Supplementary Information for

Role of internal loop dynamics in antibiotic permeability of outer membrane porins

Archit Kumar Vasan, Nandan Haloi, Rebecca Joy Ulrich, Mary Elizabeth Metcalf, Po-Chao Wen, William W. Metcalf, Paul J. Hergenrother, Diwakar Shukla, Emad Tajkhorshid

Corresponding Author name.
E-mail: emad@illinois.edu

This PDF file includes:

Extended Methods
Figs. S1 to S15
Tables S1 to S4
SI References

Extended Methods

Preparation and simulation of membrane-embedded OmpF. We used the X-ray structure of OmpF trimer (PDB ID: 3POX) (1), including all the crystal water molecules, as a starting point for our simulations. In each monomer, residues E296, D312, and D127 were protonated in accordance with previous studies(2–4). The OmpF trimer was embedded in a symmetric membrane composed of 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) lipid molecules. Usage of an OM composition containing lipopolysaccharides (LPS) is not considered, since the membrane composition is unlikely to influence the dynamics of L3 as this internal loop is fully internalized in the protein fold and not exposed to the membrane. The protein-DMPC system was then solvated with TIP3P water (5) and buffered in 0.15 M NaCl. Each step of the membrane building process was carried out using the Membrane Builder module of CHARMM-GUI (6). The final system contained ∼140,000 atoms with dimensions of $120 \times 120 \times 100$ Å^3. Then, 10 independent molecular dynamics simulations were run. In each simulation, the prepared system was minimized using the steepest descent algorithm for 2,000 steps, followed by an initial equilibration of 5 ns, during which the protein heavy atoms were harmonically restrained using a force constant of 5 kcal mol^{-1} Å^{-2}. Then, 1 μs of unrestrained production simulation was performed for each replica.

Electric field simulations. Ionic current was calculated by performing simulations with a constant electric field normal to the membrane. Five replicas of ionic current simulations were performed with an OmpF monomer, for each of the three different OmpF conformations (O, C_A, C_B) derived from MSM analysis (see below). The starting point for each simulation was a representative OmpF monomer from each protein conformation. The OmpF monomers were independently embedded in a symmetric membrane composed of 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) lipid molecules, solvated with TIP3P water (5) and buffered in 0.5 M NaCl to enhance sampling of ionic current in the simulation. For each simulation, the prepared system was minimized using the steepest descent algorithm for 2,000 steps, followed by an initial relaxation simulation of 1 ns, during which the protein heavy atoms were harmonically restrained using a force constant of 1 kcal mol^{-1} Å^{-2}. Each production simulation was then performed for 100 ns with an electric field corresponding to a membrane electric potential difference of 100 mV. During these simulations, the protein heavy atoms were harmonically restrained using a relatively weak force constant of 0.1 kcal mol^{-1} Å^{-2} to maintain the respective conformational state of the protein (O, C_A, or C_B).

Ionic current (I) was computed by counting the number of ions (Na^+ and Cl^-) that cross the porin over time, i.e., $I = N\times\frac{q}{\tau}$, where N is the number of ion crossing events over a time interval τ, and q is the charge of the ion (1.60217662×10^{-19} Coulombs for Na^+, and $-1.60217662\times10^{-19}$ Coulombs for Cl^-). The total current was simply the sum of the net Na^+ current minus the net Cl^- current. The conductance (C) was then calculated as $C = \frac{I}{V}$.”

Antibiotic permeation free energy calculations. To investigate the energetics of permeation of fosfomycin in the open and closed states, two independent sets of bias exchange umbrella sampling (BEUS) simulations (7–11) were performed starting from the O and C_A states derived from our MSM analysis (see below). The force field parameters for fosfomycin were generated using the CHARMM General Force Field (CGenFF) (12–14) with the ParamChem server.

The initial seeds for these BEUS simulations were obtained using a Monte Carlo-based pathway search (MCPS) algorithm, specifically developed to improve sampling of the position and orientation of antibiotics in OM porins (15). Briefly, MCPS determines the most likely permeation pathway through OmpF using an energetic descriptor of the system, while systematically exploring all positions and orientations of the antibiotic in the region of interest based on an initial screening. To run MCPS, we first take a representative monomer of each protein conformational state (O or C_A) and explore translational and rotational degrees of freedoms of the drug within the pore to generate datasets containing hundreds of thousands of discrete drug-protein poses. Each drug-protein pose is then minimized while fixing protein backbone. This dataset is then used to construct a multidimensional drug-protein interaction energy (IE) landscape along the translation (Z-coordinate) and two orientation angles (inclination angle, θ, and azimuthal angle, ϕ). Then, we walk through the resulting IE landscape using Monte Carlo (MC) moves to determine favorable (low energy) trajectories/pathways connecting extracellular and periplasmic spaces. The starting point within a trajectory is randomly selected from a pose in the extracellular space. To better sample putative pathways in our defined space, we generated 2,000 MCPS trajectories. These trajectories were then used to build a connected graph to be used in Dijkstra's algorithm to determine the most favorable permeation pathway. A detailed description for each step is provided in our previous study (15).

The most likely pathways for the drug, calculated independently for the two conformational states of OmpF (open and closed), were used to seed the BEUS simulations (7–11). The main part of the pathways used, namely the part within the protein, included poses with Z-values ranging from -10 to 24 Å (relative to the midplane of the membrane) of 1 Å width; to obtain reference free energies, we extended the number of windows in the extracellular and periplasmic spaces such that the terminal windows are at least 10 Å away from any atom of the protein. To ensure adequate histogram overlap in BEUS, additional windows were added in between the original windows for the CR (Z =-3 to 12 Å) such that the window width was 0.5 Å within this region. In total, 94 windows were used spanning 80 Å from the extracellular (Z=46 Å) to the periplasmic ($Z = -34$ Å) space.

To take into account the biologically relevant configuration of the system, for each window we built a trimeric, membrane-embedded, drug-bound system. To do this, we first aligned the backbone atoms of the β-barrel of the drug-bound monomer of each window to an OmpF monomer in the trimeric X-ray structure of the protein (PDB ID: 3POX) (1). Then, we merged the resulting coordinates of the aligned antibiotic-monomer system with the two additional monomers of the trimeric OmpF. In the generated trimer for each window, residues E296, D312, and D127 were protonated (2–4). The windows were then embedded in a symmetric membrane composed of 1,2-dimyristoyl-sn-glycero-3-phosphocholine (DMPC) lipid molecules in each leaflet generated using the Membrane Builder module of CHARMM-GUI (6). Each window was solvated with TIP3P water (5) and buffered in 0.15 M NaCl to generate trimeric systems containing ∼140,000 atoms with dimensions of $120 \times 120 \times 100$ Å^3.

Before performing BEUS simulations, each trimer was minimized using 10,000 steps of the steepest descent algorithm, and then the molecular system was relaxed at the center of each window during a 1-ns MD simulation while the drug and heavy atoms of the protein were restrained with a force constant of 1 kcal mol^{-1} Å^{-2}. This was followed by 30 ns of BEUS simulations (until the convergence of the free-energy) during which the protein backbone heavy atoms were restrained with a force constant of 1 kcal mol^{-1} Å^{-2} , using the distance along the membrane normal (Z-axis) between the drug's C.O.M and the C.O.M of the drug-containing monomer as the collective variable. The force constants were 2.0 kcal mol^{-1} Å^{-2} for all windows except for the windows in the CR, which had force constants of 7.0 kcal mol^{-1} Å^{-2}. Using these force constants resulted in good window overlap for each drug-protein system. The first 10 ns of each window were discarded, and the rest was used in evaluating the free energy. A non-parametric variation of the weighted histogram analysis method (WHAM) (16), proposed by Bartels (17) and implemented by Moradi and Tajkhorshid (11) was used to estimate the free-energy profile from the BEUS simulations.

Molecular dynamics (MD) simulation protocol. MD simulations in this study were performed using NAMD (18, 19) utilizing CHARMM36m (20) and CHARMM36 (21) force field parameters for proteins and lipids, respectively. A timestep of 2 fs was used in all simulations, and periodic boundary conditions were employed in all three dimensions. Bonded and short-range nonbonded interactions were calculated every timestep. The particle mesh Ewald (PME) method (22) was used to calculate long-range electrostatic interactions every 4 fs with a grid density of 1 Å^{-3}. A force-based switching function was employed for pairwise nonbonded interactions starting at a distance of 10 Å with a cutoff of 12 Å. Pairs of atoms whose interactions were evaluated were searched and updated every 20 fs. A cutoff (13.5 Å) slightly longer than the nonbonded cutoff was applied to search for the interacting atom pairs. Constant pressure was maintained at a target of 1 atm using the Nosé-Hoover Langevin piston method (23, 24). Langevin dynamics maintained a constant temperature of 310 K with a damping coefficient, γ, of 0.5 ps^{-1} applied to all atoms. Simulation trajectories were collected every 10 ps.

Markov state model construction. We used our trajectory dataset to construct a Markov state model (MSM) using pyEmma (25), which enabled us to obtain kinetic and thermodynamic information about the system. To build the MSM, first the trajectory dataset was featurized using 26 residue-residue distance pairs with significant hydrogen bonding (occupancy greater than 25% during the simulations, or a maximum lifetime greater than 50 ns in at least one of the monomers of any replica) between the highly fluctuating residues of L3 (residues 116-123) and the barrel wall. A hydrogen bond was counted between an electronegative atom with a hydrogen atom (H) covalently bound to it (the donor, D), and another electronegative atom (the acceptor, A), if the D-A distance is less than 3 Å and the D-H-A angle is greater than 120°. For these residue pairs, the minimum distance (in each frame of every trajectory) between any donated H and any A atom was used to create the MSM feature space. Since the distance pairs are uncorrelated between monomers (Fig. S3), we considered each monomer as an independent trajectory, giving us an aggregate trajectory data of 30 μs (10 independent runs $\times$ 3 monomers $\times$ 1 μs).

To remove redundant information within the feature space and identify the slowest reaction coordinates, time-structure based independent component analysis (tICA) was used to reduce the dimensionality of the feature space ($X(t)$) to the eigenvectors of an autocovariance matrix, $\langle X(t)X^T(t+\tau)\rangle$, with a lag time, τ=1 ns (26–29). It is important to choose an optimal number of tICA eigenvectors since an MSM built using too many eigenvectors would have microstates with low statistical significance due to finite sampling error (30). We found that the first seven tICA eigenvectors are sufficient to construct the MSM because only the distribution of these eigenvectors significantly differed from the normal distribution (Fig. S4). Further statistical analysis using an MSM scoring method, VAMP-2 score (31), discussed further in the next section, showed that the quality of an MSM does not significantly improve when using more than five tICA eigenvectors (Fig. S5). Thus, we chose to reduce the number of eigenvectors to five in our study.

The conformational space was then discretized into multiple microstates using k-means clustering. To choose the number of microstates to use in the model, we used the VAMP-2 score (31), to evaluate the quality of MSMs built with different numbers of microstates. The VAMP-2 score converged when using five tICA eigenvectors and 1,000 microstates (Fig. S5); thus, we used this parameter set to build our MSM.

Then, a transition probability matrix (TPM) was constructed by evaluating the probability of transitioning between each microstate within a lag time, τ. To choose an adequate lag time to construct a TPM that ensures Markovian behavior, multiple TPMs were first created using multiple maximum-likelihood MSMs with different lag times. The implied timescales ($\tau_i = \frac{\tau}{ln(\lambda_i)}$) were evaluated for each of these transition matrices, and saturation was observed at $\tau = 2$ ns (Fig. S6). Thus, we built our final TPM using a maximum likelihood MSM with a lag time of 2 ns. This final TPM is symmetrized using a maximum likelihood approach to ensure detailed balance (25). This step did not significantly change the raw TPM (Fig. S14), indicating that the initial sampling was done under dynamic equilibrium conditions.

To identify physically meaningful metrics for projecting the free energy of the conformational transitions, we used a protocol described by Pérez-Hernández *et al.* to choose the metric with greatest correlation to the second eigenvector of the TPM (29). The normalized correlation between the second eigenvector of the TPM and each of the 26 residue-residue distance pairs was evaluated as follows:

$$Corr(r_k, \tilde{\psi}_{2,s(t)}) = \frac{\langle r_k \tilde{\psi}_{2,s(t)}\rangle_t - \langle r_k\rangle_t \langle \tilde{\psi}_{2,s(t)}\rangle_t}{\langle r_k^2\rangle_t \langle \tilde{\psi}^2_{2,s(t)}\rangle_t},$$

where r_k is the k_{th} residue-residue distance, $\tilde{\psi}_2$ is the second eigenvector of the TPM, $s(t)$ is the trajectory of microstates, and $\langle\rangle_t$ is the time average. The E117-Y22 and D121-R132 distances were chosen for this purpose, as these features showed the greatest positive and negative correlations with the second eigenvector, respectively (Fig. S7). Using these features, we projected the free energy landscape weighted with the stationary distribution obtained from the MSM (Fig.2C,D). The free energy landscape using the raw trajectory data, unweighted by the stationary distribution, is very similar to the weighted landscape (Fig. 2 and Fig. S15), indicating that the initial sampling used to build the MSM was sufficient to mitigate any sampling bias. To determine the error of the free energy landscape, we first used bootstrapping to alter the TPM, and to create new free energy landscapes. The error was then determined in every bin of the landscape (Fig. S13). We used the free energy landscape to lump our microstates into 5 macrostates depending on whether the microstate physically lies within a free energy minima (defined using an energy cutoff of 1.2 kcal/mol) shown in Fig2.C,D. Macrostates are classified according to their pore bottleneck radius (Fig. 3) leading to: an open state (O), two intermediate states (I_A, and I_B) and two closed states (C_A, and C_B).

Transition path theory. To obtain kinetic information about the processes, the mean first passage times (MFPTs) for the O-C_A and O-C_B transitions were evaluated. The uncertainty in the MFPT was evaluated using a Bayesian estimated MSM, implemented in pyEmma (25). The transition path theory module in pyEmma (25) was used to identify the conformational transitions. This step was done by choosing the O and C_A/C_B states as the source and sink, respectively, and identifying the pathways connecting them.

Construction of *E. coli ompF* mutants. Strains carrying various *ompF* alleles were constructed in three steps. All strains, plasmids, primers and gBlocks used in these constructions are shown in Tables S1 to S4. Initially, a strain carrying a complete deletion of the *ompF* locus (Δ*ompF*8897::*cat*) was constructed by λ-red-mediated recombination as described elsewhere (32). To do this, BW26678 was transformed to Cm^R using a PCR product obtained using primers ompF-cloningF and ompF-cloningR with pKD3 as the template, creating WM8897. The Δ*ompF*8897::*cat* allele removes the entire *ompF* coding sequence and 323 base-pairs upstream of the start codon containing the promoter and regulatory sequences. Second, a series of plasmids carrying various *ompF* alleles were constructed using the pAH144 (33) as the vector. This plasmid encodes resistance to streptomycin and spectinomycin (Strep/Spec^R) and can be inserted into the host chromosome in single copy at the Hong Kong phage attachment site (*att*-HK). Initially we constructed pMEM501, which carries the WT *ompF* gene and 610 upstream base-pairs, including the promoter and all known regulatory sequences. Plasmid pMEM501 was then modified by replacement of an appropriate internal restriction endonuclease fragment with a synthetic DNA fragment carrying the desired mutations. Finally, each of these plasmids was inserted into the chromosome of WM8901 by selection for Strep/Spec^R. The inserts of all plasmids were verified by DNA sequencing. All strains were verified by PCR, including DNA sequencing of the PCR product, to confirm the presence of the Δ*ompF*8897::*cat* allele and the correct plasmid inserted in single copy.

Accumulation assay protocol. The accumulation assay was performed in triplicate as outlined elsewhere (34, 35). A 5 mL overnight culture was diluted into 250 mL of fresh lysogeny broth (LB) and grown at 37°C with shaking to an optical density (OD_{600}) of 0.55-0.60. Once grown to mid-log phase, 200 mL of culture was pelleted at 3,220 r.c.f. for 10 minutes (at 4°C). The supernatant was discarded and cells resuspended in 40 mL phosphate buffered saline (PBS), pelleted as before, and resuspended in 8.8 mL PBS. Cells were aliquoted into 1.7 mL Eppendorf tubes each with 875 μL and incubated with shaking at 37°C for 5 minutes to equilibrate cells. Colony forming units (CFUs) were determined by a calibration curve. These time points were short enough to minimize metabolic and growth changes (no changes in OD_{600} or CFU count observed). Cells were treated with 50 μM compound (8.75 μL of 5 mM compound stock) for 10 minutes at 37°C with shaking. After incubation, 800 μL of culture was layered over 700 μL cold silicone oil (9:1 AR20/Sigma High Temperature, cooled to -78°C) and cells pelleted at 13,000 r.c.f. for 2 minutes at room temperature to separate supernatant and extracellular compound from bacterial cells. The supernatant and oil were removed by pipette and the cell pellet was resuspended in 200 μL MilliQ water. Samples were subjected to three freeze-thaw cycles of alternating 3 minute incubation periods in liquid nitrogen (-78°C) and a 65°C water bath. Lysed cells were pelleted at 13,000 r.c.f. for 2 minutes and 180 μL of supernatant were collected. Cell pellets were washed in 100 μL methanol, vortexed, and pelleted at 13,000 r.c.f. for 2 minutes. After pelleting, 100 μL of supernatant was collected and combined with previous supernatants. Remaining debris were removed through centrifugation at 20,000 r.c.f. for 10 minutes at room temperature. Supernatants were analyzed with the QTRAP 5500 LC/MS/MS system (Sciex) in the Metabolomics Laboratory of the Roy J. Carver Biotechnology Center, University of Illinois at Urbana-Champaign. Software Analyst 1.6.2 was used for data acquisition and analysis. The 1200 Series HPLC System (Agilent Technologies) includes a degasser, an autosampler and a binary pump. The liquid chromatography separation was performed on an Agilent Zorbax SB-Aq column (4.6 mm × 50 mm; 5 μm) with mobile phase A (0.1% formic acid in water) and mobile phase B (0.1% formic acid in acetonitrile). The flow rate was 0.3 mL min^{-1}. The linear gradient was as follows: 0–3 min: 100% A; 10–15 min: 2% A; 16–20.5 min: 100% A. The autosampler was set at 15°C. The injection volume was 1 μL. Mass spectra were acquired under positive electrospray ionization with a voltage of 5,500 V. The source temperature was 450°C. The curtain gas, ion source gas 1 and ion source gas 2 were 33, 65 and 60 psi, respectively. Multiple reaction monitoring was used for quantitation with external calibration. All compounds evaluated in biological assays were $\geq$ 95% pure, assessed by NMR and LC-MS.

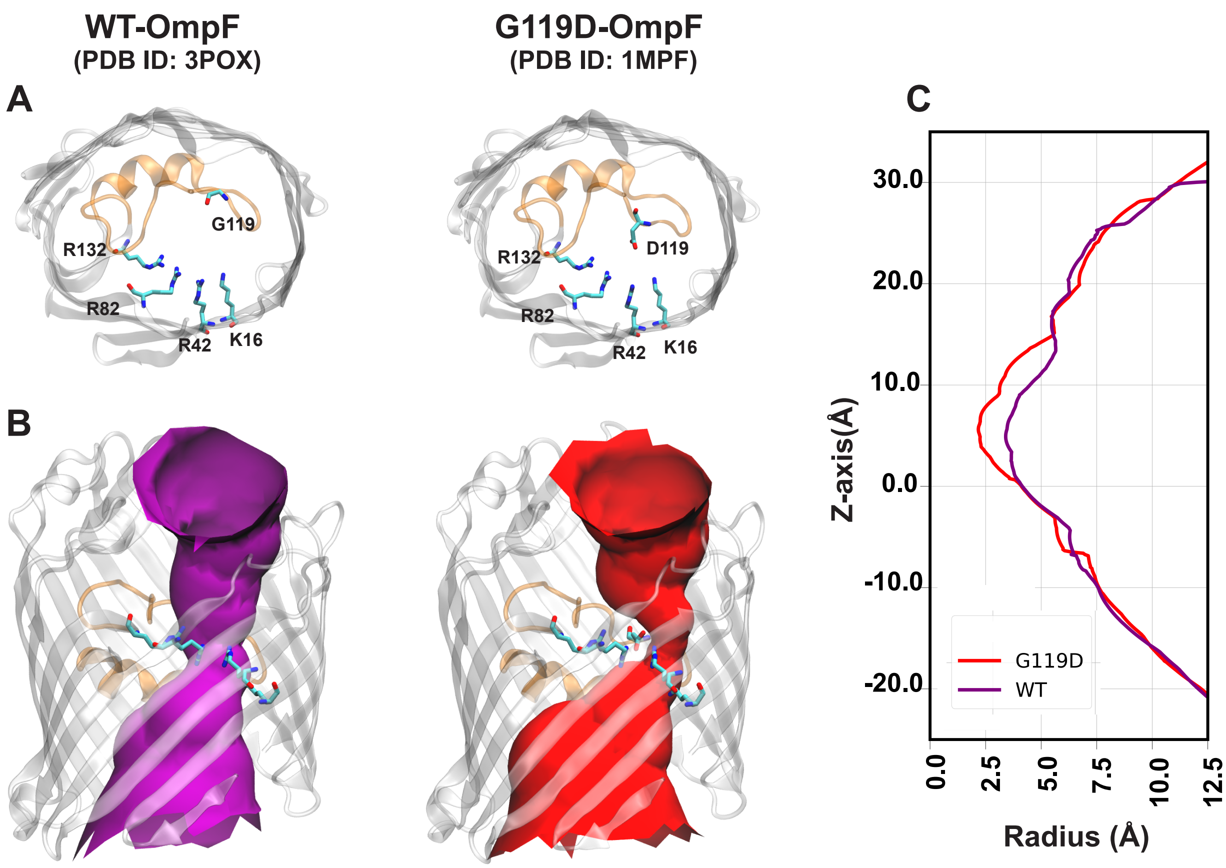

Fig. S1. (A) Top-down view of WT-OmpF and G119D-OmpF crystal structures, highlighting position 119 and the B-face residues. (B) Pore profile (determined using the program HOLE (36)) in WT-OmpF and mutant G119D-OmpF crystal structures. (C) Radius profile of the pore calculated using HOLE (36) for each structure.

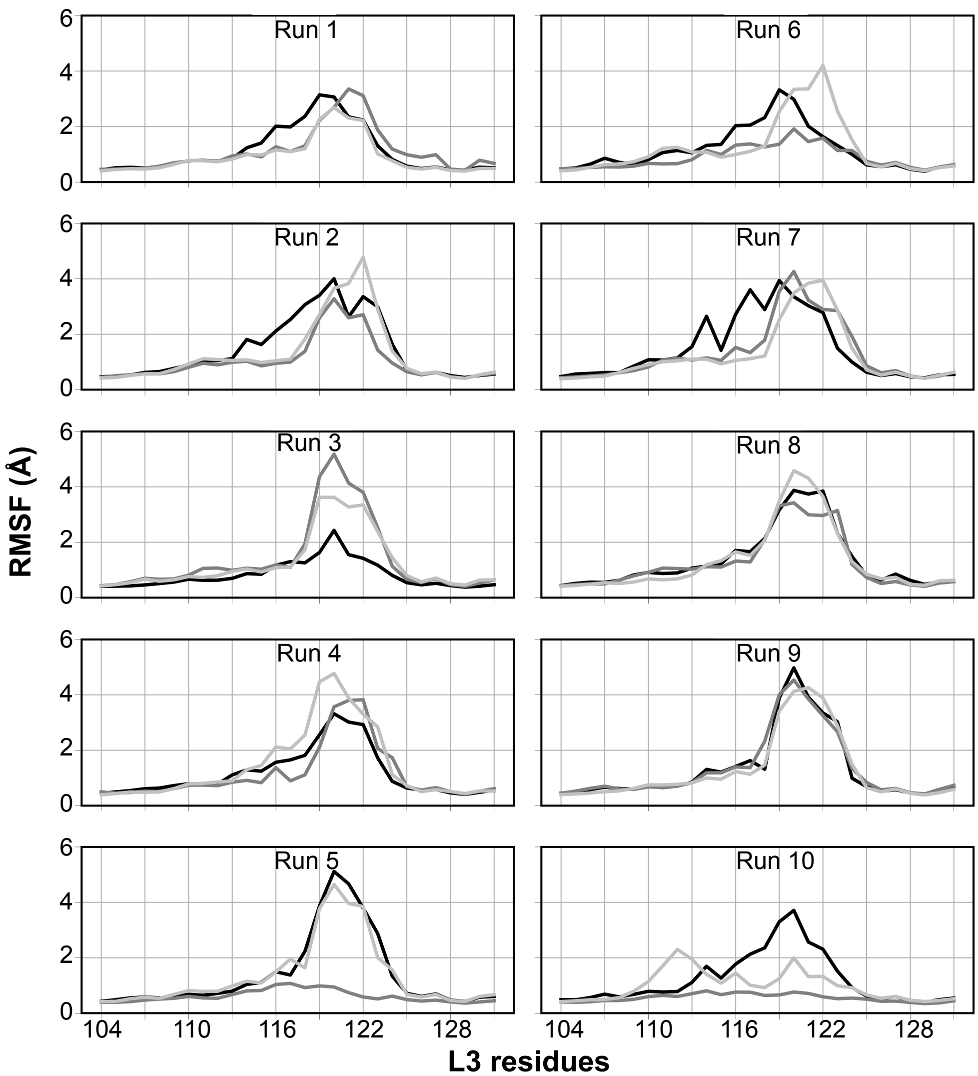

Fig. S2. Root mean-squared fluctuations of L3 residues in each monomer in all 10 replicas.

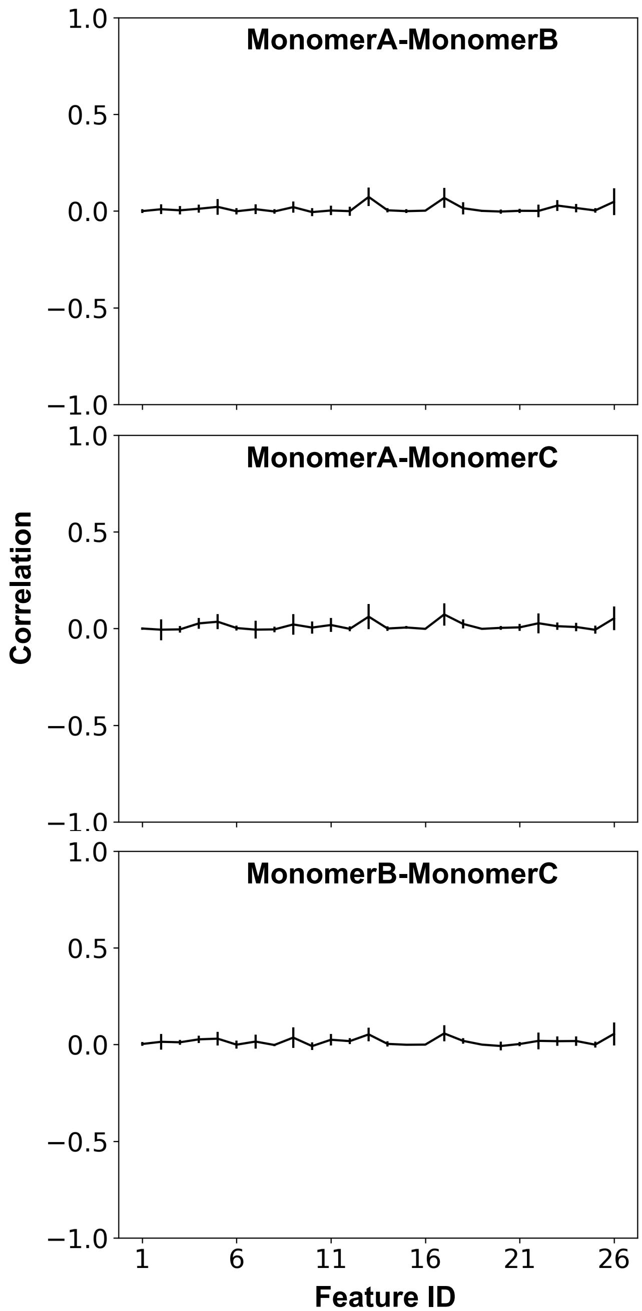

Fig. S3. The time-averaged monomer-monomer correlation coefficient for each distance pair feature. The average and standard deviation of the correlation coefficients from all simulations is shown.

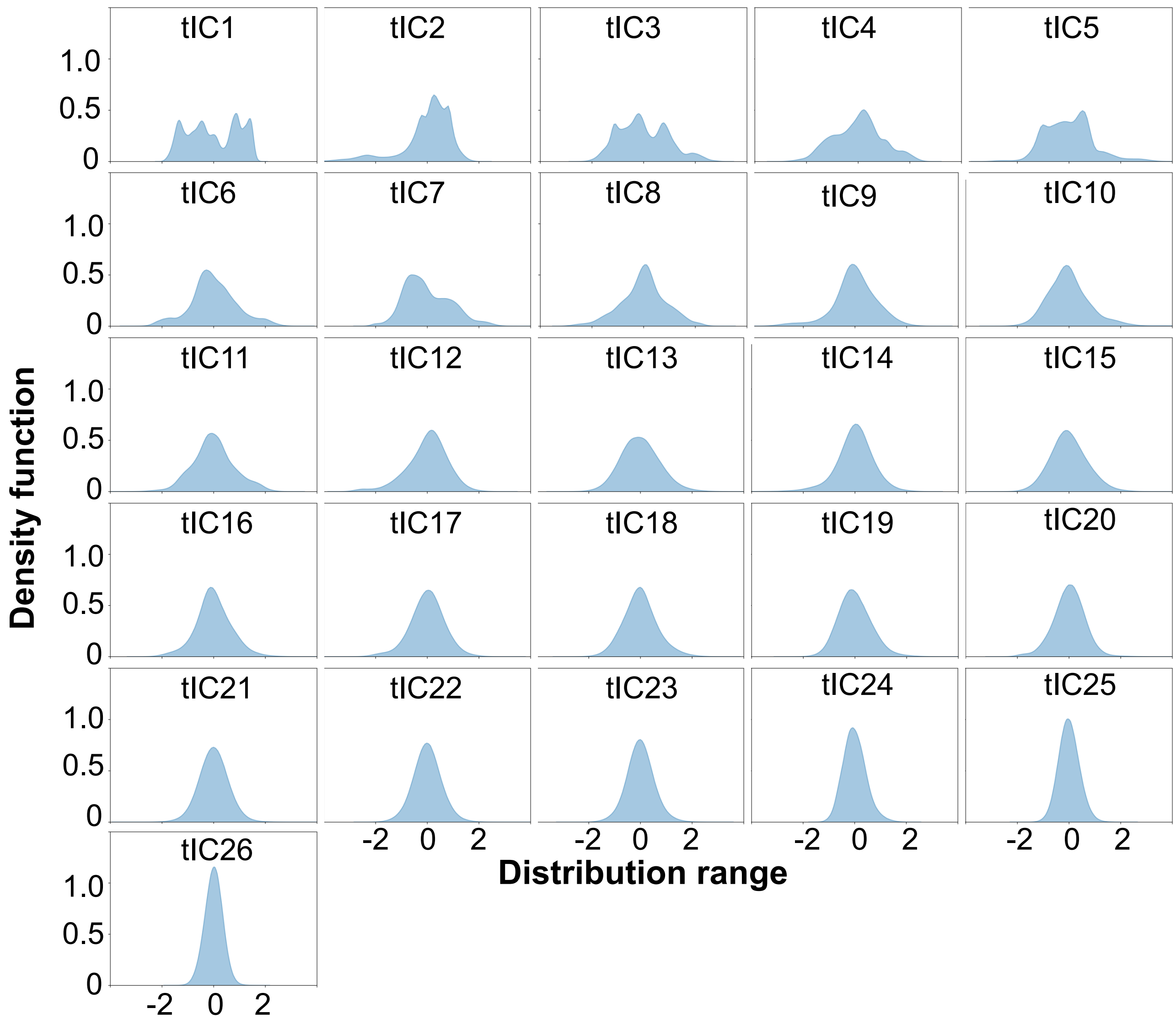

Fig. S4. Distributions of all 26 tICA eigenvectors evaluated from all of the distance pair features. The top seven eigenvectors show a non-Gaussian distribution, and are thus sufficient to build the MSM (37)

 Archit Kumar Vasan, Nandan Haloi, Rebecca Joy Ulrich, Mary Elizabeth Metcalf, Po-Chao Wen, William W. Metcalf, Paul J. Hergenrother, Diwakar Shukla, Emad Tajkhorshid

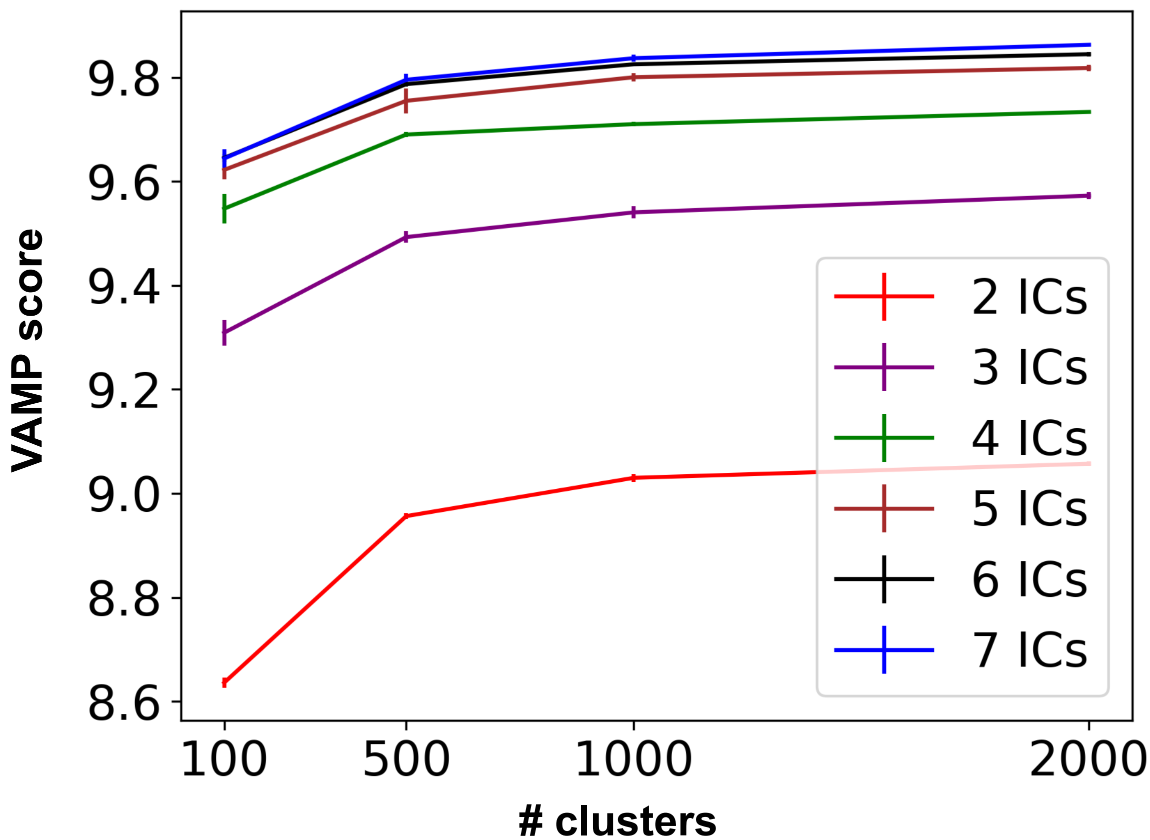

Fig. S5. Cross-validated VAMP-2 (31) score to rank the parameters (number of tICA eigenvectors and clusters) used to discretize the conformational space. Error bars are calculated by determining the VAMP-2 score after running 5 iterations of the k-means clustering algorithm. The score converges with 5 tICs and 1000 clusters, indicating an optimal parameter set.

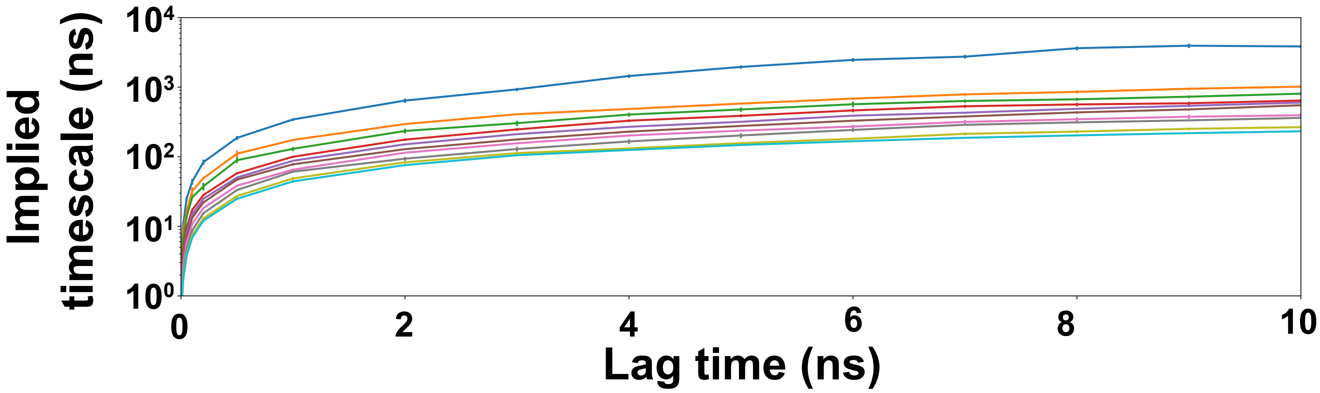

Fig. S6. Implied timescales (ITS) plot of the top 10 slowest processes using multiple lag times. Error bars indicate the uncertainty evaluated using a Bayesian estimated MSM. The lag time of 2 ns was chosen for MSM construction as the ITS curve converges, indicating Markovianity.

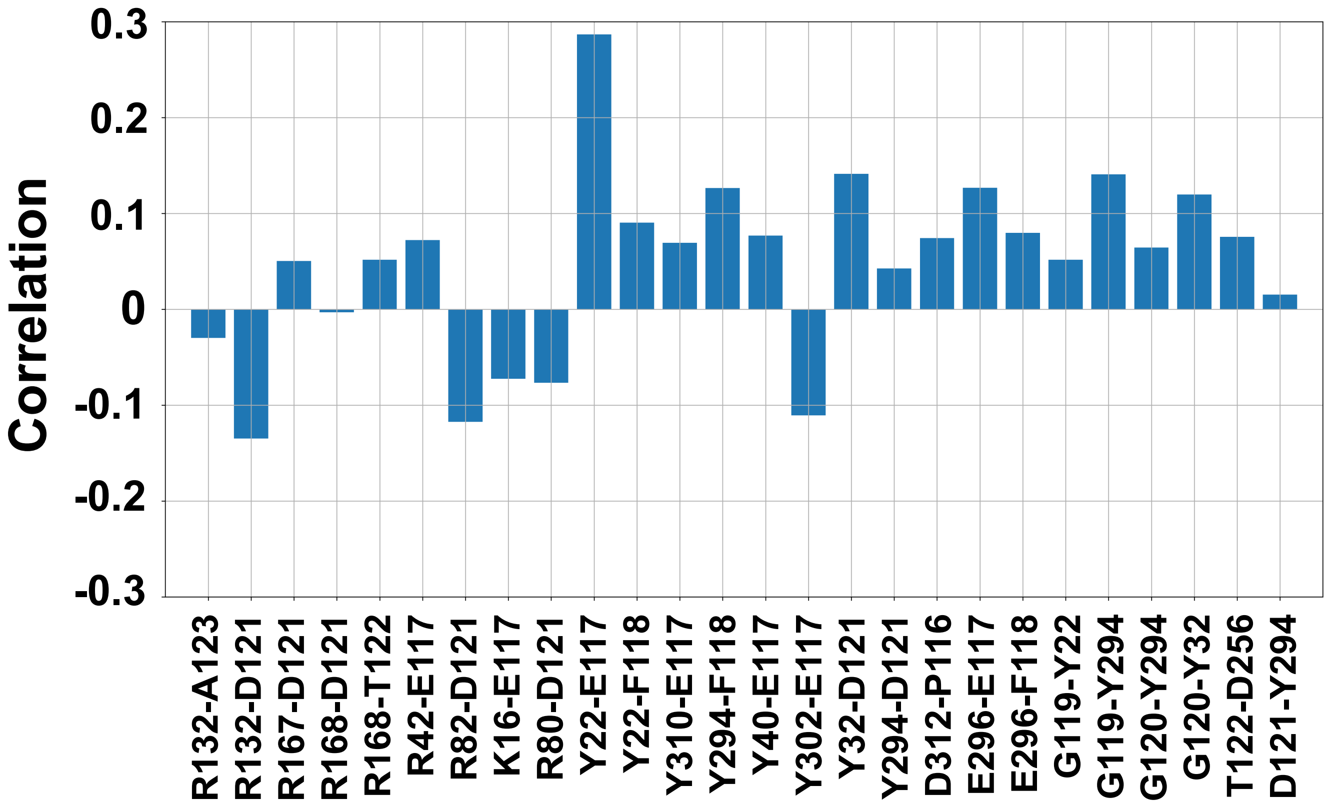

Fig. S7. Time-averaged correlation coefficient of each distance pair with the second eigenvector of the TPM to determine optimal indicators of the slowest transition. The Y22-E117 and R132-D121 distance pairs were chosen as indicators since these features had the greatest positive and negative correlation, respectively.

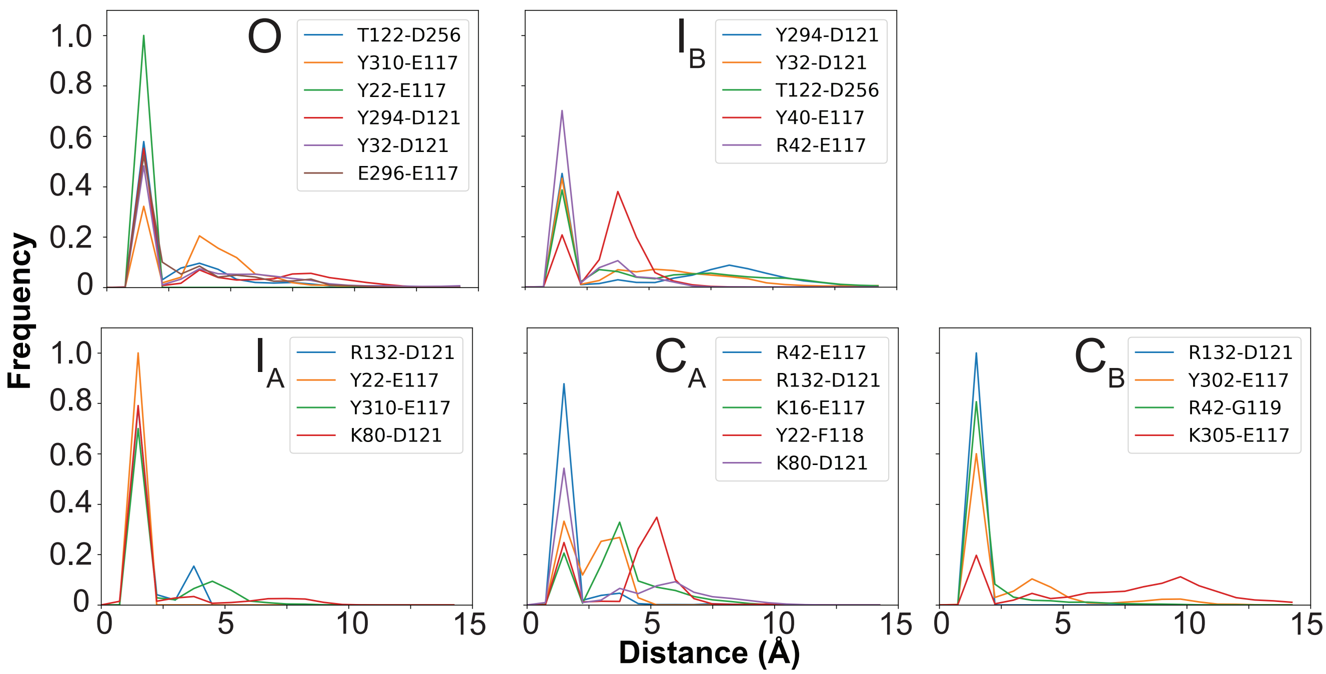

Fig. S8. The ensemble distribution of distances between each major hydrogen bond pair (>20 % occupancy) for the 5 conformational states.

Archit Kumar Vasan, Nandan Haloi, Rebecca Joy Ulrich, Mary Elizabeth Metcalf, Po-Chao Wen, William W. Metcalf, Paul J. Hergenrother, Diwakar Shukla, Emad Tajkhorshid

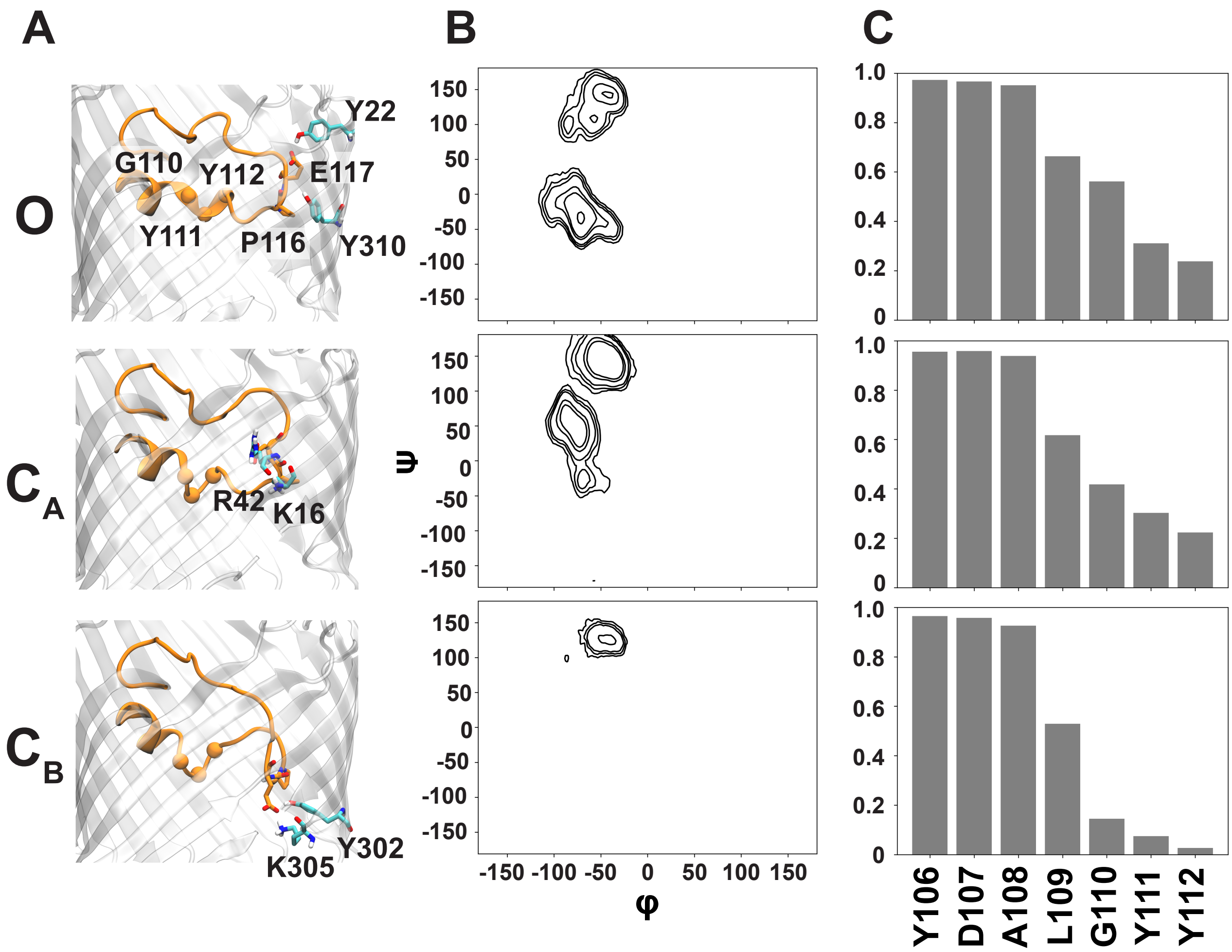

Fig. S9. Molecular explanation for the slower O-C_B kinetics than O-C_A in WT-OmpF. (A) Representative snapshots for O, C_A and C_B, highlighting the conformation of L3 (colored in orange). The C_α atom for residues belonging to the periplasmic terminal of the helical part of of L3 (G110, Y111, Y112) are shown with VDW representation. (B) Distribution of P116 ϕ/ψ angles (in degrees) for the three conformational states, highlighting the conformational restriction of P116 in state C_B. (C) The probability of residues located at the periplasmic terminal of L3 in each state to adopt a helical conformation.

```
                          16    22               40 42
   OmpF:   AEIYNKDGNKVDLYGKAVGLHYFSKGNGENSYGGNGDMTYARLGFKGETQINSDLTGYGQ 60
   OmpC:   AEVYNKDGNKLDLYGKVDGLHYFSDNKD-----VDGDQTYMRLGFKGETQVTDQLTGYGQ 55
   PhoE:   AEIYNKDGNKLDVYGKVKAMHYMSDNAS-----KDGDQSYIRFGFKGETQINDQLTGYGR 55
 OmpE36:   AEIYNKDGNKLDLYGKAVGLHYFSDNDG-----NDGDKTYARLGFKGETKINDQLTGYGQ 55
 OmpK36:   AEIYNKDGNKLDLYGKIDGLHYFSDDKS-----VDGDQTYMRVGVKGETQINDQLTGYGQ 55

                              80                                    116 120
   OmpF:   WEYNFQGNNSEGADAQTGNKTRLAFAGLKYADVGSFDYGRNYGVVYDALGYTDMLPEFGG 120
   OmpC:   WEYQIQGNSAEN-ENN--SWTRVAFAGLKFQDVGSFDYGRNYGVVYDVTSWTDVLPEFGG 112
   PhoE:   WEAEFAGNKAESDTAQ--QKTRLAFAGLKYKDLGSFDYGRNLGALYDVEAWTDMFPEFGG 113
 OmpE36:   WEYNFQGNNSEGADAQSGNKTRLAFAGLKFGDAGSFDYGRNYGLVYDAIGITDMLPEFGG 115
 OmpK36:   WEYNVQANNTESSSDQ--AWTRLAFAGLKFGDAGSFDYGRNYGVVYDVTSWTDVLPEFGG 113
                                                               L3 residues
        121 122      132
   OmpF:   DTAY-SDDFFVGRVGGVATYRNSNFFGLVDGLNFAVQYLGKNER--------------DT 165
   OmpC:   DTYG-SDNFMQQRGNGFATYRNTDFFGLVDGLNFAVQYQGKNGNPSGEGFTSGVTNNGRD 171
   PhoE:   DSSAQTDNFMTKRASGLATYRNTDFFGVIDGLNLTLQYQGKNEN--------------RD 159
 OmpE36:   DTGV-SDNFFSGRTGGLATYRNSGFFGLVDGLNFGVQYLGKNER--------------TD 160
 OmpK36:   DTYG-SDNFLQSRANGVATYRNSDFFGLVDGLNFALQYQGKNGSVSGEG----ATNNGRG 168

   OmpF:   ARRSNGDGVGGSISYE-YEG--FGIVGAYGAADRTNLQEAQ-PLGNGKKAEQWATGLKYD 221
   OmpC:   ALRQNGDGVGGSITYD-YEG--FGIGGAISSSKRTDAQNTAAYIGNGDRAETYTGGLKYD 228
   PhoE:   VKKQNGDGFGTSLTYD-FGGSDFAISGAYTNSDRTNEQNLQ-SRGTGKRAEAWATGLKYD 217
 OmpE36:   ALRSNGDGWATSLSYD-FDG--FGIVGAYGAADRTNAQQNL-QWGKGDKAEQWATGLKYD 216
 OmpK36:   WSKQNGDGFGTSLTYDIWDG--ISAGFAYSHSKRTDEQNSVPALGRGDNAETYTGGLKYD 226

   OmpF:   ANNIYLAANYGETRNATPITNKFTNTSGFANKTQDVLLVAQYQFDFGLRPSIAYTKSKAK 281
   OmpC:   ANNIYLAAQYTQTYNATRVG-----SLGWANKAQNFEAVAQYQFDFGLRPSLAYLQSKGK 283
   PhoE:   ANNIYLATFYSETRKMTPIT------GGFANKTQNFEAVAQYQFDFGLRPSLGYVLSKGK 271
 OmpE36:   ANNIYLAALYGEMRNAARLD------NGFANKTQDFSVVAQYQFDFGLRPSIAYYKSKAK 270
 OmpK36:   ANNIYLASQYTQTYNATRAG-----SLGFANKAQNFEVVAQYQFDFGLRPSVAYLQSKGK 281

                   294 296    302 305  310
   OmpF:   DVE-GIGDVDLVNYFEVGATYYFNKNMSTYVDYIINQIDSDN---KLGVGSDDTVAVGIV 337
   OmpC:   NLGRGYDDEDILKYVDVGATYYFNKNMSTYVDYKINLLDDNQFTRDAGINTDNIVALGLV 343
   PhoE:   DIE-GIGDEDLVNYIDVGATYYFNKNMSAFVDYKINQLDSDN---KLNINNDDIVAVGMT 327
 OmpE36:   DVE-GIGDEDYINYIDIGATYYFNKNMSTYVDYQINQLKDDN---KLGINNDDTVAVGLV 326
 OmpK36:   DLERGYGDQDILKYVDVGATYYFNKNMSTYVDYKINLLDDNSFTRNAGISTDDVVALGLV 341

   OmpF:   YQF 340
   OmpC:   YQF 346
   PhoE:   YQF 330
 OmpE36:   YQF 329
 OmpK36:   YQF 344
```

Fig. S10. Sequence alignment of OmpF (*E. coli*), OmpC (*E. coli*), PhoE (*E. coli*), OmpK36 (*K. pneumoniae*) and OmpE36 (*E. cloacae*). Each residue is represented by their one-letter abbreviation. Important residues in the observed gating mechanism of OmpF are highlighted with a box and colored based on residue type (black: hydrophobic, green: polar, red: acidic, and blue: basic). All other residues are colored gray. The alignment was performed on full-length sequences using the Clustal Omega (38) alignment tool within Uniprot (39).

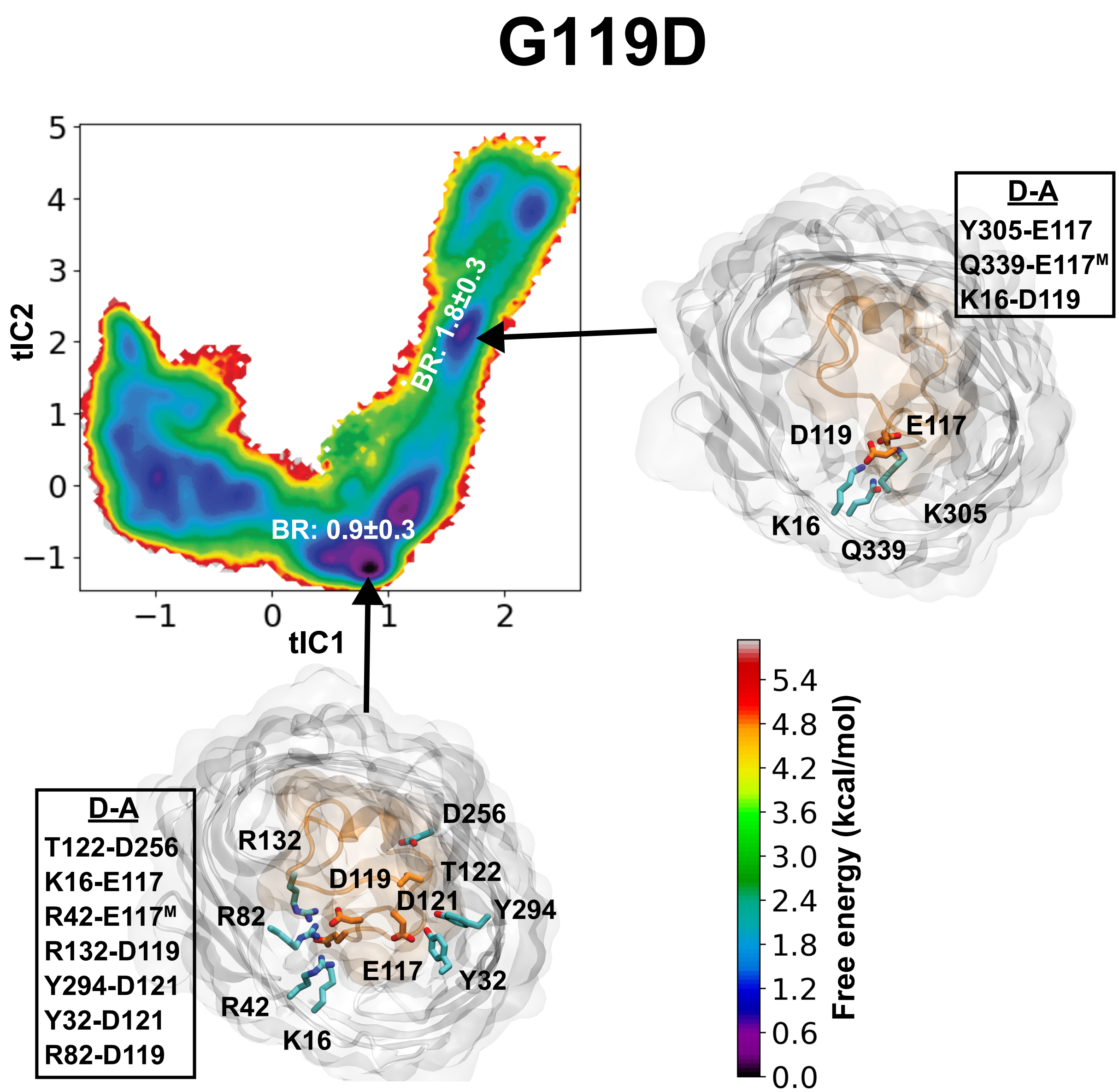

Fig. S11. Conformational landscape of L3 in G119D-OmpF mutant. Free energy landscape for dynamics of L3, reweighted by the stationary distribution, is projected onto the top two tICA eigenvectors. The pore bottleneck radii (BR) for the conformational states corresponding to energetic minima are highlighted on the free energy surface. Structural characteristics in each metastable state are depicted by the top-down snapshots of OmpF, highlighting hydrogen bonds with > 20% occurrence probability between the most fluctuating residues of L3 and the barrel residues.

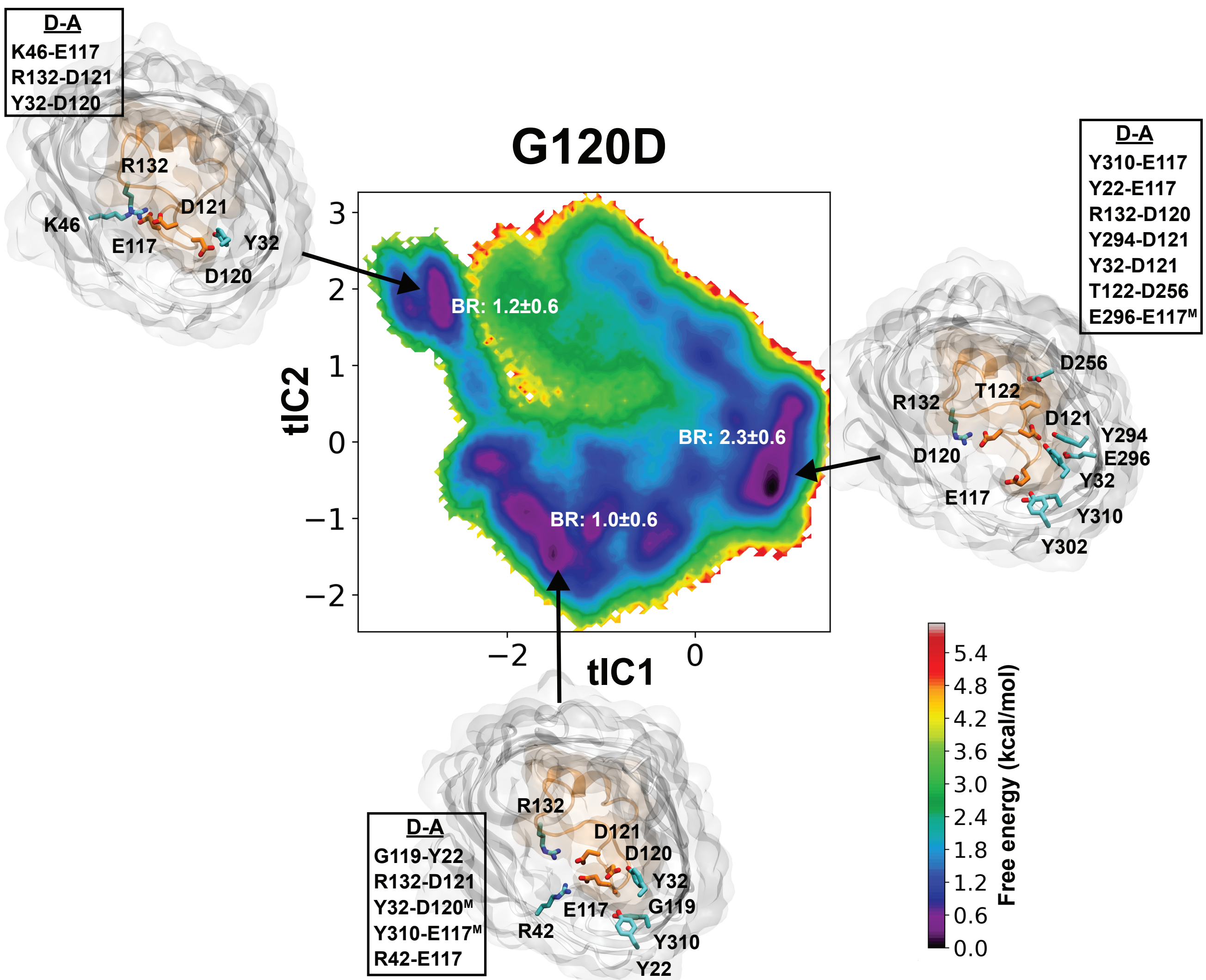

Fig. S12. Conformational landscape of L3 in G120D-OmpF mutant. Free energy landscape for dynamics of L3, reweighted by the stationary distribution, is projected onto the top two tICA eigenvectors. The pore bottleneck radii (BR) for the conformational states corresponding to energetic minima are highlighted on the free energy surface. Structural characteristics in each metastable state are depicted by the top-down snapshots of OmpF, highlighting hydrogen bonds with > 20% occurrence probability between the most fluctuating residues of L3 and the barrel residues.

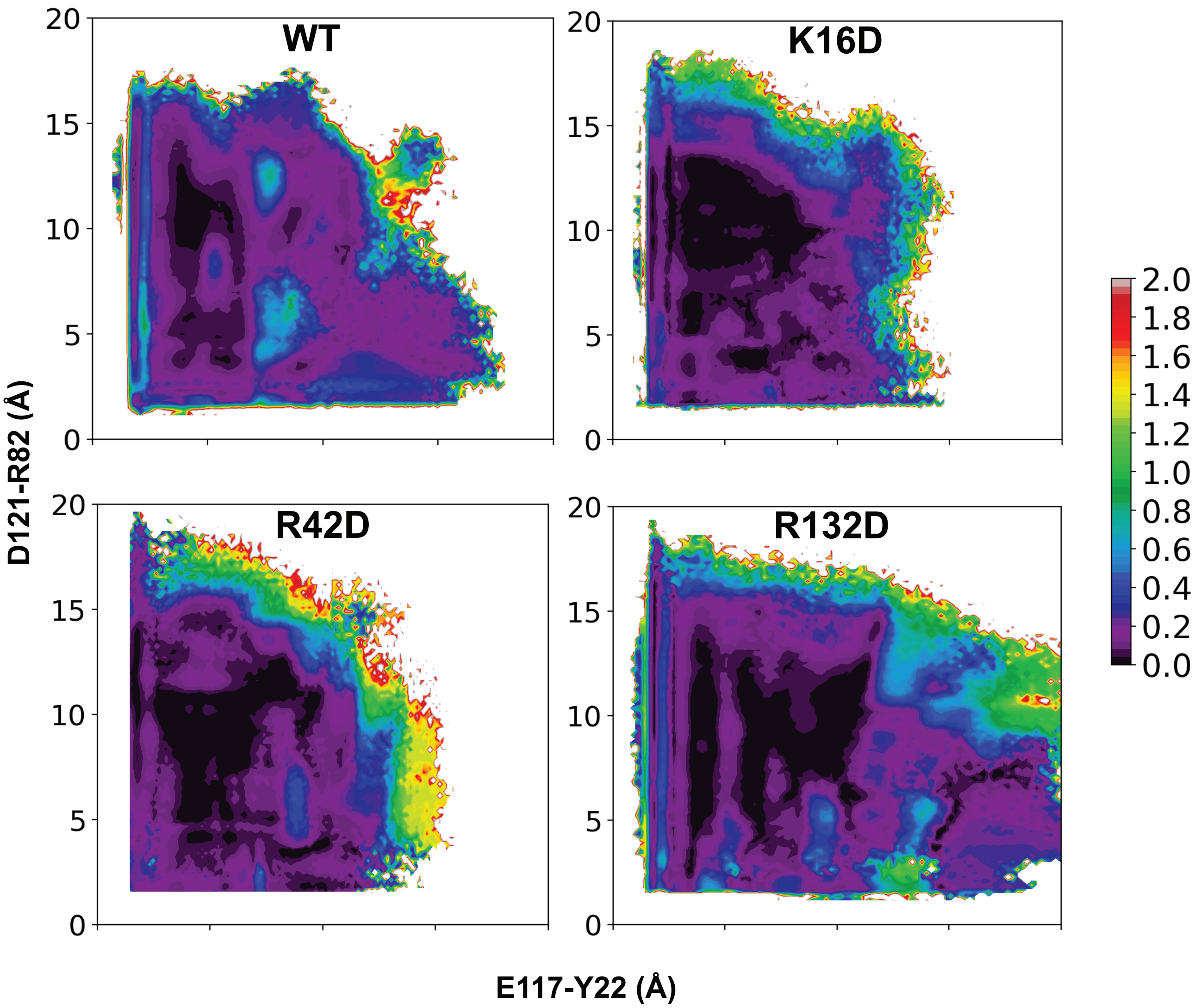

Fig. S13. Free energy error for WT, K16D, R42D, and R132D-OmpF systems. The landscape is projected onto E117-Y22 and R132-D121 for WT-OmpF and K16D-OmpF, and E117-Y22 and R82-D121 for R132D-OmpF.

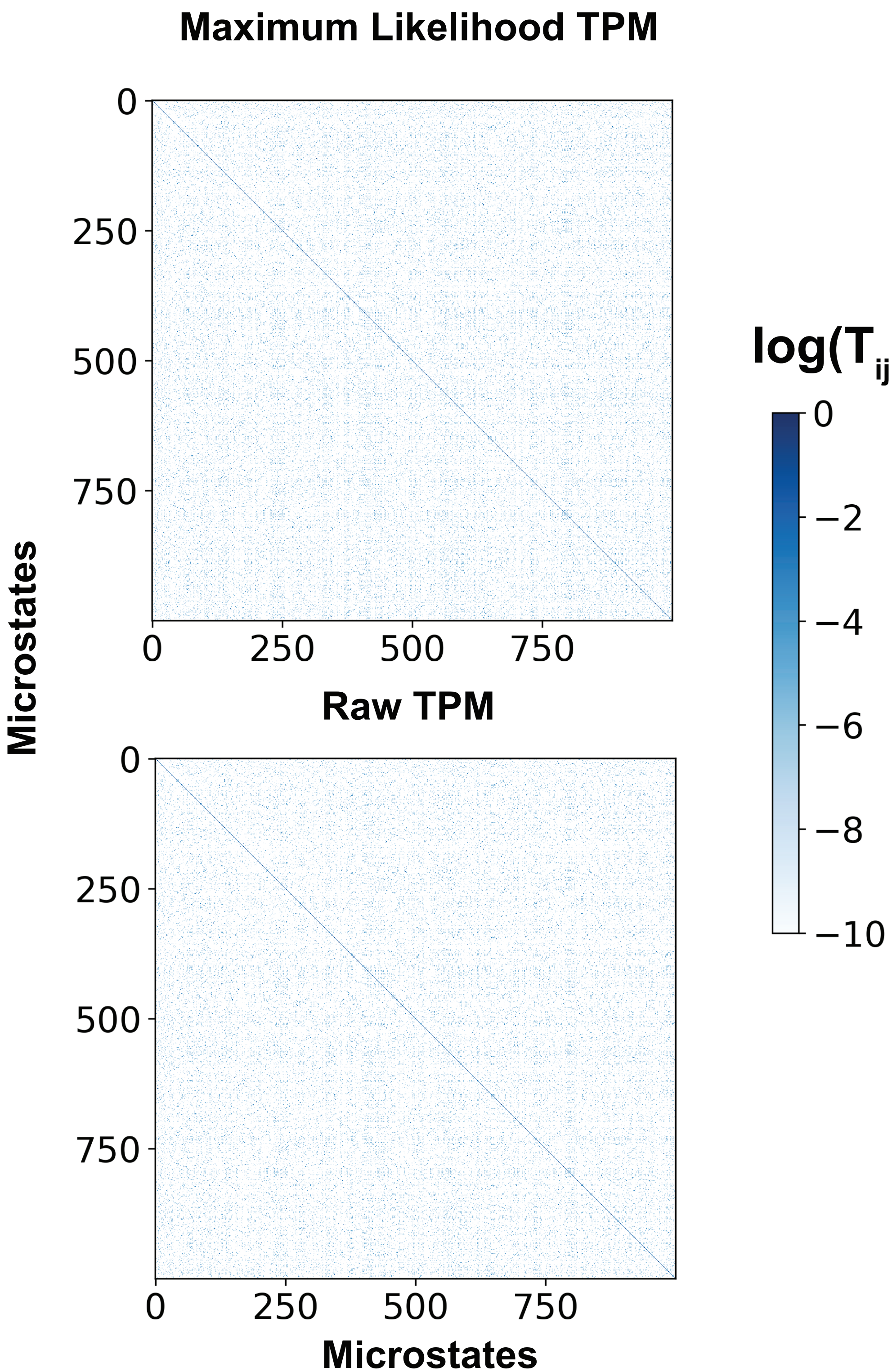

Fig. S14. Comparison of the TPM computed with maximum likelihood MSM and with the raw trajectory data. Shown is the log of each matrix element. No significant difference is observed.

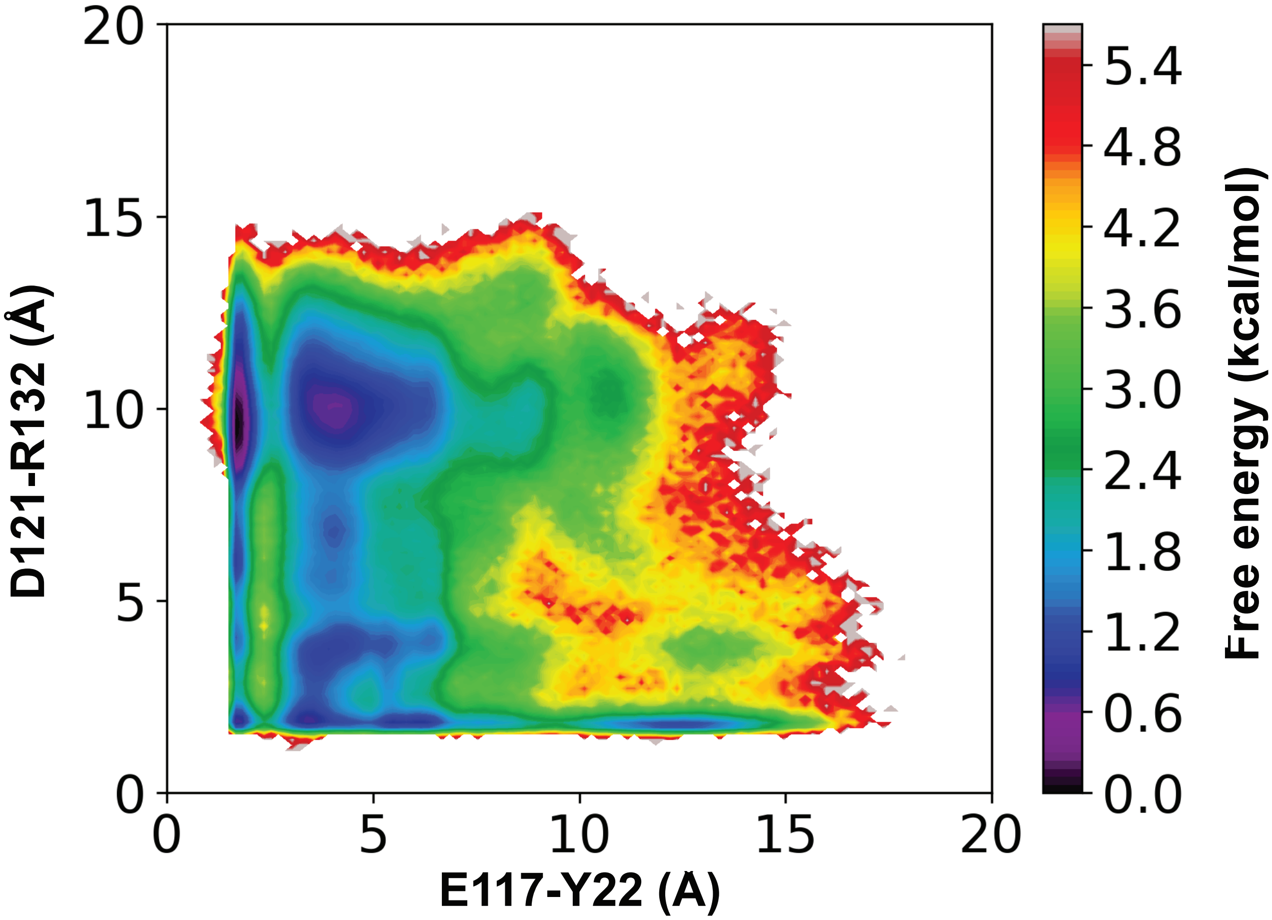

Fig. S15. The conformational landscape of L3 projected onto the E117-Y22 and D121-R132 distance features using the raw trajectory data.

Table S1. *Escherichia coli* strains used in the study

Strain	Genotype
BW26678	*lacI^q* *rrnB*3 Δ*lacZ*4787 *hsdR*514 Δ*araBAD*567 Δ*rhaBAD*568 *rph*-1/pKD46
WM8897	*lacI^q* *rrnB*3 Δ*lacZ*4787 *hsdR*514 Δ*araBAD*567 Δ*rhaBAD*568 *rph*-1 Δ*ompF*8897::*cat*
WM8901	*lacI^q* *rrnB*3 Δ*lacZ*4787 *hsdR*514 Δ*araBAD*567 Δ*rhaBAD*568 *rph*-1 Δ*ompF*8897::*cat*/pAH69
WM8819	*lacI^q* *rrnB*3 Δ*lacZ*4787 *hsdR*514 Δ*araBAD*567 Δ*rhaBAD*568 *rph*-1 Δ*ompF*8897::*cat* *att*-HK::pMEM503
WM8820	*lacI^q* *rrnB*3 Δ*lacZ*4787 *hsdR*514 Δ*araBAD*567 Δ*rhaBAD*568 *rph*-1 Δ*ompF*8897::*cat* *att*-HK::pMEM504
WM8826	*lacI^q* *rrnB*3 Δ*lacZ*4787 *hsdR*514 Δ*araBAD*567 Δ*rhaBAD*568 *rph*-1 Δ*ompF*8897::*cat* *att*-HK::pMEM501

Table S2. Plasmids, carrying various *ompF* alleles, used in this study

Plasmid	Description	Construction or Reference
pKD3	Source of *cat*-cassette used for ompF deletion	Datsenko and Wanner (32)
pAH144	*att*-HK integration plasmid, Strep/Spec^R	Haldimann and Wanner (33)
pAH69	Temperature-sensitive helper plasmid for chromosomal integration of *att*-HK plasmids	Haldimann and Wanner (33)
pMEM501	pAH144::*ompF*(wt)	Vector pAH144 cut with EcoRI-HF and SphI-HF (no SAP) ligated with PCR amplified *ompF*(wt) using primers ompF-cloningF and ompFcloningR
pMEM503	pAH144::*ompF*(G120D)	Hi Fi Assembly of MluI/PvuII-cut pMEM501 and ompF-G120D gBlock
pMEM504	pAH144::*ompF*(R132D)	Hi Fi Assembly of MluI/PvuII-cut pMEM501 and ompF-R132D gBlock

Archit Kumar Vasan, Nandan Haloi, Rebecca Joy Ulrich, Mary Elizabeth Metcalf, Po-Chao Wen, William W. Metcalf, Paul J. Hergenrother, Diwakar Shukla, Emad Tajkhorshid

Table S3. Primers used in this study

Name	Purpose	Sequence
ompF-cloningF	cloning *ompF*(wt)	GGCGCGCCGCATGCTTCCGTTCCCACGTACTCCG
ompF-cloningR	cloning *ompF*(wt)	GGCGCGCCGAATTCCAGGAGCGGCGGTAATGTTC
del-ompF-F	amplification of *cat*-cassette used for construction of Δ*ompF* mutant	AGATTTTGTGCCAGGTCGATAAA-GTTTCCATCAGAAACAAGTGTAGGCTGGAGCTGCTTC
del-ompF-R	amplification of *cat*-cassette used for construction of Δ*ompF* mutant	GTCCTGTTTTTTCGGCATTTAAC-AAAGAGGTGTGCTATTACATATGAATATCCTCCTTAG
HK022-P1	verification of single copy plasmid integration at *att*-HK	GGAATCAATGCCTGAGTG
HK022-P2	verification of single copy plasmid integration at *att*-HK	GGCATCAACAGCACATTC
HK022-P3	verification of single copy plasmid integration at *att*-HK.	ACTTAACGGCTGACATGG
HK022-P4	verification of single copy plasmid integration at *att*-HK	ACGAGTATCGAGATGGCA

Table S4. gBlocks used for construction of *att*-HK plasmids carrying mutated *ompF* alleles

Name	
ompF-G120D	CTCTGAAGGCGCTGACGCTCAAACTGGTAACAAAACGCGTCTGGCATTCGCGGGTCTTAAATACG-CTGACGTTGGTTCTTTCGATTACGGCCGTAACTACGGTGTGGTTTATGATGCACTGGGTTACACC-GATATGCTGCCAGAATTTGGTGATGATACTGCATACAGCGATGACTTCTTCGTTGGTCGTGTTGGCG-GCGTTGCTACCTATCGTAACTCCAACTTCTTTGGTCTGGTTGATGGCCTGAACTTCGCTGTTCA-GTACCTGGGTAAAAACGAGCGTGACACTGCACGCCGTTCTAACGGCGACGGTGTTGGCGGTTCTATC-AGCTACGAATACGAAGGCTTTGGTATCGTTGGTGCTTATGGTGCAGCTGACCGTACCAACCTGCAAG-AAGCTCAACCTCTT
ompF-R132D	CTCTGAAGGCGCTGACGCTCAAACTGGTAACAAAACGCGTCTGGCATTCGCGGGTCTTAAATACG-CTGACGTTGGTTCTTTCGATTACGGCCGTAACTACGGTGTGGTTTATGATGCACTGGGTTACACCGATAT-GCTGCCAGAATTTGGTGGTGATACTGCATACAGCGATGACTTCTTCGTTGGTGATGTTGGCGGCGT-TGCTACCTATCGTAACTCCAACTTCTTTGGTCTGGTTGATGGCCTGAACTTCGCTGTTCAGT-ACCTGGGTAAAAACGAGCGTGACACTGCACGCCGTTCTAACGGCGACGGTGTTGGCGGTTCTATCAGCTAC-GAATACGAAGGCTTTGGTATCGTTGGTGCTTATGGTGCAGCTGACCGTACCAACCTGCAAGAAGCT-CAACCTCTT

References

1. Efremov, R.; Sazanov, L. Structure of *Escherichia coli* OmpF porin from lipidic mesophase. *Journal of Structural Biology* **2012**, *178*, 311–318.
2. Im, W.; Roux, B. Ions and counterions in a biological channel: a molecular dynamics study of OmpF porin from *Escherichia coli* in an explicit membrane with 1 M KCl aqueous salt solution. *Journal of Molecular Biology* **2002**, *319*, 1177–1197.
3. Im, W.; Roux, B. Ion permeation and selectivity of OmpF Porin: A theoretical study based on Molecular Dynamics, Brownian Dynamics, and Continuum Electrodiffusion theory. *J. Mol. Biol.* **2002**, *322*, 851–869.
4. Varma, S.; Chiu, S.-W.; Jakobsson, E. The influence of amino acid protonation states on molecular dynamics simulations of the bacterial porin OmpF. *Biophysical Journal* **2006**, *90*, 112–123.
5. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *Journal of Chemical Physics* **1983**, *79*, 926–935.
6. Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: a Web-based Graphical User Interface for CHARMM. *Journal of Computational Chemistry* **2008**, *29*, 1859–1865.
7. Torrie, G. M.; Valleau, J. P. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella Sampling. *Journal of Computational Physics* **1977**, *23*, 187–199.
8. Sugita, Y.; Okamoto, Y. Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters* **1999**, *314*, 141–151.
9. Sugita, Y.; Kitao, A.; Okamoto, Y. Multidimensional replica-exchange method for free-energy calculations. *Journal of Chemical Physics* **2000**, *113*, 6042–6051.
10. Moradi, M.; Tajkhorshid, E. Mechanistic picture for conformational transition of a membrane transporter at atomic resolution. *Proceedings of the National Academy of Sciences, USA* **2013**, *110*, 18916–18921.
11. Moradi, M.; Tajkhorshid, E. Computational recipe for efficient description of large-scale conformational changes in biomolecular systems. *Journal of Chemical Theory and Computation* **2014**, *10*, 2866–2880.
12. Vanommeslaeghe, K.; MacKerell, Jr., A. D. Automation of the CHARMM General Force Field (CGenFF) I: bond perception and atom typing. *Journal of Chemical Information and Modeling* **2012**, *52*, 3144–3154.
13. Vanommeslaeghe, K.; Raman, E. P.; MacKerell, Jr., A. D. Automation of the CHARMM General Force Field (CGenFF) II: assignment of bonded parameters and partial atomic charges. *Journal of Chemical Information and Modeling* **2012**, *52*, 3155–3168.
14. Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; MacKerell, Jr., A. D. CHARMM General Force Field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. *Journal of Computational Chemistry* **2010**, *31*, 671–690.
15. Haloi, N.; Vasan, A. K.; Geddes, E. J.; Prasanna, A.; Wen, P.-C.; Metcalf, W. W.; Hergenrother, P. J.; Tajkhorshid, E. Rationalizing generation of broad spectrum antibiotics with the addition of a positive charge. *Chemical Science* **2021**, *12*, 15028–15044.
16. Kumar, S.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A.; Rosenberg, J. M. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry* **1992**, *13*, 1011–1021.
17. Bartels, C. Analyzing biased Monte Carlo and molecular dynamics simulations. *Chemical Physics Letters* **2000**, *331*, 446–454.
18. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **2005**, *26*, 1781–1802.
19. Phillips, J. C. et al. Scalable molecular dynamics on CPU and GPU architectures with NAMD. *Journal of Chemical Physics* **2020**, *153*, 044130.
20. Huang, J.; Rauscher, S.; Nawrocki, G.; Ran, T.; Feig, M.; de Groot, B. L.; Grubmüller, H.; MacKerell Jr, A. D. CHARMM36m: An improved force field for folded and intrinsically disordered proteins. *Nature Methods* **2017**, *14*, 71–73.
21. Klauda, J. B.; Venable, R. M.; Freites, J. A.; O'Connor, J. W.; Tobias, D. J.; Mondragon-Ramirez, C.; Vorobyov, I.; MacKerell Jr., A. D.; Pastor, R. W. Update of the CHARMM all-atom additive force field for lipids: Validation on six lipid types. *Journal of Physical Chemistry B* **2010**, *114*, 7830–7843.
22. Darden, T.; York, D.; Pedersen, L. Particle mesh Ewald: an $N \cdot \log(N)$ method for Ewald sums in large systems. *Journal of Chemical Physics* **1993**, *98*, 10089–10092.
23. Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant pressure molecular dynamics algorithms. *Journal of Chemical Physics* **1994**, *101*, 4177–4189.
24. Feller, S. E.; Zhang, Y.; Pastor, R. W. Constant pressure molecular dynamics simulation: the Langevin piston method. *Journal of Chemical Physics* **1995**, *103*, 4613–4621.
25. Scherer, M. K.; Trendelkamp-Schroer, B.; Paul, F.; Pérez-Hernández, G.; Hoffmann, M.; Plattner, N.; Wehmeyer, C.; Prinz, J.-H.; Noé, F. PyEMMA 2: A software package for estimation, validation, and analysis of Markov models. *Journal of Chemical Theory and Computation* **2015**, *11*, 5525–5542.
26. Schwantes, C. R.; Pande, V. S. Improvements in Markov state model construction reveal many non-native interactions in the folding of NTL9. *Journal of Chemical Theory and Computation* **2013**, *9*, 2000–2009.
27. Schwantes, C. R.; Shukla, D.; Pande, V. S. Markov State Models and tICA Reveal a Nonnative Folding Nucleus in Simulations of NuG2. *Biophysical Journal* **2016**, *110*, 1716–1719.
28. Molgedey, L.; Schuster, H. G. Separation of a mixture of independent signals using time delayed correlations. *Physical*

Review Letters **1994**, *72*, 3634.
29. Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. Identification of slow molecular order parameters for Markov model construction. *Journal of Chemical Physics* **2013**, *139*, 07B604_1.
30. Lyman, E.; Zuckerman, D. M. On the structural convergence of biomolecular simulations by determination of the effective sample size. *Journal of Physical Chemistry B* **2007**, *111*, 12876–12882.
31. Wu, H.; Noé, F. Variational approach for learning Markov processes from time series data. *Journal of Nonlinear Science* **2020**, *30*, 23–66.
32. Datsenko, K. A.; Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences, USA* **2000**, *97*, 6640–6645.
33. Haldimann, A.; Wanner, B. L. Conditional-replication, integration, excision, and retrieval plasmid-host systems for gene structure-function studies of bacteria. *Journal of Bacteriology* **2001**, *183*, 6384–6393.
34. Richter, M. F.; Drown, B. S.; Riley, A. P.; Garcia, A.; Shirai, T.; Svec, R. L.; Hergenrother, P. J. Predictive compound accumulation rules yield a broad-spectrum antibiotic. *Nature* **2017**, *545*, 299–304.
35. Geddes, E.; Li, Z.; Hergenrother, P. A LC-MS/MS assay and complementary web-based tool to quantify and predict compound accumulation in *E. coli*. *Nature Protocols* In Press.
36. Smart, O. S.; Neduvelil, J. G.; Wang, X.; Wallace, B.; Sansom, M. S. HOLE: a program for the analysis of the pore dimensions of ion channel structural models. *Journal of Molecular Graphics* **1996**, *14*, 354–360.
37. Razavi, A. M.; Khelashvili, G.; Weinstein, H. A Markov state-based quantitative kinetic model of sodium release from the dopamine transporter. *Scientific Reports* **2017**, *7*, 40076.
38. Sieversi, F.; Wilm, A.; Dineen, D.; Gibson, T. J.; Karplus, K.; Li, W.; Lopez, R.; McWilliam, H.; Remmert, M.; Söding, J.; Thompson, J. D.; Higgins, D. G. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Molecular Systems Biology* **2011**, *7*, 539.
39. The UniProt Consortium, UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Research* **2019**, *47*, D506–D515.