

Article

Not peer-reviewed version

Charge Neutralization During Peptide Transport in the Bacterial SecYEG Translocon

[Laura Nübl](#), [Ekaterina Sobakinskaya](#), [Frank Müh](#)*

Posted Date: 25 July 2025

doi: 10.20944/preprints202507.2180.v1

Keywords: confined water; dielectric constant; molecular dynamics simulation; Monte Carlo sampling; pKa value; Poisson–Boltzmann equation; protonation probability; titration curve; translocation



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Charge Neutralization during Peptide Transport in the Bacterial SecYEG Translocon

Laura Nübl, Ekaterina Sobakinskaya and Frank Müh *

Institute for Theoretical Physics, Department for Theoretical Biophysics, Johannes Kepler University Linz, Altenberger Strasse 69, 4040 Linz, Austria

* Correspondence: frank.mueh@jku.at

Abstract

The driving force behind protein translocation across the cell membrane is not yet fully understood. In bacteria, there is an electrochemical potential across the cell membrane, which can interact with charged residues in the translocation substrate. In this study, the protonation states of lysine and glutamate, serving as test residues in a peptide translocating across the bacterial channel SecYEG, are investigated by applying Poisson-Boltzmann continuum electrostatic free energy calculations and Monte Carlo titrations to snapshots of molecular dynamics (MD) simulations. A clear shift in protonation probability towards the uncharged state is found for both test residues as they move deeper into the channel. Thus, charge neutralization occurs irrespective of whether the original charge of the test residue is positive (lysine) or negative (glutamate). Electrostatic interactions of acidic and basic residues of SecYEG with the peptide cancel out. The main determinants of the test residue's protonation state are the dielectric properties of its surroundings and interactions with non-titrating charges in the channel. Crucially, the membrane protein – including its water-filled pore – is assigned a low dielectric constant. The results are discussed in the context of the limitations inherent to continuum electrostatics and MD simulations with fixed protonation states.

Keywords: Confined water; dielectric constant; molecular dynamics simulation; Monte Carlo sampling; pK_a value; Poisson-Boltzmann equation; protonation probability; titration curve; translocation

1. Introduction

In nature, membranes composed of lipid bilayers are used by both pro- and eukaryotic cells to contain their interior structures and protect them from the environment. The tail-to-tail arrangement of the lipids in such membranes gives the structure a hydrophobic core, while the polar head groups point towards the aqueous exterior (see Figure 1a). This provides an efficient barrier to polar and charged (or simply very large) substances on either side, which require the help of channel proteins embedded in the membrane in order to cross [1]. The transportation of proteins or polypeptide chains – which are charged, polar, and very large – is facilitated by specialized channel proteins called translocons.

The Sec (from secretion system) complex is a universally conserved, heterotrimeric membrane protein that is involved in protein transport across, and insertion into, the plasma membrane. In prokaryotes, it is made up of subunits SecY, SecE, and SecG, while the analogous complexes found in eukaryotes and archaea are known as Sec61 $\alpha\beta\gamma$ and SecYE β , respectively [2–5]. The main subunit of SecYEG is SecY, whose ten transmembrane helices (TM) form two halves of a clamshell-like structure. It has a roughly hourglass shape, perpendicular to the membrane plane, with a ring of six hydrophobic amino acids – called the pore ring (PR) – whereby four of the hydrophobic groups define the constriction. For the insertion of proteins into the membrane, the translocon can open its lateral gate (LG), formed by helices TM2b and TM7 [6]. In the resting state, the central pore is blocked by a short helix, TM2a, termed the plug, which moves to the side in the active state [5,7]. Figure 1b shows

a cartoon representation of the main features of the channel. Of the other two subunits, only SecE is essential for function by wrapping around the central unit and holding the two halves together. SecG is non-essential for cell viability. The full structure of the heterotrimeric complex is shown in Figure 1a.

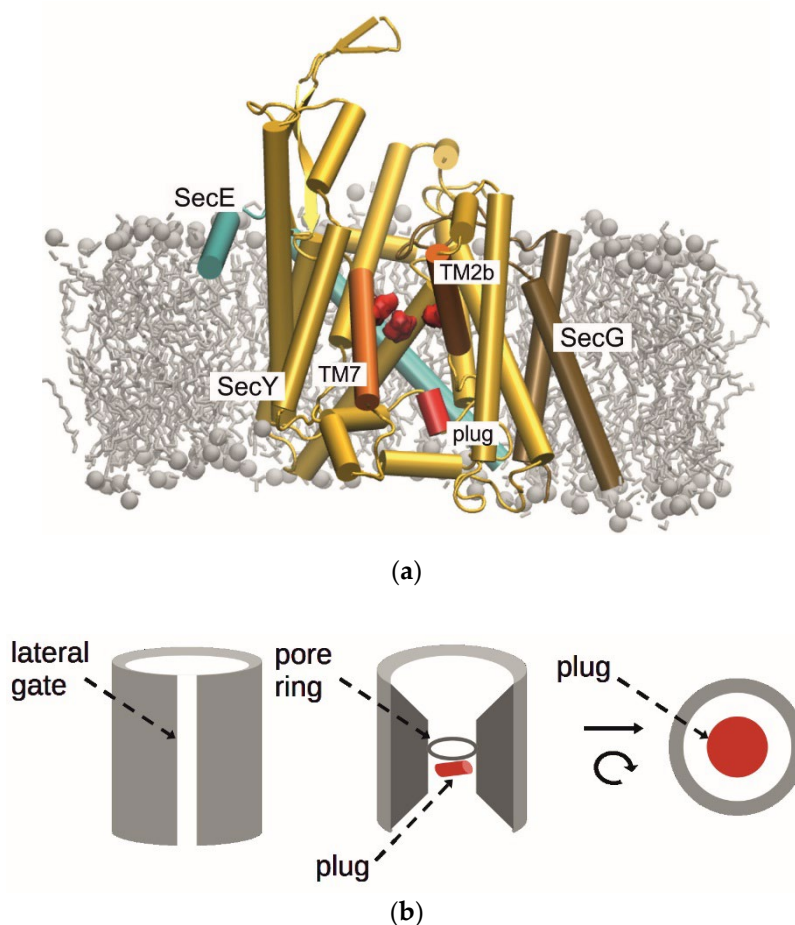


Figure 1. (a) Side view of SecYEG embedded in a lipid bilayer in the open state with the cytoplasmic side at the top and the periplasmic side at the bottom. Subunits SecY, SecE, and SecG are shown in yellow, cyan, and brown, respectively, with the exception of the lateral gate helices (TM2b, TM7), which are highlighted in orange. The pore ring (represented by four isoleucine residues) as well as the plug domain are colored red, while the membrane lipids are light grey. Figure made with VMD using the active state version of the *T. thermophilus* channel (PDB ID 5EUL [8]). (b) Cartoon representation of the SecY subunit highlighting its primary features. The hourglass shape of the channel interior and the hydrophobic pore ring can be seen in the center, while the lateral gate is shown on the left and on the right. The importance of the plug domain in sealing the pore is emphasized on the right (Figure adapted from [2]).

In order to activate the channel for translocation or insertion, it must dock to either a ribosome, in what is known as co-translational translocation, or the motor protein SecA for post-translational translocation. In the former, the protein is translocated as it emerges from the ribosome; in the latter, the protein is translated first and is then targeted to the translocon via a signal recognition particle and receptor complex [2]. Once the protein is in the active state, however, the question arises as to what the driving forces are for the subsequent translocation. For the insertion of largely hydrophobic helices into the membrane, it has been shown that the free energy gained by the partitioning into the lipid membrane causes a pulling force on the peptide chain still in the channel [9]. Regarding the translocation through the membrane, extensive research, both experimentally and theoretically, has been performed to uncover the physical mechanisms [2,4,5,10,11]. It appears that there is a kind of consensus on a general “rough” model of activation, which considers a series of steps: (i) association

of the channel with the protein to be translocated and ATP-bound SecA, (ii) opening of the LG by the signal sequence (SS) of the transported protein (via the movement of TM7 and TM2b), and (iii) plug relocation. However, details of the physical mechanism underlying activation and operation of the SecY machinery are still poorly understood. Recently, we tried to wrap the available information into a working model based on molecular dynamics (MD) simulations [5]. We came to the conclusion that binding of SecA (in the ATP- bound state) and the protein signal sequence leads to a transmembrane helix rearrangement that weakens contacts inside the hydrophobic core of SecYEG, and provides a driving force for plug opening. The conformational transitions are enabled by a delicate interplay between hydrophobic forces on one side, and a combination of forces on the other side that we termed PEES. This abbreviation stands for the proton motive force, external forces (due to binding with the translocation partners), entropic effects (e. g. due to conformational changes of the protein), and solvent-induced forces (e. g. due to bound water). In the open state, SecYEG still provides a barrier for bulky residues, meaning translocation is very likely not just a diffusion process, but pulling forces are required. In this context, it has been hypothesized earlier that the proton motive force plays a significant role in translocation [11].

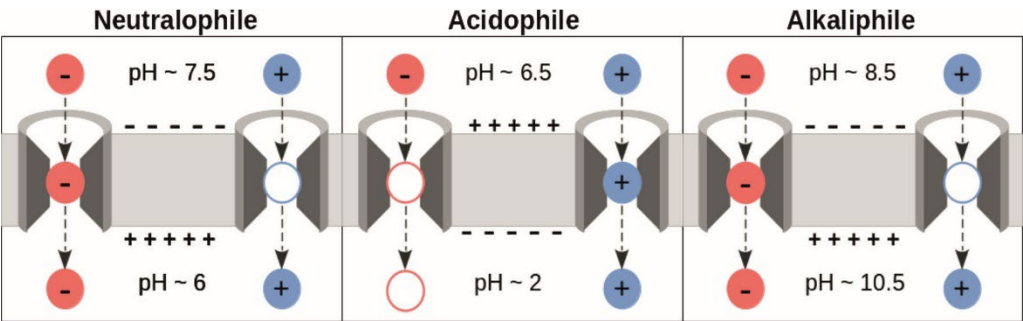


Figure 2. Summary of charge states of acidic (red) and basic (blue) residues over the course of translocation that would allow a net force to act in the direction of translocation; empty outlines imply the uncharged state of the residues. The black, dashed arrows show the direction of translocation (cytoplasm to periplasm), while the plus and minus symbols on either side of the membrane represent the transmembrane potential. The values are taken from [11] and the figure is an adapted version of the one published therein.

The proton motive force (PMF) is defined as the sum of the transmembrane potential and the transmembrane proton gradient, and it is an essential source of energy for cellular processes [12]. In *Escherichia coli* cells, the transmembrane potential is about 100-150 mV (negative inside), and has been shown to influence the translocation of model peptides containing presumably negatively charged residues [13]. However, as proteins can contain both acidic and basic amino acids – and generally have similar amounts of both – this means that, given the pK_a values of those residues at physiological pH values, it is questionable that there is a net force supporting translocation [11]. Since pK_a values of titratable groups depend on the electrostatic environment in which they are located, it is possible that shifts in pK_a values may cause a net charge of the substrate due to different charge states in acids and bases [11]. If the electrostatic interactions in the translocon stabilize negative charges, to compensate for them existing in an area of lower permittivity, while not doing the same for positively charged residues, there would be a net force acting in the direction of translocation.

Additionally, while the above considerations are based on the transmembrane potential and pH values of neutralophilic bacteria, the mechanism could still work in the direction of translocation in extremophilic bacteria [11], if the pK_a shifts experienced by acids and bases lead to the protonation states shown in Figure 2. Additionally, the electrostatic interactions with titratable residues in SecYEG could stabilize one charge over another in the different bacteria, as the specific protonation state of the translocon is pH dependent, too. However, there must be an alternative driving force involved in protein translocation in eukaryotic cells, as Sec61 is located in the endoplasmic reticulum (ER) membrane which does not have an appreciable transmembrane potential [11,14].

The goal of the present work is to computationally determine the titration behavior of amino acids during their translocation across SecYEG in order to gain insight into the driving forces behind this process.

2. Theory and Methods

2.1. Electrostatics

Different kinds of methods have been devised to assess the pK_a values of titratable groups in proteins, as well as the proteins' protonation states [15,16]. Among the physics-based computational methods, approaches combining Poisson-Boltzmann (PB) electrostatics with Monte Carlo (MC) sampling of protonation states are popular due to their cost-benefit ratio [17–20]. In the PB approach, proteins are modeled as a set of atomic partial charges (APCs) embedded in a dielectric continuum. The APCs are placed on atomic positions, which in the present work are derived from molecular dynamics (MD) simulations. The method is based on the Poisson equation

$$\nabla \cdot [\varepsilon(\mathbf{r}) \nabla \Phi(\mathbf{r})] = -\frac{\rho(\mathbf{r})}{\varepsilon_0}. \quad (1)$$

Here, ε_0 is the vacuum permittivity. The APCs q_i placed at the atomic positions $\mathbf{r}_i$ are represented by the charge density

$$\rho(\mathbf{r}) = \sum_i q_i \delta(\mathbf{r} - \mathbf{r}_i). \quad (2)$$

The electrostatic potential $\Phi(\mathbf{r})$ is obtained from a numerical solution of the differential equation (1) and used to compute interaction energies of titratable groups as detailed below and in the literature [17–19,21]. The relative dielectric permittivity $\varepsilon(\mathbf{r})$ describes the continuum in which the APCs are embedded as a heterogeneous dielectric medium with a piecewise constant electric polarizability. A distinction is made between the protein interior ($\varepsilon(\mathbf{r}) = \varepsilon_{\text{protein}}$), the lipid bilayer ($\varepsilon(\mathbf{r}) = \varepsilon_{\text{mem}}$), and the surrounding aqueous solution ($\varepsilon(\mathbf{r}) = \varepsilon_{\text{aq}}$).

For the description of proteins in a realistic environment, one has to take into account the effects of mobile ions in the aqueous phase. This is done by considering the total charge density in Equation (1) instead of $\rho(\mathbf{r})$

$$\rho_{\text{total}}(\mathbf{r}) = \rho(\mathbf{r}) + \rho_{\text{ion}}(\mathbf{r}), \quad (3)$$

where

$$\rho_{\text{ion}}(\mathbf{r}) = N_A \sum_s c_s(\mathbf{r}) q_s e^{-\beta q_s \Phi(\mathbf{r})} \quad (4)$$

represents the mobile ions. Here, β is the inverse thermal energy $k_B T$ (with k_B being Boltzmann's constant and T the absolute temperature) and N_A is Avogadro's constant, while c_s and q_s are, respectively, the molar concentration and charge of ion species s . Equation (1) with $\rho_{\text{total}}(\mathbf{r})$ at the right-hand side (rhs) is referred to as the Poisson-Boltzmann equation (PBE). For small potentials ($q_s \Phi \approx k_B T$), and preferably monovalent ions, the exponential term can be simplified by doing a Taylor expansion to the first order:

$$\sum_s c_s(\mathbf{r}) q_s e^{-\beta q_s \Phi(\mathbf{r})} \approx \sum_s c_s(\mathbf{r}) q_s - \beta \sum_s c_s(\mathbf{r}) q_s^2 \Phi(\mathbf{r}). \quad (5)$$

In Equation (5), the first term on the rhs is equal to zero due to the condition of electroneutrality. To further simplify the remaining term, the ionic strength $I(\mathbf{r})$ is introduced according to

$$I(\mathbf{r}) = \frac{1}{2} \sum_s c_s(\mathbf{r}) z_s^2, \quad (6)$$

where $z_s = q_s/e$ is the charge number of ion type s , and e is the elementary charge. Note that $I(\mathbf{r})$ depends on position, since it is nonzero only in the aqueous phase outside the protein and the membrane. Finally, by substituting the simplified terms into the PBE, the linearized Poisson-Boltzmann equation (LPBE) is obtained:

$$\epsilon_0 \nabla \cdot [\epsilon(\mathbf{r}) \nabla \Phi(\mathbf{r})] = -\rho(\mathbf{r}) + 2N_A \beta e^2 I(\mathbf{r}) \Phi(\mathbf{r}). \quad (7)$$

2.2. Protonation Patterns

The calculation of protonation states in proteins with multiple titratable residues can be done by determining the electrostatic work necessary to transfer each titratable residue in the protonated and deprotonated state from an aqueous medium into the protein environment, as well as computing the interaction energies with other titratable groups. This type of computation is based on the thermodynamic cycle shown in Figure 3, from which the free energy of deprotonation of a titratable group in the protein, $\Delta G_{\text{protein}}$, is obtained as [19,21,22]:

$$\Delta G_{\text{protein}} = \Delta G_{\text{model}} + \Delta G_{\text{transfer}}^{\text{d}} - \Delta G_{\text{transfer}}^{\text{p}}. \quad (8)$$

The free energy of deprotonation of the model compound, ΔG_{model} , is inferred from experimental $\text{p}K_a$ data [18]. Electrostatic computations are used to determine the free energy of transfer of the group A from an aqueous to a protein environment ($\Delta G_{\text{transfer}}$ with superscripts p and d for the protonated and deprotonated states, respectively).

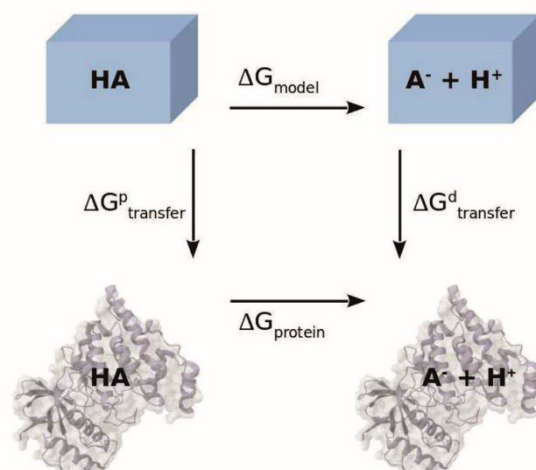


Figure 3. Thermodynamic cycle used to determine the free energy of deprotonation of a titratable group A in a protein environment. The blue cubes represent the aqueous environment in which the model compound titrations occur, while the molecular surface and cartoon structure in grey represent the group being in a protein environment.

The primary assumption of such a model is that the $\text{p}K_a$ of an ionizing group in the protein is that of a model compound (the same amino acid but isolated and with capping groups placed on its termini), shifted solely due to electrostatic interactions with the group's surroundings; any conformational changes and steric effects are neglected [23]. In the present work, we do not employ a single structure, but structural ensembles generated by MD simulations. The second assumption is that only a subset of APCs in the group change when the protonation state is changed (referred to as the titrating charges, Q_α), while the remaining APCs are constant (non-titrating charges, q_i). When only one ionizable group is considered, meaning all other residues are in their uncharged state, the $\text{p}K_a$ is the so-called intrinsic $\text{p}K_a$ [21,24], $\text{p}K_{a,\text{intr}}$:

$$\begin{aligned}
 \text{p}K_{a,\text{intr}} &= \text{p}K_{a,\text{model}} + \frac{\Delta G_{\text{transfer}}^{\text{d}} - \Delta G_{\text{transfer}}^{\text{p}}}{RT \ln 10} \\
 &= \text{p}K_{a,\text{model}} + \frac{(W_{\text{protein}}^{\text{p}} - W_{\text{protein}}^{\text{d}}) - (W_{\text{model}}^{\text{p}} - W_{\text{model}}^{\text{d}})}{RT \ln 10}.
 \end{aligned} \tag{9}$$

Here, energies are given per mole – as indicated by the use of the gas constant R . The ΔW terms, then, are the work required to create the charge distributions of those states. Writing the explicit contributions to each W term, the interaction between non-titrating charges in the protein will cancel between the W^{p} and W^{d} terms as the q_i do not change during a protonation state change. The Coulomb interactions of titrating charges will cancel between the W_{model} and W_{protein} expressions for each state, as they are identical in the protein and model compounds. This leaves only the interactions of the titrating and non-titrating charges with the electrostatic potential (ESP), which are called the Born and background terms by convention. Therefore, the $\text{p}K_{a,\text{intr}}$ can be re-written as:

$$\text{p}K_{a,\text{intr}} = \text{p}K_{a,\text{model}} + \frac{\Delta\Delta G_{\text{Born}} + \Delta\Delta G_{\text{bg}}}{RT \ln 10}. \tag{10}$$

The Born term, $\Delta\Delta G_{\text{Born}}$, results from the interaction of the titrating charges Q_α with the potentials $\Phi_{\text{protein}}^{\text{p}}$, $\Phi_{\text{protein}}^{\text{d}}$ of the protein and $\Phi_{\text{model}}^{\text{p}}$, $\Phi_{\text{model}}^{\text{d}}$ of the model compound – determined by solving the LPBE (Equation 7) with the charge distribution $\rho(\mathbf{r})$ comprising only the titrating charges Q_α^{p} or Q_α^{d} , and with the spatial dependence of $\varepsilon(\mathbf{r})$ and $I(\mathbf{r})$ being determined by the molecular surfaces of the protein and the model compound:

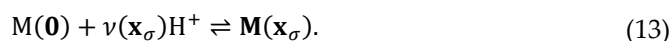
$$\begin{aligned}
 \Delta\Delta G_{\text{Born}} &= \frac{1}{2} \sum_{\alpha} \{ Q_{\alpha}^{\text{p}} [\Phi_{\text{protein}}^{\text{p}}(\mathbf{r}_{\alpha}) - \Phi_{\text{model}}^{\text{p}}(\mathbf{r}_{\alpha})] \\
 &\quad - Q_{\alpha}^{\text{d}} [\Phi_{\text{protein}}^{\text{d}}(\mathbf{r}_{\alpha}) - \Phi_{\text{model}}^{\text{d}}(\mathbf{r}_{\alpha})] \}.
 \end{aligned} \tag{11}$$

The background (bg) term represents the interaction of the non-titrating (background) charges q_i with the various potentials:

$$\begin{aligned}
 \Delta\Delta G_{\text{bg}} &= \sum_i q_{i,\text{protein}} [\Phi_{\text{protein}}^{\text{p}}(\mathbf{r}_i) - \Phi_{\text{protein}}^{\text{d}}(\mathbf{r}_i)] \\
 &\quad - \sum_k q_{k,\text{model}} [\Phi_{\text{model}}^{\text{p}}(\mathbf{r}_k) - \Phi_{\text{model}}^{\text{d}}(\mathbf{r}_k)].
 \end{aligned} \tag{12}$$

This leads to the LPBE being solved four times for each titratable site.

A protein with N titratable sites, having two different protonation states each (d or p), has a total of 2^N protonation patterns, which can be described by an N -dimensional vector $\mathbf{x}_{\sigma} = (x_1^{\sigma}, x_2^{\sigma}, \dots, x_N^{\sigma})$ of binary elements representing the charge state of each site ($x_i^{\sigma} = 0$ for deprotonated, $x_i^{\sigma} = 1$ for protonated; $i = 1, \dots, N$). In the following, we shall use the term “pattern” for such multisite states – to better distinguish them from the protonation states of individual sites – and denote the protein in such a multisite state by $\mathbf{M}(\mathbf{x}_{\sigma})$. The index $\sigma = 1, \dots, 2^N$ counts the various patterns. Let the vector $\mathbf{0}$ represent the reference pattern, in which all sites are in their neutral state. The entries in this vector are $x_i^0 = 1$ for acids and $x_i^0 = 0$ for bases. Following [17], we consider the chemical equilibrium between an arbitrary pattern $\mathbf{x}_{\sigma}$ and the reference pattern:



Here, $\nu(\mathbf{x}_{\sigma})$ is the net number of protons that would be released on going from pattern $\mathbf{x}_{\sigma}$ to the reference pattern. Note that $\nu(\mathbf{x}_{\sigma})$ may be negative, which would correspond to the situation that a net proton uptake leads from $\mathbf{M}(\mathbf{x}_{\sigma})$ to $\mathbf{M}(\mathbf{0})$. The net proton number can be written as a sum over sites:

$$\nu(\mathbf{x}_{\sigma}) = \sum_i \nu_i(x_i^{\sigma}), \tag{14}$$

where $v_i(x_i^\sigma) = x_i^\sigma - x_i^0 \in \{-1, 0, +1\}$. Then, the free energy change for going from $\mathbf{M}(\mathbf{0})$ to $\mathbf{M}(\mathbf{x})$ at some fixed pH is [17]:

$$\Delta G(\mathbf{x}_\sigma, \text{pH}) = \sum_i v_i(x_i^\sigma) 2.303 RT (\text{pH} - \text{p}K_{i,\text{intr}}) + \frac{1}{2} \sum_{\substack{ij \\ i \neq j}}^N W_{ij} v_i(x_i^\sigma) v_j(x_j^\sigma). \quad (15)$$

The first sum contains the changes in free energy due to the acid dissociation reactions of individual sites i , while the second term sums up all additional work originating from pairs of sites being in their non-reference states. Note that the entries in the double sum are non-zero only if both sites i and j are charged. The interaction energy is given by

$$W_{ij} = \sum_{\alpha}^{\text{site } i} (Q_{\alpha}^p - Q_{\alpha}^d) [\Phi_j^p(\mathbf{r}_{\alpha}) - \Phi_j^d(\mathbf{r}_{\alpha})] \quad (16)$$

with Q_{α} being the titrating charges of site i , and Φ_j the potentials due to the titrating charges of site j originating from the solution of the LPBE.

Using the free energy change, the protonation probability x_i of the individual site i as a function of pH can be expressed as a thermodynamic average over all protonation patterns of the protein [18]:

$$\langle x_i \rangle = \frac{\sum_{\sigma=1}^{2^N} x_i^{\sigma} \exp(\Delta G(\mathbf{x}_{\sigma}, \text{pH})/RT)}{\sum_{\sigma=1}^{2^N} \exp(\Delta G(\mathbf{x}_{\sigma}, \text{pH})/RT)}. \quad (17)$$

Note that x_i can be any number between 0 and 1 in contrast to the binary variable x_i^{σ} . The protonation probabilities can be plotted against the pH to yield titration curves. In general, the number 2^N is too large for a direct computation of the protonation probabilities on the basis of Equation (17). The solution to this problem is the use of a MC technique (see below).

2.3. Computational Methods

2.3.1. Preparation of Structures and General Setup of Molecular Dynamics Simulations

In order to determine the $\text{p}K_a$ changes experienced by amino acids in a translocating peptide – as a result of their position in the translocon SecYEG – constant velocity steered molecular dynamics (SMD) simulations were performed. The peptide was pulled through the channel and the titration behavior of residue 31 was investigated.

For the active state model of the Sec channel, the resting-state crystal structure of the *Thermus thermophilus* channel (PDB ID 5AWW [25]) was opened by moving its main helices and plug domain based on the coordinates of an active-state chimera structure (*Geobacillus thermodenitrificans* SecYE and *Bacillus subtilis* SecA; PDB ID 5EUL [8]) as described in our previous work [5,26]. The peptide used is composed of the signal sequence and first strand of the outer membrane protein A (OmpA) β -barrel, whose structure was created from its primary sequence (Figure 4) before being manually placed inside the channel using a CHARMM script [27]. These steps were part of a previous project (see [5] for details).

MD simulations were performed with NAMD version 2.13 [28] along with the CHARMM36m force field [29,30]. The TIP3P model was used for water molecules [31]. All simulations were run under *NPT* conditions: constant temperature control was regulated with Langevin dynamics [32] with a 5 ps^{-1} damping coefficient coupled to all heavy atoms ($T = 300 \text{ K}$); constant pressure was maintained at 1 atm by a Nosé–Hoover Langevin piston barostat [33,34] with decay period 100 fs and a damping time of 50 fs. Electrostatic interactions were computed with particle-mesh Ewald summation using a grid spacing of $1 \text{ \AA}$ and a cut-off of the short-range real-space interactions at $12 \text{ \AA}$ with a smooth switch function beginning at $10 \text{ \AA}$. The equations of motion were integrated in 2 fs or 1 fs steps using the velocity Verlet algorithm. The number of time steps between evaluations of the short-range non-bonded interactions and the full electrostatic interactions were equal to 2 and 4,

respectively. The SHAKE algorithm was employed to constrain the length of the bonds involving hydrogen atoms. Coordinates were saved every 2 ps.

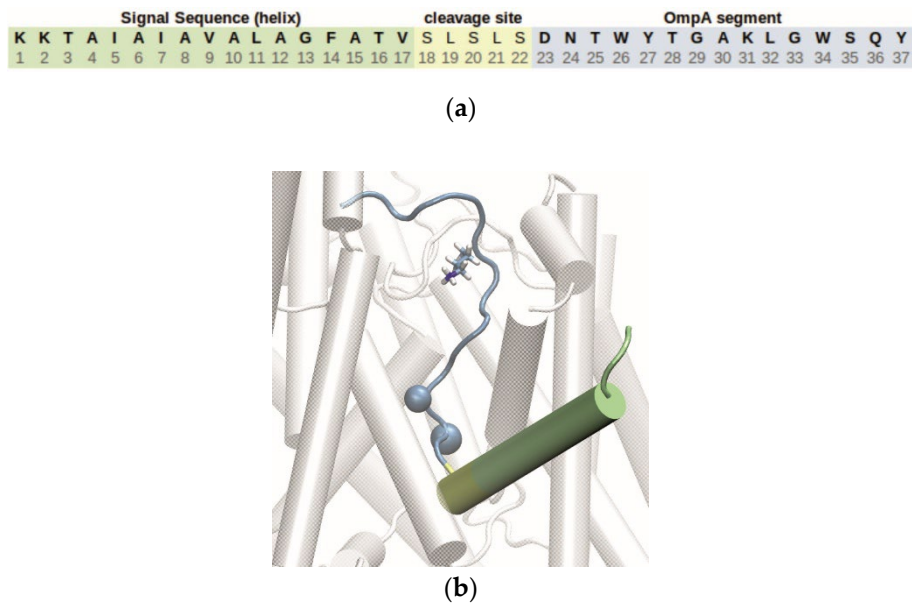


Figure 4. (a) Primary sequence of the pre-protein construct (referred to as "peptide") used as the translocation substrate. Position 31, which is a lysine in this sequence, marks the residue used to test the titration behavior. Positions 24 and 25 (C_{α} -atoms) were tagged as SMD atoms. (b) Visualization of peptide structure at the start of the SMD simulations, with color coding for the different sections as described in (a). The blue beads correspond to the SMD atoms, while the lysine shown in licorice representation is the test residue at position 31.

In our previous work [5], steered MD (SMD) with a constant velocity protocol [35] implemented in NAMD was used to pull the peptide through SecYEG in order to define various positions for umbrella sampling. Here, we used the same technique to define different positions of residue 31 of the peptide inside the channel for sampling of protonation states. The present SMD was performed with a step size of 1 fs. Harmonic constraints with a force constant of 1 kcal/(mol $\text{\AA}^2$) were placed on the protein backbone atoms as well as the plug domain and a linker segment. To pull the peptide, the C_{α} -atom of residue 24 of the peptide was tagged as the SMD atom, which was then pulled through the pore along the z -axis (set perpendicular to the membrane, see [5,26]) with a force constant of 36 kcal/(mol $\text{\AA}^2$) at a velocity of 2 $\text{\AA}/\text{ns}$. The simulation was run for 10 ns, followed by four additional runs for 7 ns each. After run 2, pulling was switched to the C_{α} -atom of residue 25 of the peptide to avoid the unwinding of the signal sequence helix. Snapshots of the system were then taken from the trajectories at regular intervals, covering various positions of residue 31 in the channel.

2.3.2. Preparation and Titration with Residue 31 being Lysine

Initial titrations were done on snapshots of the active-state structure of SecYEG with the peptide containing Lys 31 taken from our previous work [5]. A total of 8 snapshots in different stages of translocation were used as starting conformations for MD simulations. Before titrations were performed on the different snapshots (see Figure 5a for the snapshots chosen), the structures were used as starting conformations for 10 ns trajectories. The simulation parameters were kept the same as for the SMD simulation except that the constraints were modified. Harmonic constraints were placed on the C_{α} -atom of peptide residue 31 as well as the backbone of the plug domain and helices 5, 7, 8, and 9 of SecY. Force constants of 0.4 and 1 kcal/(mol $\text{\AA}^2$) were used, respectively. From these trajectories, 50 frames were taken from the final 5 ns and the structure files edited for titration by adding uncharged water and dummy atoms to the pore (see Section 2.3.4). Titrations were then performed on these frames as described in Sections 2.3.4-6.

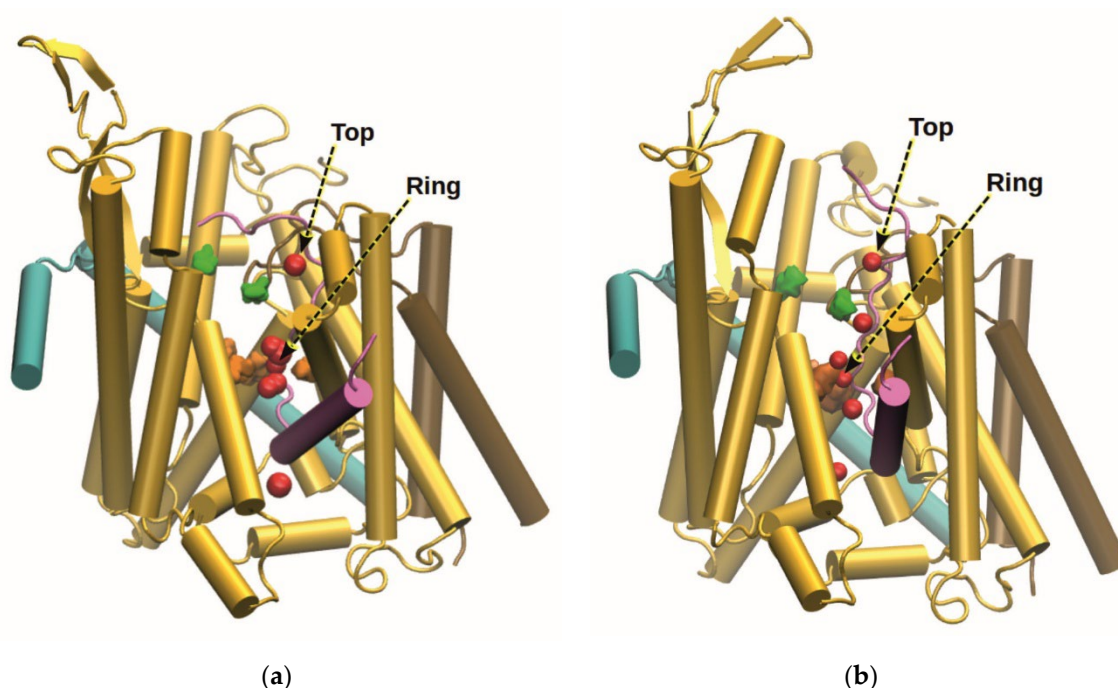


Figure 5. Visualizations of SecYEG with a translocating peptide in the cartoon representation. The protein subunits SecY, E, and G are shown in gold, teal, and brown, respectively, while the peptide and signal sequence are colored magenta. Asp Y410 on SecY and Asp G34 on SecG are displayed in green, and the pore ring is represented by the Ile residues Y81, Y188, Y275, and Y403 on SecY in orange. The red beads correspond to the positions of peptide residue 31 in the frames chosen for titration. In the case of Lys 31 (a), frames 1 and 5 correspond to the "top" and "ring" frames used for simulations with altered protonation forms. In the case of Glu 31 (b), frames 1 and 4 were chosen. Figures made with VMD [36].

2.3.3. Preparation and Titration with Residue 31 being Glutamic Acid

In order to determine the pK_a shifts experienced by a negatively charged residue, Lys 31 in the peptide was changed to Glu using CHARMM [27], before it was again placed into the active state structure of the SecYEG translocon. Once the residue was mutated, energy minimization was performed on the system according to the following protocol: (1) water only: all atoms except water were fixed; (2) membrane and water: protein atoms were fixed; (3) protein, membrane, and water: all N, CA, C, and O atoms were fixed; (4) all: no atoms were fixed. In each step, 500 steps of minimization using the steepest descent (SD, TOLG 0.1) and adopted basis Newton-Raphson (ABNR, TOLG 0.01) algorithms were performed, except for the final ABNR minimization, which was set to 1000 steps. The non-bonded list was regenerated every 20 steps.

Following the energy-minimization, the system was equilibrated with NAMD [28] in six consecutive steps, with decreasing constraints with each run. The simulation parameters are as follows: (1) 1 ns of simulation time, with force constants of 5 and 0.5 kcal/(mol Å²) on the protein backbone and membrane atoms, respectively; (2) 1 ns of simulation time with the force constants lowered to 2 and 0.2 kcal/(mol Å²); (3) 1 ns of simulation time with the force constant lowered to 1 kcal/(mol Å²) for the protein, and everything else is released. For the final three steps, more specialized constraints were placed on the following selections: (i) SecY helices 5, 7, 8, 10, parts of helix 9, and the backbone of SecE, (ii) KCl ions, (iii) the plug domain, (iv) the translocating peptide: (4) 1 ns of simulation time with force constants of 1, 0.2, 1, 0.2 kcal/(mol Å²); (5) 1 ns of simulation time with force constants of 1, 0.1, 1, 0.1 kcal/(mol Å²); (6) 5 ns of simulation time with only the constraints on SecYE in place at 1 kcal/(mol Å²). The constraints could not be removed completely as the system is based on the modified crystal structure of a closed channel – without a docked ribosome or SecA – and so SecY would have closed again if left unconstrained (see [5]).

Once the system had been equilibrated for a total of 10 ns simulation time, the final frame was taken from the trajectory and used as the starting conformation of the constant-velocity SMD simulation for the translocation of the peptide. The parameters used for SMD were the same as described for the Lys-peptide in Section 2.3.2. Snapshots of the system were then taken from the trajectory files at regular intervals. Before titrations were performed on these snapshots (see Figure 5b), the structures were used as starting conformations for 10 ns trajectories as for the Lys-case. From these trajectories, 50 frames were taken from the final 5 ns, edited by adding uncharged water and dummy atoms, and titrated as described below.

2.3.4. Poisson-Boltzmann Computations

PB computations based on the theory described in Sections 2.1 and 2.2 were performed with the solver TABPS developed by Rabenstein and Kieseritzky in the group of Knapp [37,38]. TAPBS is a front-end connected to the Adaptive Poisson-Boltzmann Solver APBS [39]. The LPBE is solved by a finite-difference technique [40] employing a grid-focusing algorithm with cubic grids. The grid spacings were 3, 1, and 0.25 Å for the titratable group in the protein environment (three grids) as well as 1 and 0.25 Å for the model compound (two grids). The ionic strength was set to that of a 0.1 M NaCl solution with an ion radius of 2 Å defining the ion exclusion layer. The solvent radius defining the boundaries between different values of $\epsilon(\mathbf{r})$ (see Section 2.1.) via the solvent-excluded surface (SES) [41–43] was set to 1.4 Å as usually assumed for water. For the PB computations, all lipids constituting the membrane, water molecules, and explicit ions in the aqueous phase were removed, and the protein part was placed in a membrane slab with $\epsilon_{\text{mem}} = 2.0$ as illustrated in Figure 6. We used $\epsilon_{\text{aq}} = 80$ and $\epsilon_{\text{protein}} = 4$. The latter value is a compromise for the protein, which is actually heterogeneous, and lies in a range suggested by MD simulations ($\epsilon = 3 - 6$) [44], theoretical estimates ($\epsilon = 2.5 - 4$) [45] as well as a recent Kramers-Kronig analysis of the spectroscopic properties of bovine serum albumin (BSA; dry powder, $\epsilon = 3.6 - 3.7$) [46].

Based on test titrations, it was found that the surface calculations to obtain $\epsilon(\mathbf{r})$ resulted in "bubble artefacts" of high dielectric permittivity inside the pore. This is due to an in-built feature of TAPBS [37] that automatically assigns ϵ_{aq} to any region inside the protein structure (cavity) able to fit a sphere of radius 1.4 Å. Therefore, in order to keep the dielectric constant inside the transmembrane parts of the protein homogeneous, the structure files were edited to place uncharged atoms inside the pore and block the probe from entering. For this purpose, scripts were written using the python library MDAnalysis [47,48]: From each trajectory file, 50 structures were extracted from the final 5 ns, and water molecules in the channel were selected based on their coordinates. PDB files were created for each frame consisting of the protein coordinates, the combined channel water coordinates from the first 20 frames of the selection, as well as the coordinates of dummy atoms (with a radius of 2 Å) placed at even intervals throughout the trans-membrane region of SecY (Figure 6). The charges of both water and dummy atoms were set to zero (cf. Supplementary Material SM1).

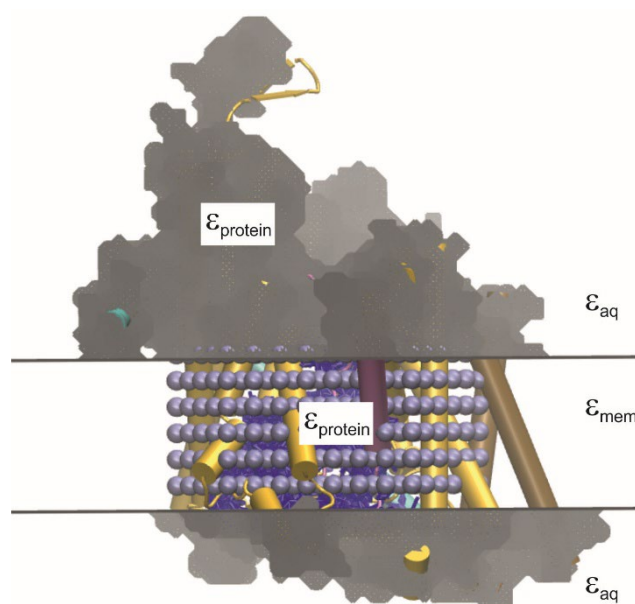


Figure 6. Visualization of the solvent-excluded surface calculated for SecYEG with uncharged dummy atoms (ice blue spheres) and water molecules (dark blue lines) placed in the channel. The grey mesh shows the interface between ϵ_{aq} and $\epsilon_{protein}$, and the horizontal lines the interface between ϵ_{aq} and the membrane slab ($\epsilon_{mem} = 2.0$).

2.3.5. Protonation Forms

PB-based approaches to protonation states in proteins, using a purely electrostatic, continuum dielectric model as described in Section 2.1, were primarily designed to predict pK_a values based on X-ray structures [23,24,49,50]. Since hydrogen atoms could not (and nowadays often still cannot) be resolved in X-ray structures, earlier calculations modeled the protonation/deprotonation by changing APCs on heavy atoms of the titrating site [23,24]. Later refinements employed explicitly represented protons [51] with proton positions inferred from modeling (e. g. with CHARMM [27]). In these models, the number of atoms in the deprotonated form is still the same as in the protonated form, i. e. deprotonation is realized implicitly by changing APCs (see Figure 7). In the present work, all atom positions are inferred from MD simulations, so that all protons are – in principle – explicitly modeled. However, in MD simulations a standard protonation state of each titratable site is usually assumed, implying that acids are deprotonated and negatively charged, and bases are protonated and positively charged. It is possible that these fixed protonation states bias the dynamics because of both the fixed APCs and the prescribed absence or presence of explicit protons.

In order to investigate this bias, we performed variants of the simulations with residue 31 of the peptide in different *protonation forms*. By protonation “form” we mean the explicit absence or presence of a hydrogen atom in the structure (Figure 7). In the following, the protonation forms will be indicated by a three-letter code with capital letters, while the amino acids in general will be referred to by a three-letter code without capitalization. For example, Glu can be in the protonation form GLU with no explicit proton at the carboxyl group, and in the form GLP with one explicit proton. Likewise, Lys can be in the protonation form LYS with three protons at the amino group, or in the form LSN with only two protons. The non-standard forms GLP and LSN were created by employing the CHARMM patches GLUP and LSN, respectively (for APCs, see SM1). Note that each protonation form can still exist in two protonation states determined by the APCs as illustrated in Figure 7. In this way, each form used in the MD simulation can be subjected to a PB-based determination of the protonation probability. As the residue state description (st) files used by TAPBS to calculate the intrinsic pK_a values and interactions of titratable residues do not include LSN or GLP forms, new files were created manually based on LYS and GLU state description files (see SM2 and SM3).

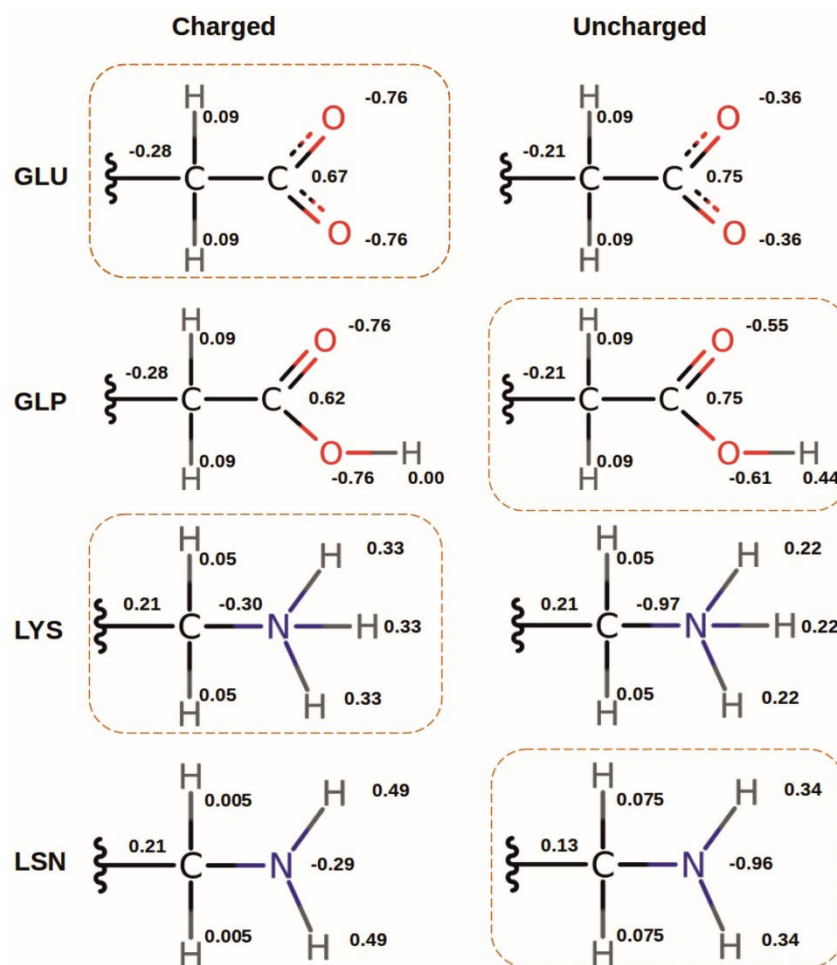


Figure 7. APCs assigned to side chain atoms of Glu and Lys in two different protonation forms. GLU and LYS are the standard forms for simulations (at physiological pH), as they correspond to the structures of the amino acids when charged, which is the most likely state based on their standard pK_a values. For the neutral forms, GLU must be protonated ($\Rightarrow$ GLP) and LYS deprotonated ($\Rightarrow$ LSN). For each protonation form, the APCs are assigned to reflect the charged and uncharged state of the residue for titration calculations. Forms with dashed frames correspond to those used in the MD simulations. Note that in the framed cases, protonation form and protonation state coincide.

Histidine is a special case, as its side chain is usually represented by three protonation states (including two uncharged tautomers [52]) with the corresponding protonation forms HSD, HSE, and HSP in CHARMM. Both protons must be included explicitly (the HSP form) in order to apply the standard titration protocol. The charged version would have to be used for the MD simulations, which would affect the dynamics. Determining the magnitude of this effect would require three separate MD simulations to be done for each of the protonation forms, which was considered too computationally expensive. Since there are only two His residues in the translocon and neither one is in the top half of the pore, we neglected His as a titratable residue in the present work.

2.3.6. Monte-Carlo Titration of Protonation Patterns

The MC titrations were performed with the program Karlsberg2 developed by Rabenstein and Kieseritzky in the Knapp group [37,38]. This program uses the intrinsic pK_a values of each titratable site and the corresponding matrix W_{ij} of charged-site interactions produced by TAPBS (cf. Section 2.3.4). Titrations were performed from pH 0 to 14 in increments of 0.5 pH units at 300 K. The minimum absolute interaction energy to consider double moves for a strongly coupled pair [18,53] was set to 2.5 pH units, and 5 pH units for triple moves. The number of scans was set to 10^5 for MC

scans with all sites, as well as for scans with the reduced set of sites, which was determined by a reduced set tolerance of 10^{-4} . The correlation limit was set to 0.1, and the maximum correlation time to 100 scans.

3. Results

Following an SMD simulation of the translocation, the pK_a shift of peptide residue 31 as a result of its position was investigated by performing continuum electrostatics calculations and MC titrations with the residue at selected positions in the channel. For this purpose, snapshots were taken from the SMD trajectory and used as starting structures for MD simulations with positional constraints on the C_α -atom of the residue. The positions chosen can be seen in Figure 5.

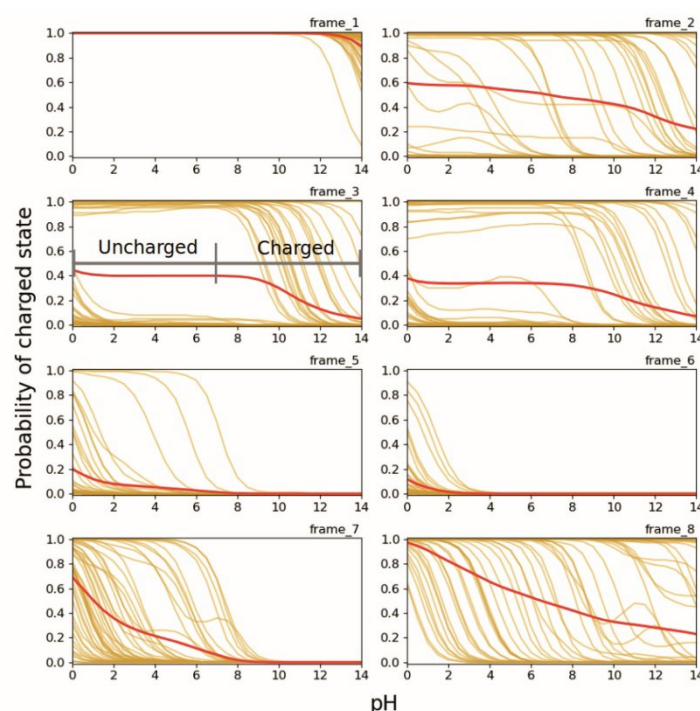


Figure 8. Titration curves of peptide residue LYS 31 over the course of translocation. The curves in frames 1-8 are each based on 50 snapshots from constrained MD simulations with the position of residue 31 determined by constant velocity SMD simulations pulling the peptide through the channel. The frames correspond to the positions indicated in Figure 5a. The golden curves are individual titration curves of the 50 snapshots, while the red lines correspond to the average of the 50 snapshots in each frame to illustrate the average protonation probability (= probability of being positively charged). The "charged" and "uncharged" labels on frame 3 highlight the binary behavior in that frame, which can also be seen in frame 4, as well as, to a lesser extent, in frame 2.

In Figure 8, the titration curves for Lys at position 31 for snapshots taken from MD simulations with Lys 31 in the LYS protonation form (cf. Figure 7) are shown for the eight different positions in the SecYEG channel. Titrations were performed on 50 snapshots for each frame. A clear trend can be seen in the average probability of the charged state decreasing as the residue moves farther into the channel (frames 1 to 5), and increasing again once it has passed the PR (frames 6 to 8). The form of some individual titration curves deviates strongly from the standard Henderson-Hasselbalch (HH) form, indicating strong coupling of Lys 31 to other titratable groups above and below the PR. The titration curves for frames 3 and 4 can be divided into two groups corresponding to fully charged (i. e. protonation probability near 1) or fully uncharged Lys 31 at physiological pH, as indicated for frame 3 in Figure 8. This binary behavior is due to strong interactions with Asp G34 on SecG, which is less than 5 Å away in the frames (measured from the NZ atom of LYS to the CD atom of Asp). This

is confirmed by the fact that in the frame 3 snapshots, Asp G34 shows the same binary titration behavior and protonation state as LYS for the individual frames, meaning when LYS 31 is charged, Asp G34 is charged, too, and when LYS 31 is uncharged, so is Asp G34. The Asp residue is located on the cytosolic loop of SecG at the top of the channel, and is thus free to interact with the peptide in multiple positions. The wide spread of the individual titration curves in other frames also shows the strong dependence of the residue's protonation probability on its orientation, as the C_{α} -atom of the residue is constrained, and so variations are due to thermal fluctuations of the side chain.

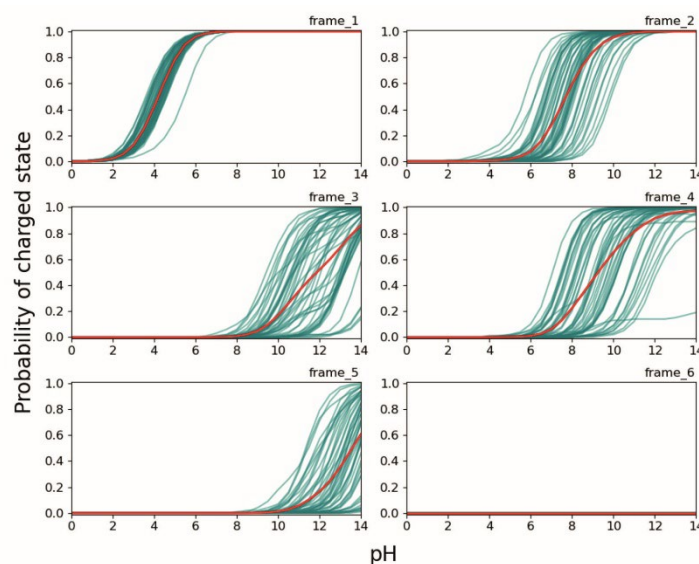


Figure 9. Titration curves of peptide residue GLU 31 over the course of translocation. The curves in frames 1-8 are each based on 50 snapshots from constrained MD simulations, with the position of residue 31 determined by constant velocity SMD simulations pulling the peptide through the channel. The frames correspond to the positions indicated in Figure 5b. The teal curves are individual titration curves of the 50 snapshots, while the red lines correspond to the average of the 50 snapshots in each frame to illustrate the average probability of being negatively charged.

In Figure 9, the titration curves for Glu at position 31 for snapshots taken from MD simulations with Glu 31 in the GLU protonation form (cf. Figure 7) are shown for the six different positions in the SecYEG channel. For the titration curves of the GLU residue (Figure 9), a general transition to the uncharged state can be seen as the residue travels further down the channel, with the exception of the small increase from frame 3 to 4. Unlike LYS 31, GLU 31 does not return to the charged state once it has passed the PR. This may be due to residues Glu 68 and Glu 195 located at the bottom of the channel – which remain almost entirely charged in frames 4-6 – though this was not tested explicitly by fixing their protonation states – as was done for two Asp residues (see below). Additionally, it appears that the protonation state of GLU 31 is less dependent on its orientation, as the individual titration curves have a much smaller spread compared to LYS 31. This observation might be related to the fact that the Glu chain is not only shorter, but also less flexible than that of Lys.

In order to determine the bias in protonation probability introduced by the choice of protonation form (cf. Section 2.3.5) used for the MD simulation, we performed additional simulations with the alternative protonation forms GLP and LSN. These simulations were done only for two frames in each case referred to as “top” (frame 1, where residue 31 is located in the cytoplasmic entrance in the channel) and “ring” (frame 5 for LSN and frame 4 for GLP, where residue 31 is located in the PR). The average titration curves of the 50 snapshots from the top and ring frames were then compared for the two protonation forms. Additionally, the effect of other titratable residues on the pK_a shift of a translocating peptide was investigated by fixing Asp G34 on SecG and Asp Y410 on SecY in different combinations of protonation states (“asp_states” curves) and titrating peptide residue 31. To

determine the effects of positive and negative charges, one run was done with all Asp and Glu residues fixed in their uncharged states ("pos_only" curves), and another fixing Arg and Lys residues in their uncharged states ("neg_only" curves). The net effect of titratable residues was determined by fixing them all in the uncharged state ("none" curves), so that any pK_a shift must be due to the dielectric environment and interaction with non-titrating charges. All titration curve averages are overlaid for comparison in Figure 10.

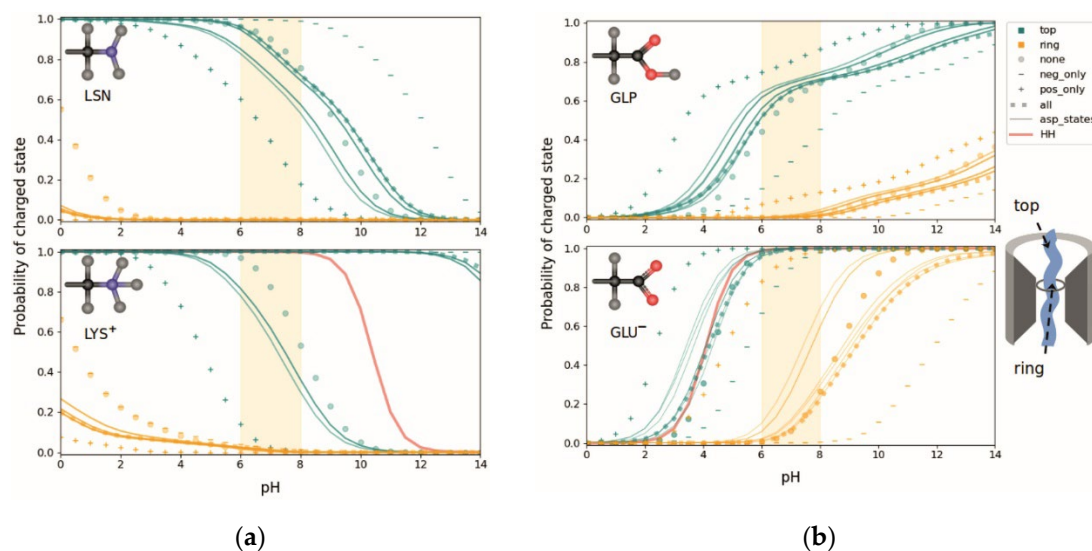


Figure 10. (a) Comparison of average titration curves of LSN (top graph) and LYS (bottom graph) at the top of the channel and in the PR. Teal curves show the probability of the residue being in the charged state while at the top of the channel; orange curves when the residue is in the PR. The rectangle marker denotes the unfixed state ("all"), in which all titratable residues are free to titrate, while the solid curves ("asp_states") are with Asp G34 and Asp Y410 fixed in various combinations of charge states (see SM4, Figures S1 and S2 for labelled graphs of the different charge states). The circle marker corresponds to all titratable residues in the protein being fixed in the uncharged state, while the "+" marker shows all negative residues, and the "-" marker all positive residues fixed in the uncharged state ("none", "pos_only" and "neg_only" curves, respectively). The salmon curve is the Henderson-Hasselbalch (HH) standard curve for the protonation probability of Lys based on a pK_a of 10.4 [54]. The physiological pH range of 6-8 is highlighted in orange. (b) Same as in (a), but for GLP (top) and GLU (bottom; see SM4, Figures S3 and S4 for labelled graphs of the different charge states of Asp G34 and Asp Y410). The HH standard curve corresponds to a pK_a of 4.1 [55].

The averaged titration curves show a largely monotonic behavior, so that an effective pK_a value of the titrating group may be defined by the pH for which the probability of the charged state is 0.5. For both LYS and LSN, a shift of the pK_a to lower values is apparent in the ring frame compared to the top frame (Figure 10a) – meaning a higher probability for residue 31 being uncharged in the PR at physiological pH. In fact, in this pH range, the probability of Lys being charged in the PR is essentially zero. However, when the deprotonated and uncharged form LSN is used in the MD simulation, the pK_a in both the top and the ring frames is lower compared to LYS, indicating a bias towards the uncharged state. Possibly titrations on structures obtained from MD simulations tend to be biased towards the protonation states of the protonation forms chosen.

The largest changes in the top frames is seen when Asp G34 is fixed in the uncharged state, resulting in a pK_a shift from >14 to about 7 for the LYS curves (Figure 10a, bottom), while the corresponding shift in the LSN curves is only about 1 pH unit (Figure 10a, top graph). The change in charge state of Asp Y410 does not result in significant shifts. This finding can be explained by the fact that Asp G34 is spatially very close to residue 31 in the top frame. Therefore, there are strong stabilizing interactions between the anionic Asp and cationic Lys, preventing the latter from giving up its proton. As LSN is uncharged during the MD simulation, there are no such interactions and the charge state of Asp G34 has a smaller effect. The influence of Asp G34 on the charge state of Lys 31 is

thus dependent on the standard charge states of the protonation forms used in the MD simulations, and is likely exaggerated in the LYS frames. Labelled graphs for the different Asp states can be found in SM4, Figures S1 and S2.

Switching off the remaining negative charges ("pos_only") in the top frames lowers the pK_a of LYS by another 4 pH units and LSN by about 2, while switching off all positive charges ("neg_only") has virtually no effect on LYS, but increases the pK_a of LSN by 2 pH units. The combined effect of all titratable residues on LSN is negligible, with the "none" and "all" states resulting in nearly identical curves. In the LYS case, the interaction with Asp G34 dominates, resulting in an ~ 7 pH unit shift between the "all" and "none" states. Any effects on Lys in the PR are negligible as it is almost entirely uncharged independent of any other titratable residues. What is noticeable, however, is that for both LYS and LSN, the "neg_only" state is identical to the "none" state, showing that negatively charged groups have no significant interactions with positively charged peptide residues when in the PR.

Similar to Lys, Glu experiences a shift towards the uncharged state as it moves through the pore (Figure 10b), as well as a difference in the protonation probabilities between the deprotonated and protonated form used for the MD simulation. Again, the frames using the protonated form (GLP) show a bias towards the protonated state, seen by an increase in pK_a . Additionally, the GLP residue's protonation probability seems to be more dependent on its orientation than that of GLU, as is evident from the "double-humped" shape of the average titration curves resulting from a large spread in the curves of individual snapshots (not shown for GLP). In all cases, though, the "all" and "none" curves are very similar, meaning that the combined effect of titratable residues on the pK_a of Glu is very small. The presence of only positive or negative charges would result in a significant shift, as is shown by the curves in the "pos_only" and "neg_only" states. Apparently, the effects of positively and negatively charged groups cancel each other. This is an interesting finding as there are 1.7 times more basic amino acids in SecYEG than acidic ones. The position of Glu in the channel allows for a more balanced interaction with basic and acidic groups, including a strong influence of Asp G34. Labelled graphs for the different Asp states can be found in SM4, Figures S3 and S4. Overall, the pK_a of Glu, like that of Lys, appears to be dictated by the dielectric properties of its environment, interaction with non-titrating charges, and its distance from the dielectric interface – rather than interactions with ionizable residues in the channel.

We have also tested the effect of the precise spatial dependence of $\epsilon(\mathbf{r})$, i. e. the position of the interface between high and low dielectric constant in the channel (Figure 11). This was done only for the ring frames, where residue 31 is in the PR. For the first test (Figure 11a), where neither a membrane slab nor dummy atoms, but an explicit membrane was used, the SES probe (sphere of radius 1.4 Å; cf. section 2.3.4.) was able to travel all the way through the pore – causing the residue to be surrounded by a medium of $\epsilon = 80$. In the test visualized in Figure 11b, there are also no dummy atoms, but an implicit membrane in the form of a slab with $\epsilon = 2$ replacing the explicit lipids bilayer (cf. Figure 6). This resulted in the positioning of a 2/80 interface almost at the position of the test residue. Additionally, due to the algorithms in TAPBS, the low permittivity area of the membrane slab was also defined in the pore region. In both of these setups, the test residue was in contact with the $\epsilon = 80$ interface and thus remained charged. The final panel (Figure 4.6 (c)) shows the setup generally used in this project: the same implicit membrane as in (b) but with the addition of uncharged dummy atoms in the channel to impede the entry of the SES probe. This procedure results in the entire protein, including the PR area, being assigned $\epsilon = 4$, and holds the $\epsilon = 80$ interface off from protruding almost to the PR. This was the only case in which the test residue changed its protonation state. Thus, if the channel water is modelled as having bulk properties, the peptide residue 31 is not neutralized.

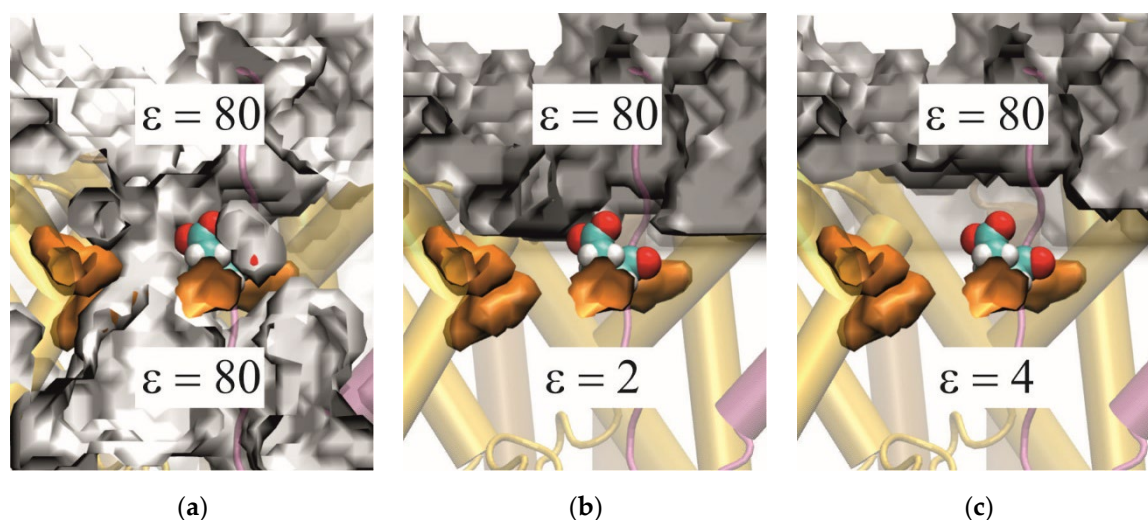


Figure 11. Comparison of $\epsilon(\mathbf{r})$ interfaces (silver iso-surfaces) of GLU ring frames (position of GLU 31 near the PR as in frame 4 of Figure 5b) for (a) an explicit membrane, (b) an implicit membrane, and (c) an implicit membrane plus dummy atoms. The pore ring is represented by four Ile residues in orange, and the GLU residue is shown in a van der Waals representation. The implicit membrane is set up as a slab of low dielectric permittivity ($\epsilon = 2$) constrained to certain z -coordinates (cf. Figure 6). Due to the algorithms in TAPBS, this slab is able to "leak" into the pore and cause an area with $\epsilon = 2$ inside the channel. (c) Placing evenly spaced, uncharged dummy atoms inside the protein blocks the SES probe from entering and allows for assigning $\epsilon = 4$ to the channel interior.

4. Discussion

The goal of the present work was to determine the probability of being charged of a test residue in a translocating peptide as it travels through the SecYEG translocon. The simulations clearly suggest that both, Lys and Glu as test residues, are charge-neutralized when they pass the PR, i.e. the constriction in the hydrophobic interior of the translocon. This result argues against a discrimination of positively and negatively charged residues that would suggest the electric component of the PMF directly contributes to the driving force behind peptide translocation in neutralophilic bacteria as previously hypothesized [11]. However, the simulation methods involve a number of approximations that must be discussed, and the results should be confronted with recent experimental studies on the translocation of charged residues [13,56].

PB-based electrostatic methods to determine pK_a values are usually applied to X-ray structures of proteins – and probably work best in this context. In the case of active SecYEG, we had neither X-ray nor cryo-EM structures of the translocon in different stages of the translocation process with a sufficiently high resolution at our disposal. So we decided to create structures of the translocating peptide by employing MD simulations. These simulations were built upon an open *T. thermophilus* structure that was constructed in previous work [5,26] from the closed channel [25] by taking into account information about the opening from an active-state chimera structure [8]. The different stages of translocation were created using an SMD protocol [35] and the test residue fixed at its backbone by a harmonic constraint. One possible way to proceed from there would have been an energy minimization of the respective stages of translocation and PB computations based upon these minimized structures. However, due to the flexibility of the test residue's side chain, it probably would have been difficult to find a single, representative conformation. Therefore, we decided to perform MD simulations to create an ensemble of structures and to perform PB/MC titrations on each structure. It should be noted that the translocation process is slow (in the range of milliseconds per residue), so that the transported peptide can in fact sample a large number of conformations in each stage of the translocation.

Since the used snapshots were taken randomly (i.e. with equal weight) from the second half of a 10-ns *NPT* simulation, the structural ensembles may be considered as representing roughly

isothermal-isobaric ensembles. Then, the ensemble-averaged protonation probabilities are obtained by just averaging the individual titration curves as shown in Figures 8 and 9. Clearly, the results may be improved by increasing the number of snapshots and the length of the *NPT* trajectories. There is always a compromise between accuracy and simulation time. A problem that cannot be solved within this framework is the bias of the MD simulations due to the explicit protonation form of the titratable residue used in these simulations. Note that any standard MD simulation is limited by this problem. Therefore, it is advisable to perform tests with different protonation forms to assess errors. In the present case, we obtain the same basic result for each protonation form: The probability of the test residue being charged is close to zero when the residue is in the PR.

A critical – and much debated [21,45,57] – aspect of PB computations is the choice of the dielectric constant of the protein. We shall not discuss here the applicability of the macroscopic concepts of medium electrodynamics to the length scale of biomolecules, as this is – to our knowledge – an unsolved fundamental problem. Instead, we work with the widely accepted presumption that PB-based electrostatics provides a meaningful physical model of proteins. The next problem is that ϵ is actually the real part of a complex function that, for analytical purposes, can be extended to the complex plane [46,58,59]. The value that we use in the computations is the zero-frequency limit, i.e. the static dielectric constant. To understand why it can be used in computations on MD snapshots, one has to note that our MD simulations do not actually serve to study dynamics, but to provide a sampling of structural variations. Each snapshot then corresponds to a static structure, and we are left with two questions: (i) What is the appropriate value of $\epsilon_{\text{protein}}$? And: (ii) What about the water in the channel?

Ad (i): As discussed by Gilson and Honig [45], dipolar groups in a protein play a dual role: They act as sources of the protein's permanent electrostatic field (represented here by APCs) and also, to the extent that they are free to rotate, contribute to enhance the protein's dielectric constant over that of electronic polarization ($\epsilon_{\text{protein}} > \epsilon_{\infty}$). When we compute on a given snapshot, the APCs represent the permanent electrostatic field in that particular conformation. However, we also have to take into account possible reorientations of polar groups that may take place on the timescale of the equilibration of de/protonation reactions starting from that protein conformation. We think that the moderately high value of $\epsilon_{\text{protein}} = 4$ corresponding to protein powders with a low water content [45,46] is suitable for this purpose, particularly for the inner region of a membrane protein.

Ad (ii): The channel in SecYEG, through which the peptide is translocated, contains a lot of water even in the presence of the peptide. Clearly, for a PB/MC computation, the explicit water has to be removed to avoid biasing of the MC sampling; the water must be represented implicitly by a dielectric continuum, and the question arises as to what value of ϵ is appropriate here. The algorithms in TABPS assign to each cavity, which is able to accommodate a sphere of radius 1.4 Å (corresponding to a water molecule) a value of $\epsilon_{\text{aq}} = 80$. Such an assignment is also made to a significant part of the SecYEG channel so that the test residue 31 still is in contact with a highly polarizable environment when being positioned at the PR (Figure 11a). From a molecular point of view, the high dielectric constant means that the water molecules in this region are free to rotate, so that they basically behave like bulk water. However, it is questionable that water in the channel behaves like that.

Earlier numerical simulations suggest that confined water should have a significantly smaller dielectric constant than bulk water [60–62]. This prediction was confirmed experimentally by Fumagalli et al. [63] for liquid water within nanochannels. Seyedi and Matyushov [64] computed the dielectric constant of the hydration shell of cytochrome *c* based on MD simulations and found a value of $\epsilon \approx 4$ at room temperature. Therefore, we think that assigning such a low dielectric constant to the whole interior of SecYEG, including the water-filled regions, is physically realistic. This assumption is crucial, as we found that only with such a low dielectric constant, it is possible to achieve charge neutralization of the test residue. So, we conclude that the low dipolar susceptibility of water confined in the channel during the translocation process critically contributes to the preference of the titratable residues Lys and Glu being transported in the uncharged state. This

behavior simply reflects the energetic cost of transferring a charge from an aqueous environment to a less polar one, which according to the approximation provided by Born's formula for ions

$$\Delta G_{\text{transfer}} = -\frac{N_A q^2}{8\pi\epsilon_0 r_0} \left(\frac{1}{\epsilon_{\text{aq}}} - \frac{1}{\epsilon_{\text{protein}}} \right), \quad (18)$$

where r_0 is the radius of the ion, does not depend on the sign of the charge q . The same is true for the more elaborate computations based on the LPBE. Thus, if the dielectric properties of the interior of SecYEG determine the protonation state changes, no discrimination between acids and bases is to be expected. The only exception could be arginine (Arg) due to its high reference pK_a (see below).

Previous MD simulations [65], performed on the basis of a SecYE structure from the hyperthermophile archaeon *Pyrococcus furiosus* [66], revealed an anomalous behavior of water inside the channel. The rotational relaxation time of water molecules (defined as the time required for the dipole autocorrelation function to decay to $1/e$), which is 1 ps in bulk water, was found to increase to 20 – 30 ps inside SecY and even above the observation time near the center of SecY. It should be noted that the structure used for these simulations is highly distorted and it is unclear, which state of the translocon it represents. Nonetheless, the simulations can be taken as a hint that the rotational dynamics of channel-bound water is retarded, which may be related to a decreased dielectric constant. Another important result from these simulations is that the water molecules are highly oriented inside SecYE. This finding, together with the retarded reorientation, hints at a problem: Oriented water molecules can stabilize the charged state of a titratable residue by virtue of their dipole. Such a behavior is not covered by a dielectric continuum model with a low dielectric constant. Rather, one would have to put explicit water molecules with non-zero APCs in the model. However, explicit water molecules are prohibitive in PB/MC titrations as they bias the MC sampling as mentioned above. A solution to this problem would be a constant-pH MD simulation [67] employing a non-equilibrium MD/MC approach [68] or λ -dynamics [69]. However, these methods are expensive, are outside the scope of the present project, and have their own problems [67]. Nevertheless, our conclusion that Lys and Glu are charge-neutralized during translocation, has to be drawn subject to the proviso that our continuum model mimics the behavior of channel water well enough.

In their extensive study on the rate-limiting transport of positively charged residues through the Sec-machinery, Allen et al. [56] showed that translocation is retarded by Arg residues in the peptide, while no such retardation is observed for Lys. Allen et al. explain the different behavior of Lys by a deprotonation, which is principally in agreement with our findings. To substantiate their proposal, Allen et al. [56] performed an in silico pK_a analysis based on MD snapshots (by successively substituting each peptide residue with Lys), but they employed the PROPKA3.1 program [70]. While the pK_a computation in PROPKA is structure-based, it is rather empirical than physics-based. In particular, the long-range electrostatic interaction of a titratable group with background charges is computed only within a radius of 6 Å, and coupled sites are not treated with MC, but by employing empirical rules [70,71]. Also, it remains unclear how the channel water is included in the computations performed by Allen et al. [56]. Another problem is that for each position in the channel, a different peptide was investigated. Remarkably, though, Allen et al. found a pK_a decrease for Lys inside the channel down to about 6.5. While this pK_a shift is probably not sufficient to ensure deprotonation at physiological pH, it qualitatively points in the same direction as our simulations.

We do not know how Arg at position 31 in the peptide would behave under our simulation conditions. On one hand, it has a high reference pK_a value of 13.8 [72], and Arg is generally considered to stay protonated even when buried inside the protein [73]. On the other hand, it could be deprotonated if it experiences a pK_a shift of the same order of magnitude as Lys in our simulations. In our opinion, the significance of positively charged residues in a translocating peptide for barriers to transport is not clear. Allen et al. [56] found a major barrier in the free energy profile for translocation (based on the coarse-grained MARTINI 2.2 force field [74,75]) that is dominated by the positively charged Lys residue at the PR and significantly reduced by de-charging the Lys. In our previous simulation of the Lys-peptide [5] (based on the all-atom CHARMM36m force field), we found several barriers that are rather determined by bulky amino acids than a single charged residue

– although we cannot exclude that de-charging Lys would alter the free energy profile. We believe that the role of net charges for the translocation barrier is overestimated. The different behavior of Arg compared to Lys could also result from the bulkiness of the guanidino group. Further simulations are necessary to clarify this point.

In line with this argumentation is the finding of Allen et al. [56] that the stimulation of translocation by the PMF does not significantly depend on the (presumed) net charge of the peptide. On the other hand, there is a stimulation of translocation by the PMF. We have shown in our previous work [5] that the transmembrane potential, which is part of the PMF, affects the orientation of transmembrane α -helices and, in this way, can contribute to an opening of the channel. Thus, the effect of the PMF is more complicated than simply dragging negatively charged residues through the PR, meaning that it is not required that a Glu residue in the peptide remains charged.

Ismail et al. [13] found evidence of a force acting on a translocating peptide containing an Asp segment (5D stretch), which they ascribed to the transmembrane potential – based on experiments with the uncoupler indole. According to our understanding of these results, the force is exerted on the 5D stretch while it is located in the cytoplasmic vestibule of the translocon. Based on our simulations, we speculate that the 5D stretch is still highly charged in this position, although not necessarily fully charged because of the mutual electrostatic interaction of the five Asp residues. It remains to be clarified, how these charges – in conjunction with the transmembrane potential – facilitate translocation of the 5D stretch.

In the present work, we concentrated on the course of the translocating residue from the cytoplasmic vestibule of SecYEG to the PR. However, some frames also cover the periplasmic region below the PR (Figures 5, 8, and 9). Here, we made the interesting observation that Lys and Glu behave differently: Lys is recharged, while Glu stays uncharged. In the light of a simple model of how the transmembrane potential acts on charged residues, this finding is counterintuitive, since a positive charge at the periplasmic side is not favored. Further work is required to clarify the titration behavior of the peptide at this side of the membrane and the role of the PMF in the later stages of the translocation.

5. Conclusions

Based on calculations performed on a titratable residue in a peptide during translocation in SecYEG, it appears that the farther into the channel a residue travels, the less likely it is to be found in the charged state. This was shown for Lys and Glu. Thus, the charge neutralization is independent of the sign of the charged state. The channel does not stabilize one charge over the other, which is expected as the low dielectric permittivity within the channel plays a decisive role. We expect Asp residues to behave like Glu under these conditions, but we cannot say yet, whether Arg would be charge neutralized, too. It was also shown that the protonation form of the residue used in the MD simulation biases the protonation probability significantly, although this artifact had no influence on our basic result. This result is visualized in Figure 12 with reference to the hypothesis depicted in Figure 2. Since there is no discrimination between acids and bases regarding charge neutralization, we conclude that the transmembrane potential plays no direct role in pulling negatively charged residues through the PR, while positively charged residues need to be deprotonated. The variant behavior observed for Arg [56] could be due to a retained charge or steric effects. We do not exclude the possibility that the transmembrane potential exerts a force on a charged residue while being located in the cytoplasmic vestibule of the translocon. Details of how the PMF facilitates translocation under these conditions and the role it plays at the periplasmic side remain to be clarified.

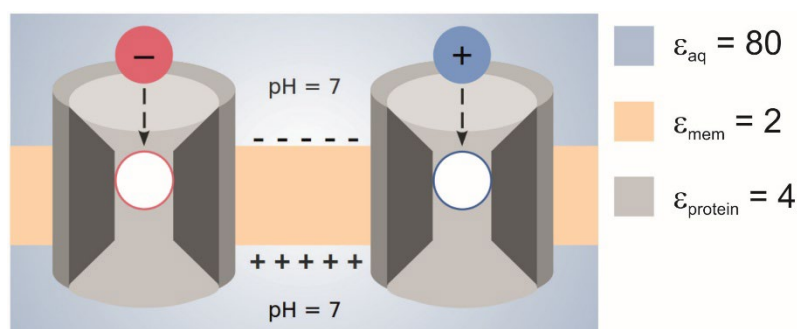


Figure 12. Visual summary of the titration behavior of acids and bases as determined in this work. The blue and red circles represent the charged states of acidic and basic amino acids, while the white circles show the uncharged states. The magnitude of the dielectric constants used in the calculations, as well as the areas in which they were defined, are shown in the legend. The figure is based on the representation used in [11].

The calculations done in this project are based on continuum electrostatics models. The simplicity, low computational cost, and rigorous testing over the years make these models an important tool in biophysics. However, the PB approach faces difficulties in describing the behavior of protein- or channel-bound water. In addition, it remains a challenge to directly include the protein's structural flexibility in the pK_a calculations. Therefore, since SecYEG is a membrane bound protein, as well as containing confined water with distinctly non-bulk properties in its pore, continuum electrostatics based models may be too blunt of a tool to accurately determine titration behavior in a system of this complexity. Simulation techniques allowing for protonation changes to be taken into account during MD simulations might be able to confirm the results obtained, or shed light on effects that hydration shells and dynamic protein flexibility may have on the translocation substrate's pK_a values.

In summary, the results obtained in this project show a clear transition of titratable residues from the charged to the uncharged state as a function of their position in the channel, independent of the sign of the charge. They are based on continuum electrostatics calculations performed on snapshots of MD simulations, and are thus based on the assumptions that (i) the interior of the protein, including the channel water, has a low dielectric permittivity, and (ii) stabilizing effects of oriented water molecules/hydration shells around the residues are not strong enough to compensate for the more hydrophobic environment. The complexity of channel proteins as a system, though, warrants the use of more elaborate models, and constant-pH MD simulations are a possible next step in investigating the titration behavior of translocating peptides.

Supplementary Materials: The following supporting information can be downloaded at: Preprints.org, SM1: PQR parameters for non-standard molecules; SM2: State description of GLP; SM3: State description of LSN; SM4: Titration curves (Figures S1 to S4).

Author Contributions: Conceptualization and methodology, E.S. and F.M.; formal analysis and investigation, L.N.; writing—original draft preparation, L.N. and F.M.; writing—review and editing, L.N., E.S. and F.M.; supervision, project administration, and funding acquisition, F.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Austrian Science Fund (FWF) in conjunction with the federal state of Upper Austria, project P 33154-B.

Data Availability Statement: No new experimental data were created in this study. All relevant theoretical data are contained in the article or in the Supplementary Material.

Acknowledgments: This work was supported by hardware grants (PI Kramer, JKU Linz) from NVIDIA and utilized NVIDIA A100 GPUs.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Yang, N. J.; Hinner, M. J., Getting Across the Cell Membrane: An Overview for Small Molecules, Peptides, and Proteins. In *Site-Specific Protein Labeling: Methods and Protocols*, Gautier, A.; Hinner, M. J., Eds. Humana Press: New York, 2015; pp 29-53.
2. Oswald, J.; Njenga, R.; Natriashvili, A.; Sarmah, P.; Koch, H. G., The Dynamic SecYEG Translocon. *Front. Mol. Biosci.* **2021**, *8*, 664241. <https://doi.org/10.3389/fmolb.2021.664241>.
3. Lycklama a Nijeholt, J. A.; Driessen, A. J. M., The bacterial Sec-translocase: structure and mechanism. *Phil. Trans. R. Soc. B* **2012**, *367*, 1016-1028. <https://doi.org/10.1098/rstb.2011.0201>.
4. Rapoport, T. A.; Li, L.; Park, E., Structural and Mechanistic Insights into Protein Translocation. *Annu. Rev. Cell Dev. Biol.* **2017**, *33*, 369-390. <https://doi.org/10.1146/annurev-cellbio-100616-060439>.
5. Sobakinskaya, E.; Müh, F., Physical mechanisms of the Sec machinery operation. *Phys. Chem. Chem. Phys.* **2024**, *26*, 27176-27188. <https://doi.org/10.1039/d4cp03201b>.
6. Tsirigotaki, A.; De Geyter, J.; Sostaric, N.; Economou, A.; Karamanou, S., Protein export through the bacterial Sec pathway. *Nat. Rev. Microbiol.* **2017**, *15*, 21-36. <https://doi.org/10.1038/nrmicro.2016.161>.
7. Tam, P. C. K.; Maillard, A. P.; Chan, K. K. Y.; Duong, F., Investigating the SecY plug movement at the SecYEG translocation channel. *EMBO J.* **2005**, *24*, 3380-3388. <https://doi.org/10.1038/sj.emboj.7600804>.
8. Li, L.; Park, E.; Ling, J.; Ingram, J.; Ploegh, H.; Rapoport, T. A., Crystal structure of a substrate-engaged SecY protein-translocation channel. *Nature* **2016**, *531*, 395-399. <https://doi.org/10.1038/nature17163>.
9. Ismail, N.; Hedman, R.; Schiller, N.; von Heijne, G., A biphasic pulling force acts on transmembrane helices during translocon-mediated membrane integration. *Nat. Struct. Mol. Biol.* **2012**, *19*, 1018-U1068. <https://doi.org/10.1038/nsmb.2376>.
10. Collinson, I.; Corey, R. A.; Allen, W. J., Channel crossing: how are proteins shipped across the bacterial plasma membrane? *Phil. Trans. R. Soc. B* **2015**, *370*, 20150025. <https://doi.org/10.1098/Rstb.2015.0025>.
11. Knyazev, D. G.; Kuttner, R.; Zimmermann, M.; Sobakinskaya, E.; Pohl, P., Driving Forces of Translocation Through Bacterial Translocon SecYEG. *J. Membr. Biol.* **2018**, *251*, 329-343. <https://doi.org/10.1007/s00232-017-0012-9>.
12. Shioi, J. I.; Matsuura, S.; Imae, Y., Quantitative measurements of proton motive force and motility in *Bacillus subtilis*. *J. Bacteriol.* **1980**, *144*, 891-897. <https://doi.org/10.1128/jb.144.3.891-897.1980>.
13. Ismail, N.; Hedman, R.; Lindén, M.; von Heijne, G., Charge-driven dynamics of nascent-chain movement through the SecYEG translocon. *Nat. Struct. Mol. Biol.* **2015**, *22*, 145-149. <https://doi.org/10.1038/nsmb.2940>.
14. Lam, A. K. M.; Galione, A., The endoplasmic reticulum and junctional membrane communication during calcium signaling. *Biochim. Biophys. Acta Mol. Cell Res.* **2013**, *1833*, 2542-2559. <https://doi.org/10.1016/j.bbamcr.2013.06.004>.
15. Alexov, E.; Mehler, E. L.; Baker, N.; Baptista, A. M.; Huang, Y.; Milletti, F.; Nielsen, J. E.; Farrell, D.; Carstensen, T.; Olsson, M. H. M.; Shen, J. K.; Warwicker, J.; Williams, S.; Word, J. M., Progress in the prediction of pK_a values in proteins. *Proteins* **2011**, *79*, 3260-3275. <https://doi.org/10.1002/prot.23189>.
16. Navo, C. D.; Jiménez-Osés, G., Computer Prediction of pK_a Values in Small Molecules and Proteins. *ACS Med. Chem. Lett.* **2021**, *12*, 1624-1628. <https://doi.org/10.1021/acsmedchemlett.1c00435>.
17. Bashford, D., Macroscopic electrostatic models for protonation states in proteins. *Front. Biosci. (Landmark Ed.)* **2004**, *9*, 1082-1099. <https://doi.org/10.2741/1187>.
18. Ullmann, G. M.; Knapp, E. W., Electrostatic models for computing protonation and redox equilibria in proteins. *Eur. Biophys. J.* **1999**, *28*, 533-551. <https://doi.org/10.1007/s002490050236>.
19. Renger, T.; Müh, F., Understanding photosynthetic light-harvesting: a bottom up theoretical approach. *Phys. Chem. Chem. Phys.* **2013**, *15*, 3348-3371. <https://doi.org/10.1039/c3cp43439g>.
20. Gunner, M. R.; Baker, N. A., Continuum Electrostatics Approaches to Calculating pK_as and E_{ms} in Proteins. *Methods Enzymol.* **2016**, *578*, 1-20. <https://doi.org/10.1016/bs.mie.2016.05.052>.
21. Schutz, C. N.; Warshel, A., What are the dielectric "constants" of proteins and how to validate electrostatic models? *Proteins* **2001**, *44*, 400-417. <https://doi.org/10.1002/Prot.1106>.
22. Warshel, A., Calculations of Enzymatic-Reactions - Calculations of pK_a, Proton-Transfer Reactions, and General Acid Catalysis Reactions in Enzymes. *Biochemistry* **1981**, *20*, 3167-3177. <https://doi.org/10.1021/Bi00514a028>.

23. Bashford, D.; Karplus, M., pK_a's of Ionizable Groups in Proteins: Atomic Detail from a Continuum Electrostatic Model. *Biochemistry* **1990**, *29*, 10219-10225. <https://doi.org/10.1021/Bi00496a010>.
24. Schaefer, M.; Van Vlijmen, H. W. T.; Karplus, M., Electrostatic contributions to molecular free energies in solution. *Adv. Prot. Chem.* **1998**, *51*, 1-57. [https://doi.org/10.1016/s0065-3233\(08\)60650-6](https://doi.org/10.1016/s0065-3233(08)60650-6)
25. Tanaka, Y.; Sugano, Y.; Takemoto, M.; Mori, T.; Furukawa, A.; Kusakizako, T.; Kumazaki, K.; Kashima, A.; Ishitani, R.; Sugita, Y.; Nureki, O.; Tsukazaki, T., Crystal Structures of SecYEG in Lipidic Cubic Phase Elucidate a Precise Resting and a Peptide-Bound State. *Cell Rep.* **2015**, *13*, 1561-1568. <https://doi.org/10.1016/j.celrep.2015.10.025>.
26. Sobakinskaya, E.; Krobath, H.; Renger, T.; Müh, F., Structural determinants of a permeation barrier of the SecYEG translocon in the active state. *Phys. Chem. Chem. Phys.* **2021**, *23*, 25830-25840. <https://doi.org/10.1039/d1cp02702f>.
27. Hwang, W.; Austin, S. L.; Blondel, A.; Boittier, E. D.; Boresch, S.; Buck, M.; Buckner, J.; Caflisch, A.; Chang, H. T.; Cheng, X.; Choi, Y. K.; Chu, J. W.; Crowley, M. F.; Cui, Q.; Damjanovic, A.; Deng, Y. Q.; Devereux, M.; Ding, X. Q.; Feig, M. F.; Gao, J. L.; Glowacki, D. R.; Gonzales, J. E.; Hamaneh, M. B.; Harder, E. D.; Hayes, R. L.; Huang, J.; Huang, Y. D.; Hudson, P. S.; Im, W.; Islam, S. M.; Jiang, W.; Jones, M. R.; Käser, S.; Kearns, F. L.; Kern, N. R.; Klauda, J. B.; Lazaridis, T.; Lee, J.; Lemkul, J. A.; Liu, X. R.; Luo, Y.; MacKerell, A. J. r.; Major, D. T.; Meuwly, M.; Nam, K.; Nilsson, L.; Ovchinnikov, V.; Paci, E.; Park, S.; Pastor, R. W.; Pittman, A. R.; Post, C. B.; Prasad, S.; Pu, J. Z.; Qi, Y. F.; Rathinavelan, T.; Roe, D. R.; Roux, B.; Rowley, C. N.; Shen, J. A.; Simonett, A. C.; Sodt, A. J.; Töpfer, K.; Upadhyay, M.; van der Vaart, A.; Vazquez-Salazar, L. I.; Venable, R. M.; Warrensford, L. C.; Woodcock, H. L.; Wu, Y. J.; Brooks, C. L.; Brooks, B. I. I.; Karplus, M., CHARMM at 45: Enhancements in Accessibility, Functionality, and Speed. *J. Phys. Chem. B* **2024**, *128*, 9976-10042. <https://doi.org/10.1021/acs.jpcc.4c04100>.
28. Phillips, J. C.; Hardy, D. J.; Maia, J. D. C.; Stone, J. E.; Ribeiro, J. V.; Bernardi, R. C.; Buch, R.; Fiorin, G.; Henin, J.; Jiang, W.; McGreevy, R.; Melo, M. C. R.; Radak, B. K.; Skeel, R. D.; Singharoy, A.; Wang, Y.; Roux, B.; Aksimentiev, A.; Luthey-Schulten, Z.; Kale, L. V.; Schulten, K.; Chipot, C.; Tajkhorshid, E., Scalable molecular dynamics on CPU and GPU architectures with NAMD. *J. Chem. Phys.* **2020**, *153*, 044130. <https://doi.org/10.1063/5.0014475>.
29. Lee, J.; Cheng, X.; Swails, J. M.; Yeom, M. S.; Eastman, P. K.; Lemkul, J. A.; Wei, S.; Buckner, J.; Jeong, J. C.; Qi, Y.; Jo, S.; Pande, V. S.; Case, D. A.; Brooks, C. L., 3rd; MacKerell, A. D., Jr.; Klauda, J. B.; Im, W., CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations Using the CHARMM36 Additive Force Field. *J. Chem. Theory Comput.* **2016**, *12*, 405-413. <https://doi.org/10.1021/acs.jctc.5b00935>.
30. Zhu, X.; Lopes, P. E.; Mackerell, A. D., Jr., Recent Developments and Applications of the CHARMM force fields. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **2012**, *2*, 167-185. <https://doi.org/10.1002/wcms.74>.
31. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926-935. <https://doi.org/10.1063/1.445869>.
32. Schlick, T., *Molecular Modeling and Simulation: An Interdisciplinary Guide*. Springer: New York, 2010.
33. Martyna, G. J.; Tobias, D. J.; Klein, M. L., Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys.* **1994**, *101*, 4177-4189. <https://doi.org/10.1063/1.467468>.
34. Feller, S. E.; Zhang, Y. H.; Pastor, R. W.; Brooks, B. R., Constant Pressure Molecular Dynamics Simulation: The Langevin Piston Method. *J. Chem. Phys.* **1995**, *103*, 4613-4621. <https://doi.org/10.1063/1.470648>.
35. Lu, H.; Schulten, K., Steered molecular dynamics simulations of force-induced protein domain unfolding. *Proteins* **1999**, *35*, 453-463. [https://doi.org/10.1002/\(Sici\)1097-0134\(19990601\)35:4<453::Aid-Prot9>3.0.Co;2-M](https://doi.org/10.1002/(Sici)1097-0134(19990601)35:4<453::Aid-Prot9>3.0.Co;2-M).
36. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J. Mol. Graphics* **1996**, *14*, 33-38, 27-38. [10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5).
37. Kieseritzky, G.; Knapp, E. W., Optimizing pK_a computation in proteins with pH adapted conformations. *Proteins* **2008**, *71*, 1335-1348. <https://doi.org/10.1002/prot.21820>.
38. Rabenstein, B.; Knapp, E. W., Calculated pH-dependent population and protonation of carbon-monoxymyoglobin conformers. *Biophys. J.* **2001**, *80*, 1141-1150. [https://doi.org/10.1016/S0006-3495\(01\)76091-2](https://doi.org/10.1016/S0006-3495(01)76091-2).

39. Baker, N. A.; Sept, D.; Joseph, S.; Holst, M. J.; McCammon, J. A., Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98*, 10037-10041. <https://doi.org/10.1073/pnas.181342398>.
40. Klapper, I.; Hagstrom, R.; Fine, R.; Sharp, K.; Honig, B., Focusing of electric fields in the active site of Cu-Zn superoxide dismutase: effects of ionic strength and amino-acid modification. *Proteins* **1986**, *1*, 47-59. <https://doi.org/10.1002/prot.340010109>.
41. Connolly, M. L., Analytical Molecular-Surface Calculation. *J. Appl. Cryst.* **1983**, *16*, 548-558. <https://doi.org/10.1107/S0021889883010985>.
42. Sanner, M. F.; Olson, A. J.; Spehner, J. C., Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers* **1996**, *38*, 305-320. [https://doi.org/10.1002/\(Sici\)1097-0282\(199603\)38:3<305::Aid-Bip4>3.3.Co;2-8](https://doi.org/10.1002/(Sici)1097-0282(199603)38:3<305::Aid-Bip4>3.3.Co;2-8).
43. Müh, F., The Entropy of Mixing in Self-Assembly and the Role of Surface Tension in Modeling the Critical Micelle Concentration. *Colloids Interfaces* **2024**, *8*, 60. <https://doi.org/10.3390/Colloids8060060>.
44. Simonson, T.; Brooks, C. L., Charge screening and the dielectric constant of proteins: Insights from molecular dynamics. *J. Am. Chem. Soc.* **1996**, *118*, 8452-8458. <https://doi.org/10.1021/Ja960884f>.
45. Gilson, M. K.; Honig, B. H., The Dielectric Constant of a Folded Protein. *Biopolymers* **1986**, *25*, 2097-2119. <https://doi.org/10.1002/bip.360251106>.
46. Nakov, S.; Sobakinskaya, E.; Müh, F., Novel method to analyse and reconstruct optical constants of biological substances with application of Kramers-Kronig relations. *Measurement* **2025**, *242*, 115925. <https://doi.org/10.1016/j.measurement.2024.115925>.
47. Gowers, R. J.; Linke, M.; Barnoud, J.; Reddy, T. J. E.; Melo, M. N.; Seyler, S. L.; Domanski, J.; Dotson, D. L.; Buchoux, S.; Kenney, I. M.; Beckstein, O., MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations. *Proc. of the 15th Python in Science Conf.* **2016**. <https://proceedings.scipy.org/articles/Majora-629e541a-00e>.
48. Michaud-Agrawal, N.; Denning, E. J.; Woolf, T. B.; Beckstein, O., Software News and Updates MDAnalysis: A Toolkit for the Analysis of Molecular Dynamics Simulations. *J. Comput. Chem.* **2011**, *32*, 2319-2327. <https://doi.org/10.1002/jcc.21787>.
49. Antosiewicz, J.; McCammon, J. A.; Gilson, M. K., Prediction of pH-Dependent Properties of Proteins. *J. Mol. Biol.* **1994**, *238*, 415-436. <https://doi.org/10.1006/jmbi.1994.1301>.
50. Demchuk, E.; Wade, R. C., Improving the continuum dielectric approach to calculating pK_as of ionizable groups in proteins. *J. Phys. Chem.* **1996**, *100*, 17373-17387. <https://doi.org/10.1021/Jp960111d>.
51. Simonson, T.; Carlsson, J.; Case, D. A., Proton binding to proteins: pK_a calculations with explicit and implicit solvent models. *J. Am. Chem. Soc.* **2004**, *126*, 4167-4180. <https://doi.org/10.1021/ja039788m>.
52. Couch, V.; Stuchebrukhov, A., Histidine in continuum electrostatics protonation state calculations. *Proteins* **2011**, *79*, 3410-3419. <https://doi.org/10.1002/prot.23114>.
53. Beroza, P.; Fredkin, D. R.; Okamura, M. Y.; Feher, G., Protonation of Interacting Residues in a Protein by a Monte Carlo Method: Application to Lysozyme and the Photosynthetic Reaction Center of *Rhodobacter sphaeroides*. *Proc. Natl. Acad. Sci. U. S. A.* **1991**, *88*, 5804-5808. <https://doi.org/10.1073/pnas.88.13.5804>.
54. André, I.; Linse, S.; Mulder, F. A. A., Residue-specific pK_a determination of lysine and arginine side chains by indirect ¹⁵N and ¹³C NMR spectroscopy:: Application to *apo* calmodulin. *J. Am. Chem. Soc.* **2007**, *129*, 15805-15813. <https://doi.org/10.1021/ja0721824>.
55. Forsyth, W. R.; Antosiewicz, J. M.; Robertson, A. D., Empirical relationships between protein structure and carboxyl pK_a values in proteins. *Proteins* **2002**, *48*, 388-403. <https://doi.org/10.1002/prot.10174>.
56. Allen, W. J.; Corey, R. A.; Watkins, D. W.; Oliveira, A. S. F.; Hards, K.; Cook, G. M.; Collinson, I., Rate-limiting transport of positively charged arginine residues through the Sec-machinery is integral to the mechanism of protein secretion. *eLife* **2022**, *11*, e77586. <https://doi.org/10.7554/eLife.77586>.
57. Warshel, A.; Sharma, P. K.; Kato, M.; Parson, W. W., Modeling electrostatic effects in proteins. *Biochim. Biophys. Acta Proteins Proteom.* **2006**, *1764*, 1647-1676. <https://doi.org/10.1016/j.bbapap.2006.08.007>.
58. Brau, C. A., *Modern Problems in Classical Electrodynamics*. Oxford University Press: New York, 2004.
59. Landau, L. D.; Lifshitz, E. M., *Electrodynamics of Continuous Media*. Pergamon Press: Oxford, 1963.
60. Senapati, S.; Chandra, A., Dielectric constant of water confined in a nanocavity. *J. Phys. Chem. B* **2001**, *105*, 5106-5109. <https://doi.org/10.1021/jp011058i>.

61. Schlaich, A.; Knapp, E. W.; Netz, R. R., Water Dielectric Effects in Planar Confinement. *Phys. Rev. Lett.* **2016**, 117. <https://doi.org/10.1103/Physrevlett.117.048001>.
62. De Luca, S.; Kannam, S. K.; Todd, B. D.; Frascoli, F.; Hansen, J. S.; Davis, P. J., Effects of Confinement on the Dielectric Response of Water Extends up to Mesoscale Dimensions. *Langmuir* **2016**, 32, 4765-4773. <https://doi.org/10.1021/acs.langmuir.6b00791>.
63. Fumagalli, L.; Esfandiar, A.; Fabregas, R.; Hu, S.; Ares, P.; Janardanan, A.; Yang, Q.; Radha, B.; Taniguchi, T.; Watanabe, K.; Gomila, G.; Novoselov, K. S.; Geim, A. K., Anomalous low dielectric constant of confined water. *Science* **2018**, 360, 1339-1342. <https://doi.org/10.1126/science.aat4191>.
64. Seyed, S.; Matyushov, D. V., Dipolar susceptibility of protein hydration shells. *Chem. Phys. Lett.* **2018**, 713, 210-214. <https://doi.org/10.1016/j.cplett.2018.10.045>.
65. Capponi, S.; Heyden, M.; Bondar, A. N.; Tobias, D. J.; White, S. H., Anomalous behavior of water inside the SecY translocon. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, 112, 9016-9021. <https://doi.org/10.1073/pnas.1424483112>.
66. Egea, P. F.; Stroud, R. M., Lateral opening of a translocon upon entry of protein suggests the mechanism of insertion into membranes. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, 107, 17182-17187. <https://doi.org/10.1073/pnas.1012556107>.
67. de Oliveira, V. M.; Liu, R. B.; Shen, J. N., Constant pH molecular dynamics simulations: Current status and recent applications. *Curr. Opin. Struct. Biol.* **2022**, 77. <https://doi.org/10.1016/J.Sbi.2022.102498>.
68. Stern, H. A., Molecular simulation with variable protonation states at constant pH. *J. Chem. Phys.* **2007**, 126, 164112. <https://doi.org/10.1063/1.2731781>.
69. Kong, X. J.; Brooks, C. L., λ -dynamics: A new approach to free energy calculations. *J. Chem. Phys.* **1996**, 105, 2414-2423. <https://doi.org/10.1063/1.472109>.
70. Sondergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H., Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pK_a Values. *J. Chem. Theory Comput.* **2011**, 7, 2284-2295. <https://doi.org/10.1021/ct200133y>.
71. Olsson, M. H. M.; Sondergaard, C. R.; Rostkowski, M.; Jensen, J. H., PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pK_a Predictions. *J. Chem. Theory Comput.* **2011**, 7, 525-537. <https://doi.org/10.1021/ct100578z>.
72. Fitch, C. A.; Platzer, G.; Okon, M.; Garcia-Moreno, B.; McIntosh, L. P., Arginine: Its pK_a value revisited. *Protein Sci.* **2015**, 24, 752-761. <https://doi.org/10.1002/pro.2647>.
73. Harms, M. J.; Schlessman, J. L.; Sue, G. R.; Garcia-Moreno, B., Arginine residues at internal positions in a protein are always charged. *Proc. Natl. Acad. Sci. U. S. A.* **2011**, 108, 18954-18959. <https://doi.org/10.1073/pnas.1104808108>.
74. 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* **2007**, 111, 7812-7824. <https://doi.org/10.1021/jp071097f>.
75. Monticelli, L.; Kandasamy, S. K.; Periole, X.; Larson, R. G.; Tieleman, D. P.; Marrink, S. J., The MARTINI coarse-grained force field: Extension to proteins. *J. Chem. Theory Comput.* **2008**, 4, 819-834. <https://doi.org/10.1021/ct700324x>.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.