Supporting Information for manuscript:

The Impact of Protonation on Early Translocation of Anthrax Lethal Factor: Kinetics From Molecular Dynamics Simulations and Milestoning Theory

Piao Ma[†], Alfredo E. Cardenas[‡], Mangesh I. Chaudhari[o], Ron Elber[†,‡,*], and Susan B. Rempe[o]

†Department of Chemistry, University of Texas at Austin, Austin, TX, 78712

[‡]*Institute for Computational Engineering and Sciences, University of Texas at Austin, Austin, TX, 78712*

[o]Biological and Engineering Sciences, Sandia National Laboratories, Albuquerque, NM 87185

II. Methods in detail

Atomically detailed simulations supplant experimental data and provide a comprehensive picture of molecular mechanisms. Simulations of large systems like LF translocation through the membrane channel (about 680,000 atoms of the solvated systems) can be challenging because an ensemble of trajectories at long times (milliseconds and seconds) is necessary [1]. We consider the translocation of the first 30 residues of N terminal segment of LF, which we also call LF_N. Direct MD simulations of the problem at hand are not feasible with current technology. Milestoning provides a less direct approach to simulate the system kinetics and thermodynamics at extended time scales. In Milestoning, many short trajectories (picoseconds to nanoseconds) are conducted. By biasing their starting positions, we can compute the overall mean first passage time (MFPT) and the free energy landscape [2].

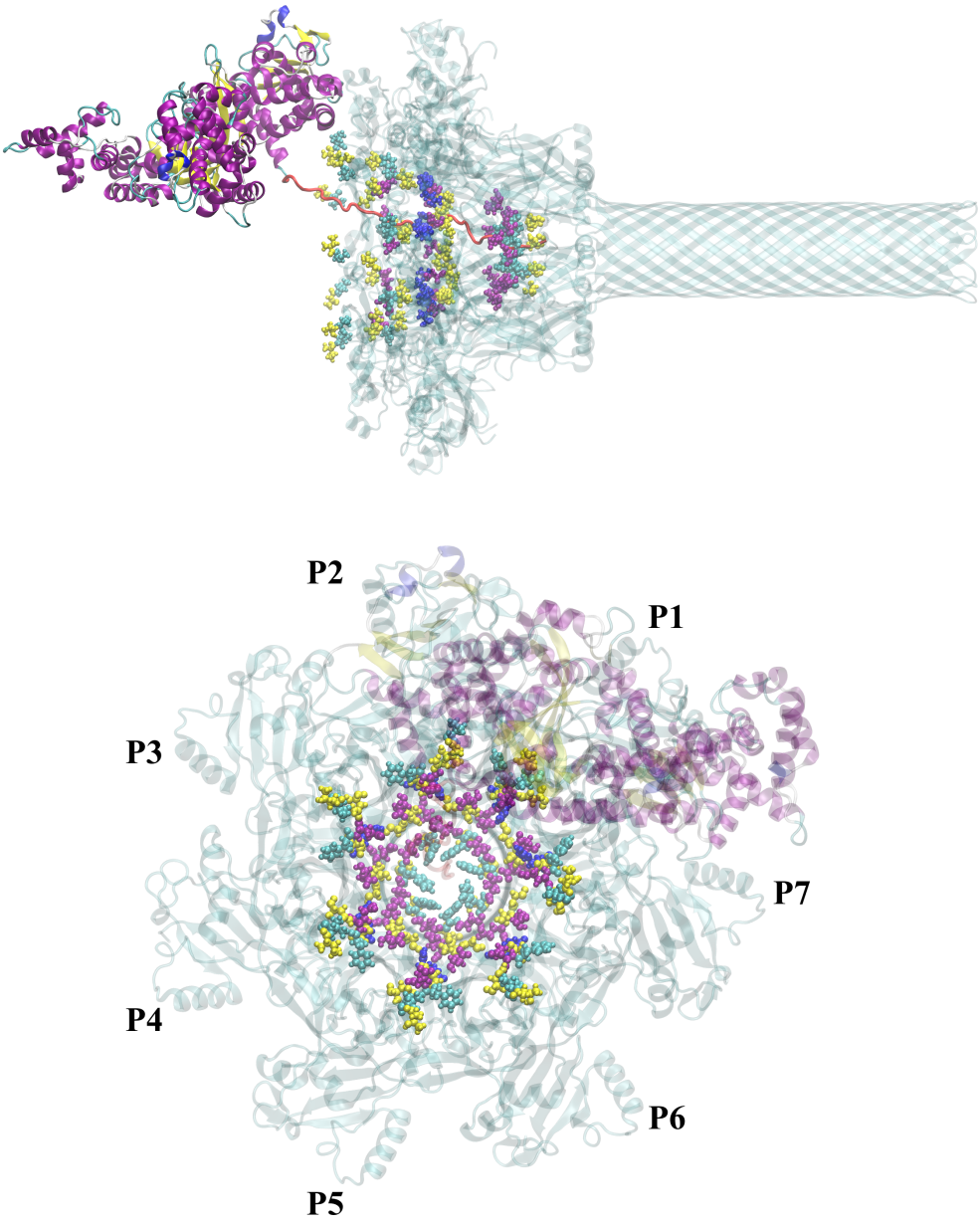

Figure 1. The positions of several important residues are illustrated (hydrophobic residues F427, F464, V175 are in light blue; non-acidic non-basic hydrophilic residues T473, G474, T174, S429, N463, N180 are in yellow; acidic residues E465, D472, D426, E398 are in purple; basic residue R470 is in blue.). The results are from the highly protonated simulations.

The atomic model of the protective antigen pore (PA channel) that was used in the simulation is based on the CryoEM structure by Jiang, et al. [3] (PDB: 3J9C) at a resolution of 2.9 Å. The channel has 7 monomers, and the structure was built by rotating one

monomer 7 times. The residues are numbered starting from the first residues available in the crystal structure.

An atomically detailed model of the complex of LF and the channel is required as a starting point for simulations of the transport process. The LF and the channel were docked to match the experimental structure of LF binding to the PA channel (PDB: 3KWV [4]). The parallel version of the NAMD program was used in the calculations [5]. The cut-off distance for long-range forces was 12Å and the time step was 1fs. About 720 core hours are required to run 1ns of a solvated system in the Texas Advanced Computing Center (TACC), which generates about 15GB of data for 1000 configurations. The proteins were solvated in a cylindrical volume of aqueous solution. The radius of the cylinder is 85Å and its height 310Å. The aqueous solution has physiological ion concentration of 100mM NaCl. The simulations were conducted with a weak restraining force (force constant of 0.1 kcal/mol $Å^2$) on the atoms of the β-sheet structure of the channel.

A Milestoning simulation is typically conducted in three steps. In the first step, anchors or centers of Voronoi cells in coarse space are generated [6-7]. In the second step, initial conditions of the short trajectories are constructed at the milestones, which are the interfaces separating the Voronoi cells [6-7]. These initial conditions can be obtained by a direct sampling from the canonical ensemble at the milestone [8] (as done here), or sampled iteratively to obtain the exact distributions at the milestones [9]. In the third step, short trajectories are run between the interfaces of the Voronoi cells, providing statistics of

transitions between cell boundaries. Finally, the short trajectories create a network between the Voronoi cells that can be analyzed in detail to obtain kinetic and thermodynamic descriptions of the system [10]. In the present system, the network is trivial since the reaction occurs along a one-dimensional coordinate.

The first step: Generating milestones. The translocation of the N-terminal domain of LF is modeled with milestones along a one-dimensional reaction coordinate, which is the distance along the channel axis between the N-terminus of LF and the center of mass of the ϕ-clamp. A 4 ns MD trajectory with steering force constant equals to 1 kcal/(molÅ) and 6 ns MD trajectory with steering force constant equals to 2 kcal/(molÅ) were conducted to sample centers of Voronoi cells along reaction coordinate.

The second step: Sampling configurations at the milestones. To sample configurations at the milestones, a harmonic restraining potential was added with a force constant of 100 kcal/mol $Å^2$, confining the simulations to the interface between the Voronoi cells. If better sampling was required (for example, when the fraction of trajectories going forward (or backward) was less than 5 percent), additional sampling runs were conducted with a stiffer force constant of 500 kcal/mol $Å^2$.

The third step: Running short trajectories between the milestones. Once initial conditions for trajectories at the milestone are generated in the second step, the next step is to run short trajectories until they hit for the first time another milestone. The number

of trajectories initiated at each milestone was typically 100 or larger depending on the accuracy of the estimates (see **Error bars**).

The trajectory data were analyzed with the Milestoning theory [2,9], which we summarize below. The complete simulation time, including the steered molecular dynamics simulations to generate the reaction coordinate, the sampling in the milestones, and the short trajectories, was about 225 nanoseconds (ns). The short trajectories required only 17 nanoseconds of accumulated time when the milestones were separated by 0.5Å, and 37ns when separated by 1Å. These estimate does not include the calculations of different protonation states. The time was sufficient to estimate the overall mean first passage time (MFPT) and the free energy landscape. The accumulated time is much shorter than the MFPT, which can reach milliseconds. This difference in time illustrates the efficiency of the Milestoning approach for kinetic calculations.

The last step: Analysis. We considered a stationary or a steady state process for the calculations of the MFPT, which is frequently used in the study of kinetics. Reflecting boundary conditions were used for the equilibrium study. More general non-equilibrium and time-dependent processes can be considered as well; however, these are unnecessary for the present study. In a stationary process, the number of trajectories that arrive at the product state and are terminated is equal to the number of trajectories that are initiated at the reactant state. When reflecting boundary conditions are used, trajectories that arrive at the final milestone are reflected back with probability one. Under these conditions, we computed the following functions:

(1) The probability of a transition from milestone i to milestone j, or the kernel K_{ij}, which is a time-independent function for a stationary process. The kernel is estimated from short trajectories initiated at milestone i as

$$K_{ij} \cong n_{ij} / n_i \,, \tag{1}$$

where n_i is the number of trajectories initiated at milestone i and n_{ij} is the number of trajectories initiated at milestone i that reach milestone j ($i \neq j$) before reaching any other milestone.

(2) The average lifetime of a milestone is estimated as

$$t_i = \left(1/n_i\right) \sum_{l=1}^{n_i} t_{il} \,, \tag{2}$$

where t_{il} is the time that it takes a specific trajectory l, initiated at milestone i, to pass for the first time a milestone different from i.

With the above two functions for all milestones at hand, we estimate thermodynamic and kinetic parameters. A core function of Milestoning is the stationary flux, $q_{i,}$ which describes the number of trajectories that pass through milestone i in a unit time under steady state conditions. The stationary flux is the eigenvector of the transition matrix **K** with an eigenvalue of one. Hence, we solve the following vector-matrix equation to determine **q** (where **K** is computed from the Milestoning trajectories as described above)

$$\mathbf{q}^{\mathbf{t}}\mathbf{K}=\mathbf{q}^{\mathbf{t}} \,. \tag{3}$$

The bold face denotes a vector or a matrix.

The matrix elements K_{ij} are non-negative. The transition kernel is normalized as $\sum_j K_{ij} = 1$. To ensure uniqueness of the stationary flux vector **q** we also require that the matrix **K** has only one eigenvector with eigenvalue equal to one [10].

With the stationary flux q_i and the local average lifetime t_i, we compute the mean first passage time $\langle \tau \rangle$ (MFPT) and the free energy F_i for each milestone state i. Note that a milestone state is defined as the set of trajectories that pass milestone i last.

$$\sum_i q_i K_{ij} = q_j \,, \tag{4}$$

$$\langle \tau \rangle = \frac{\sum_i q_i t_i}{q_f} \quad , \tag{5}$$

$$F_i = -k_B T \ln\left(q_i t_i\right) \,, \tag{6}$$

where q_f is the flux into the product state which is assumed to be absorbing. We note that if the local first passage times, t_{il}, are distributed exponentially, then a Markovian model can be used to analyze the data [11]. Markovian analysis implies that the first moment of the local first passage time is sufficient to characterize the kinetics. In Figure 2, we show a distribution of first passage times and illustrate that the exponential fit is good, provided that sufficient sampling was collected.

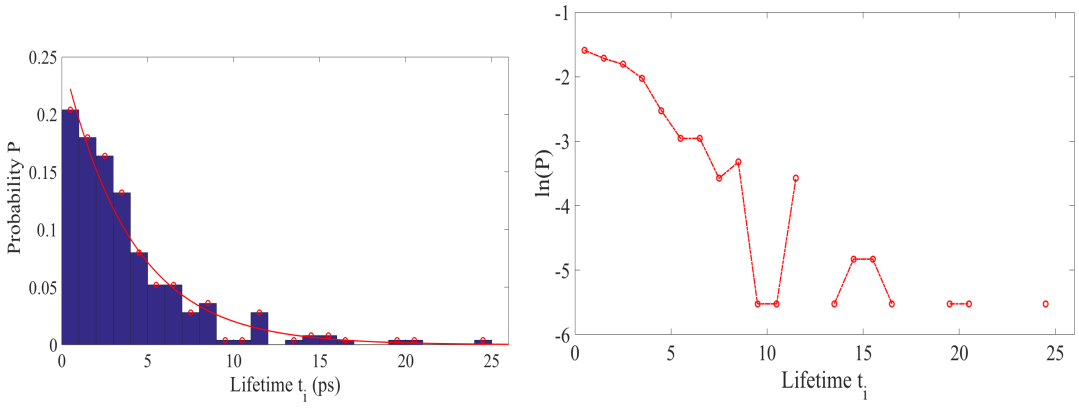

Fig. 2 An exponential fit for a distribution of local mean first passage times at milestone 66. The distribution is fitted to $P(t_{il}) = k\exp(-kt_{il})$ with k=0.25±0.03 at 95% confidence level. At longer times in which the statistics are poor, significant noise is observed. Note the rapid relaxation of the distributions in which significant transition times are sampled in the picosecond time scale.

Error bars

We estimated the error bars for the free energies and the MFPT (Eqs. (6) and (5)) using the following statistical models for the transition kernel K_{ij} and the life time t_i. For both observables we must ensure that negative values are not sampled from the model distributions. For the average life time t_i, we assume that it is distributed normally and estimate the variance and mean values from the sampled statistics. Indeed, the variance was much smaller than the mean and negative times were never sampled in practice. Hence, the use of a normal distribution in this case is adequate.

We cannot, however, use the Central Limit Theorem for the limiting distribution of the elements of the kernel. The K_{ij} are averages of independent events with values of zeroes and ones. Each independent event is a trajectory initiated at milestone i, which can be of

value one (if the trajectory cross milestone j before any other milestone) or of value zero (if the trajectory crosses any other milestone). The trajectories at the milestone are assumed to be statistically independent. The magnitudes of the variances are comparable to the value of the averages of the elements of the kernel. The probability must be greater or equal to zero, and the distribution of the kernel elements cannot be distributed normally and still be guaranteed to be non-negative. Therefore, elements are sampled from the β distribution

$$P\left(K_{ij}\right)=\frac{\Gamma\left(\alpha+\beta\right)}{\Gamma\left(\alpha\right)\Gamma\left(\beta\right)}\left(K_{ij}\right)^{\alpha-1}\left(1-K_{ij}\right)^{\beta-1} \quad , \tag{7}$$

where $\alpha=n_{ij}$ and $\beta=\left(n_i-n_{ij}\right)$ use the estimated n_{ij} and n_i from the trajectories.

We sampled 1000 transition matrices from the β distribution and average local life times from the normal distributions and computed the free energy profiles and the MFPTs. From the statistics, we estimated the mean and the standard deviations and the results are reported in Figures 2 and 4 of the main manuscript.

To further check the convergence of the Milestoning calculations, we computed the free energy by varying the number of milestones for both, the maximal and minimal protonated states. In principle, changing the number of milestones along the reaction coordinate should not affect the free energy or the MFPT. A smaller number of milestones, however, makes the calculation less efficient. The calculations were repeated for maximally and minimally protonated states each. For the maximally protonated state

we consider 75 and 38 milestones. For the minimally protonated case we had 75 and 34 milestones. The number of milestone was different as we attempted to improve the statistics. The results were very similar and therefore we report only the simulations with 75 milestones with a distance between the milestone of 0.5Å.

Boundary conditions

In Milestoning, we impose different boundary conditions by adjusting the **K** matrix. The boundaries represent equilibrium or non-equilibrium states. One boundary condition, which is convenient to determine the MFPT and used here, is of a stationary flux under cyclic or periodic boundary. That condition means that the flux into the first milestone is made identical to the flux leaving the last milestone. This condition is enforced by setting $K_{f,1}$=1 where f is the index of the last (final) milestone. This condition describes a non-equilibrium state in which there is a net flux passing through the system. This picture is consistent with the irreversibility of protein translocation and the ratchet model that was proposed for this process [12]. To compute the free energy, we used reflecting boundary conditions in which trajectories that reached the final milestone are reflected back to the previous milestone. That is, $K_{f,f-1}$=1.

Generating protonation conditions

To grasp the contribution of protonation states to permeation we consider first extreme states of a fully protonated and unprotonated states of the system as two limiting cases. The limiting cases, while less realistic than a full calculation allowed us to grasp the maximum variation we can see in the system. The fully protonated case is particularly

intriguing since it supports the ratchet model proposed earlier [12]. It is also the model that shows better agreement with mutation studies (see below). So, even though the simulation with pH=5 is closer to the physiological conditions, certain features of the extreme model are in accord with experiment.

To determine intermediate protonation states we consider first a simple model in which only the histidine residues are protonated. The free energy and the MFPT curves were between the two extreme curves mentioned above (Fig. 2 & 4 of the main text).

Of course a protonation state can be determined only statistically according to the pKa value of the amino acid under consideration. Accurate simulation of the dynamics will require simulating the kinetics of proton hopping between different groups in addition to the kinetics of the permeation of the N terminal. Atomically detailed simulations of the pKa, is an alternative and was recently shown to provide significant insight [17] to equilibrium protonation states of histidine residues inside a channel. However, the coupling of proton *dynamics* to conformational *dynamics* is expensive and its accuracy uncertain.

We therefore considered a cheaper approach that allows us to estimate dynamic pH changes along the reaction coordinate. We estimated the pKa of the different groups of the N terminal segment of the Anthrax LF protein along the reaction coordinate. For each milestone along the reaction coordinate we use all the sampled structures to compute pKa's using the algorithm PROPKA 3.0 [15] as implemented in PDB2PQR 2.1.1 program [16]. This algorithm is based on empirical theory. However, it was shown to be efficient and quite accurate. With the pKa values estimated (Fig. 3) we sampled protonation states at pH=5 for the residues that can be protonated (glutamic and aspartic

acids and histidines). We estimate the probability for protonation, P as $P = \frac{e^{pKa-pH}}{e^{pKa-pH}+1}$.

The protonation state was fixed for a given milestone but it is resampled for each milestone and therefore changes as we move along the reaction coordinate.

We comment that the protonation states sampled are quite close to our simple model that assumes protonation of only histidine residues. A large fraction of the stem of the anthrax channel is in water. It is hydrophilic and therefore supports histidine protonation. It is hydrophobic only near the ϕ clamp. For values of the reaction coordinate that are larger than zero (beyond the ϕ clamp), we observe a polar to hydrophobic shift in the environment and glutamic acid 13 (for example) is likely to get protonated.

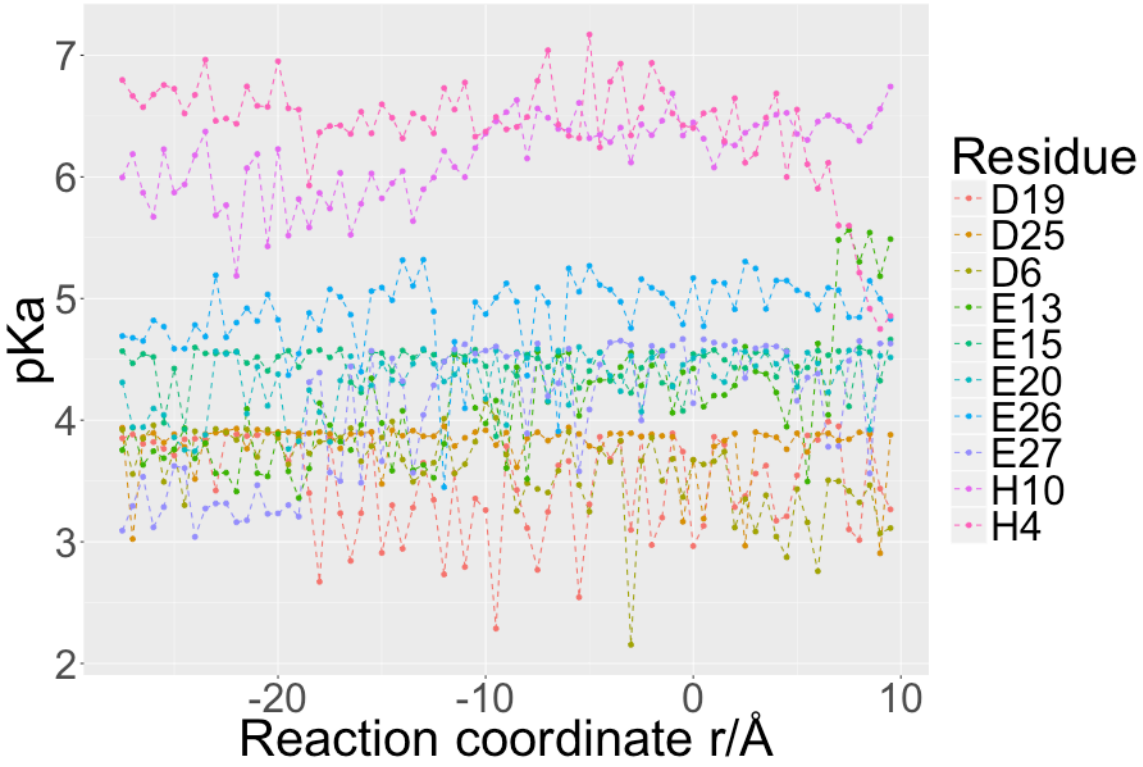

Figure 3. Calculations of the pKa along the reaction coordinate for different residues. Note that for most of the reaction coordinate the pKa of the histidine residues is larger than of glutamic and aspartic acids. See text more details.

With the pKa at hand we can sample protonation state for different milestone. In figure 4 we show sampled protonation states as a function of the milestone number. There are two histidine residues in the N terminal segments and they are protonated most of the time along the reaction path. There are eight aspartic or glutamic acids and their protonation state is less than a half. Hence the more elaborate calculation of the protonation state support the model that emphasize histidine protonation. We recomputed the MFPT and the free energy for pH=5 using Milestoning. The results are presented as the black curves in Fig. 2 and 4 in the main text and are similar to the curve with only the histidine protonated.

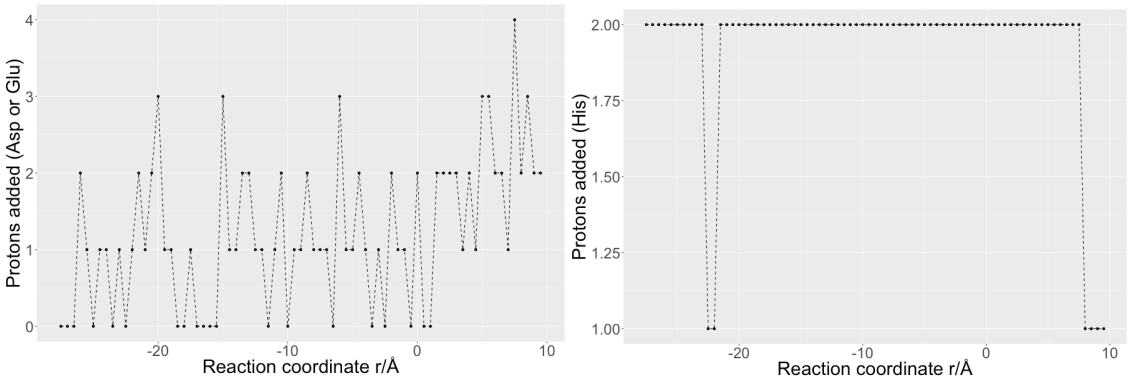

Figure 4. Protonation states along the reaction coordinate of the N terminal segment sampled according to the pKa calculations reported in figure 3 and for pH=5. The left side reports the protonation state of the eight aspartic and glutamic acids as a function of the milestone number and the right side provides the protonation state of the two histidine residues. Note that the histidine protonation is more complete. See text for more details on the calculations of the pKa and of the protonation probability.

Additional results

The Committor function

Another useful function to characterize the reaction progress is the committor function, **C** [13], which is the probability that the trajectories initiated at a particular milestone will make it first to the product before the reactant. It is considered to be the "ultimate" reaction coordinate. The committor function is shown in Figure 5. Note the striking commitment of the N-terminus (LF_N) to the ϕ-clamp under acidic conditions at early stages of the reaction coordinate, supporting the model in which protonated states permeate more readily. We use a method developed recently[14] to estimate the committor function from the transition matrix, K_{ij}. The committor of a milestone is given by $\mathbf{KC} = \mathbf{C}$. Numerically, we consider $\lim_{n\to\infty} \mathbf{K}^n\mathbf{C} = \mathbf{C}$. **C** is a vector with the committor values for different milestones. We increase n, the power of **K**, until the change in the committor values (and the error) is smaller than 0.02. (Note that **K** is an asymmetric matrix and the above equation is different from $\mathbf{q}^t\mathbf{K} = \mathbf{q}^t$ - Eq. (3)).

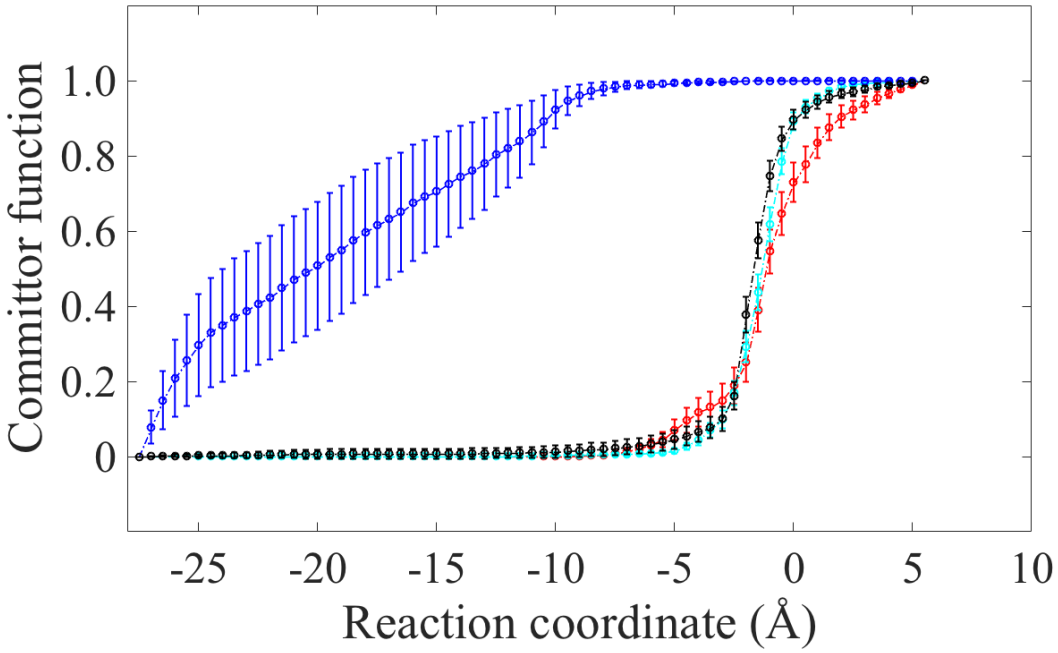

Figure 5: The committor as a function of the position of the milestone along the reaction coordinate. The product is defined at +5.5Å Note the sharp difference between the dark blue curve (acidic conditions) that is committed early to the ϕ-clamp and the three other curves. The other three curves represent less acidic conditions that are committed late in the process. See Fig. 2 in the main text for coloring information.

Structural features of permeation

Finally, we examine some structural features of the reaction. We collected all the saved frames from the Milestoning trajectories (saved every 0.2ps) and computed collision weight of channel residue i , CW_i, with any residue of the N terminal segment. A collision is defined if a pair of heavy atoms, one of the channel, and one of the N terminal segment, is within a distance of 5 Å. We weight the collision number by the Milestoning flux to obtain the correct statistics of the conformations. The collision weight of residue i is given by $CW_i = \sum_{j=1}^{N-1} \frac{C_{ij}}{n_j} q_j$, where C_{ij} is the number of times residue i of the channel collides with the N terminal segment at milestone j. The total number of structures in milestone j is n_j. The flux at milestone j is q_j and the total number of milestone is N. In Fig. 6 we show the top colliding residues detected for the high protonation limit.

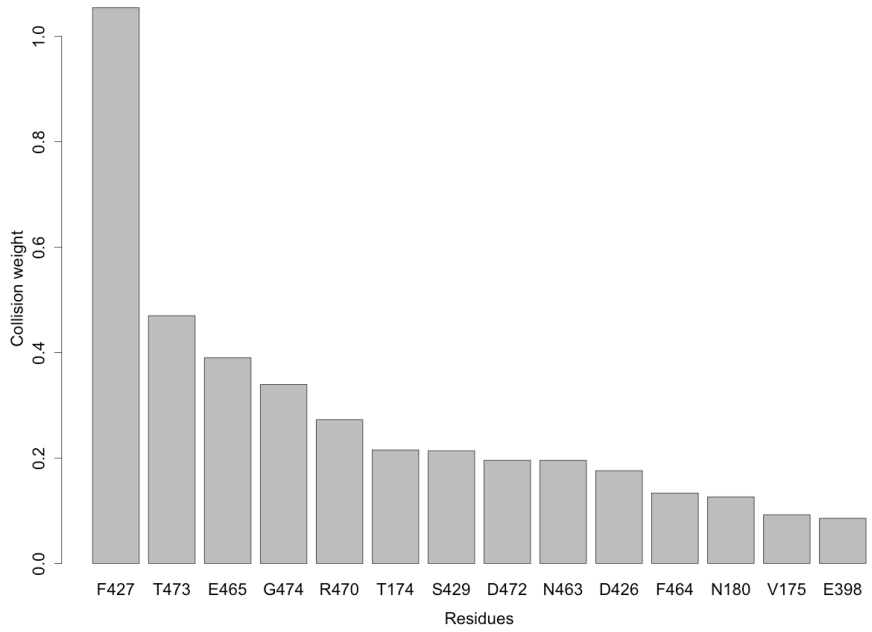

Figure 7. The 14 channel residues with the largest collision weights with the N terminal fragment at highly acidic conditions.

In references [18-19] different mutations were examined experimentally for their impact on translocation and toxicity. Some of the residues found by the collision analysis are in agreement with experiment. These residues include F427, D426 and E398. We did not identify K397 and D425 that were shown to be important experimentally. However, nearby residues (E398 and D426/F427) were detected by us. We predict that E465 G474 R470 and T174 could have significant impact on the process, which experimentally have not been tested yet. A structural view of these residues is provided in Fig. 1 of this SI.

Note that the comparison to residues mutated by experiment is qualitative at best. Our results are confined to early events of the process and do not probe translocation of the

whole protein. Nevertheless, they may motivate new experiments with a focus on the first step of establishing a hook in the channel mouth.

The same collision analysis was applied to other simulations with different protonation conditions. In figure 8 we show the collision weight as a function of the milestone number for the Milestoning calculation with constant pH (with protonation states that can vary as a function of milestone number).

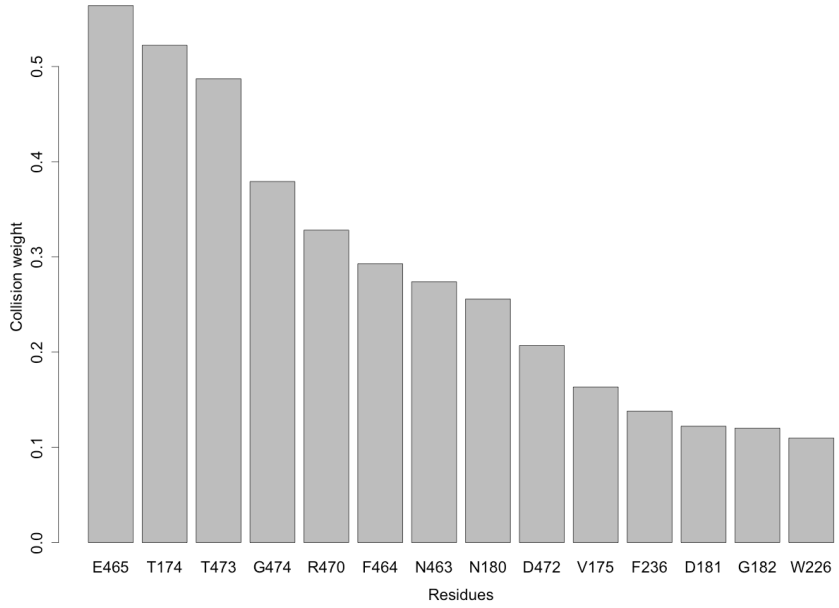

Figure 8. Collision weights evaluated for Milestoning trajectories at pH=5. See text for more details.

Note that the colliding residues are different in Fig. 7 and Fig. 8. This is a reflection of different committors and free energy profiles. For example, the ϕ clamp residues are hardly "touched" at pH=5 simulation since it is a location of a free energy barrier under these conditions according to our simulations (see Fig. 2 in the main text). For that reason it is a rare event to observe collisions between F427 and LF_N, for example.

The channel consists of seven monomers. The permeating N terminal segment interacts with these monomers differently. In Fig. 9 we show the monomer collision weight as a function of the reaction coordinate for the simulation at pH=5. Pathways with other protonation states provide similar weight. The monomer collision weight is computed in the same way as the residue weight.

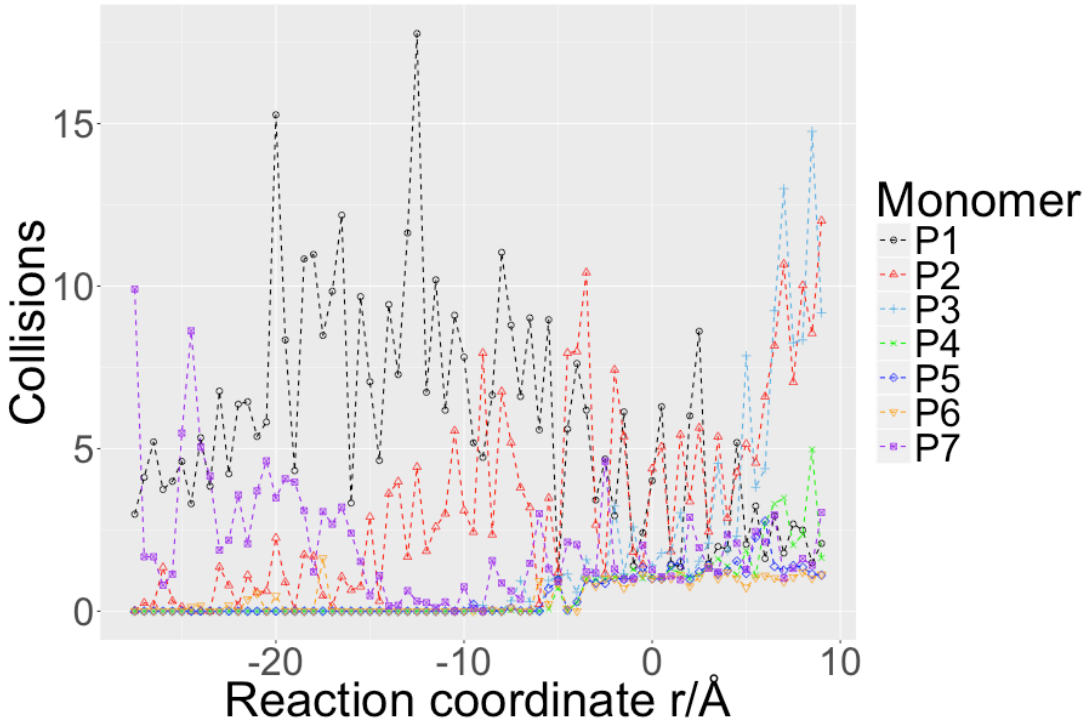

Figure 9: The monomer collision weight computed along the reaction coordinate for the Milestoning calculations at pH=5.

Another measure of the reaction progress is to follow the center of mass position of each amino acid as a function of the milestone number (Fig. 10). The N-terminal residue is constrained to be at the milestone, and therefore its center of mass progression is linear with the number of milestone. However, the rest of the chain is not constrained. The centers of mass of other amino acids move in more slowly as the chain stretches itself.

Also of interest is that at some milestones (e.g., milestone 36) there are significant non-monotonic jumps towards the ϕ-clamp that exceeds the average position of the next milestone. These excessive displacements are collective and include ~10 amino acids.

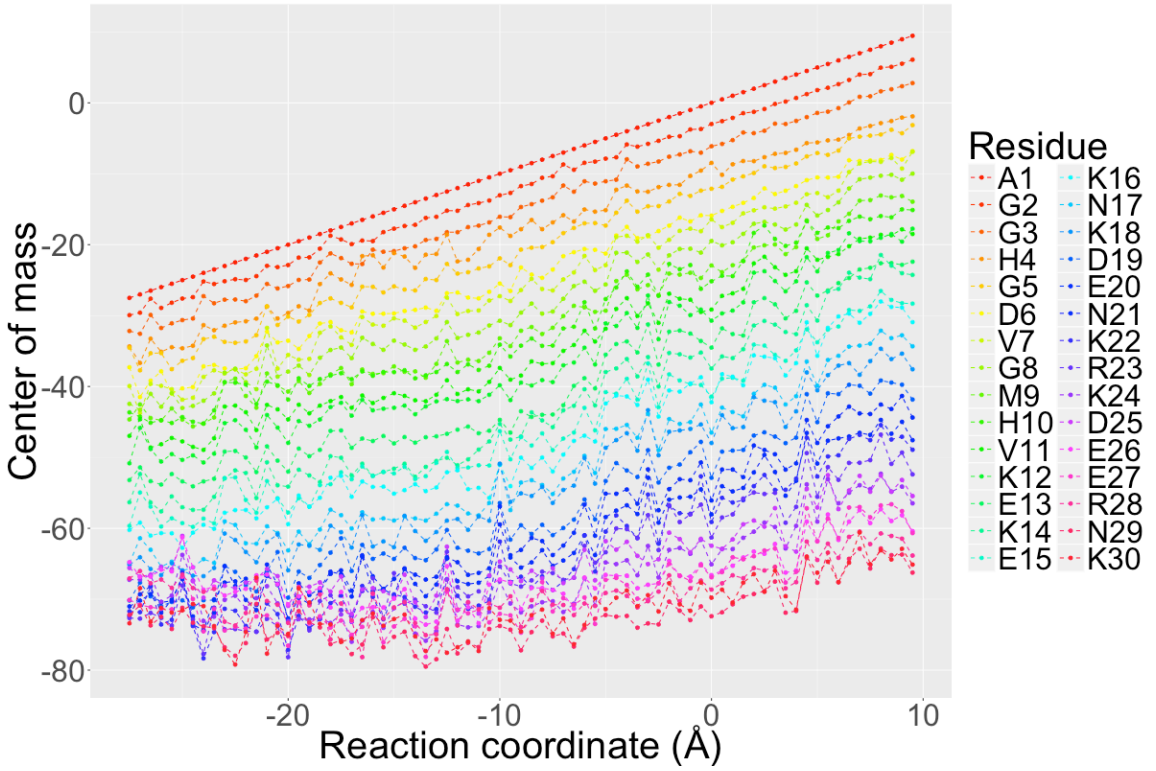

Figure 10. Progression of the centers of mass of different amino acids of the LF N-terminal segment as a function of the milestone number. The displacement of the N-terminal residue is linear with the milestone number by construction. However, the rest of the chain progresses slower, in a non-linear fashion, as the system changes from a collapsed state to a more extended configuration. Note that the displacements overall are monotonic and the chain crosses itself in the channel only rarely. Nevertheless, some spikes and collective displacements of a few amino acids are observed, especially for the last 10 amino acids. The results for different protonation states are similar and therefore not shown.

References for supporting material

1. Elber, R., Perspective: Computer simulations of long time dynamics. *J. Chem. Phys.* **2016,** *144,* 060901
2. Kirmizialtin, S.; Elber, R., Revisiting and Computing Reaction Coordinates with Directional Milestoning. *J. Phys. Chem. A* **2011,** *115*, 6137-6148
3. Jiang, J. S.; Pentelute, B. L.; Collier, R. J.; Zhou, Z. H., *Nature* **2015,** *521* (7553), 545-U323.
4. Feld, G. K.; Thoren, K. L.; Kintzer, A. F.; Sterling, H. J.; Tang, II; Greenberg, S. G.; Williams, E. R.; Krantz, B. A., *Nat. Struct. & Mol. Biol.* **2010,** *17*, 1383-U245.
5. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., *J. Comp. Chem.* **2005,** *26*, 1781-1802.
6. Vanden-Eijnden, E.; Venturoli, M., *J. Chem. Phys.* **2009,** *130*, 13
7. Majek, P.; Elber, R., *J. Chem. Theory and Comp.* **2010,** *6*, 1805-1817.
8. Faradjian, A. K.; Elber, R., *J. Chem. Phys.* **2004,** *120*, 10880-10889.
9. Bello-Rivas, J. M.; Elber, R., *J. Chem. Phys.* **2015,** *142,094102*
10. Viswanath, S.; Kreuzer, S. M.; Cardenas, A. E.; Elber, R., *J. Chem. Phys.* **2013,** *139*,174105
11. Anthony M.A. West, Ron Elber, and David Shalloway, *J. Chem. Phys.* **2007**,126,145104
12. Feld, G. K.; Brown, M. J.; Krantz, B. A., *Protein Sci.* **2012,** *21*, 606-624.
13. Peters, B.; Bolhuis, P. G.; Mullen, R. G.; Shea, J. E., *J. Chem. Phys.* **2013,** *138,054106*
14. Elber, R.; Bello-Rivas, M. J.; Ma, P.; Cardenas, A. E.; Fathizadeh, A., *Entropy* **2017,** *19*, 219
15. Olsson, M.H., Sondergaard, C.R.,Rostkowski, M., Jensen, J.H., J. Chem. Theory Comput., **2011**,*8*,525-537
16. Dolinsky T.J., Nielsen, J.E., McCammon, J.A., Baker, N.A., Nucleic Acids Res., **2004**,*32*,W665-W667
17. Chen, W., Huang, Y., Shen, J., J. Phys. Chem. Lett. **2016**,*7*,3961-3966
18. Sellman, B. R.; Mourez, M.; Collier, R. J., *Science* **2001,** *292*, 695-697
19. Sellman, B. R.; Nassi, S.; Collier, R. J., *J. Biol. Chem.* **2001,** *276*, 8371-8376.