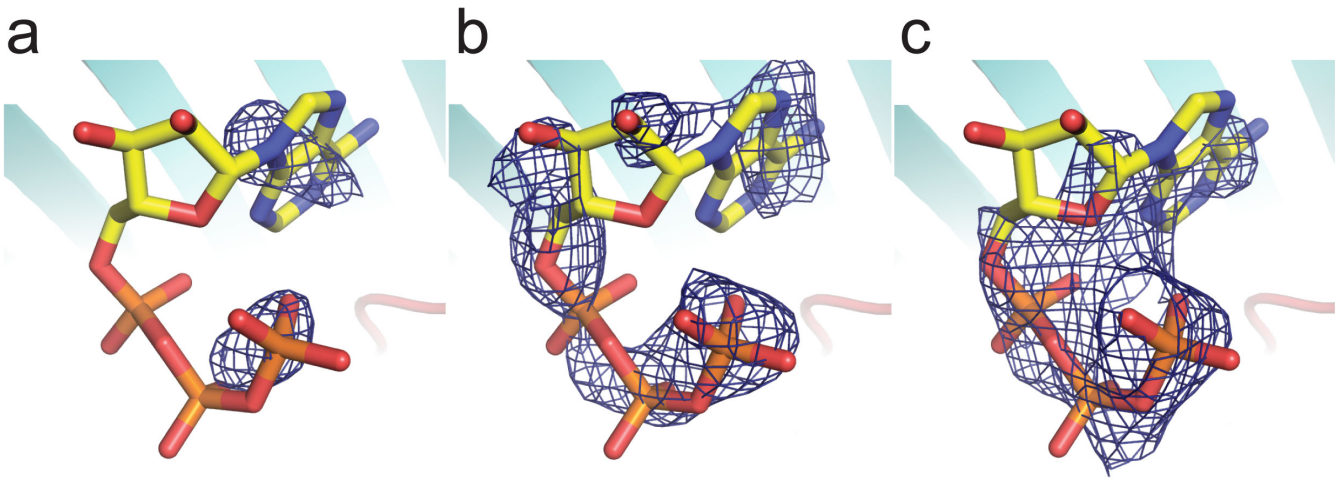
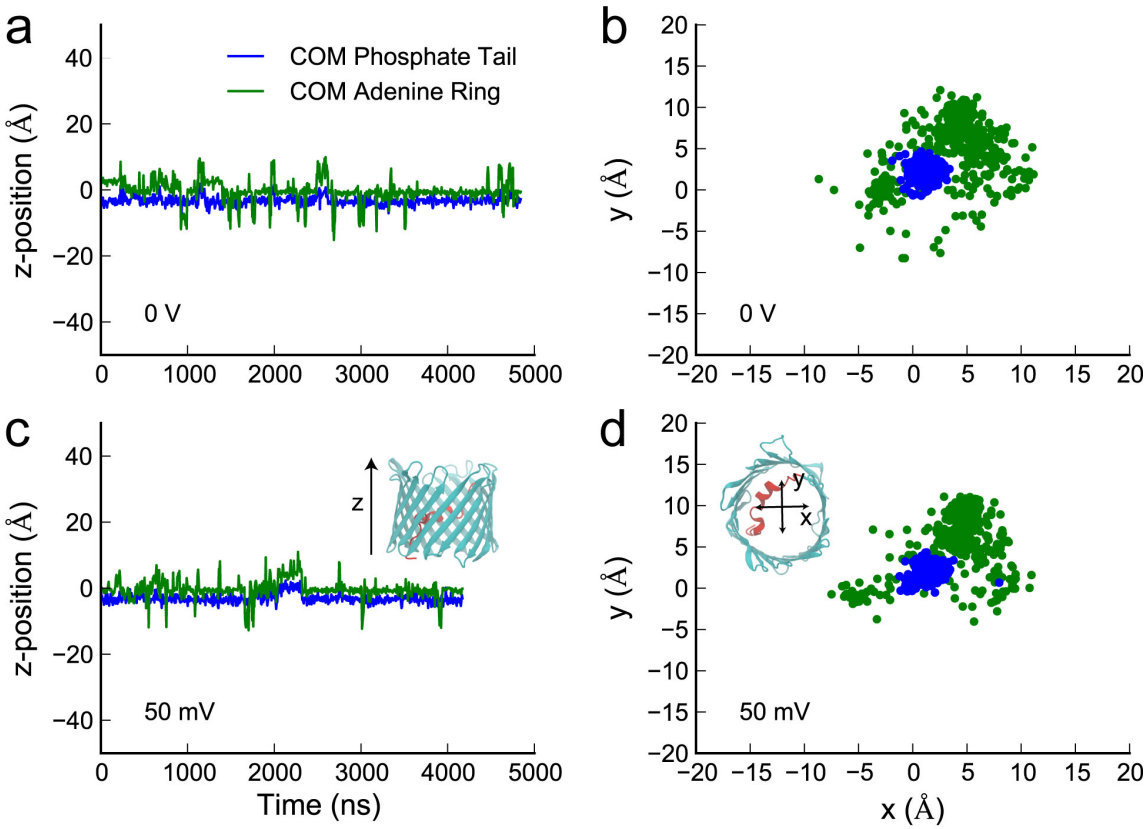


Supplementary Figure 1

Electron density maps of the mVDAC1–ATP complex.

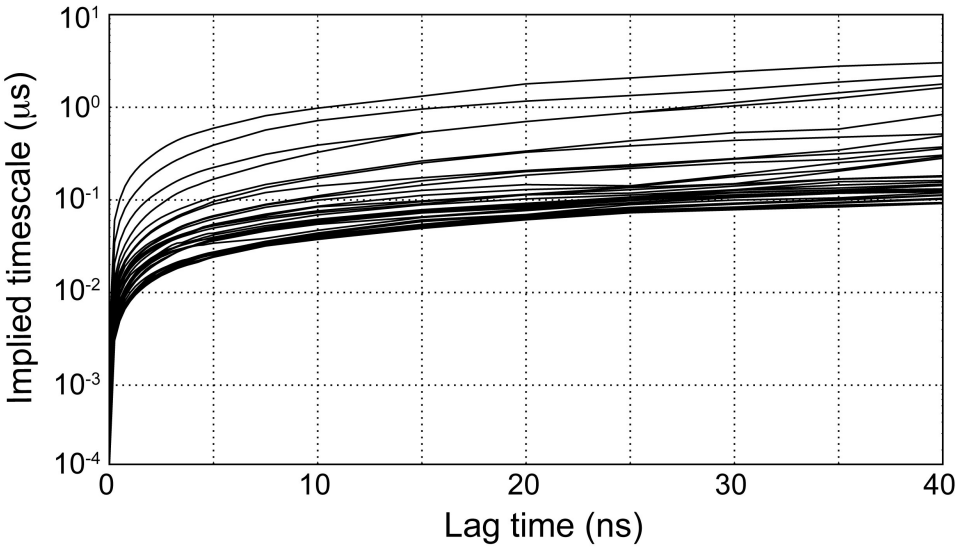
(**a**) A Fo^{ATP}-Fo^{Native} difference Fourier map calculated between mVDAC1 ATP co-crystal and an apo crystal contoured at 3.1 σ displayed on the refined complex coordinates. (**b**) A solvent annealed omit map for the ATP molecule contoured at 3.3 σ. (**c**) A 2Fo-Fc electron density map of the refined model contoured at 1 σ. Color scheme is the same as in Fig. 1 of the main text. See Supporting Online Material for details.



Supplementary Figure 2

ATP fails to permeate mVDAC1 on the microsecond timescale.

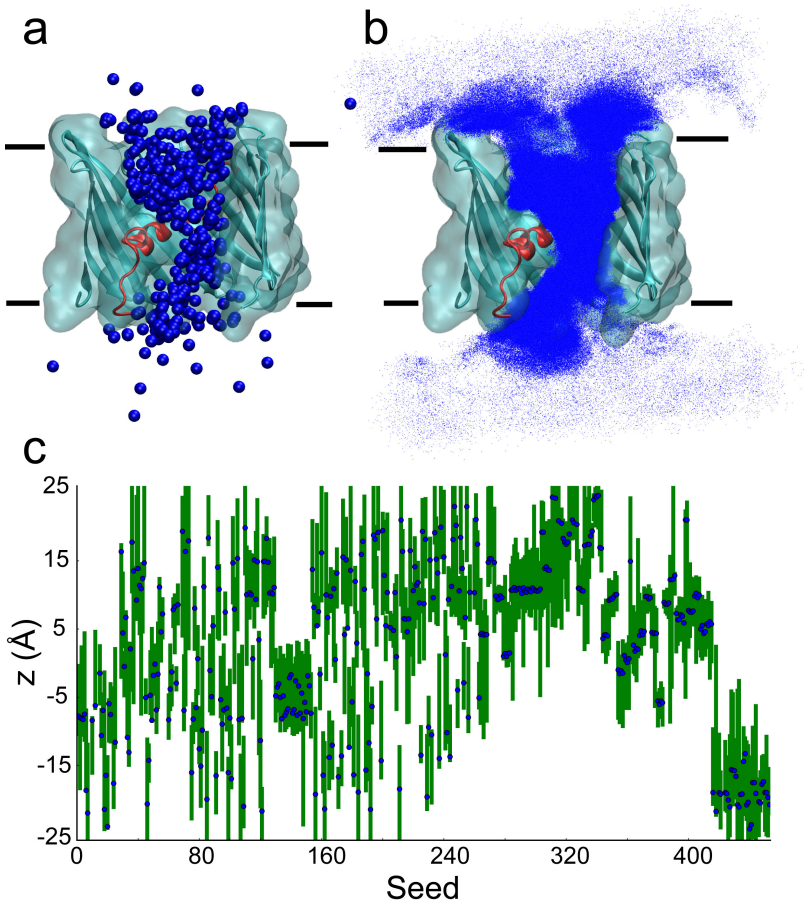
(**a,c**) The *z*-position of the COM of the triphosphate tail and adenine ring are plotted versus time for two simulations run with either a 0 or 50 mV electric field. The channel was centered at $z = 0$ and aligned along the *z*-axis. (**b,d**) The same trajectories, projected onto the *x*-*y* plane. For the simulation performed with a 50 mV field, the ATP was initiated in the upper bath; the simulation at 0 mV was initiated in the pore. Despite their different initial positions within the simulation box and applied voltage, the ATP in both simulations bind to the N-terminal helix and sample a similar set of conformations. The phosphate tail remains largely immobile over the full length of the simulation, whereas the adenine ring interacts with a number of moieties within the channel.



Supplementary Figure 3

Implied timescales for the MSM.

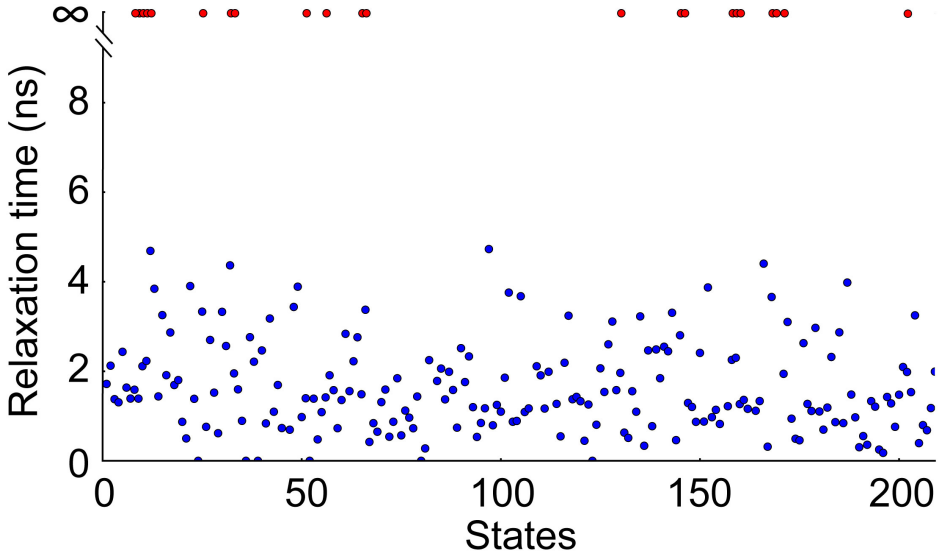
The 50 slowest implied timescales are plotted on a log scale for a range of lag times, τ. Implied timescales were computed using Eq. 7, and the transition probability matrix obtained from the full 40 μs dataset coupled to the continuum bath as described in Supplementary Note. The lag time chosen for all analysis in the main text was 5 ns, since the timescales plateau at this point indicating that the model is Markovian.



Supplementary Figure 4

Initial placement and ATP coverage of the channel for the MSM construction.

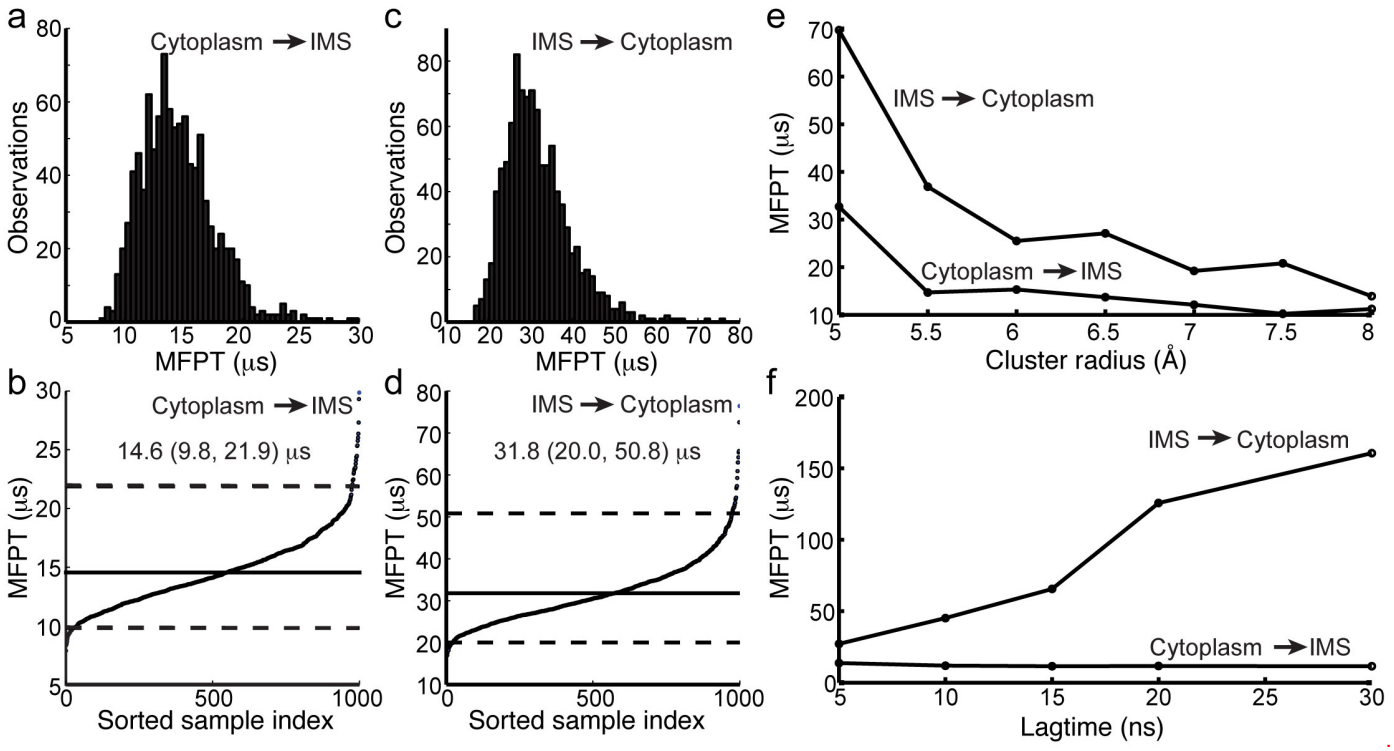
(**a**) Blue spheres represent the β-phosphate atom of all 453 ATP configurations used to initiate MD simulations. The mVDAC1 channel is represented with the β-barrel in cyan and the N-terminal helix in red. The majority of simulations start with ATP in the pore domain. Membrane boundaries are indicated by black bars, but lipid and water are not shown for clarity. (**b**) Blue dots represent the β-phosphate of ATP plotted every 20 ps from all MD simulations initiated in panel a. The aggregate simulation time was 40 μs, and there are 10,000,000 ATP configurations represented. The densest regions are in the channel around the helix indicating that configurations in the pore domain were highly sampled. (**c**) Extent of ATP motion during each individual simulation. The starting *z* position (blue dot) and range of motion (green bar) of the COM of the ATP for each of the 453 MD simulations. The ATP undergoes 10–20 Å movements in many simulations, but never crosses the entire channel.



Supplementary Figure 5

States in the MSM are well equilibrated.

Each state in the MSM was separately analyzed for internal barriers. To do this, all configurations in each state were isolated and clustered into two substates using the *k*-centers *k*-mediods algorithm. The transition probability matrix at a 5 ns lag time was constructed, and the second eigenvalue was used to identify the relaxation time between the substates according to Eq. 7. Of the 210 states in the full MSM, 187 states (89%) relaxed faster than the 5 ns lag time used to construct the full MSM (blue circles), while 23 states (11%) relaxed much slower or failed to relax at all (red circles). Dropping these later states from the MFPT analysis had little effect on the results (data not shown).



Supplementary Figure 6

Error analysis of MFPT calculations.

(**a,c**) Distributions of the MFPT of ATP permeation from the cytoplasm to the IMS, and IMS to the cytoplasm, respectively, were calculated from 1000 bootstrap samples of the simulation trajectory set. (**b,d**) The sorted MFPTs calculated from each bootstrap sample. The mean MFPT of the distributions in a and c are indicated by a black horizontal line, while the bounds of the 95% confidence intervals are shown as dashed lines. Varying the cluster radii and lag time used to generate the MSM illustrates the sensitivity of the model to these parameters. (**e**) MFPTs as a function of cluster radius size for the original data set of trajectories calculated at a fixed lag time of 5 ns. (**f**) MFPTs as a function of lag time for the original data set of trajectories calculated at a fixed cluster radius size of 6.5 Å.

Data Set for Markov State Model Construction

A kinetic model of the permeation of ATP through VDAC was constructed using a Markov State Model (MSM) built from 453 independent simulations each lasting 40–130 ns (40 μs aggregate simulation time). These independent simulations were initiated from a non-equilibrium distribution of starting conformations, selected from either:

1. An equilibrated model of the pre-refined crystallographic ATP-bound VDAC structure.

2. Conformations selected from temperature-accelerated MD simulations.

3. Initial analysis of earlier simulations using adaptive sampling methods [1, 2, 3].

4. Heuristic measures used to identify poorly sampled regions of conformational space.

Each starting conformation was assigned a seed number, and its velocities were drawn from the Boltzmann distribution at $303\,K$ to ensure divergent trajectories, even when starting from the same initial coordinates. Coordinates were saved every 4 ps, yielding over 10 million conformations. The length of the channel traversed by individual simulations along the z-axis and the total coverage of the channel are shown in Supplementary Fig. 4.

ATP placement with Temperature-Accelerated Molecular Dynamics. Our initial simulations revealed that ATP remained tightly associated with the N-terminal helix when simulations were initiated from a pre-refined co-crystal. Therefore, we used temperature-accelerated molecular dynamics (TAMD) [4] to rapidly sample alternative conformations of ATP throughout the channel, starting from this state. Previously, TAMD was successfully used to generate a broad coverage of conformational space for other biomolecular systems [5, 6]. Briefly, TAMD involves simulating the full molecular system (channel, ATP, etc.) at room temperature, $T = 303\,K$, using the standard molecular force field described in the Online Methods, while the ATP is coupled to a high temperature bath $\bar{T} = 5\,T$. The coupling occurs by harmonically restraining the the center-of-mass (COM) of the phosphate tail of the ATP and the COM of the purine ring to a fictitious particle that evolves according to Langevin dynamics at $\bar{T}$. Thus, as the fictitious particle aggressively explores phase space, it applies a force on the ATP molecule in the true system causing it to escape from local energy minima. For a full theoretical description of this method please see to Ref. [4].

Nature Structural and Molecular Biology: doi:10.1038/nsmb.2841

We carried out two 25 ns TAMD simulations starting from the pre-refined co-crystal state. The extended system in the collective coordinates space was simulated with a 50 kcal/mol·Å^2 coupling constant and a damping coefficient of 250 ps^{-1}. The ATP rapidly moved from its starting conformation and sampled a large volume of the channel escaping to both baths. Importantly, the channel root-mean-squared deviation (RMSD) was indistinguishable from our standard non-TAMD simulations. Snapshots from these simulations were clustered, and 40 distinct ATP configurations were selected to equilibrate and initiate conventional MD production runs to be used in the Markov State Model construction; however, no dynamics from the TAMD simulations were actually used in the construction of the MSM.

Markov State Model Coupled to a Continuum Bath

The MSMBuilder2 software [7, 8] was used to construct a Markov State Model of ATP permeation through mVDAC1 using the following protocol:

1. Cluster molecular dynamics trajectories using a hybrid k-centers k-mediods algorithm.

2. Restrict clusters to those inside of the channel that form the maximal ergodic subset of states.

3. Estimate transition and count matrices using a maximum likelihood reversible estimator.

4. Couple atomistic MSM to a continuum bath of ATP.

Snapshots from all trajectories were aligned to the β-barrel of the mVDAC1 apo structure, and then clustering was carried out on the ATP coordinates using the RMSD between ATP molecules as a metric. The channel, including the N-terminal tail, remained quite stable during all simulations, and based on this observation, we chose not to include the protein degrees of freedom in the state space. We first subsampled the data set by a factor of 500 (2 ns separation between trajectory snapshots), and clustering with a cut-off radius of 6.5 Å resulted in 836 states, each defined by a representative 'generator' ATP conformation. The remaining snapshots were then assigned to these initial states.

In order to efficiently focus our computational efforts on ATP movement in the channel, we chose to initiate more ATP simulations in the pore than in aqueous solution. This leads to a poor estimate of the influx of ATP into the channel from

Nature Structural and Molecular Biology: doi:10.1038/nsmb.2841

solution due to incomplete sampling. To overcome this issue and accurately capture ATP movement into and out of the channel, we developed a method for coupling an atomistic MSM to a continuum description of free ATP diffusion in solution. First, we defined the channel pore as a cylinder running from z = -18 Å to +20 Å. All states whose COM fell outside the channel were lumped into a single state corresponding to the empty channel. The empty channel is labeled state N, and states in the pore are labeled 1 through $N-1$.

Next, the dynamical trajectories were projected onto the state space, and we constructed the count matrix $C_{ij}(\tau)$, whose elements correspond to the number of observed transitions of ATP from state i to state j after a lag time, τ. Once an initial count matrix at a given τ was constructed, states were trimmed from the MSM to produce a maximal ergodic subset of states [8, 9]. These ergodically trimmed MSMs typically had 200 states. From the count matrix, we used a maximum likelihood reversible estimator (MLE) to obtained the transition probability matrix, $P_{ij}(\tau)$ [8]:

$$\boldsymbol{P} = \begin{pmatrix} P_{1,1} & P_{1,2} & \cdots & P_{1,N-1} & P_{1,N} \\ P_{2,1} & P_{2,2} & \cdots & P_{2,N-1} & P_{2,N} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ P_{N-1,1} & P_{N-1,2} & \cdots & P_{N-1,N-1} & P_{N-1,N} \\ P_{N,1} & P_{N,2} & \cdots & P_{N-1,N} & P_{N,N} \end{pmatrix}, \tag{1}$$

The element $P_{ij}(\tau)$ is the conditional probability that the system is in state j at time $t+\tau$ given that it was in state i at time t. More generally, the transition probability matrix determines the time evolution of the system according to:

$$\boldsymbol{p}(t+\tau) = \boldsymbol{p}(t)\boldsymbol{P}, \tag{2}$$

where $\mathbf{p}$(t) is a vector representing the probability of each state in the MSM at time t and $\mathbf{p}$(t+τ) is the distribution after a time τ. The MLE algorithm ensures that equilibrium dynamics generated from $\boldsymbol{P}(\tau)$ obey detailed balance. The MSM is coupled to a continuum bath through elements $P_{i,N}$ (ATP transitions to bath), $P_{N,i}$ (ATP transitions into the channel), and $P_{N,N}$ (self transitions for empty channel), which must all be considered carefully.

$P_{N,i}$: ATP transitions into the empty channel. First, we consider the elements of $P_{N,i}$ that involve ATP *entering* the empty channel. Let the total flux of ATP into the channel from the upper bath (A) be J_A, and the total flux into the empty channel from the lower bath (B) is J_B. We use Fick's law [10] to estimate this flux assuming that the channel presents a flat surface of radius R to solution:

Nature Structural and Molecular Biology: doi:10.1038/nsmb.2841

$$J = 4\pi RcD_{ATP}, \tag{3}$$

where c is the concentration of ATP in bath A or B and D_{ATP} is the diffusion coefficient of ATP in water, 20 $\text{Å}^2/ns$ [11]. There are N_A upper boundary states that can exchange with bath A, and N_B lower boundary states that can exchange with bath B. Assuming that the incoming flux is equally distributed among all boundary states, the incoming rate into any state adjacent to bath A is J_A/N_A, and the incoming rate for states adjacent to bath B is J_B/N_B. For our analysis, we assume that all states within 4 Å of the upper boundary of the MSM can exchange with bath A and all states within 4 Å of the lower MSM boundary can exchange with bath B. The rate (number per unit time) of entry for ATP into state i from bath A is then $\rho_N \frac{J_A}{N_A}$, where ρ_N is the probability that the channel is empty so that an incoming ATP can be received in state i. Finally, the one step probability value is $P_{N,i}$:

$$P_{N,i} = \frac{4\pi Rc_A D_{ATP}}{N_A}\tau, \tag{4}$$

where τ must be sufficiently small such that $P_{N,i}$ is less than 1. In practice, $P_{N,i}$ is on the order of 1%, since our MSM is constructed at a lag time that is much smaller than the mean time for ATP entry into the channel. We see here that the incoming flux has an explicit dependence on the concentration of ATP in bath A, and the flux into states adjacent to bath B have a similar concentration dependence.

$P_{N,N}$: self transitions of the empty channel. The requirement that the row elements normalize to 1 determines the value of the self transition probability for the empty channel:

$$P_{N,N} = 1 - \left(\sum_{\{i\}_A} \frac{J_A}{N_A} + \sum_{\{i\}_B} \frac{J_B}{N_B} \right) \tau = 1 - (J_A + J_B)\tau. \tag{5}$$

$P_{i,N}$: ATP transitions out of the channel. Next, we consider the elements $P_{i,N}$ that involve ATP *exiting* the channel, leaving it empty. Since we initiated many simulations from within the pore domain, we observed tens of thousands of exit events out of the channel. Therefore, we believe that $P_{i,N}$ elements in Eq. 1 determined directly from the MLE step in MSMBuilder2 are a good estimate of this state dependent jump probability. To test this assumption, we computed the average jump probability out of the channel, where we averaged $P_{i,N}$ over all states in 2 Å windows along the z axis. The analysis revealed a smooth decrease in the jump probability as the states penetrate deeper into the channel from either the upper or lower face of the channel, which is expected since it is harder to escape the channel the deeper the

Nature Structural and Molecular Biology: doi:10.1038/nsmb.2841

penetration. ATP in states near the upper bath never escape to the lower bath, and states near the lower bath never exhibit transitions to the upper bath. Also, states at the upper face of the channel have a nearly 50% probability of jumping out of the channel, which is intuitive if the ATP is only weakly interacting with the pore at the edge of the channel. At the lower mouth, which is closer to the N-terminal helix, the probability of exiting the channel is smaller, which reflects that these states are still under some electrostatic influence of the pore, which we described in an earlier manuscript [12]. Based on these results, we believe that we have sampled these exit transitions well.

Stationary State Distribution and Implied Timescales

The left eigenvalues and eigenvectors of $\boldsymbol{P}(\tau)$ are related to the natural relaxation timescales and transitions between states in the MSM, respectively. For an ergodic and irreducible Markov State Model, there is a single eigenvector with an eigenvalue of 1:

$$\boldsymbol{\pi}\boldsymbol{P}(\tau) = \boldsymbol{\pi}, \tag{6}$$

where $\boldsymbol{\pi}$ is the stationary distribution vector plotted in Fig. 3 of the main text. This vector determines the equilibrium probability of finding ATP in a given state.

The relaxation times, or implied timescales, $\hat{t}$, are computed from the eigenvalues as follows:

$$\hat{t}_i = -\frac{\tau}{\ln \lambda_i(\tau)}, \tag{7}$$

where $\lambda_i(\tau)$ is the i^{th} eigenvalue of $\boldsymbol{P}(\tau)$. In general, the implied timescales plateau and become constant at long lag times in which the model obeys Markovian dynamics. To determine an appropriate lag time to construct our MSM, we computed the implied timescales for τ ranging from 0 to 40 ns, and plotted the 50 slowest transitions in Supplementary Fig. 3. The resulting spectrum is well behaved, and based on the asymptotic behavior, we chose a lag time of 5 ns to construct our MSM for all subsequent analysis. This MSM has 210 states.

Calculation of the Mean First Passage Times

The mean first passage time (MFPT) from state i to state j, $m_{i,j}$, is defined as the average time taken to reach state j for the first time given that the system was

Nature Structural and Molecular Biology: doi:10.1038/nsmb.2841

initially in state i. The MFPT between two states can be determined by solving the following linear system of equations [13]:

$$m_{i,j} = P_{i,j}\tau + \sum_{k \neq j} P_{i,k} \left(m_{k,j} + \tau \right). \tag{8}$$

We are interested in the MFPT for ATP movement from the upper bath to the lower bath, and vice versa. Unfortunately, Eq. 8 cannot be applied directly to our MSM, since transitions to and from the empty channel, N, do not distinguish whether ATP exited or entered to the upper or lower bath. However, we can modify our MSM to take the separate baths into account since we know the exchange probabilities with each bath.

We split state N into two separate states: A where ATP is in the upper bath, and B where ATP in the lower bath. We then self-consistently computed all MFPTs including $m_{A,B}$ and $m_{B,A}$. However, there are several transition probability elements that must be reconsidered: $P_{A,A}$, $P_{B,B}$, $P_{i,A}$, $P_{i,B}$, $P_{A,i}$, and $P_{B,i}$. Both $P_{A,B}$ and $P_{B,A}$ are zero, as we never observe movement from one bath to the other in one step. $P_{A,i}$ and $P_{B,i}$ are given by Eq. 4 with the corresponding ATP concentration values, c_A or c_B, respectively. $P_{i,A}$ and $P_{i,B}$ are determined directly from the simulation data as described in the section *Markov State Model Coupled to a Continuum Bath*. Based on this data, we know ATP exiting states at the top of the channel only enter bath A and ATP exiting states at the bottom of the channel only enter bath B. This uniquely determines $P_{i,A/B}$ from $P_{i,N}$ based on the location of state i. The last terms to be determined are the self transition probabilities for ATP in states A and B, which are given by the normalization condition:

$$P_{A,A} = 1 - \sum_{i \neq A} P_{A,i}. \tag{9}$$

With these terms identified, the MFPT for arriving in bath B starting from all other states can be determined from the following set of self-consistent equations:

$$\begin{pmatrix} P_{A,A} - 1 & P_{A,1} & \cdots & P_{A,N-1} \\ P_{1,A} & P_{1,1} - 1 & \cdots & P_{1,N-1} \\ \vdots & \vdots & \ddots & \vdots \\ P_{N-1,A} & P_{N-1,1} & \cdots & P_{N-1,N-1} - 1 \end{pmatrix} \begin{pmatrix} m_{A,B} \\ m_{1,B} \\ \vdots \\ m_{N-1,B} \end{pmatrix} = \begin{pmatrix} -\tau \\ -\tau \\ \vdots \\ -\tau \end{pmatrix}, \tag{10}$$

where the element $m_{A,B}$ is the quantity of interest, and the MFPT from B to A, $m_{B,A}$, can be determined from a similar set of equations.

Nature Structural and Molecular Biology: doi:10.1038/nsmb.2841

Transition Path Theory Analysis

We computed the highest flux pathways taken by ATP as it passes through the channel using transition path theory [14, 15] as implemented in MSMBuilder2 [7, 8, 16]. We started with the original MSM defined on the states that fall between -18 and +20 Å along the z axis, but rather than coupling to a continuum bath, we extended the atomistic MSM 4 Å in the positive, Ω^+ ($+20\,\text{Å} < z \leq +24\,\text{Å}$), and negative directions, Ω^- ($-22\,\text{Å} \geq z < -18\,\text{Å}$). We then computed all the pathways from all states in Ω^- to all states in Ω^+ (Fig. 5 in the main text), and vice versa. However, we removed pathways from the analysis which contained transitions with zero counts in the count matrix, $C_{ij}(\tau)$. This occasionally occurs because the MLE method for computing the transition probability matrix will assign a finite probability value to a transition with zero counts if there are counts in the reverse direction. Finally, we analyzed the top 143 pathways, which account for 70% of the total flux. In Fig. 5b in the main text, we grouped the pathways into four categories (paths 1–4) based on which residue the ATP interacts with just prior to escape to the cytoplasm. We included a 5^{th} path, which does not interact with the central portion of the N-terminal helix during permeation.

Local Equilibration Times Within States

To ensure that there are no kinetic barriers within MSM states, we calculated the time required for ATP to relax within each state. Following the approach in Ref. [15], we partitioned each state into two sub-states using the same hybrid k-centers k-medoids algorithm used to construct the full MSM, as shown in Supplementary Fig. 5. We then computed the 2×2 transition probability matrix between the two sub-states for a range of lag times. The relaxation time between the two sub-states, $\hat{t}$, is related to the second eigenvalue, λ_2, of the sub-state transition probability matrix:

$$\hat{t}_2 = -\frac{\tau}{\ln\lambda_2(\tau)}. \tag{11}$$

As seen in Supplementary Fig. 5, the internal relaxation times of all but 23 of the 210 generators were less than the lag time used in the construction of the full MSM. Since the majority of the states relax faster than the lag time used to construct the MSM, it indicates that there are no internal barriers within states and that they are well equilibrated. Moreover, removing these 23 states from our analysis had no significant change in the MFPT for ATP permeation.

Nature Structural and Molecular Biology: doi:10.1038/nsmb.2841

MFPT Error Analysis

The statistical uncertainties in the MFPT were calculated using the percentile bootstrap method [17]. From the original data set of 453 simulations, 1000 bootstrap samples were generated, each by randomly selecting 453 trajectories with replacement. For each bootstrap sample, a new MSM was constructed and the MFPTs were calculated as in section *Calculation of the Mean First Passage Times*. The 95% confidence intervals for the MFPT in each direction were then estimated from the distributions shown in Supplementary Fig. 6a–d. Additionally, we examined the sensitivity of the MFPTs calculated from the model to our choice of lag time and cluster radius. In Supplementary Fig. 6e, the MFPT for permeation in each direction through VDAC is shown as a function of cluster radius. As the radius of each cluster increases the general trend is for the MFPT to decrease as the states begin to subsume energetic barriers that separate the true metastable basins. While decreasing the cluster radius cutoff reduces the presence of slow relaxation times within a state, the number of states increases, decreasing the sampling of state-to-state transitions for a finite data set. Our choice of cluster radius, was selected to minimized slow relaxation times within each state, as assessed in section *Local Equilibration Times Within States*, while ensuring that the MSM represented a well-connected state space with a robust number of observed transitions along the permeation pathways. Supplementary Fig. 6f shows the dependence of the MFPTs on the lag time used to construct the MSM. While the MFPT for transitions from the IMS to the cytoplasm was robust over the full range of examined lag times (5–30 ns), the MFPT for the reverse transition increases with lag time. This increase is expected in general since data is removed from the analysis when the lag time is increased. We expect that this rise is due to reduced transitions counts between a particular bottle neck state. We note that there is a jump from 15 to 20 ns, but from 5 to 15 ns the increase in the MFPT is in line with the 95% confidence error provided by our analysis.

Simulation	**Time (μs)**	**Applied Voltage (mV)**	**Initial ATP position**
1	4.84	0	pore
2	0.61	0	upper bath
3	1.99	50	upper bath
4	1.99	100	upper bath
5	4.17	50	upper bath
6	3.11	-50	lower bath

Supplementary Table 1: VDAC dataset at Anton.

Nature Structural and Molecular Biology: doi:10.1038/nsmb.2841

References

[1] Huang, X., Bowman, G., Bacallado, S. & Pande, V. Rapid equilibrium sampling initiated from nonequilibrium data. *Proceedings of the National Academy of Sciences* **106**, 19765–19769 (2009).

[2] Singhal, N. & Pande, V. Error analysis and efficient sampling in markovian state models for molecular dynamics. *The Journal of Chemical Physics* **123**, 204909 (2005).

[3] Hinrichs, N. & Pande, V. Calculation of the distribution of eigenvalues and eigenvectors in markovian state models for molecular dynamics. *The Journal of Chemical Physics* **126**, 244101 (2007).

[4] Maragliano, L. & Vanden-Eijnden, E. A temperature accelerated method for sampling free energy and determining reaction pathways in rare events simulations. *Chemical Physics Letters* **426**, 168–175 (2006).

[5] Maragliano, L., Cottone, G., Ciccotti, G. & Vanden-Eijnden, E. Mapping the network of pathways of co diffusion in myoglobin. *Journal of the American Chemical Society* **132**, 1010–7 (2010).

[6] Vashisth, H., Maragliano, L. & Abrams, C. F. "DFG-flip" in the insulin receptor kinase is facilitated by a helical intermediate state of the activation loop. *Biophysical Journal* **102**, 1979–87 (2012).

[7] Bowman, G., Huang, X. & Pande, V. Using generalized ensemble simulations and markov state models to identify conformational states. *Methods* **49**, 197–201 (2009).

[8] Beauchamp, K. A. *et al.* MSMBuilder2: Modeling conformational dynamics at the picosecond to millisecond scale. *Journal of Chemical Theory and Computation* **7**, 3412–3419 (2011).

[9] Scalco, R. & Caflisch, A. Equilibrium distribution from distributed computing (simulations of protein folding). *The Journal of Physical Chemistry B* **115**, 6358–65 (2011).

[10] Berg, H. C. *Random walks in biology* (Princeton University Press, 1993).

Nature Structural and Molecular Biology: doi:10.1038/nsmb.2841

[11] Yoshizaki, K., Watari, H. & Radda, G. K. Role of phosphocreatine in energy transport in skeletal muscle of bullfrog studied by ^{31}P-NMR. *Biochimica et Biophysica Acta (BBA)-Molecular Cell Research* **1051**, 144–150 (1990).

[12] Choudhary, O. *et al.* The electrostatics of VDAC: implications for selectivity and gating. *Journal of Molecular Biology* **396**, 580–592 (2010).

[13] Singhal, N., Snow, C. & Pande, V. Using path sampling to build better markovian state models: Predicting the folding rate and mechanism of a tryptophan zipper beta hairpin. *The Journal of Chemical Physics* **121**, 415 (2004).

[14] Metzner, P., Schütte, C. & Vanden-Eijnden, E. Transition path theory for markov jump processes. *Multiscale Modeling & Simulation* **7**, 1192–1219 (2009).

[15] Noé, F., Schütte, C., Vanden-Eijnden, E., Reich, L. & Weikl, T. Constructing the equilibrium ensemble of folding pathways from short off-equilibrium simulations. *Proceedings of the National Academy of Sciences* **106**, 19011–19016 (2009).

[16] Voelz, V. A., Bowman, G. R., Beauchamp, K. & Pande, V. S. Molecular simulation of ab initio protein folding for a millisecond folder NTL9(1-39). *Journal of the American Chemical Society* **132**, 1526–8 (2010).

[17] Efron, B. & Tibshirani, R. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science* 54–75 (1986).

Nature Structural and Molecular Biology: doi:10.1038/nsmb.2841