# Identification and functional analyses of cardiolipin binding sites on the bacterial Sec translocase

Robin A. Corey[a], Euan Pyle[b,c], William J. Allen[a], Bruno Miroux[d], Ignacio Arechaga[e], Argyris Politis[b], Ian Collinson[a,1]

[a] School of Biochemistry, University of Bristol, University Walk, Bristol, BS8 1TD

[b] Department of Chemistry, King's College London, 7 Trinity Street, SE1 1DB

[c] Department of Chemistry, Imperial College London, Imperial College Road, London SW7 2AZ, UK

[d] Laboratoire de Biologie Physico-Chimique des Protéines Membranaires, Institut de Biologie Physico-Chimique, CNRS, Univ Paris Diderot, Sorbonne Paris Cité, PSL Research University, Paris, France

[e] Departamento de Biología Molecular and Instituto de Biomedicina y Biotecnología de Cantabria (IBBTEC), Universidad de Cantabria – CSIC - SODERCAN, Santander, Spain

[1] Corresponding author: ian.collinson@bristol.ac.uk

Keywords: SecYEG, protein translocation, molecular dynamics, cardiolipin, native mass spectrometry

## Abstract

The transport of proteins across or into membranes is a vital biological process, achieved in every cell by the conserved Sec machinery. In bacteria, SecYEG associates with the SecA motor protein to carry out secretion of pre-proteins across the plasma membrane, in a process powered by ATP and the trans-membrane proton-motive force. The activities of SecYEG and SecA are modulated by membrane lipids, particularly by the bioenergetically-important phospholipid cardiolipin. Here, we identify two specific cardiolipin binding sites on the *Thermotoga maritima* SecA-SecYEG complex, through application of coarse grained molecular dynamics simulations. We validate the simulation data and demonstrate the conserved nature of the binding sites using *in vitro* mutagenesis of *Escherichia coli* SecYEG, followed by native mass spectroscopy and biochemical analyses. Our data reveal that the two sites account for the preponderance of functional cardiolipin binding to SecYEG, and mediate its roles in the ATPase cycle and protein transport. This study demonstrates the power of coupling coarse grained simulation and experimental analyses towards investigation of both the nature and functional implications of protein-lipid interactions.

## Introduction

The translocation of proteins across and into membranes is an essential cellular process, acting as a key step in the biogenesis of a diverse array of proteins. In every cell, a major portion of this process is handled by the structurally-conserved Sec translocon – SecYEG in prokaryotes and archaea and Sec61 in eukaryotes. In the plasma membrane of gram-negative bacteria, SecYEG can carry out translocation either post-translationally, using the SecA motor protein (Figure 1A), or co-translationally, interacting directly with the ribosome-nascent-chain complex.

The post-translational route is primarily involved in protein secretion, providing the trafficking route for the bulk of periplasmic, outer membrane and secreted proteins. Pre-proteins – still containing their targeting signal sequence – are passed through a pore in the centre of SecY (1, 2) in a process powered by both ATP and the membrane proton-motive-force (PMF). The mechanism of ATP-powered translocation is currently debated (3-5), whereas the mechanism of PMF-powered translocation is almost entirely unresolved. In bacteria, co-translational protein secretion is generally reserved for the insertion of membrane proteins into the inner membrane (6, 7). The insertion process also requires the action of the membrane 'insertase' YidC (8), which associates with the SecYEG core complex along with the accessory sub-complex SecDF and YajC. Together they form the 'holo-translocon' capable of protein secretion and membrane protein insertion (9, 10).

Like many integral membrane proteins, SecYEG is functionally modulated by interactions with specific lipids (11, 12). Our understanding of the nature of these interactions is restricted by the fact that much of the published structural and functional data on SecYEG is in the presence of detergent, limiting the amount of native lipids present. Various studies have aimed at addressing this, with the most prominent findings being that anionic phospholipids, such as cardiolipin (CL) and phosphatidylglycerol (PG), can stimulate SecA ATPase activity both alone and in complex with SecYEG (13, 14) and are important for normal levels of translocation (15, 16). In addition, CL binding has been demonstrated to drive SecYEG dimer formation between adjacent SecE amphipathic helices (14), as well as stabilising the holo-translocon (9).

SecYEG is not unusual in its interaction with CL, which co-isolates with a considerable number of prokaryotic and eukaryotic energy-transducing protein complexes (17). CL has a distinctive structure, comprising two PG molecules joined by a glycerol head group, giving it

two phosphate head groups and four acyl chains (Figure 1B). It has been proposed that each phosphate on CL has a separate $pK_a$ (18), resulting in a bicyclic resonance structure at pH 7, allowing CL to act as a reservoir for protons and buffer against localised shifts in pH (19). However, more recent analyses support similar $pK_a$ values for both phosphates (20, 21), meaning CL would carry a -2 charge at pH 7.

CL binding sites on proteins are typified by the presence of one or more basic residues that form pockets of positive charge on the surface of the protein (12). This has been observed in structural data (22, 23), as well as in simulation data (24, 25) gathered using the MARTINI coarse-grained molecular dynamics (CGMD) force field (26, 27). Although of lower resolution than atomistic modelling, CGMD is particularly well suited to the study of protein-lipid interactions, primarily as a reduced particle count and longer step size produces significant improvement in computational efficiency. This permits both longer and larger systems for the same computational resource: a notable advantage when attempting to simulate the free diffusion of lipids within a bilayer system.

Here, we adapt and extend these analyses towards identification of CL binding sites in *Thermotoga maritima* SecA-SecYEG. The data reveals the presence of two distinct CL binding sites in SecY, as well as supporting a key role in lipid binding for the N-terminus of SecA. We then validate the putative SecY CL binding sites and demonstrate the conserved nature of the SecY-CL interaction using biochemical and native mass spectrometry analyses of *Escherichia coli* SecYEG variants. In addition, we establish the importance of specific CL binding at these sites to heightened SecA-SecYEG activity. Finally, we demonstrate that whilst CL drives SecYEG dimerization, it is probably not *via* specific interaction at these sites.

## Results

### *Coarse-grained analyses of Sec CL interaction using self-assembling membranes*

We applied the MARTINI CG force field (26, 27) to predict Sec-CL interactions, using the crystal structure of SecA-SecYEG as an input model (28). In MARTINI, 4 heavy atoms (*i.e.* C, P, O, N *etc*.) and associated hydrogens are represented by a single bead, allowing CL to be represented by 25 beads (Figure 1C), including two phosphate beads ('PO1' and 'PO2') and a linking glycerol bead ('GL0'). Amino acids are represented with a single bead for the backbone ('BB') and up to 4 beads for the side chain ('SC1', 'SC2', 'SC3' and 'SC4') (Figure 1C; blue beads).

Our initial analyses followed a 'self-assembly' approach, whereby free lipids (with approximately 10% CL, as per a typical bacterial plasma membrane (29, 30)) were allowed to form membranes around a positionally-restrained protein over 1 µs simulations. Bilayer formation was verified using visual analysis with VMD (31), and the post-bilayer time frames from 10 independent simulations were combined into one trajectory. The occupancy of the CL head groups along the plane of the membrane was then computed using density analysis in VMD, with the data revealing key regions of increased CL density (Figure 2A), suggesting specific Sec-CL interactions.

To probe these interactions in more detail, the distance between basic residues in SecA-SecYEG and nearby CL head groups was calculated using tools available in Gromacs (32) for each separate simulation. The data reveal multiple residues implicated in CL binding in SecA-SecYEG (Figure 2B; red data), including several residues with >90% binding occupancy (Figure S1; red data).

### *Coarse-grained analyses of Sec CL interaction using preformed membranes*

Next, we built five different mixed bilayers containing ~1% CL around SecA-SecYEG using the insane.py program (33). This lower concentration was chosen to reduce the high binding apparent from in the self-assembly dataset (Figure S1; red data), allowing us to better identify specific Sec-CL interactions, and model the kinetics of CL binding and release. In addition, the restraints were removed from the protein beads to allow SecA-SecYEG to structurally react to the presence of the membrane. Each bilayer was simulated for an extended period of between 5 and 55 µs (sampling over 100 µs in total), with RMSD analyses of the protein BB beads confirming that the simulations were stable (Figure S2). As before, we determined the CL occupancy for each basic residue in the system, finding a range of 0-11% occupancy over the course of the simulation (Figure 2B; blue data).

Our MD data highlighted several residues at which CL binding occurred. Those displaying the highest CL occupancy were K103, K115, K118, R122, R177 and K181 in SecY, and K54 in SecG (Figure 2C and Table 1). Of these, K103, K118, R122 and R177 in SecY are very poorly conserved across different bacterial species (Figure S3), and SecG is itself a dispensable subunit of the complex (34), with its removal not affecting anionic lipid interactions with the translocon (35). This led us to the conclusion that SecY residues K115 and K181 were of particular importance for CL binding. In the case of CL bound to K181, additional contribution to binding appears to come from the acyl tails, as can be seen by the perturbed acyl tail order parameters (Figure S4).

In addition to these residues, we observe considerable CL binding to the 25 N-terminal residues of SecA ($SecA_N$ Figure 2 and Figure S1), which have previously been shown as important in lipid binding (4, 36).

### *Kinetics of CL binding*

Our longer (5-55 µs) simulations permit insight into the binding kinetics of CL to SecY. Distance analyses between K115/K181 and CL head groups reveals several clear binding events (Figure S5), where CL-residue distance is less than ~1 nm. A more detailed analysis of the 55 µs simulation demonstrates that hundreds of binding events occur at both K115 and K181, lasting from ~100 ns to 2-3 µs, with a very broad sampling of binding times (Figure S6). This indicates that CL freely binds and releases SecY on the ns to µs timescale, similar to analyses of CL binding to other proteins (24, 25).

### *Validation of CL binding sites by native mass spectrometry*

Plotting a heat map of binding onto basic residues in *T. maritima* SecA-SecYEG reveals that the residues identified above form two distinct sites, termed here ‘Site 1’ and ‘Site 2’ (Figure 3). Site 1 comprises the higher CL-occupancy residues K103, K115, K118 and R122, which are all located on trans-membrane helices (TMHs) 2-3 of SecY, between the SecY lateral gate (LG), the SecA helical scaffold domain (HSD) and SecG (Figure 3B). Site 2 comprises K181 and R177 on SecY TMH 4, along with the mid CL-occupancy R15 and R17 on TMH 1. This site is close to the lipid-binding $SecA_N$.

To validate the ability of these sites as *bona fide* CL binding sites, and to ascertain whether they are conserved across bacterial species, we designed *E. coli* SecYEG variants to abolish all their associated positive charges: $SecY^{R113Q,R114Q,K115E,R121Q}EG$ for Site 1 and $SecY^{K20Q,R21Q,R22Q,R181E}EG$ for Site 2 (Figure S7). We analysed the variants using native mass spectrometry (nMS (37-39)), observing peaks for SecE, SecG, SecEG, SecY, SecYG and

SecYE (Figure S8A). The intact SecYEG complex was not detected due to the application of strong collisional activation energy in the mass spectrometer, which is required to remove the DDM micelle from the protein, resulting in the partial dissociation of full complex (40).

Substantial CL binding can be observed to SecY, SecYG and SecYE. We see no CL binding to SecE, SecG, or SecEG, supporting the notion that CL binding occurs mainly to SecY. When using our two variants, we see a significant decrease in CL binding to SecY (Figure 4A), validating the assignment of the two CL binding sites from the CGMD.

For the SecYG and SecYE peaks, no significant difference is seen (Figure S8B) between the WT and variant samples. This indicates that the presence of SecE and SecG are able to compensate for the loss of the binding sites on SecY in our Site 1 and 2 variants. This is supported by higher CL binding to SecYG and SecYE than SecY across all of the samples (Tables 3-4). Part of this effect is most likely through a direct contribution of SecE and SecG to the CL binding sites – particularly through provision of a hydrophobic surface for the acyl tails, as can be seen for SecG in Figure 3B. In addition, part of the effect may come through a stabilisation of the correct SecY conformation, which would explain the higher abundancies of SecYE and SecYG than SecY for all three samples (Table 3). It should also be noted that there is considerable difference between the SecEs of *E. coli* and *T. maritima*, with an additional two TMHs in *E. coli*, which could be having an effect not observable in the *T. maritima* CGMD data.

### *CL binds SecY at two distinct sites*

For each variant, we analysed the effect of CL on SecA stimulation. CL binding to SecYEG is known to stimulate SecA (41), so following SecA ATPase activity provides a readout of this process. We used CL titrations over a range of 0-40 µM, as CL appears to have an inhibitory effect on SecA ATPase activity above this concentration (Figure S9A). When fitting the data to tight binding equations (Figure 9A, dotted lines) the inability to fit the full curve precludes accurate extraction of $K_d$ or $B_{max}$. Therefore, to compare the sites qualitatively, we only fitted the linear concentration range (0-40 µM; Figure 9A, solid lines). The slopes from these fits reveal that CL stimulation of the Site 1 (0.60 ± 0.02 $\mu M^{-1}$ $min^{-1}$) and Site 2 (0.62 ± 0.02 $\mu M^{-1}$ $min^{-1}$) variants are reduced by almost half as compared to wild-type SecYEG (WT; 1.08 ± 0.05 $\mu M^{-1}$ $min^{-1}$) (Figure 4B). The $k_{cat}$ in the absence of CL is also elevated in the Site 1 mutation (18.5 ± 0.31 $min^{-1}$) relative to WT and Site 2 (8.6 ± 0.88 and 7.5 ± 0.30 $min^{-1}$, respectively) (Figure 9A).

Between them, the two sites thus account for nearly 100% of CL binding to SecY, and for most of its stimulatory effect on ATPase activity. Therefore, these analyses strongly support the existence of two conserved CL-binding sites on SecY, as predicted from CGMD.

### *The role of specific CL binding in SecYEG function*

Next, we addressed the physiological effects of CL binding at these sites on translocation efficiency (15, 16) and dimer formation (14). We carried out translocation assays using inverted membrane vesicles (IMVs) over-expressing variant or WT SecYEG, which revealed that the two variants generally have a lower translocation efficiency than WT SecYEG (Figure 4C & S9B, “+CL”).

To test whether this was due to perturbation of CL binding or a consequence of structural changes in the complex conferred by the residue substitutions alone, we repeated the assay

using IMVs from a strain of *E. coli* C43 in which all three enzymes involved in CL biosynthesis had been knocked out (29). The data reveal both variants behave in the same way as the WT in the absence of CL (Figure 4C & S9B, "-CL"). Thus, the action of CL binding at these two sites stimulates translocation.

To investigate the role of CL on SecYEG dimerization, further CGMD simulations were run on a system containing 16 copies of the *Thermus thermophilus* SecYEG translocon (42) in a membrane with either 10% or 0% CL. Visual analysis of the systems at 1.1 µs seems to support a role for CL in SecYEG dimerization (Figure 5A), although the dimeric forms observed are highly heterogenous (Figure 5A: SecE amphipathic helix shown in blue for context, as per Figure 5B). Analysing the inter-protomer distances across the last 250 ns reveals 4 dimers are present for >75 % of the time in the CL-containing membrane, but only 2 in the non-CL membrane (Figure 5C). This produces an average time spent in dimer form of 45 % for SecYEG in CL membranes but only 25 % for non-CL membranes. These data suggest that CL has an important role in stabilising SecYEG dimerisation, as previously noted (14). However, whilst CL binding is observed at the equivalent Sites 1 and 2, in most cases this does not form part of the dimer binding interface. This suggests that CL acts in a more general manner, either stabilising each protomer or changing the biophysical properties of the membrane to encourage dimerisation.

Together, the results demonstrate that specific binding of CL to the sites identified by CGMD analyses is important to SecYEG activity, having a stimulatory effect on both SecA ATPase and pre-protein translocation. The importance of CL in formation of SecYEG dimers however, is likely not driven by CL binding at these sites.

## Discussion

Integral membrane proteins are strongly affected by the lipid environment in which they are localised, through both specific and nonspecific interactions (11, 12). Whilst structural data can occasionally reveal the binding of specific phospholipids to membrane proteins, more often than not the use of harsh detergents during purification precludes this. As a result, biophysical analyses of lipid binding can be a powerful complement to structural data.

Here, we have used the combined powers of computational (CGMD) and experimental (biochemical assays and nMS) approaches to identify and characterise specific CL binding sites on the bacterial Sec translocon.

The positions of the two identified CL-binding sites (termed here Site 1 and Site 2) are striking when considered in the context of SecA-SecYEG function. Site 1 is positioned on the edge of the LG, so that the bound CL will contact 2 of the 3 LG helices. Previous data have shown that the LG can adopt an open or closed state (2, 5, 28). The transition between these two states has been implicated as an important feature of the protein translocation process (5). Thus, the heightened transport activity induced by specific binding of CL might relate to its effect on the equilibrium between these conformations. In addition, CL binding at Site 1 will position the phosphate head group immediately adjacent to residues on the SecA HSD that have previously been shown to respond to lipid binding (43). This may be important in the activation of SecA, and likely explains the raised basal ATPase rate of the Site 1 variant (Figure 4A).

Site 2 is located on the other side of the channel, formed between SecY and the edge of SecG. This site is in direct contact with the N-terminus of SecA, previously shown to be important for SecA-membrane binding in the *E.coli* system (4), particularly residues 23-25 (36). The CGMD data support this observation, with considerable CL binding observed for

this region, including residue Lys-16 (Figure 3B “K16”), the *T. maritima* equivalent of the *E. coli* residue 23. A CL molecule bound to Site 2, therefore, will likely have an important role in SecA activation.

While SecYEG dimerization does not appear to be driven by specific interactions at Sites 1 or 2, this does not preclude their involvement in other protein-protein interactions. The preliminary structure of the holo-translocon (44) predicts that Sites 1 and 2 are proximal to the bacterial membrane insertase YidC in this complex; indeed, the CGMD data position the CL head group directly on the C1 region of YidC (45). The C1 region contains multiple highly conserved basic residues, including R396, less than 1 nm from R114 of Site 1 (both *E. coli* numbering) when complexed in the holo-translocon (44). As with SecYEG alone, CL has been implicated in mediating the interaction of SecA with the HTL (9). Together, these results demonstrate the importance of specific protein-lipid interactions to membrane protein structure and function, and the power of CGMD to identify and probe these interactions in other energy transducing membrane embedded machines.

## Materials and methods

*Coarse grained simulations*

Simulations of SecA-SecYEG were built using the *T. maritima* crystal structure (PDB 3DIN (28)), with the nucleotide removed and missing loops remodelled in Modeller (46). Following this, two distinct setups were used, each described below.

Firstly, for the self-assembly method, atoms were converted to CG beads using the Martinize method (47), using MARTINI 2.2. Next, the protein beads were built in a 15x15x15 nm cubic box, and 460 distearoyl phosphoethanolamine (DSPE) and 52 CL molecules were added before solvation and addition of counter ions. The CL parameters used are of bacterial di-1-palmitoyl-2-oleoyl-phosphatidylglycerol in a -2 charge state (48, 49). The systems were minimized using steepest descents for 2500 steps of 20 fs, before two rounds of equilibration: firstly 100 ps of NPT (constant-temperature, constant-pressure) equilibration at 300K with 2 fs time steps, then 5 ns of NPT equilibration at 340K with 20 fs time steps. Finally, simulations were run for 1 µs at 340K using 20 fs time steps, with positional restraints applied to the protein beads. The frames in which the bilayer had not yet formed – or in one case formed in the yz axis – were removed, and the simulations were analysed for CL density in VMD (Theoretical and Computational Biophysics group, University of Illinois at Urbana-Champaign), to identify possible CL interaction sites.

Alternatively, for the insane.py method, symmetrical membranes of 9 CL and 823 dipalmitoyl phosphatidylcholine (DPPC) molecules were built around the protein beads using insane.py (33), modified here for the inclusion of CL. The protein beads were described using MARTINI 2.2, with additional harmonic bonds applied to supplement the non-bonded interactions and help preserve the higher-order structure of the input models, with omission of these bonds leading to considerable deformation of the protein (data not shown). The bonds had a force constant of 500 kJ $mol^{-1}$ $nm^{-2}$ and an upper bond length cut-off of 0.9 nm. Next, the systems were minimized and equilibrated as described above, before production simulation. For these simulations, RMSD analyses were carried to ensure stability of the protein beads (Figure S2). These simulations were largely run on EPCC’s Cirrus HPC Service and the ARCHER UK National Supercomputing Service.

As well as modelling of the SecA-SecYEG systems, models were built from *T. thermophilus* SecYEG molecule (PDB 5AWW (42)). For this, 16 protomers were aligned evenly across a

40x40x12 nm box, and built into a membrane of either 3499 DPPC and 388 CL molecules ("CL") or 3888 DPPC molecules with no CL ("NoCL") using the insane.py method (33). The systems were minimized and equilibrated as described above, before further equilibration for 200 ns and finally a 1.1 µs production simulation. Dimer formation was carried out through iterative minimum distance analyses for each protomer, where a distance of >0.9 nm was considered to be bound in a dimer.

All simulations were run in Gromacs 5.1.2 (32). Non-bonded interactions were treated with a switch function from 0-1.2 nm and 0.9-1.2 nm for the Coulomb and Lennard-Jones interactions, respectively. Temperature coupling was achieved with the Bussi-Donadio-Parrinello thermostat (50), and semi-isotropic pressure coupling was by the Parrinello-Rahman barostat (51, 52).

Images of proteins were made in PyMOL (53) or VMD (31). Graphs were plotted using Microsoft Excel, GraphPad Prism or Matplotlib.

*Protein expression and purification*

Mutations were introduced using the QuikChange protocol (Stratagene) to *secEYG* in a pBAD expression plasmid. The SecYEG variants were expressed as previously described (54); briefly, each variant was expressed in C43 *E. coli*, which were lysed and the membranes isolated by ultracentrifugation. The SecYEG protein was solubilised from the membrane fraction using 1% DDM (n-Dodecyl β-D-maltoside) and purified using nickel chromatography (in 0.5% C12E9 for the CL-titration experiments or in 0.1 % DDM for the nMS) followed by size exclusion chromatography and reverse ion-exchange chromatography (in 0.1% C12E9 for the CL-titration experiments or in 0.02 % DDM for the nMS). The C12E9 purification has previously been shown to remove all bound CL, while the DDM purification allows some CL to remain bound (14). The purified proteins were concentrated and aliquoted for storage at -80$^{o}$C.

SecA and the model substrate proOmpA were expressed and purified as described previously (55). To facilitate quantification by western blot, a C-terminal minimal V5 tag (IPNPLLGL) was added to the proOmpA gene by PCR, and confirmed by DNA sequencing.

*ATPase assays*

*In vitro* ATPase assays were performed as described previously (55), with 0.3 µM SecA and 2.4 µM SecYEG in 0.1% C12E9, in $TK_{150}M$ buffer (20 mM Tris pH 7.5, 150 mM KCl, 2 mM $MgCl_2$) with 0.03 % C12E9. Reactions were followed by coupling ATP hydrolysis to NADH oxidation, measuring absorbance at 340 nm. To each reaction, a set amount of CL in 0.6 % C12E9 was added to produce a final concentration of between 0 and 80 µM. Note that above 80 µM, CL appears to have an inhibitory effect on SecA ATPase. Fixed volumes of each reagent were used to ensure that identical C12E9 concentrations were present in each reaction (0.1 %). Data were initially fitted to a tight-binding equation (Figure 9A; dotted line (5)), but the quality of fit was too low to extract meaningful information from. Data from 0-40 µM were therefore fitted by linear regression instead.

*Native mass spectrometry*

Capillaries for nMS were produced using a Model P-97 capillary puller (Sutter Instruments) and gold coated using a Q150RS sputter coater (Quorum). Purified SecYEG in 0.02% DDM was exchanged into MS suitable buffer (250mM ethylenediamine diacetate (EDDA), 0.014% DDM (v/v)) to a final concentration of 9 µM (SecYEG WT) or 12µM (SecYEG variants) using Micro Bio-Spin 6 columns (Bio-Rad). 1.5 µL protein sample was then loaded into the gold-

coated capillaries and introduced into a Synapt G2-Si (Waters) mass spectrometer by nano-electrospray ionisation. The following conditions were used in the mass spectrometer: capillary voltage +1.2 kV, sampling cone voltage 20-50V, backing pressure 3.88 mbar, trap and transfer pressure (argon) $1.72e^{-2}$ mbar, ion mobility cell pressure (nitrogen) 2.58 mbar. Trap and transfer cell collision energies (CE) were adjusted to remove the DDM detergent micelle, allowing the spectra to be resolved. Spectra were recorded at trap CE 200V and transfer CE 200V. Mass measurements were calibrated using caesium iodide (100mg/ml). Spectra were recorded and smoothed using Masslynx 4.1 (Waters) software.

The relative abundances of each oligomeric/lipid bound state of SecYEG was calculated with a Bayesian deconvolution software, UniDec (56). Detector efficiency was accounted for before quantification (57). The spectra were analysed between 3,000-8,000 m/z to allow the quantification of the SecY based complexes. SecE and SecG monomers were found in the 1,000-3,000 m/z range and were discounted from our analysis as they were detected in vast excess to the other oligomeric species. The mass range for peak detection was 45,000-77,000 Da and species were identified under the following masses: SecY 48,325 ±50 Da; SecY + CL 49,700±50 Da; SecYG 59,725 ± 50 Da; SecYG + CL 61,125 ±50 Da, SecYE 63,050 ±50 Da; SecYE + CL 64,425 ±50 Da; SecYEG 74,550 ±50Da. The comparison of the apo:CL bound relative abundances assumes lipid bound species have similar ionisation efficiencies.

*Preparation of inverted membrane vesicles*

Inverted membrane vesicles (IMVs) were made from C43 *E. coli* cells, as well as a C43 strain in which the three CL synthase genes (*clsA*, *clsB* and *clsC*) have been knocked out (29). The cells were lysed, and centrifuged at 13,000 rpm for 20 mins to clear the lysate, followed by centrifugation on a bed of 20% sucrose at 35,000 rpm for 2 hours. The pellets were resuspended in $TK_{150}M$, homogenised and run on a stepped sucrose gradient of 0-1.6 M sucrose for ~16 hours at 32,000 rpm. The IMV band was detected visually, extracted and diluted in $TK_{150}M$, and centrifuged a 60,000 rpm for 90 minutes. The pellets were resuspended in $TK_{150}M$ and aliquoted for storage at -80$^{o}$C.

*Translocation assays*

Transport assays were performed using a modified version of the standard protocol (58). IMVs at a concentration of 0.4 µM SecY were mixed with 1 µM SecA and an ATP regenerating system (0.1 mg $ml^{-1}$ creatine kinase and 5 mM creatine phosphate) in $TK_{150}M$. Under these conditions, ATP both powers translocation, and generates a PMF *via* F-ATPase.

Reactions were prewarmed to 25 ˚C then started by the addition of 1 mM ATP. After 3 minutes, reactions were quenched by the addition of the same volume of proteinase K at 0.6 mg/ml in 50 mM HEPES pH 8, incubated on ice for 30 mins. The proteinase step ensures that any untranslocated material will be digested. The samples were then precipitated with TCA added to 20%, before centrifugation for 10 minutes at 13,000 rpm. Pellets were dried and resuspended in 4X LDS gel loading buffer, before analysis with SDS-PAGE and western blotting against the V5 tag using a V5 tag antibody (SV5-Pk1, GeneTex) followed by a DyLight 800 anti-mouse secondary (SA5-10172, Thermo). Gels were imaged in a LI-COR Odyssey, and bands were quantified in Image Studio.

## Acknowledgements

This work was funded by the BBSRC: BB/M003604/1, BB/I008675/1 and BB/N015126/1 and Wellcome Trust: 104632 and 109854/Z/15/Z. EP is the recipient of an Imperial College

London Institute of Chemical Biology EPSRC CDT studentship. This work was carried out using the computational facilities of the Advanced Computing Research Centre, University of Bristol (http://www.bris.ac.uk/acrc/). Additional simulations were carried out using computer time on EPCC's Cirrus HPC Service (https://www.epcc.ed.ac.uk/cirrus) and on the ARCHER UK National Supercomputing Service (http://www.archer.ac.uk), provided by HECBioSim, the UK High End Computing Consortium for Biomolecular Simulation (hecbiosim.ac.uk), supported by the EPSRC.

## References

1. Cannon KS, Or E, Clemons WM, Shibata Y, & Rapoport TA (2005) Disulfide bridge formation between SecY and a translocating polypeptide localizes the translocation pore to the center of SecY. *The Journal of Cell Biology* 169(2):219-225.
2. Van den Berg B*, et al.* (2004) X-ray structure of a protein-conducting channel. *Nature* 427(6969):36-44.
3. Collinson I, Corey RA, & Allen WJ (2015) Channel crossing: how are proteins shipped across the bacterial plasma membrane? *Philos Trans R Soc Lond B Biol Sci* 370(1679).
4. Bauer BW, Shemesh T, Chen Y, & Rapoport TA (2014) A "Push and Slide" Mechanism Allows Sequence-Insensitive Translocation of Secretory Proteins by the SecA ATPase. *Cell* 157(6):1416-1429.
5. Allen WJ*, et al.* (2016) Two-way communication between SecY and SecA suggests a Brownian ratchet mechanism for protein translocation. *eLife* 5:e15598.
6. Muller M, Koch HG, Beck K, & Schaefer U (2001) Protein traffic in bacteria: Multiple routes from the ribosome to and across the membrane. *Progress in Nucleic Acid Research and Molecular Biology, Vol 66* 66:107-157.
7. Ulbrandt ND, Newitt JA, & Bernstein HD (1997) The E. coli signal recognition particle is required for the insertion of a subset of inner membrane proteins. *Cell* 88(2):187-196.
8. Samuelson JC*, et al.* (2000) YidC mediates membrane protein insertion in bacteria. *Nature* 406(6796):637-641.
9. Schulze RJ*, et al.* (2014) Membrane protein insertion and proton-motive-force-dependent secretion through the bacterial holo-translocon SecYEG-SecDF-YajC-YidC. *Proceedings of the National Academy of Sciences* 111(13):4844-4849.
10. Komar J*, et al.* (2016) Membrane protein insertion and assembly by the bacterial holo-translocon SecYEG–SecDF–YajC–YidC. *Biochemical Journal* 473(19):3341-3354.
11. Lee A (2003) Lipid–protein interactions in biological membranes: a structural perspective. *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1612(1):1-40.
12. Palsdottir H & Hunte C (2004) Lipids in membrane protein structures. *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1666(1):2-18.
13. Lill R, Dowhan W, & Wickner W (1990) The Atpase Activity of Seca Is Regulated by Acidic Phospholipids, Secy, and the Leader and Mature Domains of Precursor Proteins. *Cell* 60(2):271-280.
14. Gold VAM*, et al.* (2010) The action of cardiolipin on the bacterial translocon. *Proceedings of the National Academy of Sciences*.
15. de Vrije T, de Swart RL, Dowhan W, Tommassen J, & de Kruijff B (1988) Phosphatidylglycerol is involved in protein translocation across Escherichia coli inner membranes. *Nature* 334(6178):173-175.
16. Kusters R, Dowhan W, & de Kruijff B (1991) Negatively charged phospholipids restore prePhoE translocation across phosphatidylglycerol-depleted Escherichia coli inner membranes. *J Biol Chem* 266(14):8659-8662.
17. Planas-Iglesias J*, et al.* (2015) Cardiolipin interactions with proteins. *Biophysical journal* 109(6):1282-1294.

18. Kates M, Syz J-Y, Gosser D, & Haines TH (1993) pH-dissociation characteristics of cardiolipin and its 2′-deoxy analogue. *Lipids* 28(10):877-882.
19. Haines TH & Dencher NA (2002) Cardiolipin: a proton trap for oxidative phosphorylation. *FEBS Lett* 528(1-3):35-39.
20. Olofsson G & Sparr E (2013) Ionization Constants pK a of Cardiolipin. *PloS one* 8(9):e73040.
21. Sathappa M & Alder NN (2016) The ionization properties of cardiolipin and its variants in model bilayers. *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1858(6):1362-1372.
22. Ruprecht JJ*, et al.* (2014) Structures of yeast mitochondrial ADP/ATP carriers support a domain-based alternating-access transport mechanism. *Proc Natl Acad Sci U S A* 111(4):E426-434.
23. McAuley KE*, et al.* (1999) Structural details of an interaction between cardiolipin and an integral membrane protein. *Proceedings of the National Academy of Sciences* 96(26):14706-14711.
24. Arnarez C, Marrink SJ, & Periole X (2013) Identification of cardiolipin binding sites on cytochrome c oxidase at the entrance of proton channels. *Sci Rep* 3:1263.
25. Duncan AL, Robinson AJ, & Walker JE (2016) Cardiolipin binds selectively but transiently to conserved lysine residues in the rotor of metazoan ATP synthases. *Proceedings of the National Academy of Sciences*:201608396.
26. de Jong DH*, et al.* (2013) Improved Parameters for the Martini Coarse-Grained Protein Force Field. *J Chem Theory Comput* 9(1):687-697.
27. Marrink SJ, Risselada HJ, Yefimov S, Tieleman DP, & de Vries AH (2007) The MARTINI force field: coarse grained model for biomolecular simulations. *J Phys Chem B* 111(27):7812-7824.
28. Zimmer J, Nam Y, & Rapoport TA (2008) Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature* 455(7215):936-943.
29. Carranza G*, et al.* (2017) Cardiolipin plays an essential role in the formation of intracellular membranes in Escherichia coli. *Biochimica et Biophysica Acta (BBA)-Biomembranes* 1859(6):1124-1132.
30. Tan BK*, et al.* (2012) Discovery of a cardiolipin synthase utilizing phosphatidylethanolamine and phosphatidylglycerol as substrates. *Proceedings of the National Academy of Sciences* 109(41):16504-16509.
31. Humphrey W, Dalke A, & Schulten K (1996) VMD: visual molecular dynamics. *Journal of molecular graphics* 14(1):33-38.
32. Berendsen HJC, van der Spoel D, & van Drunen R (1995) GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications* 91(1–3):43-56.
33. Wassenaar TA, Ingólfsson HI, Böckmann RA, Tieleman DP, & Marrink SJ (2015) Computational Lipidomics with insane: A Versatile Tool for Generating Custom Membranes for Molecular Simulations. *J Chem Theory Comput* 11(5):2144-2155.
34. Brundage L, Hendrick JP, Schiebel E, Driessen AJM, & Wickner W (1990) The purified E. coli integral membrane protein SecYE is sufficient for reconstitution of SecA-dependent precursor protein translocation. *Cell* 62(4):649-657.
35. van der Does C, Swaving J, van Klompenburg W, & Driessen AJ (2000) Non-bilayer lipids stimulate the activity of the reconstituted bacterial protein translocase. *Journal of Biological Chemistry* 275(4):2472-2478.
36. Floyd JH*, et al.* (2014) The dispensability and requirement of SecA N-terminal aminoacyl residues for complementation, membrane binding, lipid-specific domains and channel activities. *Biochemical and biophysical research communications* 453(1):138-142.
37. Zhou M*, et al.* (2011) Mass spectrometry of intact V-type ATPases reveals bound lipids and the effects of nucleotide binding. *Science* 334(6054):380-385.
38. Hernández H & Robinson CV (2007) Determining the stoichiometry and interactions of macromolecular assemblies from mass spectrometry. *Nature protocols* 2(3):715.

39. Heck AJ (2008) Native mass spectrometry: a bridge between interactomics and structural biology. *Nature methods* 5(11):927-933.
40. Laganowsky A, Reading E, Hopper JT, & Robinson CV (2013) Mass spectrometry of intact membrane protein complexes. *Nature protocols* 8(4):639.
41. Robson A, Gold VAM, Hodson S, Clarke AR, & Collinson I (2009) Energy transduction in protein transport and the ATP hydrolytic cycle of SecA. *Proceedings of the National Academy of Sciences* 106(13):5111-5116.
42. Tanaka Y*, et al.* (2015) Crystal Structures of SecYEG in Lipidic Cubic Phase Elucidate a Precise Resting and a Peptide-Bound State. *Cell Rep* 13(8):1561-1568.
43. Cooper DB*, et al.* (2008) SecA, the motor of the secretion machine, binds diverse partners on one interactive surface. *Journal of molecular biology* 382(1):74-87.
44. Botte M & Zaccai NR (2016) A central cavity within the holo-translocon suggests a mechanism for membrane protein insertion. *Scientific Reports* 6.
45. Kumazaki K*, et al.* (2014) Structural basis of Sec-independent membrane protein insertion by YidC.
46. Fiser A, Do RK, & Sali A (2000) Modeling of loops in protein structures. *Protein Sci* 9(9):1753-1773.
47. Monticelli L*, et al.* (2008) The MARTINI Coarse-Grained Force Field: Extension to Proteins. *Journal of Chemical Theory and Computation* 4(5):819-834.
48. Dahlberg M & Maliniak A (2010) Mechanical Properties of Coarse-Grained Bilayers Formed by Cardiolipin and Zwitterionic Lipids. *J Chem Theory Comput* 6(5):1638-1649.
49. Dahlberg M (2007) Polymorphic phase behavior of cardiolipin derivatives studied by coarse-grained molecular dynamics. *The Journal of Physical Chemistry B* 111(25):7194-7200.
50. Bussi G, Donadio D, & Parrinello M (2007) Canonical sampling through velocity rescaling. *J Chem Phys* 126(1):014101.
51. Parrinello M & Rahman A (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics* 52(12):7182-7190.
52. Nosé S & Klein ML (1983) Constant pressure molecular dynamics for molecular systems. *Molecular Physics* 50(5):1055-1076.
53. Delano WL (2002) The PyMOL Molecular Graphics System.
54. Collinson I*, et al.* (2001) Projection structure and oligomeric properties of a bacterial core protein translocase. *EMBO J* 20(10):2462-2471.
55. Gold VAM, Robson A, Clarke AR, & Collinson I (2007) Allosteric Regulation of SecA. *Journal of Biological Chemistry* 282(24):17424-17432.
56. Marty MT*, et al.* (2015) Bayesian deconvolution of mass and ion mobility spectra: from binary interactions to polydisperse ensembles. *Analytical chemistry* 87(8):4370-4376.
57. Fraser G (2002) The ion detection efficiency of microchannel plates (MCPs). *International Journal of Mass Spectrometry* 215(1):13-30.
58. Deville K*, et al.* (2011) The oligomeric state and arrangement of the active bacterial translocon. *J Biol Chem* 286(6):4659-4669.
59. Fahy E, Sud M, Cotter D, & Subramaniam S (2007) LIPID MAPS online tools for lipid research. *Nucleic acids research* 35(suppl_2):W606-W612.
60. Wassenaar TA, Pluhackova K, Bo□ckmann RA, Marrink SJ, & Tieleman DP (2014) Going backward: a flexible geometric approach to reverse transformation from coarse grained to atomistic models. *Journal of chemical theory and computation* 10(2):676-690.

## Figures

Figure 1

A) Cartoon of SecYEG bound to SecA, from PDB 3DIN (28). The Sec subunits are shown as cartoon with mesh overlay, with SecY in light pink, SecE orange, SecG green and SecA light-blue. The ATP analogue (ADP-BeFx) is coloured as orange, blue and red spheres. A small region of lipid bilayer is represented in grey.

B) Chemical structure of a common cardiolipin (CL) molecule, in this case di-1-palmitoyl-2-oleoyl-phosphatidylglycerol. CL are defined by having four acyl tails and two phosphate head groups. Structure made using LipidMAPS online structure drawing tool (59).

C) MARTINI representation of the CL molecule in panel B, with phosphate beads in orange, glycerol beads in red, and acyl tail beads in yellow and white (for beads containing a double bond). Below, a lysine molecule with side chain beads in light blue and the backbone bead (BB) in dark blue. In MARTINI, each bead represents four heavy atoms plus associated hydrogens. Bonds connecting the beads are shown as black lines.

Figure 2

A) Density plot of CL head groups (‘PO1’, ‘PO2’ and ‘GL0’ beads; see Figure 1C) in the *xy* plane at the cytoplasmic face of the membrane. Data computed for the 10 simulations using the self-assembly method. The protein is shown in the centre of the plot in black. Distinct CL hotspots can be observed.

B) Computed occupancies for CL binding to basic residues in each subunit of the Sec complex, with membranes formed using a self-assembly approach (red) and insane.py (blue). Data here have been normalised from 0-1, and error bars represent the s.e.m of 10 repeats (red) or five repeats (blue). All residues with less than 0.2 occupancy and the 25 N-terminal residues of SecA, previously implicated in lipid binding (36), have been omitted for clarity. The full data set can be seen in Figure S1.

C) The highest binding basic Sec residues are shown on SecA-SecYEG (PDB 3DIN (28)), with the CB atom of each residue coloured according to CL occupancy from the datasets in panel B. For clarity, only the residues with occupancy above 0.5 are labelled. The Sec subunits are approximately demarcated with black dotted lines. The position of the membrane is marked with black lines.

Figure 3

A) Views of SecA-SecYEG in MARTINI representation, shown as surfaces and coloured as per Figure 1A. CL-binding hotspots are shown as spheres and coloured according to normalised CL occupancy, with yellow beads representing occupancies of 0.2-0.5, orange 0.5-0.8 and red 0.8-1.0. Visual analysis reveals two distinct CL binding sites on the cytoplasmic face of SecY, labelled ‘Site 1’ and ‘Site 2’.

B) Two frames taken from one of the CG simulation trajectories, with the protein atoms mapped back from CG to atomistic using backward.py (60). SecA-SecYEG is coloured as panel A. The CL molecule is shown in MARTINI representation, and coloured as per Figure 1C. Site 1 and 2 residues are shown as sticks and red mesh. Important SecY TMH are marked with numbers, with the SecY lateral gate (LG) and the SecA helical scaffold domain (HSD), N-terminus (N) and nucleotide binding site (ATP) labelled.

Figure 4

A) Ratios of CL-bound to non-bound (apo) SecY, as measured using nMS, where a value of 1.0 would be equal abundancies of both states. WT SecY has a significantly higher ratio of CL-bound SecY than either the Site 1 or Site 2 variants ($p$ = 0.0194 and $p$ = 0.0048 respectively, using a one-tailed *t*-test). Note that the presence of SecE or SecG can ameliorate this effect (see Figure S8B). For raw data see Table 3 and Figure S8.

B) ATPase analyses of SecA-SecYEG with increasing concentrations of CL. Shown are the rates of ATPase increase upon CL addition, as determined from fitting the titration data to linear slopes. Error bars are reported errors for the linear regression. See Figure S9A for titration data.

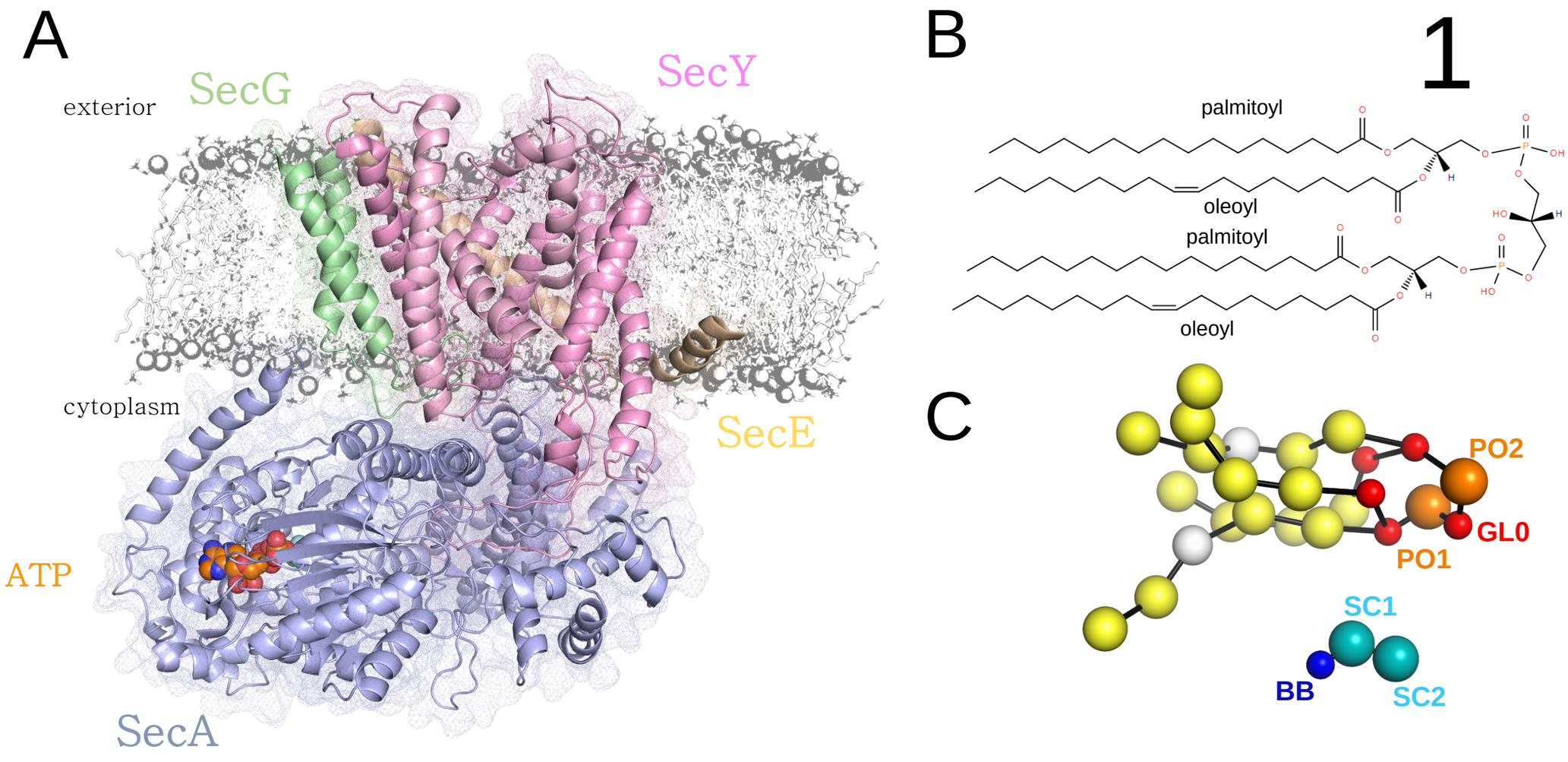
C) *In vitro* translocation data for IMVs containing overexpressed WT or variant SecYEG. On the left are data for IMVs containing normal levels of CL (+CL), which demonstrate a lower translocation efficiency in the Site 1 and Site 2 variants. On the right are data from IMVs with CL biosynthesis knocked out (-Cl), which reveal all of the samples to be the same. This validates the identification of the CL binding sites, and demonstrates that CL binding to both Site 1 and Site 2 is needed for WT levels of translocation efficiency. Error bars are s.e.m, and reported statistical analyses are from one-tailed *t*-tests, where $p$ = 0.0006, 0.0045, 0.384 and 0.229.

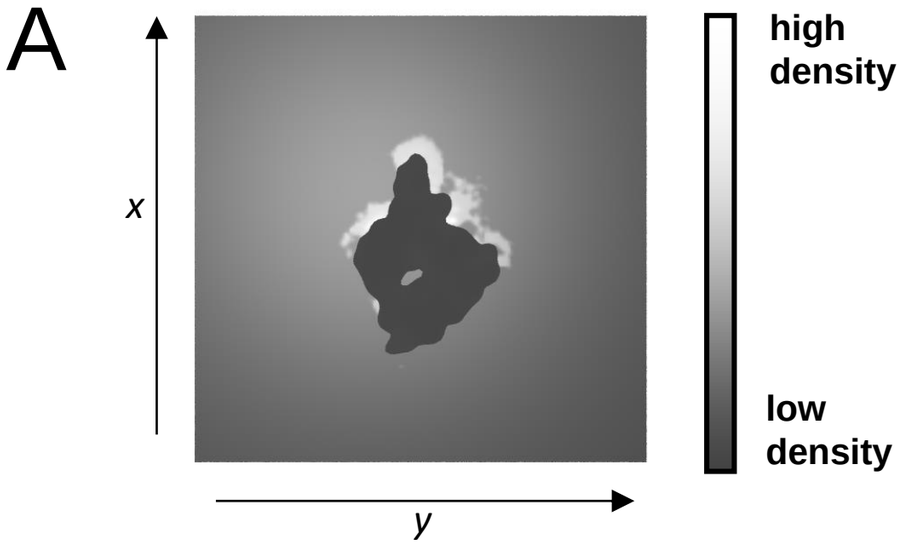
Figure 5

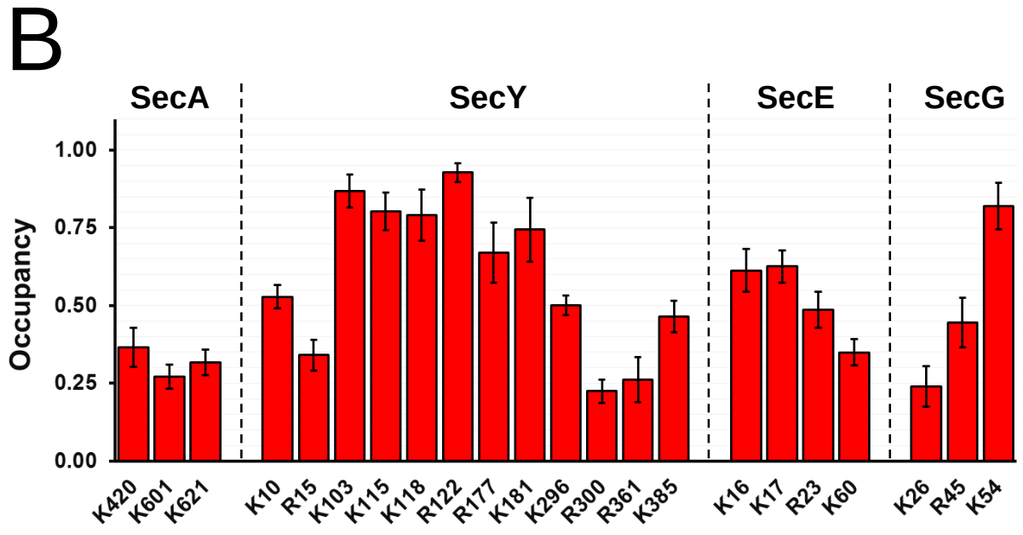
A) Snapshots taken at 1.1 µs from simulations of 16 *T. thermophilus* SecYEG molecules (PDB 5AWW (42)) in a DPPC membrane with either 10% or 0% CL. DPPC molecules are shown in grey, CL molecules in grey with the head group phosphates as orange spheres, SecYEG beads are shown in green, except for the amphipathic helix of SecE in blue (as panel B). The periodic box is shown in blue, such that the SecYEG molecules marked with a black star are same molecule. Approximate dimers are shown with a red connecting line. Visual analysis suggests that the presence of CL promotes dimerization, with 5 dimers present as oppose to 2.

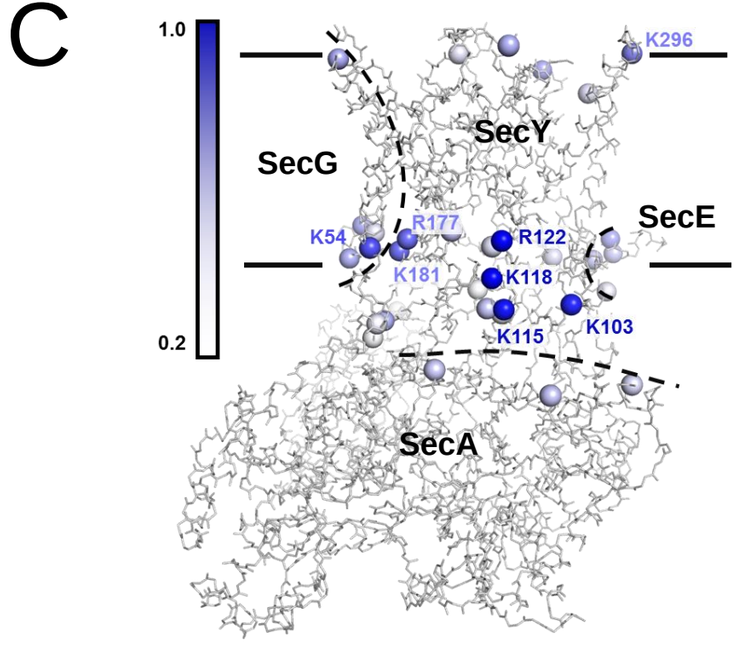
B) Molecule of SecYEG used in panel A showing the amphipathic helix of SecE. The most commonly accepted model of SecYEG dimerization involves the amphipathic helix of each protomer interacting. Within the sampling carried out here, none of the SecYEG dimers adopt this dimer state.

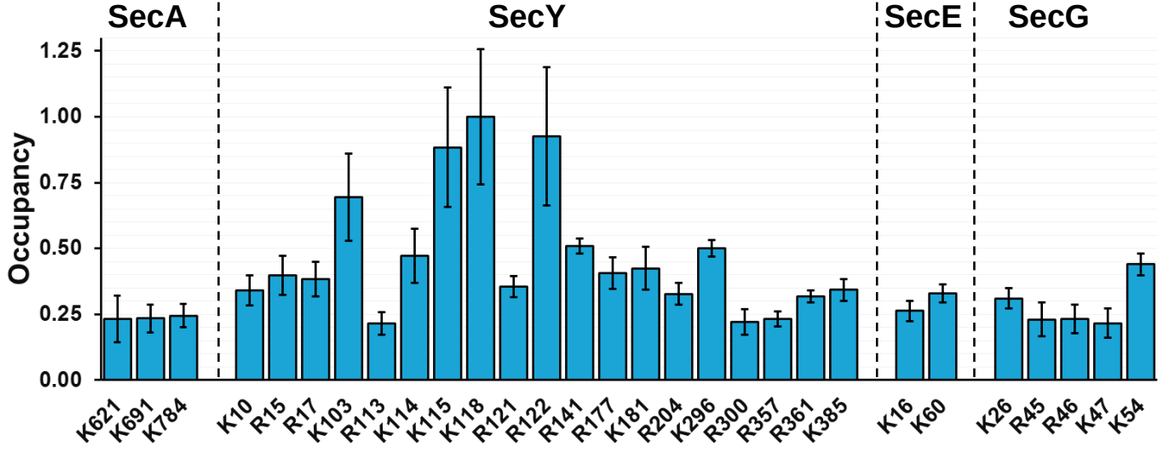
C) Distance analyses were carried out between each SecYEG protomer and the nearest neighbouring SecYEG. The molecules were considered as dimers if the minimum distance to the next protomer was less than 0.9 nm. The percentage of time each protomer spent in a dimer was computed for the last 250 ns of simulation, and the data plotted. For the CL membrane, 4 dimers (*i.e.* 8 protomers) were present more than 75% of the time, whereas only 2 dimers are seen in the membrane with CL (NoCL). This suggests that CL has a role in promoting SecYEG dimer formation, although not one mediated by CL binding to Site 1 or Site 2.


A
exterior
SecG
SecY
cytoplasm
SecE
ATP
SecA
B
1
palmitoyl
oleoyl
palmitoyl
oleoyl
C
PO2
GL0
PO1
SC1
BB
SC2

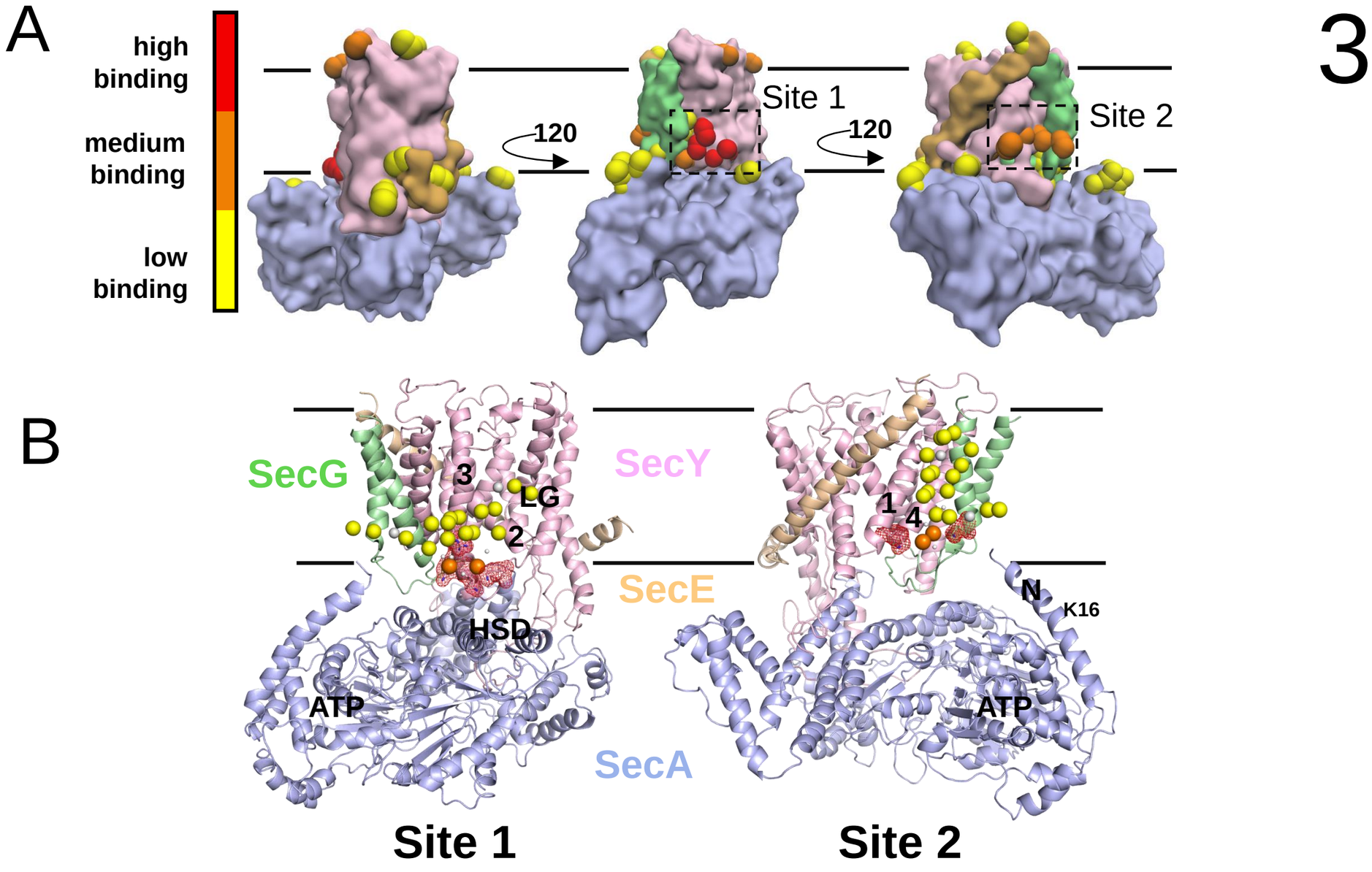

A
x
y
high density
low density

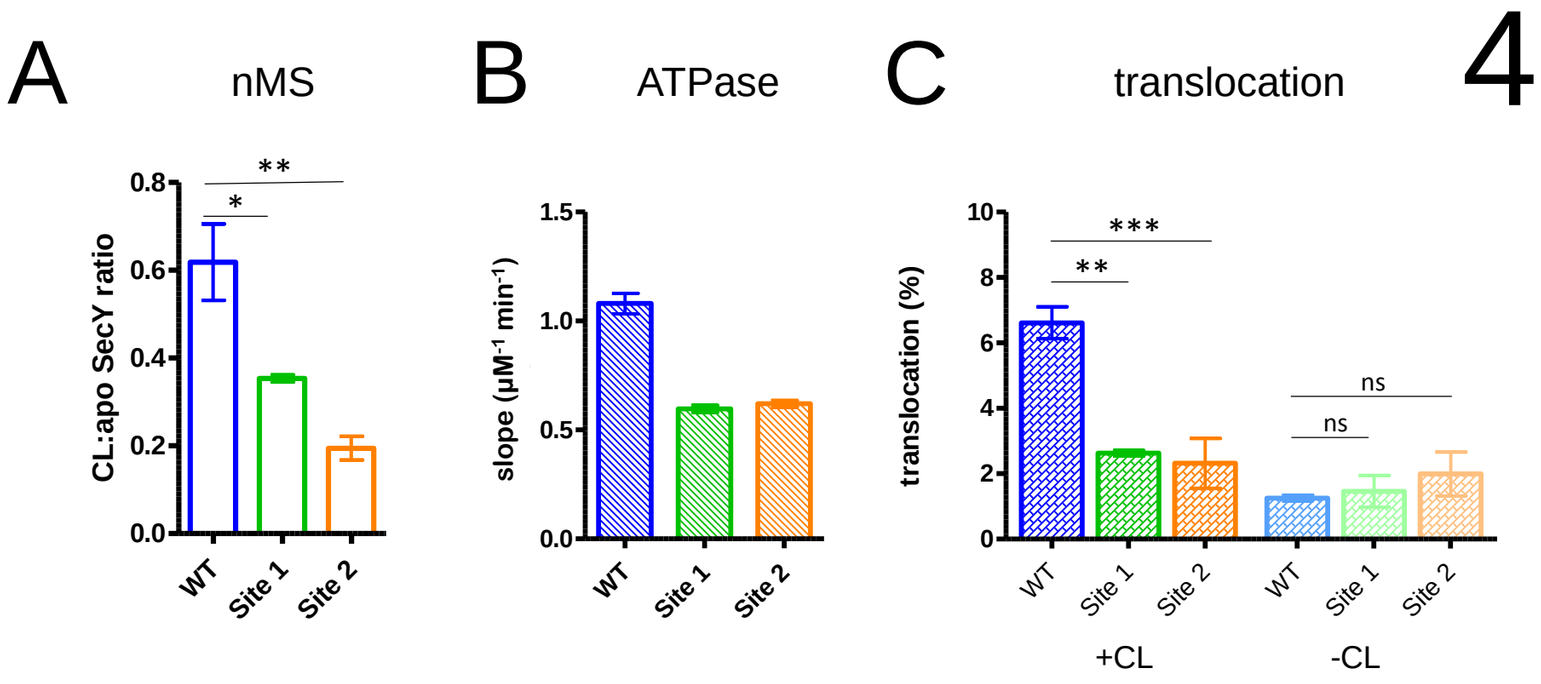

B
SecA
SecY
SecE
SecG
Occupancy
1.00
0.75
0.50
0.25
0.00
K420 K601 K621
K10 R15 K103 K115 K118 R122 R177 K181 K296 R300 R361 K385
K16 K17 R23 K60
K26 R45 K54


C
1.0
0.2
K296
SecG
SecY
SecE
K54
R177
R122
K181
K118
K115
K103
SecA

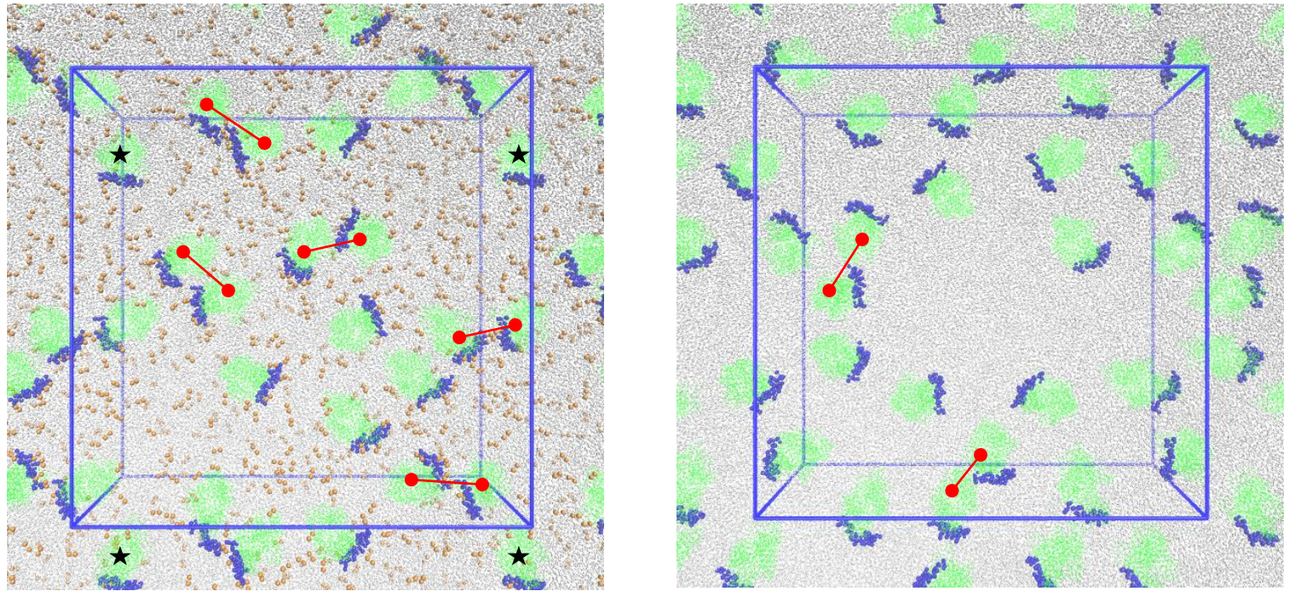

SecA
SecY
SecE
SecG
Occupancy
1.25
1.00
0.75
0.50
0.25
0.00
K621 K691 K784
K10 R15 R17 K103 R113 K114 K115 K118 R121 R122 R141 R177 K181 R204 K296 R300 R357 R361 K385
K16 K60
K26 R45 R46 K47 K54

2

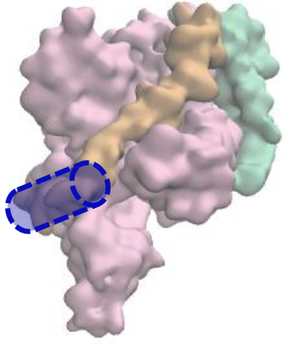

A
high binding
medium binding
low binding
120
Site 1
120
Site 2
B
SecG
3
LG
2
SecY
1
4
SecE
N
K16
HSD
ATP
ATP
SecA
Site 1
Site 2


A
nMS
B
ATPase
C
translocation
4
CL:apo SecY ratio
0.8
0.6
0.4
0.2
0.0
**
*
WT
Site 1
Site 2
slope (μM-1 min-1)
1.5
1.0
0.5
0.0
translocation (%)
10
8
6
4
2
0
***
**
ns
ns
+CL
-CL

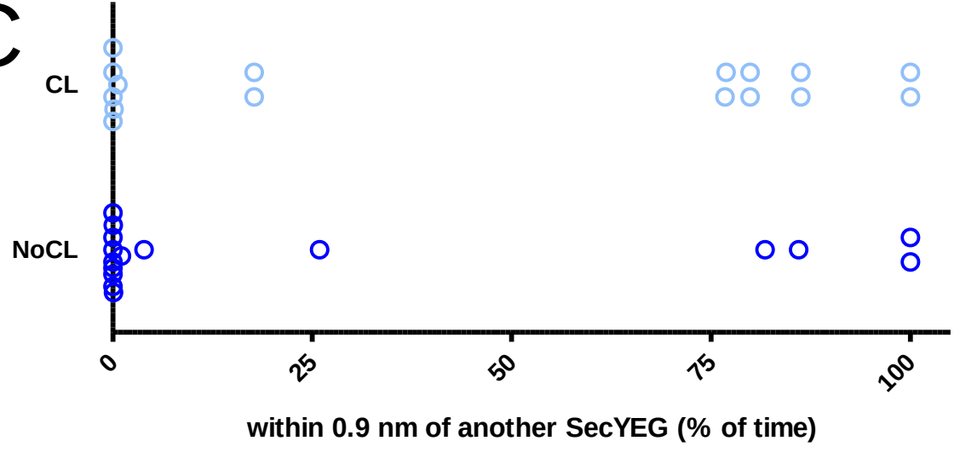
A



B



C




5