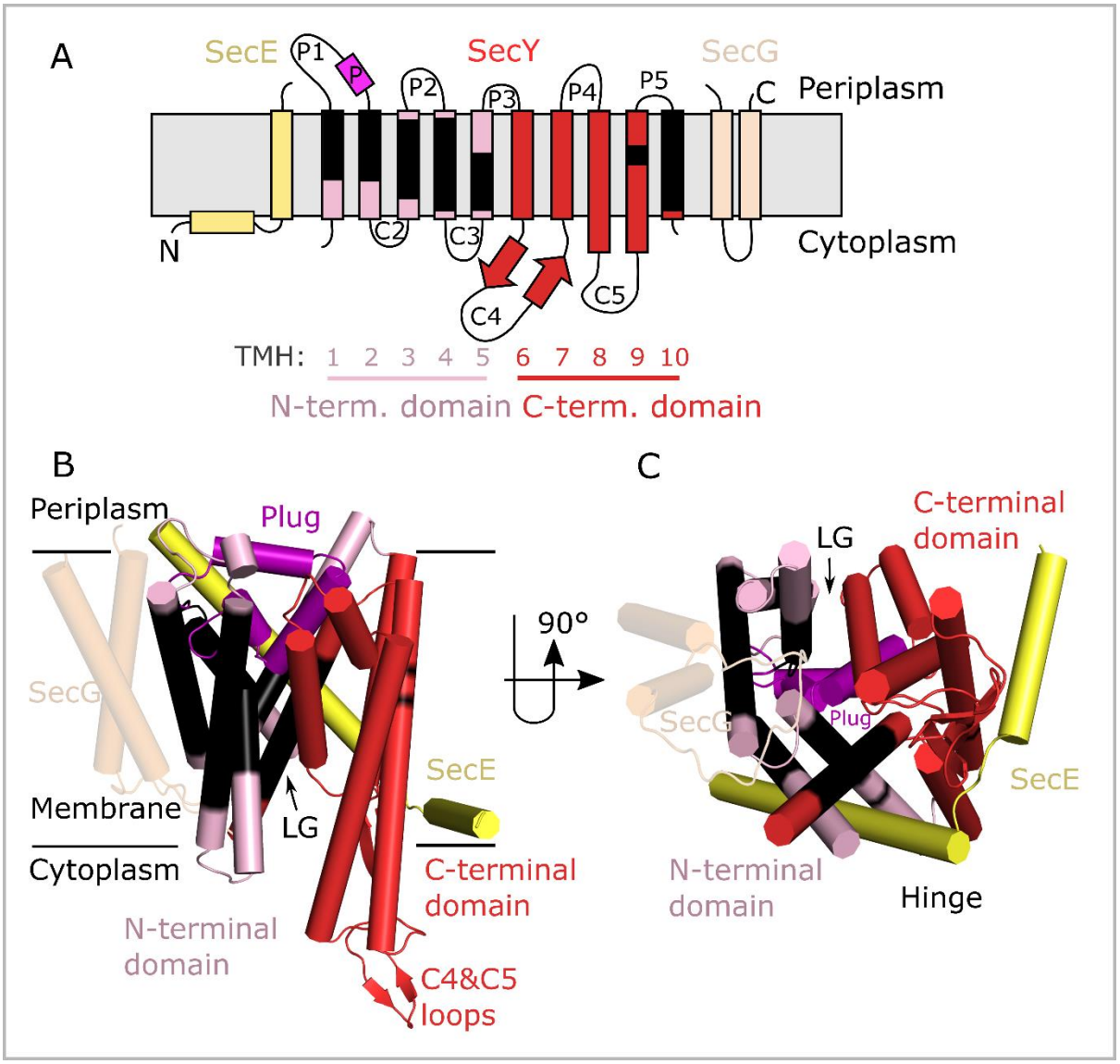

A
SecE
P1
P
SecY
P2
P3
P4
P5
SecG
C
Periplasm
N
C2
C3
C4
C5
Cytoplasm
TMH: 1 2 3 4 5 6 7 8 9 10
N-term. domain
C-term. domain
B
Periplasm
Plug
SecG
Membrane
Cytoplasm
LG
SecE
C-terminal domain
N-terminal domain
C4&C5 loops
90°
C
LG
C-terminal domain
SecG
Plug
SecE
N-terminal domain
Hinge

Figure 3

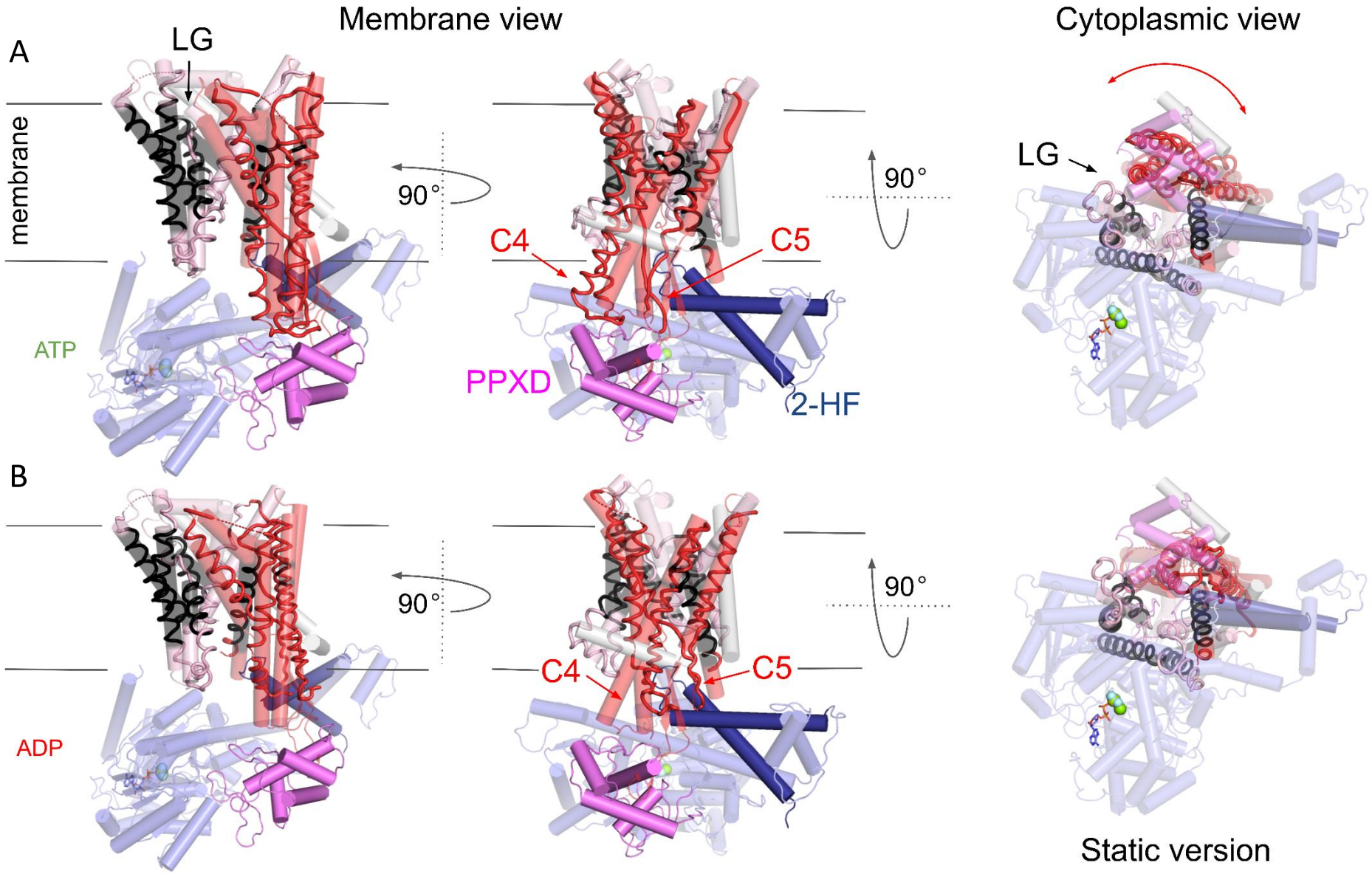

Membrane view
Cytoplasmic view
A
LG
membrane
90°
C4
C5
90°
LG
ATP
PPXD
2-HF
B
90°
90°
C4
C5
ADP
Static version

Figure 4

Structures used for PCA	
PDB ID	Citation
3J46_Y	**DOI:** 10.1038/nature12720 (Park et al., 2014)
5MG3_Y	**DOI:** 10.1038/srep38399 (Botte et al., 2016)
3DL8_G, 3DL8_H	**DOI:** 10.1038/nature07335 (Zimmer et al., 2008)
3DIN_C, 3DIN_F	**DOI:** 10.1038/nature07335 (Zimmer et al., 2008)
3J45_Y	**DOI:** 10.1038/nature12720 (Park et al., 2014)
5GAE_g	**DOI:** 10.1038/ncomms10471 (Jomaa et al., 2016)
5ABB_A	**DOI:** 10.1038/ncomms5103 (Bischoff et al., 2014)
2ZJS_Y	**DOI:** 10.1038/nature07421 (Tsukazaki et al., 2008)
5EUL_Y	**DOI:** 10.1038/nature17163 (Li et al., 2016)
2ZQP_Y	**DOI:** 10.1038/nature07421 (Tsukazaki et al., 2008)
5AWW_Y	**DOI:** 10.1016/j.celrep.2015.10.025 (Tanaka et al., 2015)
5CH4_Y	**DOI:** 10.1016/j.celrep.2015.10.025 (Tanaka et al., 2015)

Figure 4 - table supplement 1

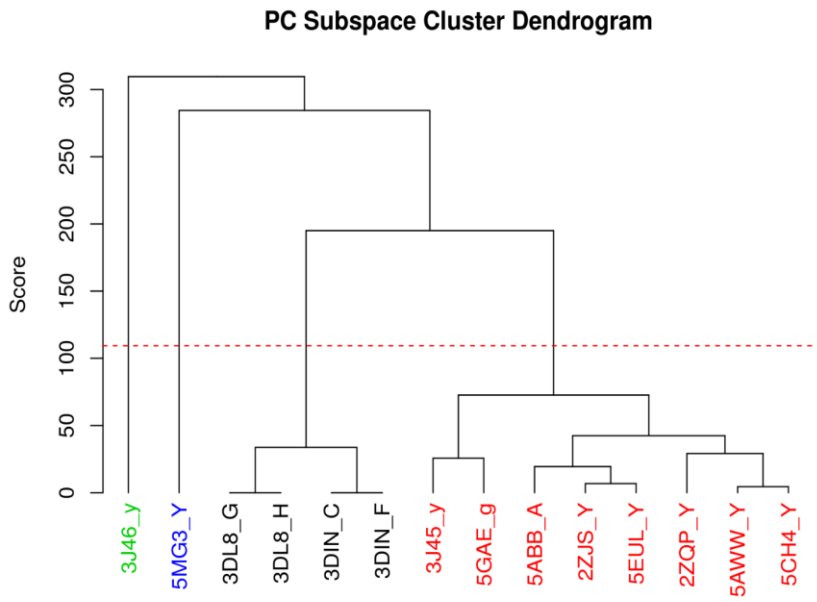


Figure 4 figure supplement 2

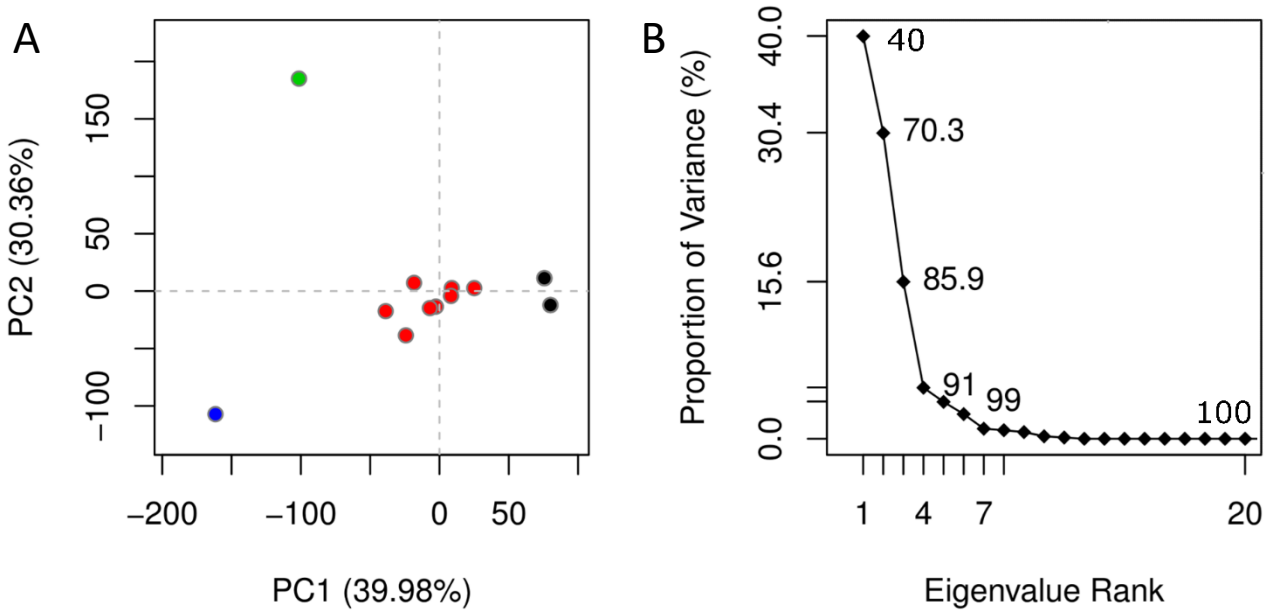


Figure 4 figure supplement 3

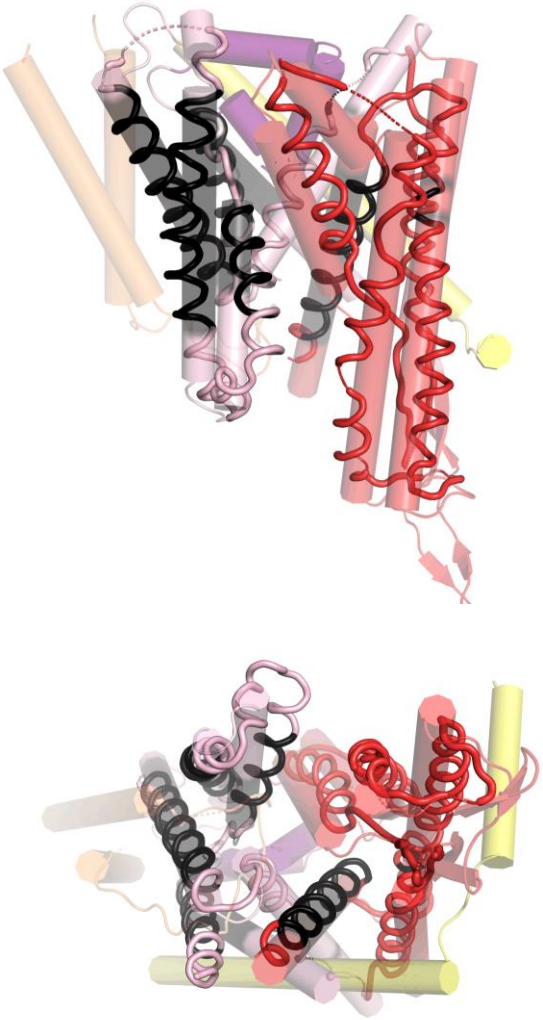


Figure 4 - movie supplement 4

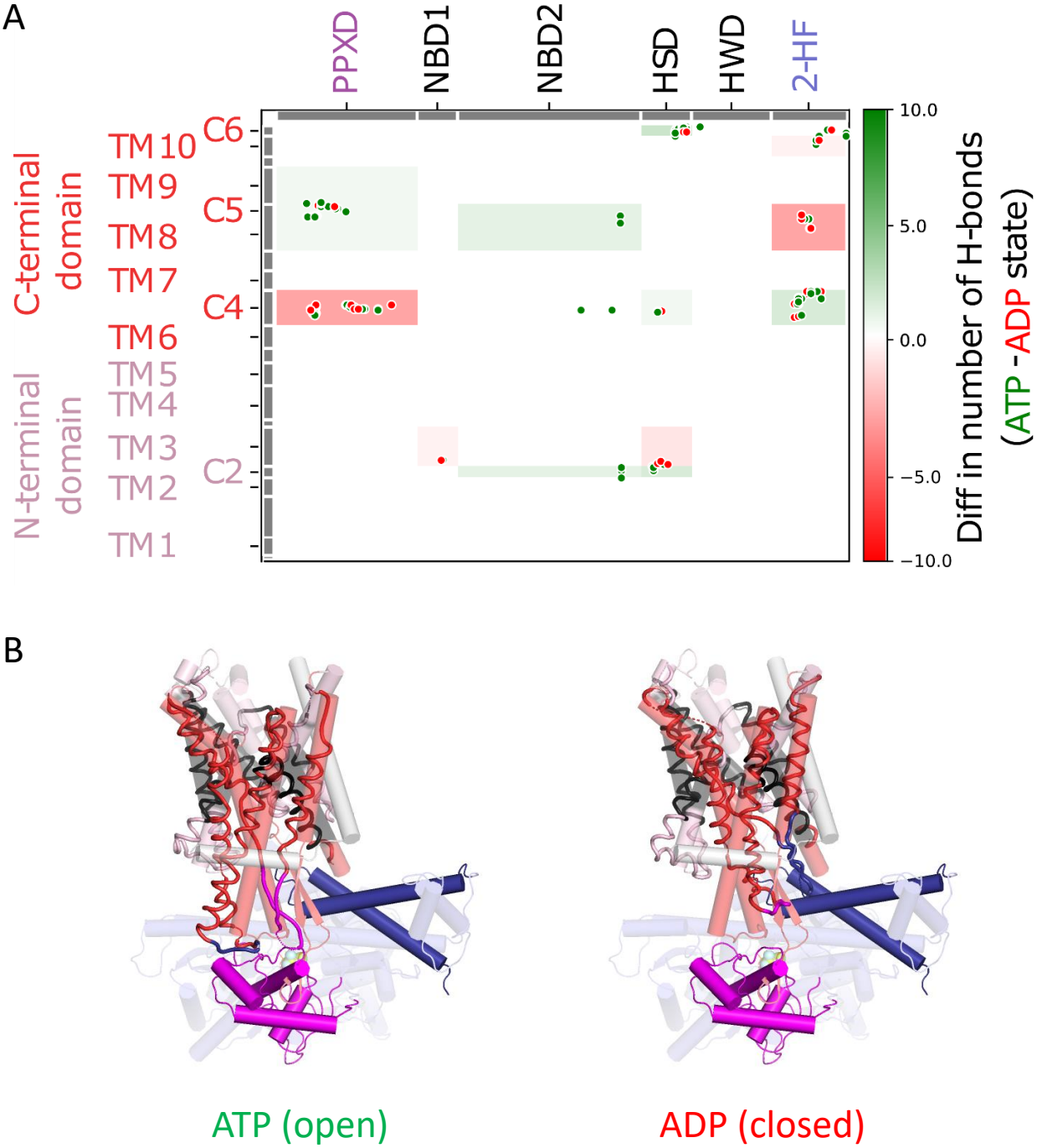

A
PPXD
NBD1
NBD2
HSD
HWD
2-HF
C-terminal domain
N-terminal domain
TM10
TM9
TM8
TM7
TM6
TM5
TM4
TM3
TM2
TM1
C6
C5
C4
C2
10.0
5.0
0.0
−5.0
−10.0
Diff in number of H-bonds
(ATP -ADP state)
B
ATP (open)
ADP (closed)

Figure 5

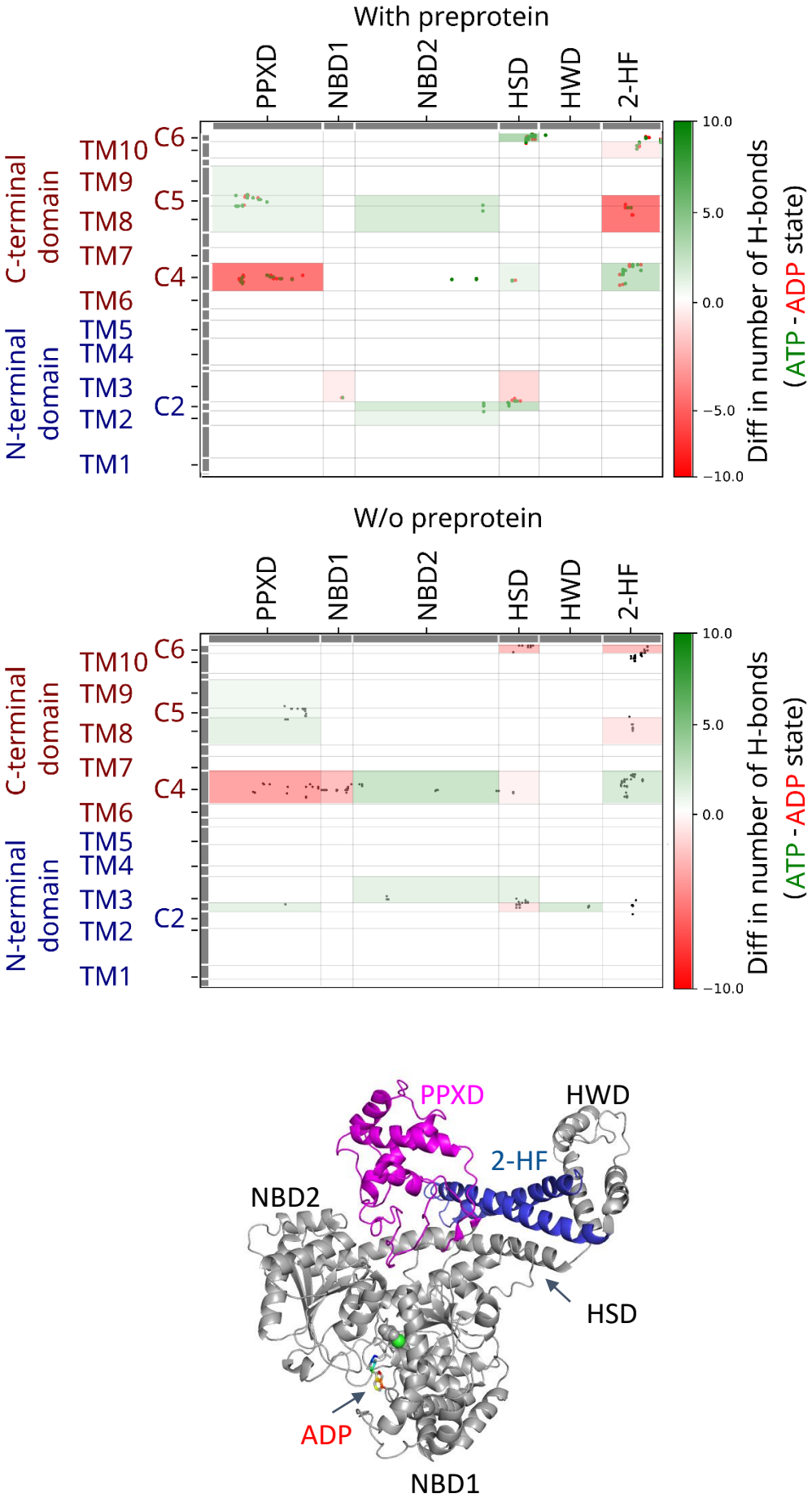
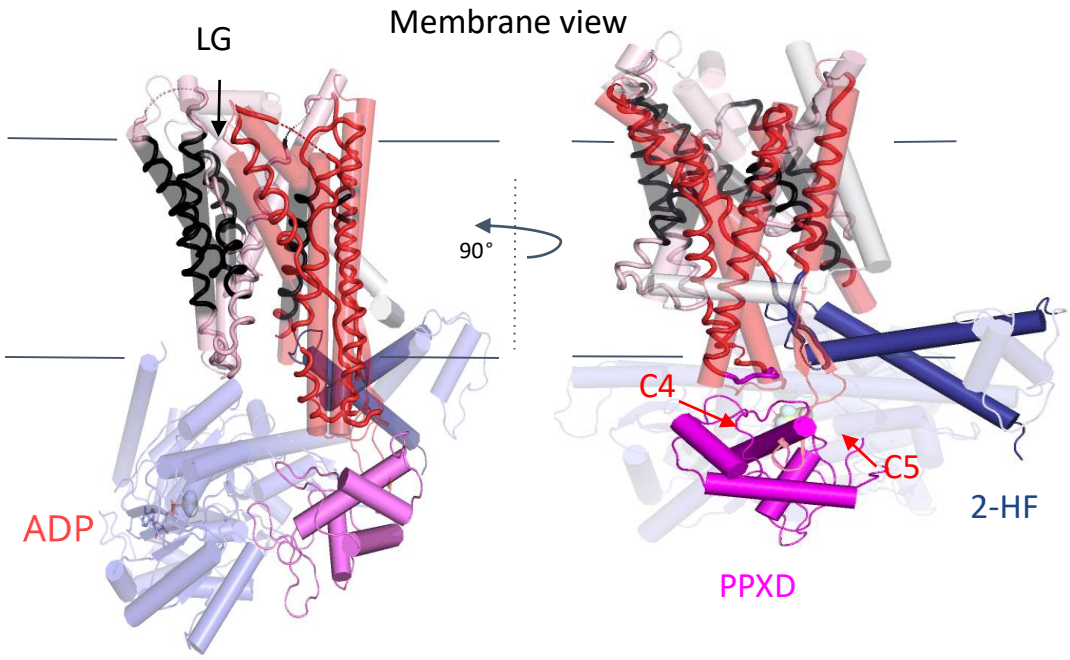
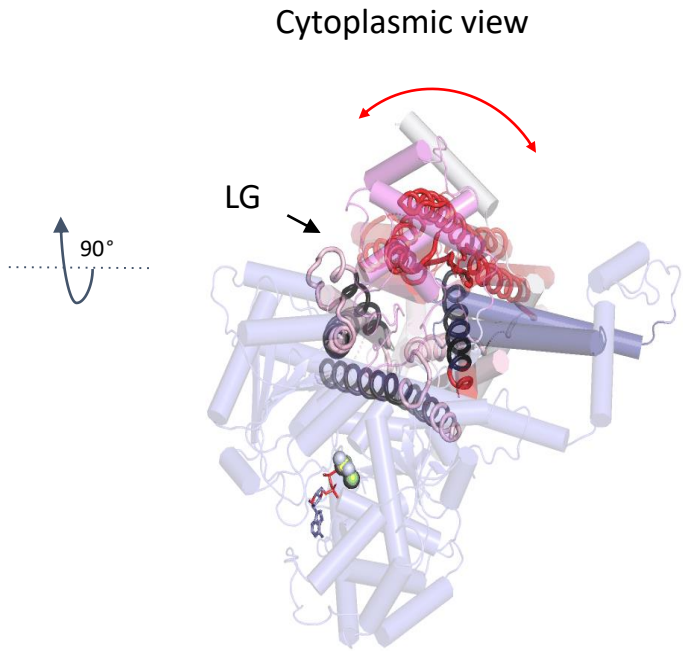


Figure 5 - figure supplement 1





Animated version

Figure 5 movie supplement 2

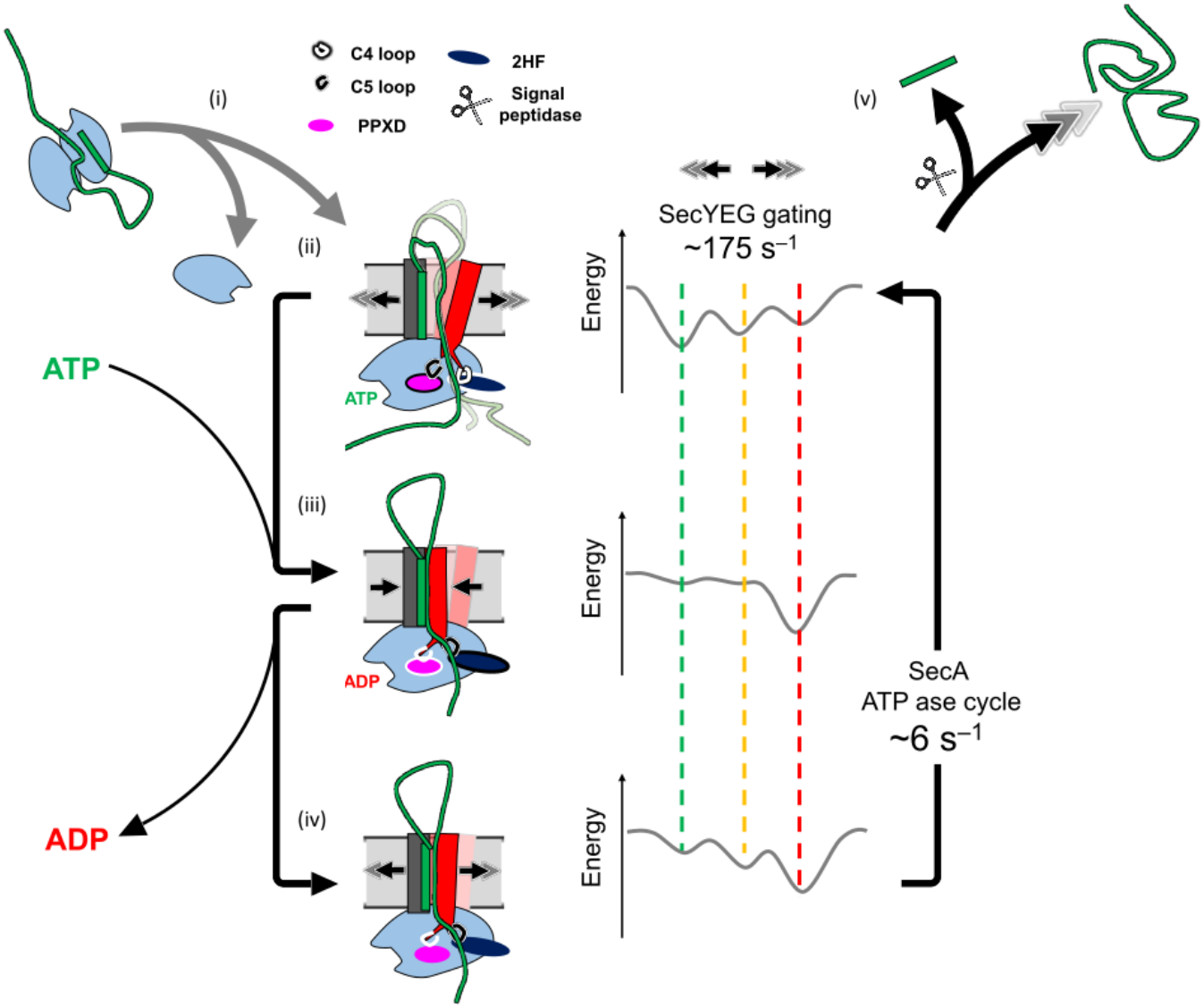

C4 loop
C5 loop
PPXD
2HF
Signal peptidase
(i)
(ii)
(iii)
(iv)
(v)
ATP
ADP
SecYEG gating
~175 s−1
Energy
SecA
ATP ase cycle
~6 s−1

Figure 6