A 'proton ratchet' for coupling the membrane potential to protein transport

William J. Allen[1,*], Robin A. Corey[1,$,*], Daniel W. Watkins[1,*], Gonçalo C. Pereira[1], A. Sofia F. Oliveira[1,2] & Ian Collinson[1,†]

Affiliations:

[1]School of Biochemistry and [2]School of Chemistry, University of Bristol, University Walk, Bristol BS8 1TD, UK.

*Contributed equally to this work

†Correspondence to: ian.collinson@bristol.ac.uk

$Present address: Department of Biochemistry, University of Oxford, Oxford OX1 3QU, UK

Abstract

The proton-motive force (PMF) – the electrochemical gradient of protons across energy-conserving membranes – powers protein transport in bacteria, mitochondria and chloroplasts. Here, we propose a 'proton ratchet' mechanism for this process. In the Sec system of bacteria, protons are stripped from lysine side chains of the pre-protein at the cytosolic face of the plasma membrane, then replaced on the exterior, aided by the pH component of the PMF (ΔpH; acidic outside). This gives the translocating region of pre-protein a net negative charge, promoting electrophoretic diffusion across the membrane driven by membrane-potential (Δψ; positive outside). For mitochondrial import (through the TIM23 complex) the proton ratchet acts in the opposite direction, with negatively charged residues protonated for passage across the inner membrane into the negative matrix. The proton ratchet is an elegant solution for coupling the PMF to transport, likely to be used by a range of other transporters of charged molecules.

Main text

Introduction

The transport of proteins across membranes is a ubiquitous feature of cells. In energy-conserving membranes – i.e. those where an electrochemical gradient is present – the PMF is generally harnessed to drive this process: this is known to be the case for the Sec and Tat complexes of bacteria[1-4] and chloroplasts[5,6], and the TIM complexes of mitochondria[7,8]. Studies using specific ionophores suggest that both the electrical charge difference (Δψ; knocked out by valinomycin) and the chemical gradient (ΔpH; knocked out by nigericin) contribute to protein transport through Sec in bacteria[9]. In mitochondria, import of model substrates seems mostly to depend on Δψ[7], while ΔpH predominates in chloroplasts[6]. Yet, despite these observations having been made decades ago, precisely how the PMF is coupled to protein transport remains a mystery.

In bacteria, most protein transport is achieved by the hetero-trimeric membrane protein complex SecYEG, and the cytosolic motor ATPase SecA. Pre-proteins with a cleavable N-terminal signal sequence (SS) are targeted to the Sec machinery[2,10], whereupon transport is initiated by intercalation of the secretory precursor into SecY (Fig. 1a). The mature region of the pre-protein is then transported in an unfolded state through a channel in SecY[11], followed by SS removal by proteolysis[12]. ATP hydrolysis by SecA and the PMF are both required for efficient secretion to occur[2]: transport can be achieved driven only by ATP[1], but it is far too slow to account for all the protein that must be secreted by a dividing cell[13,14]. Part of this PMF effect is thought to be mediated by the ancillary complex SecDF[3]; however, PMF stimulation also occurs in the absence of SecDF[2,15].

Previously, we and others have proposed models for the ATP-driven component of secretion[16-18] but the mechanism of PMF-stimulated transport has remained elusive. One attractive hypothesis is that negatively charged residues (aspartate and glutamate) experience Δψ and are pulled across the membrane electrophoretically[19,20] (Fig. 1b). This is supported by

the strong inhibitory effect of valinomycin, which selectively knocks out Δψ[21]. However, the same effect would equally inhibit transport of positively charged residues (lysines and arginines). And since even transport of positively charged proteins is promoted by the PMF (e.g. Spy, see below), electrophoresis is obviously not the sole factor involved.

Therefore, we set out to test whether basic residues are deprotonated at the entrance of the channel – eliminating their positive charge – and re-protonated as they exit (Fig. 1c), assisted by the ΔpH. This would serve both *(i)* to imbue the translocating pre-protein with a net negative charge, aiding the electrophoretic effect attracting it to the positive outside, and *(ii)* to prevent 'backsliding' of already transported regions, as the channel is known to provide a strong barrier to the movement of positive charges[22,23]. We then consider this mechanism in the context of other transporters by extending the model to the mitochondrial TIM23 importer. The results suggest a unified model for PMF-driven protein transport across a range of different membranes.

Results

PMF stimulation is inhibited by pre-protein arginines and low pH

Direct observation of individual, highly transient protonation/deprotonation events in a large, dynamic system is extremely challenging with current technology. However, the model in Fig. 1b-c does make some testable predictions, which we set out to investigate. One of these is that both transport and PMF stimulation should be less effective at lower pH, where deprotonation is disfavoured (Fig. 2a). A second prediction is that lysine will be much easier to export than arginine: the amine group of lysine has a solution pK_a of ~10, and examples abound of deprotonated lysines (i.e. with low pK_as) in protein structures[24]. By contrast, the delocalised charge of the guanidinium group of arginine (pK_a ~13.8[25]) ensures it is likely never deprotonated under physiological conditions[26]. Therefore, we designed a series of highly conservative variants of a standard translocation substrate (proOmpA), differing only in

numbers of arginine and lysine, and measured their transport with and without a PMF, at a range of pHs.

From the results, it is immediately apparent that substrates with more lysine export better than those with more arginine (Fig. 2b) – consistent with the proposed model. This is true even in the absence of a PMF: because the channel provides a strong barrier to positive charges (see below); neutralising a positive charge should aid transport irrespective of the presence of Δψ or ΔpH. The effect of the pH also supports the model: at pH 6, where all lysine and arginine residues will be fully protonated, the PMF has little or no effect. However, as pH increases above plausible pK_a values for lysine[24] – potentially allowing their deprotonation – both PMF stimulation and the Lys/Arg effects are amplified (Fig. 2b).

We performed the same experiments on a proOmpA variant with the normal number of lysine and arginine residues (10 and 8, respectively), but their positions swapped. The results confirm that it is indeed the number of lysines, not their position, that influences transport (Fig. 2b, open markers).

It is particularly worth noting that the substrate with the fewest lysines is barely affected by pH at all (Fig. 2b). This points to a specific effect on the protonation state of substrate lysines, rather than a general pH-dependent reduction in Sec activity. Indeed, while there are numerous possible reasons why lysine might transport more easily than arginine, or why a transport reaction might be pH-dependent, the correlation between the two is strong evidence specifically for substrate lysine deprotonation.

Δψ has a direct effect on import kinetics

While the classical import assay above is useful in that it is compatible with low pH, it does not provide much detail on the kinetics of protein import. For subsequent import assays we therefore switched to our newly developed NanoLuc import assay[27] (Fig. 2b). In this assay, the large fragment of a split luciferase (11S) is tethered inside an inverted membrane vesicle

(IMV), while a tight-binding variant of the small fragment (Pep86) is fused to the C-terminus of the pre-protein import substrate. Successful translocation produces a luminescence signal, allowing continuous monitoring of protein import (Fig. 2c).

Import of proOmpA under these conditions is inhibited by valinomycin (a K^+ ionophore, which selectively knocks out Δψ), but not by nigericin (a K^+/H^+ antiporter, which selectively knocks out ΔpH; Fig. 2d). Such results have previously been taken to mean that Δψ is driving transport[7]. However, this interpretation has two important caveats. Consider the inverted bacterial system, *in vitro*: firstly, the deprotonation event is dependent on external pH, not ΔpH. Indeed, if protein import causes the pH inside the vesicle to rise (as predicted by the proton ratchet), then both ΔpH and nigericin would alleviate this and facilitate transport. Secondly, the strength of the PMF in an *in vitro* system is likely to be much weaker than within a living cell; and while a reduced Δψ retains some effect, a diminishingly small ΔpH could be eliminated by the buffering capacity of the system. Therefore, while Δψ clearly does promote protein transport, the effects of nigericin on transport should be interpreted cautiously.

The PMF drives transport of lysines in a positively charged pre-protein

Because proOmpA has a net negative charge, its transport would be aided by Δψ even in the absence of a proton ratchet. To control for this, and to confirm that the effect is not substrate-specific we created a similar Lys/Arg series for *E. coli* spheroplast protein Y (Spy), a periplasmic stress response chaperone with a pI of 10.2 (i.e. positively charged at physiological pH). As with proOmpA, increasing the Lys:Arg ratio in proSpy at pH 8 promotes transport, in a strong and approximately linear relationship (Fig. 3a, Fig. S1a-c). Also similar to proOmpA, this effect is more pronounced in the presence of a PMF. Indeed, PMF stimulation of Spy transport is only observed when there are lysines present; this is consistent with an absolute requirement for lysine deprotonation in order to couple the PMF to the transport of positively charged proteins.

The PMF mainly affects transport, not the initiation step

While the PMF is generally thought to act at the later stages of import[1], it has also been proposed to aid SS insertion[28], i.e. initiation of transport (Fig. 3b). To distinguish these possibilities, we created variants of proSpy with only the N- and/or C-terminal half of the mature polypeptide enriched with lysine or arginine. Changes near the N-terminus of the protein (grey in Fig. 3b) will mainly influence the initiation step, while those closer to the C-terminus (magenta in Fig. 3b) should be specific to the subsequent transport stages.

Transport results support the notion that the PMF drives transport of the main polypeptide chain rather than simply promoting initiation (Fig. 3c, Fig. S1d): while increasing the number of arginine residues slows transport in all sequence contexts, its effect is much more pronounced for the C-terminal half. These data do not rule out the possibility that the PMF also contributes to the early stages of transport; but if it does, lysine deprotonation is unlikely to be an important factor.

The proton ratchet exerts a selection pressure on secreted proteins

If the proton ratchet is indeed important for the mechanism of secretion *in vivo*, one might expect that Sec substrates experience a selective pressure to replace arginine with lysine, where possible. To investigate this, we compared the Lys/Arg composition of the mature domain of all known *E. coli* Sec substrates with those that remain in the cytosol. As predicted, secreted proteins have a strong preference for lysine over arginine relative to those in the cytosol (Fig. 4a). We also looked at Tat substrates – exported independently of Sec – and found their Lys/Arg ratio appears to fall somewhere between Sec substrates and unsecreted proteins (Fig. 4a). This confirms that the Lys/Arg effect is at least in part a product of transport pathway, rather than the extra-cytosolic environment.

Because *E. coli* only uses Tat to export a very small number of proteins, we carried out the same analysis on a model organism with a large number of annotated Tat substrates (*Sinorhizobium meliloti*[29]; Fig. 4b). The results both confirm the *E. coli* observations, and show that they hold true across different classes of Gram-negative bacteria. We speculate that this may be relevant to the mechanism of protein export through Tat.

To ensure that this effect is connected to the proton ratcheting model, we carried out the same analysis on an organism with an inverted ΔpH, alkalophilic bacterium *Bacillus halodurans*[30]. Consistent with the model, lysine is no longer favoured over arginine for secreted proteins – indeed, the reverse appears to be true (Fig. 4c). The related neutrophilic bacterium *Bacillus subtilis*, meanwhile, shows the expected increase in lysine for secreted substrates (Fig. 4d).

Protons are removed from lysine in a pK_a 'shifter' zone at the SecY:SecA interface

In order to facilitate the removal of a proton, the local environment at the entrance to the channel must lower the pK_a of the lysine to near physiological pH. We investigated this computationally, using the pK_a prediction algorithm propKa31[31] to analyse conformational ensembles generated from molecular dynamics simulation. The data clearly show a strongly reduced pK_a (to below physiological pH) for lysine residues exactly as they are about to pass through the channel (Fig. 4e; Fig. S2a-e). The effect seems to be independent of the nucleotide occupancy (Fig. S2a-e), suggesting that ATP- and PMF-driven transport are independent of one another.

The SecY channel provides a barrier to positive charges

To explore the ease with which protonated (Lys^+) and deprotonated (Lys^0) lysine, and arginine pass through SecY, we performed atomistic steered molecular dynamics (MD) simulations, whereby regions of pre-protein containing different residues were pulled through

the channel (Fig. 4f). As expected[22,23], the data show that the force required to pull charged residues is higher than non-charged residues. To better quantify this, we used the Martini force field[32,33] to construct 1D free energy profiles for Lys^+ and Lys^0, using potential of mean force calculations (Fig. 4g, Fig. S2f) – these reveal a ~10 kJ mol^{-1} penalty for lysine crossing the channel in a protonated state. Intriguingly, the steered MD simulations suggest that arginine is easier to transport than protonated lysine (Fig. 5b), possibly due to the delocalisation of the positive charge of the guanidinium group, as well as the conformational flexibility of the arginine side chain. Given that, empirically, lysine is easier to transport than arginine (Fig. 2b, Fig. 3a,c), these results reinforce the notion that lysine does not normally traverse the channel with a positive charge.

The proton ratchet may be reversed for mitochondrial protein import through TIM23

In the case of Sec, ΔpH acts on positive charges and Δψ on negative: but this ratcheting mechanism could equally function in the opposite direction, with specific protonation of acidic residues prior to passage across the membrane. A prime candidate for this would be the TIM23 complex, which imports proteins into the acidic and negatively charged mitochondrial matrix. While a structure of the TIM23 complex remains to be determined, and many aspects of its function are unclear, transport is known to use Δψ[7], and the channel through the mitochondrial inner membrane has been shown to be selective against negative charges[34].

The pK_as of the naturally occurring negatively charged amino acids (aspartate and glutamate) are very similar (3.9 and 4.2 respectively), both well within the range where they could be protonated at near neutral pH by a favourable local environment. It is therefore not possible to produce a fixed negative charge by site-directed mutagenesis alone. Instead, we emulated the Lys/Arg effect by labelling a model TIM23 substrate with two artificial carboxylic acids. One of these has a pK_a slightly higher than aspartate or glutamate (Fig. 5a; iodoacetic

acid (IA)), while a single fluorine substitution in the other reduces its pK_a by 2 units (Fig. 5a; iodofluoroacetic acid (IFA)), giving it a highly stable negative charge that cannot easily be neutralised by protonation.

Once again, we exploited our high-resolution protein transport assay to measure mitochondrial import of these substrates. The results show that the lowered pK_a of IFA significantly inhibits transport (Fig. 5b, red vs blue data) – consistent with a proton ratchet mechanism. Interestingly, the chemical addition of a neutralisable negative charge stimulates transport relative to the native cysteine (Fig. 5b, blue vs black data). This might reflect a beneficial effect of having an easily ratcheted side chain present.

Discussion

It has been known for decades that the PMF is an important driver of protein transport across membranes, but mechanisms have so far proved elusive. Here, we provide evidence for a model in which the electrical and chemical components of the PMF cooperate. In the Sec system of bacteria (and presumably the thylakoid Sec machinery of chloroplasts), lysine residues are deprotonated just prior to passage through SecY, giving the region of pre-protein in the channel a net negative charge and allowing it to be pulled across by Δψ. The cytosolic acidification that this would create is countered by ΔpH. For mitochondrial protein import through TIM23, aspartate and glutamate residues are instead protonated in the intermembrane space, allowing Δψ to act on positive charges. In both cases, the equilibrium effect of ΔpH can also be performed by the K^+/H^+ antiporter nigericin – so in this context, the use of nigericin is not a faithful diagnostic of ΔpH-dependence.

Because the deprotonation effect is responsive to pH changes close to physiological pH (Fig. 2b, Fig. 4e), pre-proteins will have an asymmetric charge distribution when ΔpH is present: regions on the outside (*i.e.* those already transported) will be more positively charged than those on the inside. Coupled with the impermeability of the channel to positive

charges[22,23] (Fig. 4f-g), this will produce a proton ratchet (Fig. 1c) provided the pre-protein is allowed to diffuse in the channel, as we and others have previously proposed[16,17]. Therefore, we conclude that ΔpH contributes to PMF stimulation both kinetically (by ratcheting) and to the equilibrium, by providing a counterflow of protons.

When a proton is removed from the pre-protein at the cytosolic face of SecY it must be dispersed rapidly, or a sudden the drop in local pH would immediately reprotonate the lysine. We have recently identified two specific binding sites for cardiolipin (CL) near to the pK_a-perturbed area, which are required for PMF stimulation of transport[35] (Fig. S3). CL is thought to buffer pH locally[36], and simulations suggest that it is bound and released from SecYEG-SecA on the μs timescale[35]. Therefore, it seems plausible that protons extracted from lysine are transferred to CL for rapid dissipation (Fig. 6, 'Sec system'). Presumably, the transport pathway shields the pre-protein from direct contact with bulk solvent, ensuring that lysine residues are not reprotonated until they reach the periplasm.

Interestingly, CL has also been shown to bind to TIM44 – part of the TIM23 complex[37] – and indeed other transporters of charged molecules like the ammonia transporter AmtB[38] (see below). The ATP synthase also associates with CL, at the proton transporting F_O membrane sector[39]. Perhaps the proton shuttling abilities of CL are key to the energy coupling functions throughout biology.

The proton ratchet concept elegantly couples both complementary components of the PMF to transport (Δψ and ΔpH), so it seems reasonable that it might be more widely applicable for the transport of ionisable molecules across energised membranes. A search through the literature identifies several PMF-driven transport processes in which specific substrate protonation or deprotonation events are already implicated (Fig. 6). Both formate[40] (and other carboxylates) and ammonia[41] transporters have been shown to neutralise their substrates for passage across the membrane. The anthrax protective antigen pore, meanwhile, exploits the internal acidification of endosomes to protonate glutamate and aspartate residues in its

substrate[42]. However it seems likely that more examples remain to be discovered: the unique, potentially pK_a-modulating chemical environment is an often overlooked feature of ligand binding sites.

Methods

Materials

Transport reaction components were produced as described previously[27]. To make inverted membrane vesicles (IMVs) at different pHs, a standard protocol was followed[35], but substituting 5 mM MES-KOH pH 6.0, 5 mM HEPES-KOH pH 7.0 or 5 mM HEPES-KOH pH 8.0 for the normal buffer. IMVs overexpressing SecYEG and containing membrane-tethered 11S were prepared at pH 8 as described in ref 27, using membranes derived either from *E. coli* strain BL21(DE3), which produces a PMF in the presence of ATP, or HB1(DE3), which lack ATP synthase, and therefore does not[43].

Genes encoding translocation substrates were purchased from GeneArt (Thermo Fisher Scientific), cloned between the NcoI and HindIII sites of pBAD/*Myc*-HisC, then expressed and purified as described[35]. A complete list of protein sequences is shown in Table S1.

Transport assays

Transport assays of proOmpA with inhibitors (Fig. 2c) were performed in a 1 ml cuvette using a split-luciferase-based method, as described in ref 27 (see Fig. 2B for an overview). Import of full length proOmpA tagged with Pep86 into BL21-derived IMVs containing 11S was followed in real time in the presence of: no inhibitor (a PMF is produced from ATP by the F-ATPase); 2 μM nigericin (which selectively knocks out ΔpH); 1 μM valinomycin (which selectively knocks out Δψ); or both (no PMF). Inhibitors were added in DMSO to 1% final, which has no effect on transport kinetics (Fig. 2c).

Because NanoLuc is sensitive to pH, transport assays at different pHs (Fig. 2a) were performed using the classical western blotting method, as described previously[35], with reactions pre-treated with DMSO (+PMF) or 2 μM nigericin and 1 μM valinomycin in DMSO (-PMF). Note that we tried quenching transport with and without 500 mM urea, and observed that, while urea reduces the background signal, it has no effect on measured transport. The results were therefore merged to produce the final data set.

For the transport of the proSpy Lys/Arg series (Fig. 3a), assays were conducted at pH 8.0 in a 96-well plate with a plate reader (Synergy Neo2, Biotek), as described[27], using either BL21- or HB1-derived IMVs. For the control dataset in Fig. S1c, 50 pM 11S was mixed with 125 nM proSpy in the presence of furimazine (NanoGlo Luciferase Assay substrate, Promega) at the manufacturer's recommended concentration (1/100). From the raw traces, the time required for transport to reach 50% completion was determined ($t_{50\%}$), and its reciprocal used as a measure of transport rate (k_{app}; see Fig. 2d).

The proSpy variants in Fig. 3c were designed by dividing the mature region of proSpy into two halves, the first comprising residues 24-92 ('N-terminus', containing 7 lysines and 5 arginines) and the second residues 93-161 ('C-terminus', containing 10 lysines and 3 arginines). A series of mutants were then designed where each half was either kept with its native sequence (WT), higher Lys/Arg ratio, or lower Lys/Arg ratio. Tranport assays were conducted with BL21-derived IMVs exactly as for the proSpy Lys/Arg series.

Bioinformatics

The complete proteomes of *E. coli* (strain K12) and *S. meliloti* (strain 1021) were downloaded from UniProt[44] on 1st May 2018 and sorted according whether they are secreted by Sec or Tat system, or unsecreted. For *E. coli*, which is well annotated by UniProt, this was done using information in the 'Signal Peptide' column. For *S. meliloti*, secreted substrates were selected according to UniProt 'Signal Peptide', then classified as Tat if validated as such in ref

29, or Sec otherwise. *B. halodurans* and *B. subtilis* sequences were downloaded on 13th March 2019 and analysed as for *E. coli*, omitting signal peptides flagged as 'Tat-type'. For secretion substrates, the residues corresponding to the SS were removed prior to analysis. The proportion of positive residues that are lysine (Lys/(Lys+Arg)) was calculated for each protein, then plotted as a histogram for each data set (with sample size n): the number of bins was determined by Sturge's rule (number of bins = 1 + 3.322 * log(n)), and frequency calculated by dividing the population of that bin by n. Fits to single Gaussian curves were performed using Pro Fit (Quansoft)

Modelling of an engaged pre-protein in the SecA-SecYEG complex

Initial coordinates for SecA-SecYEG were taken from chains A, C, D and E of PDB 3DIN[45], with the ADP-BeF_x molecule replaced with ATP[46]. Simulations were run over 1 μs in a POPC membrane as in ref 16. Taking a 1 μs snapshot as a fully equilibrated starting model, a region of pre-protein 76 residues long was built in an extended configuration through the SecA-SecYEG complex, as described previously[47]. Briefly, the pre-protein was modelled in an extended configuration through a channel going through both SecYEG and SecA, contacting previously identified crosslinking sites in both[47,48]. The N-terminal SS was modelled as a helix and sited in the SecY lateral gate a per cryo-EM density[48].

The SecYEG-SecA-pOA models with either an ATP or ADP molecule bound (giving two different channel conformations[16]) were then embedded in a POPC membrane, solvated with explicit waters and ions to 0.15 M, and subjected to 1 μs MD simulation. Simulations were carried out as previously described[16], using the OPLS-AA force field[49], in Gromacs 5.0.4[50].

pK_a scanning pipeline

To ascertain the pK_a of charged residues as they traverse the SecA-SecYEG complex, we constructed a computational pipeline. For 20 different structural snapshots over the final

500 ns of the SecYEG-SecA-pOA simulations, each of the 76 residues in the pre-protein were substituted to a titratable residue in turn, using Scwrl4[51]; 3,000 individual structures per residue type), before pK_a analysis using the propKa31 program[31].

Linear response approximation (LRA) pK_a calculations

Using the speed of the above pathway to identify likely regions of pre-protein with perturbed pK_a values, we then selected 4 residues (postions 27, 45, 49, 51 along the pre-protein) sampling different environments within the channel for a more detailed analysis. Post 1 μs MD snapshots were taken of the SecA-SecYEG-pOA.ATP complex, and lysines were introduced in either a protonated or deprotonated state at each of the positions. MD simulations were run for ~100 ns using the charmm36 force field[52], and 50 structural snapshots were taken over the final 50 ns. These were analysed using propKa31, with a subset analysed using complementary continuum electrostatics calculations with the MEAD[53] and PETIT[54,55] softwares. Here, a dielectric constant of 80 was used for the solvent and 6 for the protein and membrane. A three-step focusing procedure was used for the protein/membrane system with the first grid formed by 251 points, each spaced 1.0 Å, a second grid composed of 81 points, each 0.5 Å apart and the third formed by 81 points, each spaced 0.25 Å. Proton isomerism was included as tautomers for the neutral forms of the titratable sites, as described elsewhere[54-56]. The calculations were done at 310 K with an ionic strength of 0.15 M, a molecular surface defined with a solvent probe radius of 1.4 Å, and a Stern (ion exclusion) layer of 2.0 Å. Each Monte Carlo run consisted in 10^3 equilibration steps followed by 10^5 production steps. In both cases, pK_a values were determined for each residue position using the linear response approximation (LRA[57-59]).

Steered MD

1 μs snapshots were taken from simulations of a pre-protein engaged SecA-SecYE complex, modelled based on PDB 5eul[60]; see ref 18 for details of the modelling. Two pre-protein residues near the cytoplasmic face of SecY were then substituted to either lysine, arginine, glutamate or tryptophan using Scwrl4[51]. Next, 9 independent steered MD simulations were run for each pre-protein, where the substrate was pulled from the substituted residues in a z-axis direction (up through the channel) at a rate of 0.5 $nm.ns^{-1}$, using a maximum force constant of 1000 $kJ.mol^{-1}.nm^{-1}$. Simulations were run in the NPT ensemble, as described previously[18].

For each repeat, the force was measured while the substituted residues were in contact with the central pore, with contact taken as 0.3 nm or less.

Construction of 1D free energy profiles

To provide a more detailed view of the energetic cost of transporting protonated lysine, we constructed free energy profiles of protonated and deprotonated lysine residues through the channel, using the Martini 2.2 force field[32,33]. Following conversion to Martini, elastic bonds of 1000 kJ mol^{-1} were applied between all backbone beads within 1 nm. Electrostatics were described using the reaction field method, with a cut-off of 1.1 nm using a potential shift modifier, and van der Waals interactions were shifted from 0.9-1.2 nm. Simulations were run in the NPT ensemble, with V-rescale temperature coupling at 323 K and semi-isotropic Parrinello-Rahman pressure coupling.

We used steered MD to construct a 1D reaction coordinate for an Ala-Lys-Ala peptide through the SecY channel, with the collective variable constructed between each of the backbone beads in the tripeptide and 5 backbone beads forming the SecY pore (Met 75, Ile 78, Ile 183, Val 278 and Ser 401 in *G. thermodentrifinicans numbering*). We performed 200 ns umbrella sampling MD, using a z-axis umbrella force constant of 2000 kJ mol^{-1} nm^{-2}, in 0.1 nm windows along this coordinate, with the lysine either protonated or deprotonated.

Construction of the 1D free-energy profile was achieved for the last 150 ns of each window using the weighted histogram analysis method, implemented in the gmx wham program[61]. Convergence was determined through analysis histogram overlap (Fig. S2f).

Mitochondrial import experiments

The model mitochondrial import substrate $\mathrm{B2}^{220}_{\Delta 43-65}$-Pep86 was expressed and purified as described previously[27]. The three cysteines were then fully reduced by incubating with 100 mM DTT for 30 min at room temperature, then the protein loaded onto a Source-S column in a buffer consisting of 20 mM HEPES pH 7.0, 6 M urea and 1 mM DTT, then eluted over a gradient of 0-1 M KCl. Labelling was performed by incubating the samples for 1 h in the dark, at room temperature, at pH 8.0 (adjusted with Tris base) with either 10 mM sodium iodoacetate (IA; Generon) or 20 mM sodium iodofluoroacetate (IFA; Apollo Scientific). These concentrations were chosen to give comparable labelling efficiencies, with at least one label per pre-protein (2.07 for IA, 1.57 for IFA, see below). Labelling reactions were quenched with 100 mM DTT, and excess label and DTT removed using a Superose-12 column. Labelling efficiencies were estimated by reacting the free cysteines with 800 μM Atto-610 maleimide for 1 h on ice, then running an SDS-PAGE gel, quantifying the bands on a Typhoon FLA 9500, and comparing the values to unlabelled protein (3 cysteines) and a similar protein with no cysteines ($\mathrm{B2}^{158}_{\Delta 43-65}$-Pep86[27]).

Mitochondrial import assays were performed with 1 μM pre-protein as described in ref 27. For each mitochondrial preparation (n=4), experiments were performed in duplicate and the averaged results quantified as for Spy (above). Differences between groups were assessed by a repeated-measurements one-way ANOVA, with Geisser-Greenhouse correction, where samples were paired to exclude variability associated with mitochondrial isolation. The p value of planned contrasts were corrected using a Holm-Sidak test. Plotting and analysis was carried on GraphPad Prism V8.0.0

Acknowledgments

We thank Gail Bartlett for help with the bioinformatics analysis. This work was funded by Wellcome (104632) and the BBSRC (BB/S008349/1; BB/N015126/1; BB/M003604/1; BB/I008675/1). Extended simulations were run 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.

Author Contributions

W.J.A., R.A.C., D.W.W and I.C. conceived and designed the experiments, and wrote the manuscript. W.J.A. performed and analysed the proOmpA transport, bioinformatics and mitochondrial substrate labelling experiments, R.A.C. performed and analysed the molecular dynamics and pKa analyses, the latter together with S.O. D.W.W. performed and analysed the Spy transport assays. G.C.P. performed and analysed the mitochondrial import experiments. I.C. supervised all the above work.

Dedication

We dedicate this paper to the late Prof A.R. Clarke (1959 – 2016) with whom we had many fruitful discussions about the concepts described here.

Competing Interests Statement

The authors have no competing interests.

References

1. Schiebel, E., Driessen, A. J., Hartl, F. U. & Wickner, W. Delta mu H+ and ATP function at different steps of the catalytic cycle of preprotein translocase. *Cell* **64,** 927–939 (1991).
2. Brundage, L., Hendrick, J. P., Schiebel, E., Driessen, A. J. & Wickner, W. The purified E. coli integral membrane protein SecY/E is sufficient for reconstitution of SecA-dependent precursor protein translocation. *Cell* **62,** 649–657 (1990).

3. Arkowitz, R. A. & Wickner, W. SecD and SecF are required for the proton electrochemical gradient stimulation of preprotein translocation. *EMBO J.* **13,** 954–963 (1994).
4. Yahr, T. L. & Wickner, W. T. Functional reconstitution of bacterial Tat translocation in vitro. *EMBO J.* **20,** 2472–2479 (2001).
5. Alder, N. N. & Theg, S. M. Energetics of protein transport across biological membranes. a study of the thylakoid DeltapH-dependent/cpTat pathway. *Cell* **112,** 231–242 (2003).
6. Mould, R. M. & Robinson, C. A proton gradient is required for the transport of two lumenal oxygen-evolving proteins across the thylakoid membrane. *J. Biol. Chem.* **266,** 12189–12193 (1991).
7. Schleyer, M., Schmidt, B. & Neupert, W. Requirement of a membrane potential for the posttranslational transfer of proteins into mitochondria. *Eur. J. Biochem.* **125,** 109–116 (1982).
8. Pfanner, N. & Neupert, W. Transport of F1-ATPase subunit beta into mitochondria depends on both a membrane potential and nucleoside triphosphates. *FEBS Lett.* **209,** 152–156 (1986).
9. Driessen, A. J. Precursor protein translocation by the Escherichia coli translocase is directed by the protonmotive force. *EMBO J.* **11,** 847–853 (1992).
10. Ito, K. *et al.* A temperature-sensitive mutant of E. coli exhibiting slow processing of exported proteins. *Cell* **32,** 789–797 (1983).
11. Arkowitz, R. A., Joly, J. C. & Wickner, W. Translocation can drive the unfolding of a preprotein domain. *EMBO J.* **12,** 243–253 (1993).
12. Josefsson, L. G. & Randall, L. L. Different exported proteins in E. coli show differences in the temporal mode of processing in vivo. *Cell* **25,** 151–157 (1981).
13. Collinson, I., Corey, R. A. & Allen, W. J. Channel crossing: how are proteins shipped across the bacterial plasma membrane? *Philos. Trans. R. Soc. Lond., B, Biol. Sci.* **370,** (2015).
14. Cranford-Smith, T. & Huber, D. The way is the goal: how SecA transports proteins across the cytoplasmic membrane in bacteria. *FEMS Microbiol Lett* **365,** (2018).
15. Schulze, R. J. *et al.* 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 (2014).
16. Allen, W. J. *et al.* Two-way communication between SecY and SecA suggests a Brownian ratchet mechanism for protein translocation. *Elife* **5,** (2016).
17. Bauer, B. W., Shemesh, T., Chen, Y. & Rapoport, T. A. A 'Push and Slide' Mechanism Allows Sequence-Insensitive Translocation of Secretory Proteins by the SecA ATPase. *Cell* **157,** 1416–1429 (2014).
18. Corey, R. A. *et al.* ATP-induced asymmetric pre-protein folding as a driver of protein translocation through the Sec machinery. *Elife* **8,** (2019).
19. Cao, G., Kuhn, A. & Dalbey, R. E. The translocation of negatively charged residues across the membrane is driven by the electrochemical potential: evidence for an electrophoresis-like membrane transfer mechanism. *EMBO J.* **14,** 866–875 (1995).
20. Ismail, N., Hedman, R., Lindén, M. & Heijne, von, G. Charge-driven dynamics of nascent-chain movement through the SecYEG translocon. *Nat. Struct. Mol. Biol.* **22,** 145–149 (2015).
21. Driessen, A. J. & Wickner, W. Proton transfer is rate-limiting for translocation of precursor proteins by the Escherichia coli translocase. *Proc. Natl. Acad. Sci. U.S.A.* **88,** 2471–2475 (1991).
22. Dalal, K. & Duong, F. The SecY complex forms a channel capable of ionic discrimination. *EMBO Rep.* **10,** 762–768 (2009).

23. Nouwen, N., Berrelkamp, G. & Driessen, A. J. M. Charged amino acids in a preprotein inhibit SecA-dependent protein translocation. *J. Mol. Biol.* **386,** 1000–1010 (2009).
24. Isom, D. G., Castañeda, C. A., Cannon, B. R. & García-Moreno, B. Large shifts in pKa values of lysine residues buried inside a protein. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 5260–5265 (2011).
25. Fitch, C. A., Platzer, G., Okon, M., Garcia-Moreno, B. E. & McIntosh, L. P. Arginine: Its pKa value revisited. *Protein Sci.* **24,** 752–761 (2015).
26. Harms, M. J., Schlessman, J. L., Sue, G. R. & García-Moreno, B. Arginine residues at internal positions in a protein are always charged. *Proc. Natl. Acad. Sci. U.S.A.* **108,** 18954–18959 (2011).
27. Pereira, G. C. *et al.* A high-resolution luminescent assay for rapid and continuous monitoring of protein translocation across biological membranes. *J. Mol. Biol.* in press, https://doi.org/10.1016/j.jmb.2019.03.007
28. van Dalen, A., Killian, A. & de Kruijff, B. Delta psi stimulates membrane translocation of the C-terminal part of a signal sequence. *J. Biol. Chem.* **274,** 19913–19918 (1999).
29. Pickering, B. S., Yudistira, H. & Oresnik, I. J. Characterization of the twin-arginine transport secretome in Sinorhizobium meliloti and evidence for host-dependent phenotypes. *Appl. Environ. Microbiol.* **78,** 7141–7144 (2012).
30. Takami, H. *et al.* Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and genomic sequence comparison with Bacillus subtilis. *Nucleic Acids Research* **28,** 4317–4331 (2000).
31. Søndergaard, C. R., Olsson, M. H. M., Rostkowski, M. & Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J Chem Theory Comput* **7,** 2284–2295 (2011).
32. Marrink, S. J., Risselada, H. J., Yefimov, S., Tieleman, D. P. & de Vries, A. H. The MARTINI Force Field: Coarse Grained Model for Biomolecular Simulations. *J. Phys. Chem. B* **111,** 7812–7824 (2007).
33. Monticelli, L. *et al.* The MARTINI Coarse-Grained Force Field: Extension to Proteins. *J Chem Theory Comput* **4,** 819–834 (2008).
34. Denkert, N. *et al.* Cation selectivity of the presequence translocase channel Tim23 is crucial for efficient protein import. *Elife* **6,** e28324 (2017).
35. Corey, R. A. *et al.* Specific cardiolipin-SecY interactions are required for proton-motive force stimulation of protein secretion. *Proc. Natl. Acad. Sci. U.S.A.* **115,** 7967–7972 (2018).
36. Haines, T. H. & Dencher, N. A. Cardiolipin: a proton trap for oxidative phosphorylation. *FEBS Lett.* **528,** 35–39 (2002).
37. Marom, M. *et al.* Interaction of the Tim44 C-terminal domain with negatively charged phospholipids. *Biochemistry* **48,** 11185–11195 (2009).
38. Patrick, J. W. *et al.* Allostery revealed within lipid binding events to membrane proteins. *Proc. Natl. Acad. Sci. U.S.A.* **115,** 2976–2981 (2018).
39. Duncan, A. L., Robinson, A. J. & Walker, J. E. Cardiolipin binds selectively but transiently to conserved lysine residues in the rotor of metazoan ATP synthases. *Proc. Natl. Acad. Sci. U.S.A.* **113,** 8687–8692 (2016).
40. Wiechert, M. & Beitz, E. Mechanism of formate–nitrite transporters by dielectric shift of substrate acidity. *EMBO J.* **36,** 949–958 (2017).
41. Khademi, S. *et al.* Mechanism of ammonia transport by Amt/MEP/Rh: structure of AmtB at 1.35 A. *Science* **305,** 1587–1594 (2004).
42. Jiang, J., Pentelute, B. L., Collier, R. J. & Zhou, Z. H. Atomic structure of anthrax protective antigen pore elucidates toxin translocation. *Nature* **521,** 545–549 (2015).
43. Ballhausen, B., Altendorf, K. & Deckers-Hebestreit, G. Constant c10 ring stoichiometry in the Escherichia coli ATP synthase analyzed by cross-linking. *J. Bacteriol.* **191,** 2400–2404 (2009).

44. The UniProt Consortium. UniProt: the universal protein knowledgebase. *Nucleic Acids Research* **45,** D158–D169 (2017).
45. Zimmer, J., Nam, Y. & Rapoport, T. A. Structure of a complex of the ATPase SecA and the protein-translocation channel. *Nature* **455,** 936–943 (2008).
46. Piggot, T. J., Sessions, R. B. & Burston, S. G. Toward a Detailed Description of the Pathways of Allosteric Communication in the GroEL Chaperonin through Atomistic Simulation. *Biochemistry* **51,** 1707–1718 (2012).
47. Corey, R. A., Allen, W. J. & Collinson, I. Protein translocation: what's the problem? *Biochem. Soc. Trans.* **44,** 753–759 (2016).
48. Park, E. *et al.* Structure of the SecY channel during initiation of protein translocation. *Nature* **506,** 102–106 (2013).
49. Jorgensen, W. L., Maxwell, D. S. & TiradoRives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **118,** 11225–11236 (1996).
50. Berendsen, H., Vanderspoel, D. & Vandrunen, R. Gromacs - a Message-Passing Parallel Molecular-Dynamics Implementation. *Computer Physics Communications* **91,** 43–56 (1995).
51. Krivov, G. G., Shapovalov, M. V. & Dunbrack, R. L. J. Improved prediction of protein side-chain conformations with SCWRL4. *Proteins* **77,** 778–795 (2009).
52. Huang, J. & MacKerell, A. D. CHARMM36 all-atom additive protein force field: validation based on comparison to NMR data. *J Comput Chem* **34,** 2135–2145 (2013).
53. Bashford, D. An object-oriented programming suite for electrostatic effects in biological molecules: An experience report on the MEAD project. in **1343,** 233–240 (Springer Berlin Heidelberg, 1997).
54. Baptista, A. M. & Soares, C. M. Some Theoretical and Computational Aspects of the Inclusion of Proton Isomerism in the Protonation Equilibrium of Proteins. *J. Phys. Chem. B* **105,** 293–309 (2001).
55. Teixeira, V. H., Soares, C. M. & Baptista, A. M. Studies of the reduction and protonation behavior of tetraheme cytochromes using atomic detail. *J. Biol. Inorg. Chem.* **7,** 200–216 (2002).
56. Teixeira, V. H. *et al.* On the use of different dielectric constants for computing individual and pairwise terms in poisson-boltzmann studies of protein ionization equilibrium. *J. Phys. Chem. B* **109,** 14691–14706 (2005).
57. Lee, F. S., Chu, Z. T., Bolger, M. B. & Warshel, A. Calculations of antibody-antigen interactions: microscopic and semi-microscopic evaluation of the free energies of binding of phosphorylcholine analogs to McPC603. *Protein Eng.* **5,** 215–228 (1992).
58. Åqvist, J. & Hansson, T. On the Validity of Electrostatic Linear Response in Polar Solvents. *J. Phys. Chem.* **100,** 9512–9521 (1996).
59. Hummer, G. & Szabo, A. Calculation of free-energy differences from computer simulations of initial and final states. *The Journal of Chemical Physics* **105,** 2004–2010 (1996).
60. Li, L. *et al.* Crystal structure of a substrate-engaged SecY protein-translocation channel. *Nature* **531,** 395–399 (2016).
61. Hub, J. S., de Groot, B. L. & van der Spoel, D. g_wham—A Free Weighted Histogram Analysis Implementation Including Robust Error and Autocorrelation Estimates. *J Chem Theory Comput* **6,** 3713–3720 (2010).

Figure legends

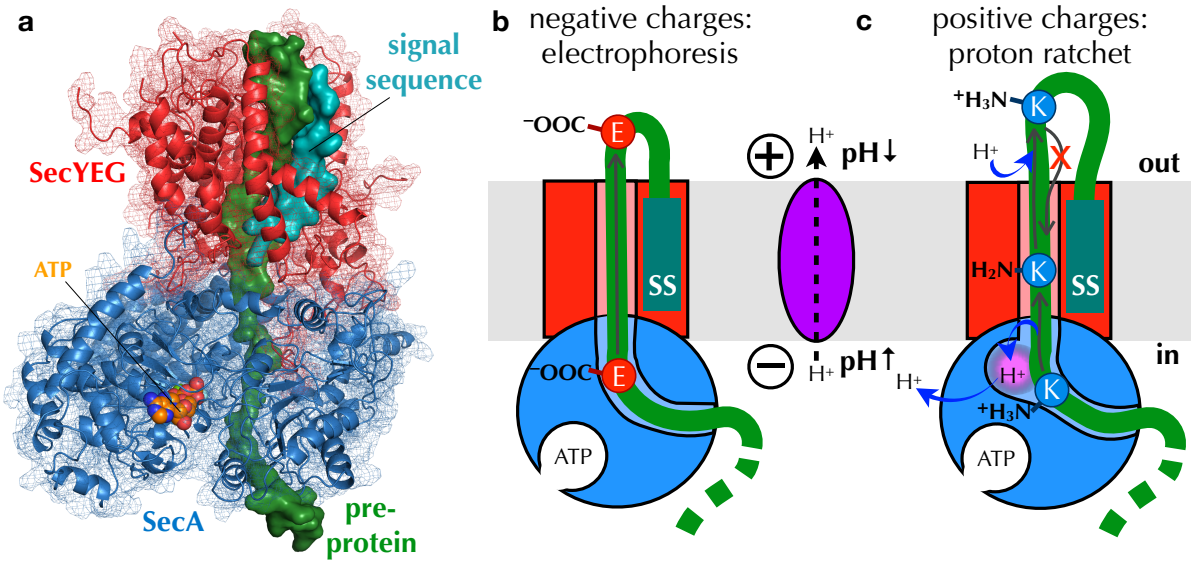


Figure 1. The proton ratchet model

a) Structure of SecYEG (red) bound to SecA (blue)[45], with a pre-protein (green, with SS in turquoise) modelled in, engaged for transport[47]. The ATP analogue bound to SecA is shown in orange spheres.

b-c) Proposed model for PMF-driven transport. The electron transport chain (purple) produces a charge (Δψ) and pH (ΔpH) gradient across the membrane (grey). Export of negatively charged residues (glutamate, aspartate) is aided electrophoretically by Δψ (panel b). Lysines are deprotonated at the cytosolic surface – neutralising their positive charge – then re-protonated on the periplasmic side, preventing backsliding through the channel (panel c).

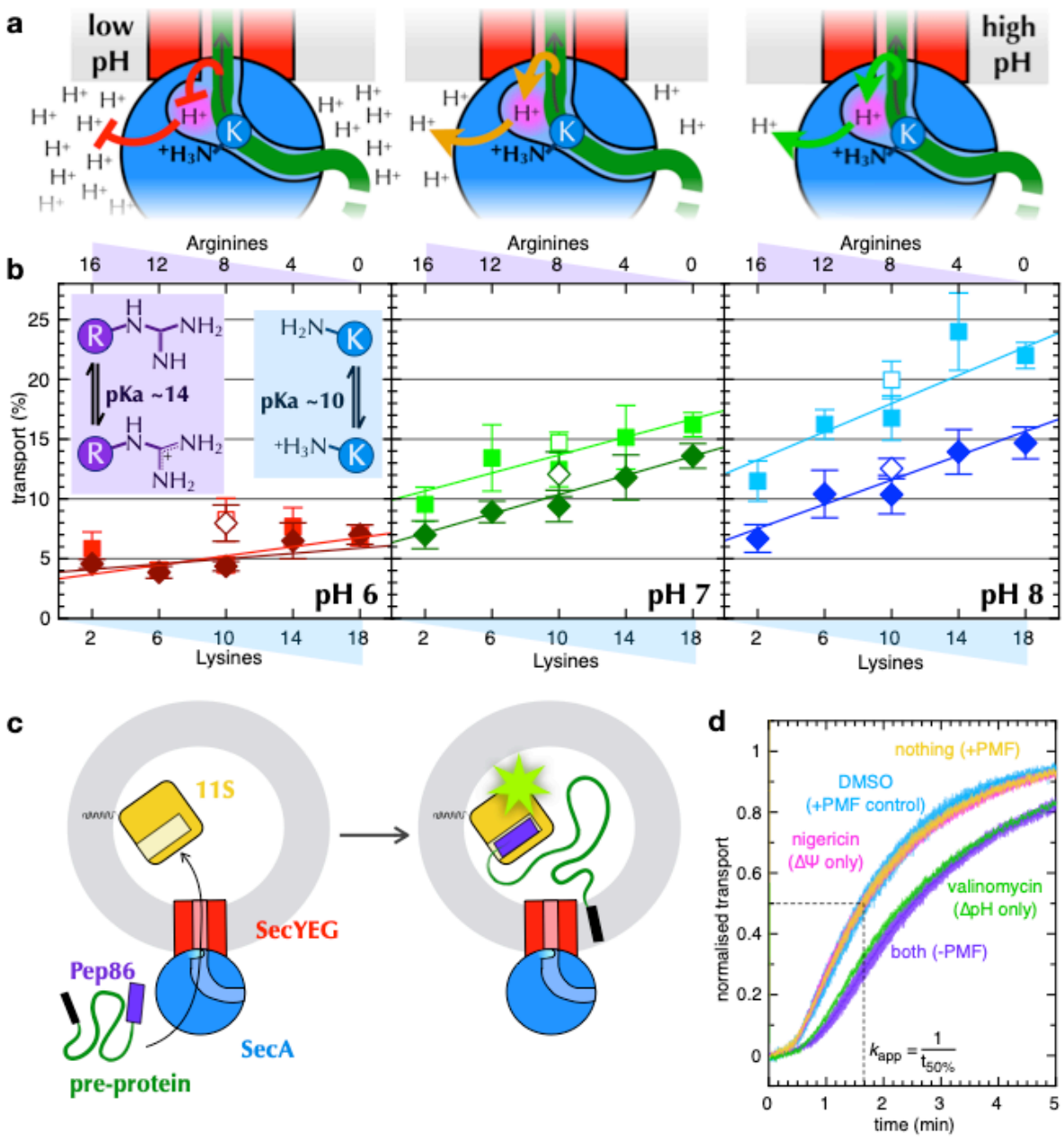


Figure 2. Transport of proOmpA and variants

a) Transport as a percent of input material after 30 s for six different Lys/Arg variants of pOA, at pH 6 (red), pH 7 (green) and pH 8 (blue) in the presence (lighter shade squares) or absence (darker shade diamonds) of a PMF. A control with all arginine and lysine positions swapped is depicted by empty shapes. Data and errors represent the average and SEM from 7 independent experiments. Inset are the structures and approximate solution pK_as of arginine (purple) and lysine (blue).

b) Schematic of the NanoLuc-based transport assay. Pre-protein with Pep86 fused to its C-terminus is imported into IMVs containing tethered 11S. As transport occurs, 11S and Pep86 associate rapidly to produce a real-time luminescence signal.

c) Normalised *in vitro* transport time courses of proOmpA-Pep86 alone (yellow), with DMSO (cyan), and with inhibitors (in DMSO) that selectively knock out individual components of the PMF: valinomycin (green), nigericin (magenta) and both (purple). Trace widths represent the SEM from three replicates. Apparent rates (k_{app}) are determined from the time required to transport half the total material (dotted black lines)

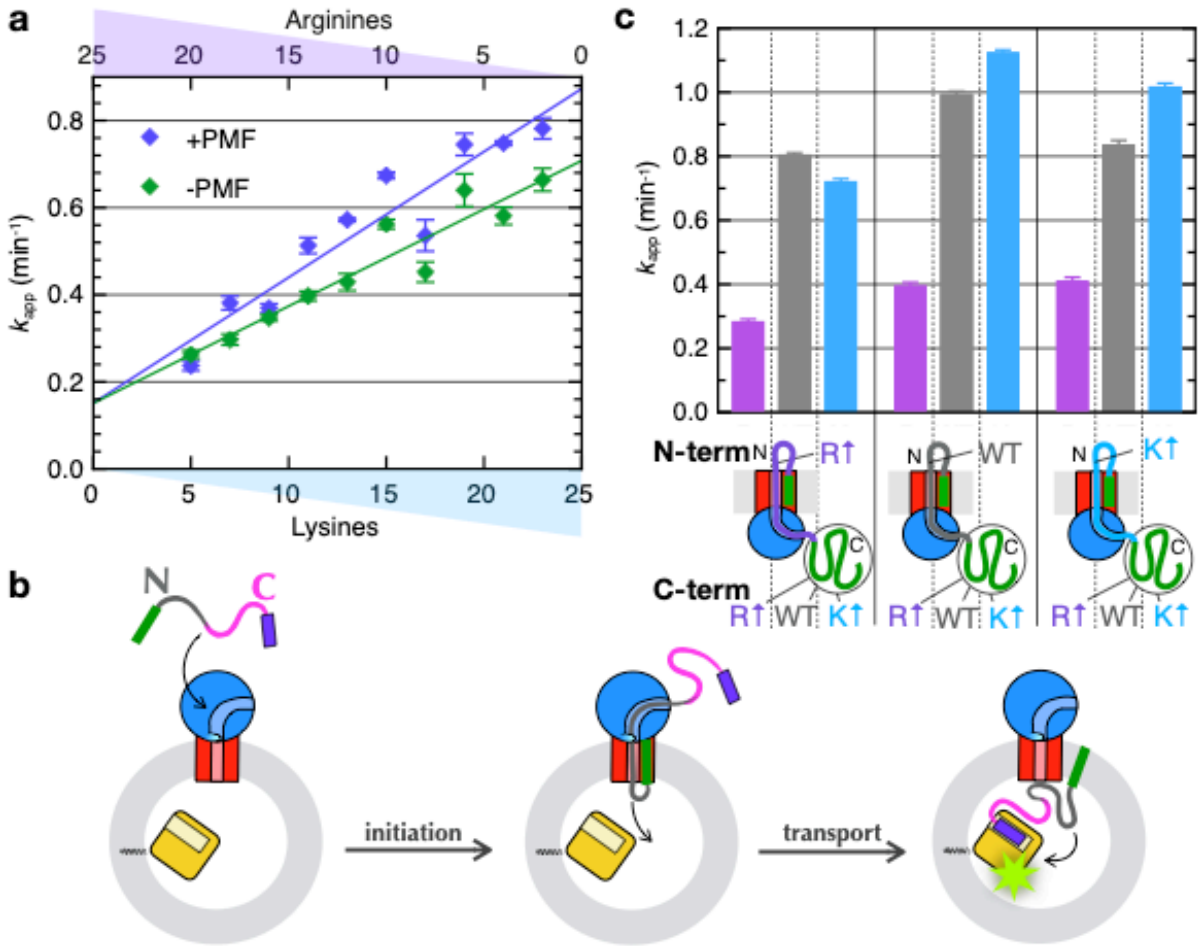


Figure 3. Transport of Spy

a) Apparent transport rates of proSpy-Pep86 as a function of the number of lysine residues, in the presence (blue) and absence (green) of a PMF. Errors represent the SEM from 3 replicates.

b) Schematic of translocation, showing both initiation and transport, using the same colour scheme as in Fig. 2B (except for the pre-protein). During initiation, the SS (green) and N-terminal part of pre-protein (grey) are inserted into the channel. In the subsequent transport step, the C-terminal portion of pre-protein (magenta) passes through the channel. Therefore, altering the N-terminus will affect both initiation and transport, while altering the C-terminus will affect only transport.

c) PMF-driven transport of proSpy-Pep86 variants with the sequences of the N- and/or C-termini either unaltered (WT, black), or with higher (blue) or lower (purple) Lys/Arg ratio (see

panel B for an explanation). Points and error bars represent the mean and SEM from 3 replicates.

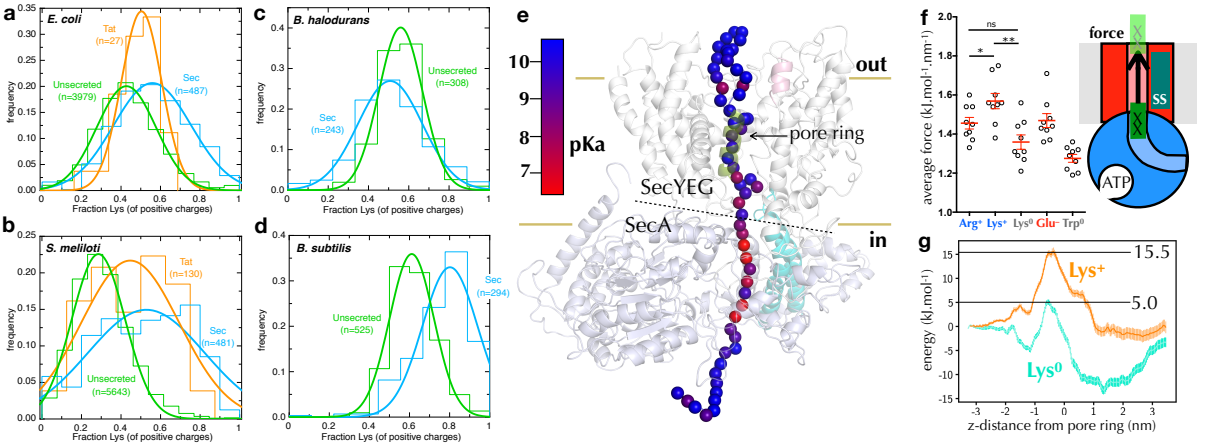


Figure 4. Computational analyses

a-d) Histograms (fine lines) showing proportion of positive residues that are lysine in cytosolic proteins (green), Sec substrates (blue) and Tat substrates (orange; a-b only), for **(a)** *E. coli,* **(b)** *S. meliloti*, **(c)** *B. halodurans* and **(d)** *B. subtilis*. Best fit single Gaussians are also shown (thick lines) for clarity.

e) Structure of SecYEG (white, plug in pink) and SecA (light blue; 2-helix finger in cyan) with a pre-protein (proOmpA) modelled through the channel (as in Fig. 1A). The SS of proOmpA is shown in transparent green cartoon, and the rest as Cα-beads, coloured according to the calculated pK_a a lysine at that position would have. The dashed line shows the interface between SecYEG and SecA.

f) Force required to pull a short region of pre-protein containing specified residues through the channel, as in the inset schematic. Raw data for each repeat are shown (n=9), along with bars for the mean and SEM, with significance determined using a two-tailed t-test.

g) 1D free energy profiles for an Ala-Lys-Ala peptide through the SecY pore, either protonated (Lys^+; orange) or deprotonated (Lys^0; aquamarine). Errors were computed using 200 rounds of bootstrap analysis.

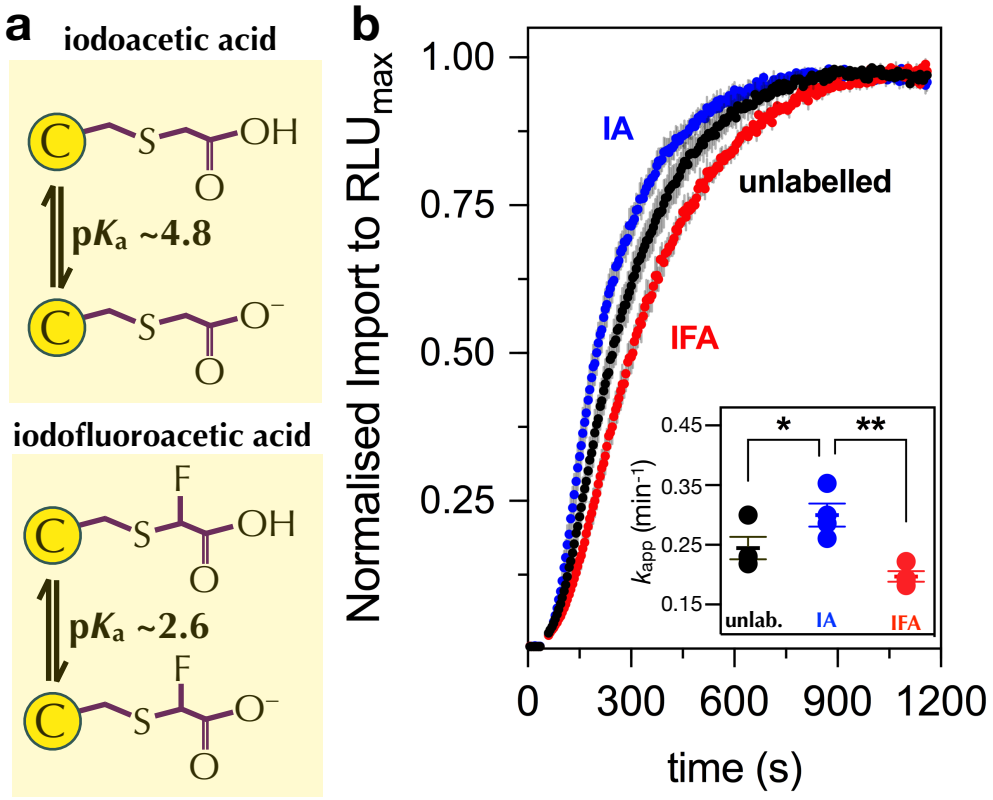


Figure 5.

Proton ratcheting in mitochondrial import

a) Structures and pK_as of cysteine residues labelled with iodoacetic acid (IA) or iodofluoroacetic acid (IFA).

b) Rate of mitochondrial import for a model TIM substrate ($B2^{220}_{\Delta 43-65}$-Pep86) either unlabelled (black), or labelled with iodoacetic acid (blue; pK_a ~ 4.8) or iodofluoroacetic acid (red; pK_a ~ 2.6). Import traces are shown normalised to maximum signal intensity (1,038,393 for unlabelled, 692,383 for IFA, 939,077 for IA), with quantified rates inset. Error bars represent the SEM from 4 independent experiments.

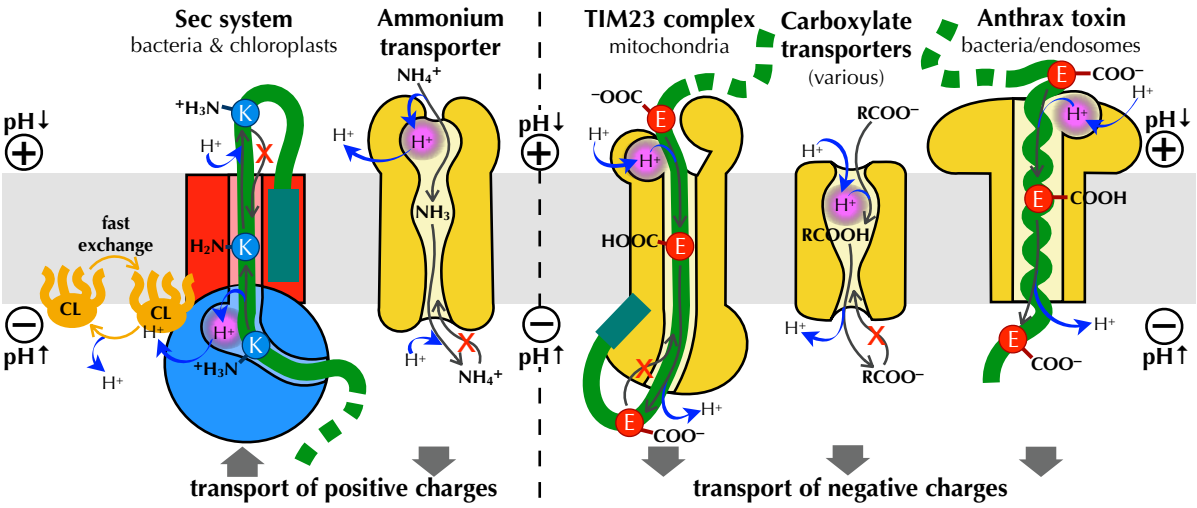


Figure 6. A universal mechanism

Diverse transporters are likely to harness pK_a perturbation to drive transport, including: the Sec system (investigated here); the ammonium transporter, which acts in reverse, generating a PMF as it imports NH_4^+ [41]; the TIM complex of mitochondria (tested here); carboxylate transporters, e.g. the formate transporter FocA[40]; and the anthrax protective antigen pore[42].

Supplemental Information

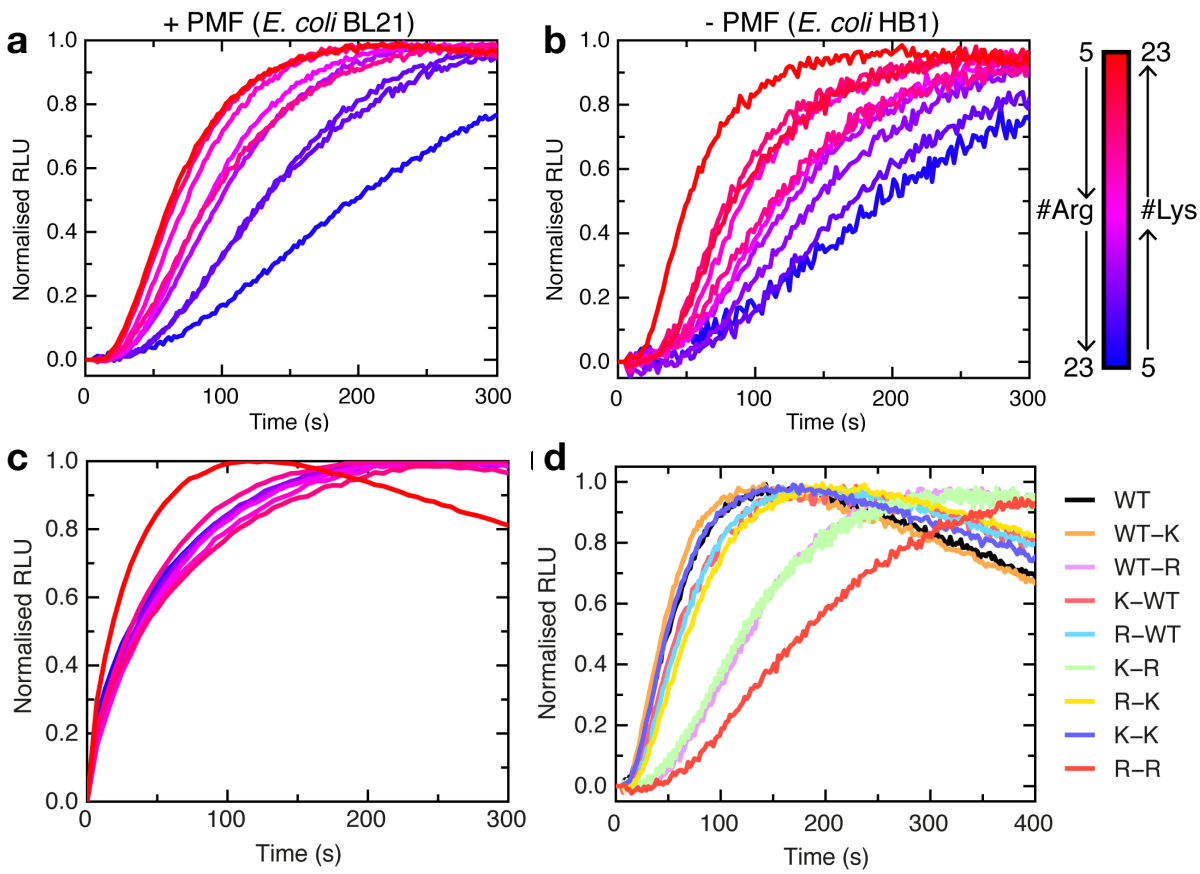
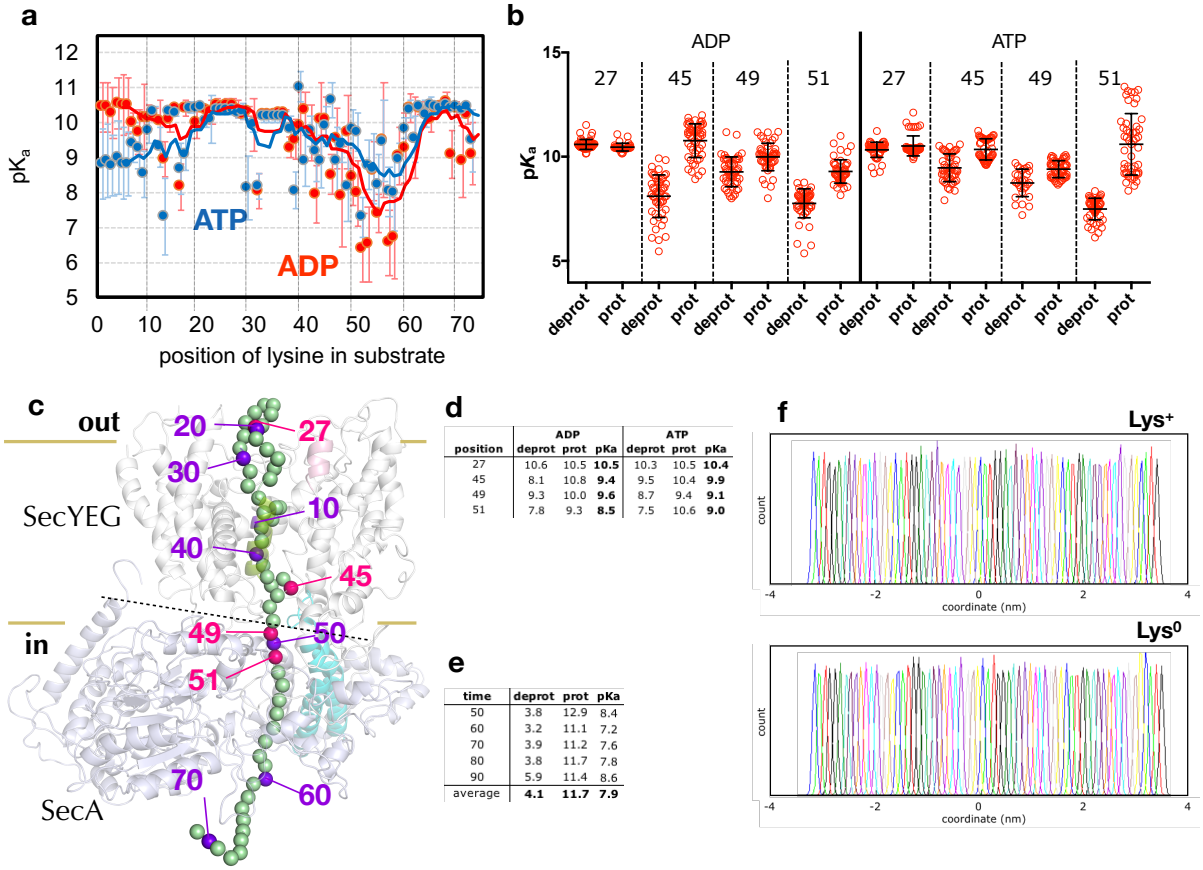


Figure S1. Raw Spy transport traces

a-b) Raw data from proSpy-Pep86 Lys/Arg transport assays (Fig. 3a), **(a)** with and **(b)** without a PMF. The traces are averaged from 3 repeats and are colour coded to the key (right), where increasingly red or blue corresponds to more lysine or arginine respectively.

c) Binding of proSpy-Pep86 Lys/Arg variants to 11S in solution, colour coded as in panel a-b. The traces confirm that the Lys/Arg effect reflects the kinetics of transport, not Pep86-11S association.

d) Raw data from the proSpy-Pep86 variant series shown in Fig. 3C. The traces are averaged from 3 repeats.



position	ADP deprot	ADP prot	ADP pKa	ATP deprot	ATP prot	ATP pKa
27	10.6	10.5	**10.5**	10.3	10.5	**10.4**
45	8.1	10.8	**9.4**	9.5	10.4	**9.9**
49	9.3	10.0	**9.6**	8.7	9.4	**9.1**
51	7.8	9.3	**8.5**	7.5	10.6	**9.0**

time	deprot	prot	pKa
50	3.8	12.9	8.4
60	3.2	11.1	7.2
70	3.9	11.2	7.6
80	3.8	11.7	7.8
90	5.9	11.4	8.6
average	**4.1**	**11.7**	**7.9**

Figure S2. Computational supporting data

a) Data from propKa31 scanning analysis, for the ADP (red) and ATP (blue) simulations. Mean pK_a values are shown for lysine residues in each position (see panel C for numbering reference), averaged over 20 snapshots taken from the final 500 ns of each simulation. Error bars are standard deviation. A 5 window moving average is plotted through the data.

b) Recorded pK_a values (raw data as red circles, mean and SD as black lines), as determined using propKa31, for lysines at selected positions (see panel c) in 50 snapshots taken from 100 ns of simulation. To eliminate bias, these were run with lysine starting both in a protonated ('prot') and deprotonated ('deprot') state. Both the ATP- and ADP-bound states were simulated.

c) Structural model of SecYEG-SecA with pre-protein, as in Fig. 4e, but with the pre-protein coloured green. For ease of reference, every 10th residue is coloured purple, and the residues selected for detailed analysis are coloured hotpink.

d) Time-averaged pK_a values from the data in panel b, for each lysine position, protonation state and nucleotide state. For each position, an average pK_a is calculated (in bold), as per the LRA technique.

e) as per panel D, but for one lysine position (51) and the ATP state only, and determined using MEAD. Shown are individual pK_a values at 50, 60, 70 80 and 90 ns. The data predict a pK_a of 7.9.

f) Histograms of the umbrella sampling energies for windows along the z-axis reaction coordinate (i.e. perpendicular to the membrane), showing sufficient overlap between histograms. Top is for the Lys^+ data, bottom for the Lys^0 data.

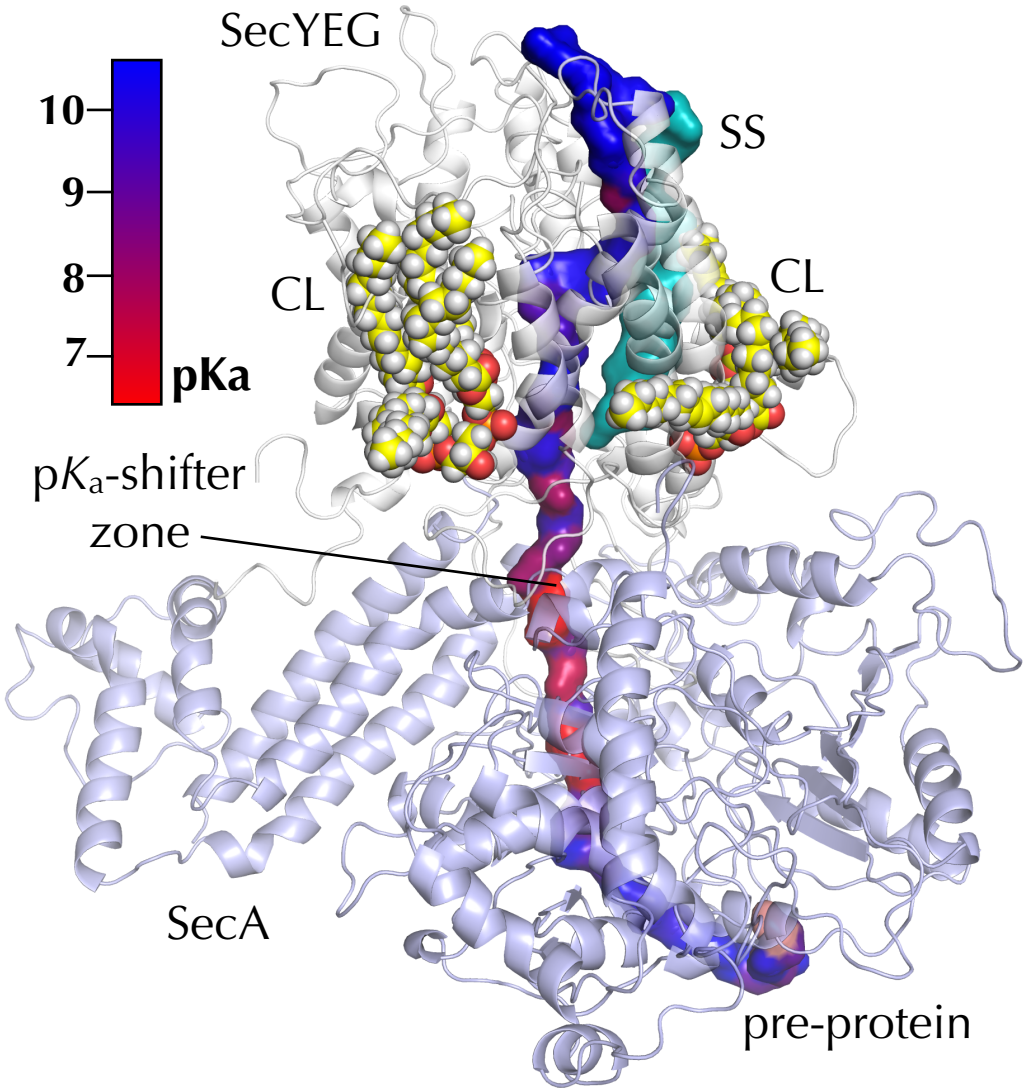


Figure S3. Cardiolipin binding sites on the Sec complex

Structural model of SecYEG-SecA-proOmpA, coloured as in Fig. 4e. The pre-protein backbone is shown in surface representation, with lysine-pK_a colour coding inset, for reference. Bound CL molecules[35] are shown as spheres with carbon in yellow, hydrogen white, phosphorus orange and oxygen red.

Table S1. Complete list of pre-protein sequences.
Colour coding: Orange = signal sequence ; **R** = Arginines; **K** = Lysines; Brown = V5 epitope; Green = linker and Pep86 sequence; Purple = TEV site/His tag.

proOmpA Lys/Arg series (Fig. 2A)	
$K_{18}R_0$	MGKKTAIAIAVALAGFATVAQAAPKDNTWYTMAKLGWSQYHDTGFINNNGPTHENQLGAGAFGGYQVNPYVGFEMG YDWLGKMPYKGSVENGAYKAQGVQLTAKLGYPITDDLDIYTKLGGMVWKADTKSNVYGKNHDTGVSPVFAGGVEYAI TPEIATKLEYQWTNNIGDAHTISAKGMGESNPVTGNTSDNVKQKAALIDSLAPDKKVEIEVKGIKDVVTQPQAIPNPLLG L*

$K_{14}R_4$	MGKKTAIAIAVALAGFATVAQAAPKDNTWYTMAKLGWSQYHDTGFINNNGPTHENQLGAGAFGGYQVNPYVGFEMG YDWLGKMPYKGSVENGAYKAQGVQLTAKLGYPITDDLDIYTRLGGMVWKADTKSNVYGKNHDTGVSPVFAGGVEYAI TPEIATRLEYQWTNNIGDAHTISAKGMGESNPVTGNTSDNVKQRAALIDSLAPDKRVEIEVKGIKDVVTQPQAIPNPLLG L*
$K_{10}R_8$ (wild-type)	MGKKTAIAIAVALAGFATVAQAAPKDNTWYTMAKLGWSQYHDTGFINNNGPTHENQLGAGAFGGYQVNPYVGFEMG YDWLGRMPYKGSVENGAYKAQGVQLTAKLGYPITDDLDIYTRLGGMVWRADTKSNVYGKNHDTGVSPVFAGGVEYAI TPEIATRLEYQWTNNIGDAHTISARGMGESNPVTGNTSDNVKQRAALIDSLAPDRRVEIEVKGIKDVVTQPQAIPNPLLG L*
K_6R_{12}	MGKKTAIAIAVALAGFATVAQAAPKDNTWYTMAKLGWSQYHDTGFINNNGPTHENQLGAGAFGGYQVNPYVGFEMG YDWLGRMPYRGSVENGAYKAQGVQLTARLGYPITDDLDIYTRLGGMVWRADTKSNVYGRNHDTGVSPVFAGGVEYA ITPEIATRLEYQWTNNIGDAHTISARGMGESNPVTGNTSDNVKQRAALIDSLAPDRRVEIEVRGIKDVVTQPQAIPNPLL GL*
K_2R_{16}	MGKKTAIAIAVALAGFATVAQAAPKDNTWYTMAKLGWSQYHDTGFINNNGPTHENQLGAGAFGGYQVNPYVGFEMG YDWLGRMPYRGSVENGAYRAQGVQLTARLGYPITDDLDIYTRLGGMVWRADTRSNVYGRNHDTGVSPVFAGGVEYA ITPEIATRLEYQWTNNIGDAHTISARGMGESNPVTGNTSDNVRQRAALIDSLAPDRRVEIEVRGIRDVVTQPQAIPNPLL GL*
R_8K_{10} (swap)	MGKKTAIAIAVALAGFATVAQAAPKDNTWYTMAKLGWSQYHDTGFINNNGPTHENQLGAGAFGGYQVNPYVGFEMG YDWLGKMPYRGSVENGAYRAQGVQLTARLGYPITDDLDIYTKLGGMVWKADTRSNVYGRNHDTGVSPVFAGGVEYA ITPEIATKLEYQWTNNIGDAHTISAKGMGESNPVTGNTSDNVRQKAALIDSLAPDKKVEIEVRGIRDVVTQPQAIPNPLLG L*
pSpy Lys/Arg series variant (Fig. 3A)	
K5	MRKLTALFVASTLALGAANLAHAADTTTAAPADAKPMMHHKGKFGPHQDMMFRDLNLTDAQRQQIREIMRGQRDQM RRPPLEERRAMHDIIASDTFDRVRAEAQIARMEEQRRANMLAHMETQNRIYNILTPEQRRQFNANFERRLTERPAAKG KMPATAEIPNPLLGLGSGVSGWRLFKKISENLYFQGHHHHHH
K7	MRKLTALFVASTLALGAANLAHAADTTTAAPADAKPMMHHKGKFGPHQDMMFRDLNLTDAQKQQIREIMRGQRDQM RRPPLEERRAMHDIIASDTFDRVRAEAQIARMEEQRRANMLAHMETQNRIYNILTPEQKRQFNANFERRLTERPAAKG KMPATAEIPNPLLGLGSGVSGWRLFKKISENLYFQGHHHHHH
K9	MRKLTALFVASTLALGAANLAHAADTTTAAPADAKPMMHHKGKFGPHQDMMFRDLNLTDAQKQQIREIMRGQRDQM KRPPLEERRAMHDIIASDTFDRVRAEAQIARMEEQRKANMLAHMETQNRIYNILTPEQKRQFNANFERRLTERPAAKG KMPATAEIPNPLLGLGSGVSGWRLFKKISENLYFQGHHHHHH
K11	MRKLTALFVASTLALGAANLAHAADTTTAAPADAKPMMHHKGKFGPHQDMMFRDLNLTDAQKQQIREIMRGQRDQM KRPPLEERRAMHDIIASDTFDRVKAEAQIARMEEQRKANMLAHMETQNRIYNILTPEQKRQFNANFEKRLTERPAAKG KMPATAEIPNPLLGLGSGVSGWRLFKKISENLYFQGHHHHHH
K13	MRKLTALFVASTLALGAANLAHAADTTTAAPADAKPMMHHKGKFGPHQDMMFKDLNLTDAQKQQIREIMKGQRDQM KRPPLEERRAMHDIIASDTFDRVKAEAQIARMEEQRKANMLAHMETQNRIYNILTPEQKRQFNANFEKRLTERPAAKG KMPATAEIPNPLLGLGSGVSGWRLFKKISENLYFQGHHHHHH
K15	MRKLTALFVASTLALGAANLAHAADTTTAAPADAKPMMHHKGKFGPHQDMMFKDLNLTDAQKQQIREIMKGQRDQM KRPPLEERRAMHDIIASDTFDKVKAEAQIARMEEQRKANMLAHMETQNKIYNILTPEQKRQFNANFEKRLTERPAAKG KMPATAEIPNPLLGLGSGVSGWRLFKKISENLYFQGHHHHHH
K17 (wild-type)	MRKLTALFVASTLALGAANLAHAADTTTAAPADAKPMMHHKGKFGPHQDMMFKDLNLTDAQKQQIREIMKGQRDQM KRPPLEERRAMHDIIASDTFDKVKAEAQIAKMEEQRKANMLAHMETQNKIYNILTPEQKKQFNANFEKRLTERPAAKG KMPATAEIPNPLLGLGSGVSGWRLFKKISENLYFQGHHHHHH
K19	MRKLTALFVASTLALGAANLAHAADTTTAAPADAKPMMHHKGKFGPHQDMMFKDLNLTDAQKQQIKEIMKGQRDQM KRPPLEEKRAMHDIIASDTFDKVKAEAQIAKMEEQRKANMLAHMETQNKIYNILTPEQKKQFNANFEKRLTERPAAKG KMPATAEIPNPLLGLGSGVSGWRLFKKISENLYFQGHHHHHH
K21	MRKLTALFVASTLALGAANLAHAADTTTAAPADAKPMMHHKGKFGPHQDMMFKDLNLTDAQKQQIKEIMKGQKDQM KRPPLEEKRAMHDIIASDTFDKVKAEAQIAKMEEQRKANMLAHMETQNKIYNILTPEQKKQFNANFEKKLTERPAAKG KMPATAEIPNPLLGLGSGVSGWRLFKKISENLYFQGHHHHHH

K23	MRKLTALFVASTLALGAANLAHAADTTTAAPADAKPMMHHKGKFGPHQDMMFKDLNLTDAQKQQIKEIMKGQKDQM KKPPLEEKRAMHDIIASDTFDKVKAEAQIAKMEEQKKANMLAHMETQNKIYNILTPEQKKQFNANFEKKLTERPAAKG KMPATAEIPNPLLGLGSGVSGWRLFKKISENLYFQGHHHHHH
pSpy variant (Fig. 3C)	
WT-WT	MRKLTALFVASTLALGAANLAHAADTTTAAPADAKPMMHHKGKFGPHQDMMFKDLNLTDAQKQQIREIMKGQRDQM KRPPLEERRAMHDIIASDTFDKVKAEAQIAKMEEQRKANMLAHMETQNKIYNILTPEQKKQFNANFEKRLTERPAAKG KMPATAEIPNPLLGLGSGVSGWRLFKKISENLYFQGHHHHHH
WT-K	MRKLTALFVASTLALGAANLAHAADTTTAAPADAKPMMHHKGKFGPHQDMMFKDLNLTDAQKQQIREIMKGQRDQM KRPPLEERRAMHDIIASDTFDKVKAEAQIAKMEEQKKANMLAHMETQNKIYNILTPEQKKQFNANFEKKLTEKPAAKG KMPATAEIPNPLLGLGSGVSGWRLFKKISENLYFQGHHHHHH
WT-R	MRKLTALFVASTLALGAANLAHAADTTTAAPADAKPMMHHKGKFGPHQDMMFKDLNLTDAQKQQIREIMKGQRDQM KRPPLEERRAMHDIIASDTFDRVRAEAQIARMEEQRRANMLAHMETQNRIYNILTPEQRRQFNANFERRLTERPAARG RMPATAEIPNPLLGLGSGVSGWRLFKKISENLYFQGHHHHHH
K-WT	MRKLTALFVASTLALGAANLAHAADTTTAAPADAKPMMHHKGKFGPHQDMMFKDLNLTDAQKQQIKEIMKGQKDQM KKPPLEEKKAMHDIIASDTFDKVKAEAQIAKMEEQRKANMLAHMETQNKIYNILTPEQKKQFNANFEKRLTERPAAKG KMPATAEIPNPLLGLGSGVSGWRLFKKISENLYFQGHHHHHH
K-K	MRKLTALFVASTLALGAANLAHAADTTTAAPADAKPMMHHKGKFGPHQDMMFKDLNLTDAQKQQIKEIMKGQKDQM KKPPLEEKKAMHDIIASDTFDKVKAEAQIAKMEEQKKANMLAHMETQNKIYNILTPEQKKQFNANFEKKLTEKPAAKG KMPATAEIPNPLLGLGSGVSGWRLFKKISENLYFQGHHHHHH
K-R	MRKLTALFVASTLALGAANLAHAADTTTAAPADAKPMMHHKGKFGPHQDMMFKDLNLTDAQKQQIKEIMKGQKDQM KKPPLEEKKAMHDIIASDTFDRVRAEAQIARMEEQRRANMLAHMETQNRIYNILTPEQRRQFNANFERRLTERPAARG RMPATAEIPNPLLGLGSGVSGWRLFKKISENLYFQGHHHHHH
R-WT	MRKLTALFVASTLALGAANLAHAADTTTAAPADARPMMHHRGRFGPHQDMMFRDLNLTDAQRQQIREIMRGQRDQM RRPPLEERRAMHDIIASDTFDKVKAEAQIAKMEEQRKANMLAHMETQNKIYNILTPEQKKQFNANFEKRLTERPAAKG KMPATAEIPNPLLGLGSGVSGWRLFKKISENLYFQGHHHHHH
R-K	MRKLTALFVASTLALGAANLAHAADTTTAAPADARPMMHHRGRFGPHQDMMFRDLNLTDAQRQQIREIMRGQRDQM RRPPLEERRAMHDIIASDTFDKVKAEAQIAKMEEQKKANMLAHMETQNKIYNILTPEQKKQFNANFEKKLTEKPAAKG KMPATAEIPNPLLGLGSGVSGWRLFKKISENLYFQGHHHHHH
R-R	MRKLTALFVASTLALGAANLAHAADTTTAAPADARPMMHHRGRFGPHQDMMFRDLNLTDAQRQQIREIMRGQRDQM RRPPLEERRAMHDIIASDTFDRVRAEAQIARMEEQRRANMLAHMETQNRIYNILTPEQRRQFNANFERRLTERPAARG RMPATAEIPNPLLGLGSGVSGWRLFKKISENLYFQGHHHHHH