

SCIENTIFIC REPORTS



OPEN

A Markov State-based Quantitative Kinetic Model of Sodium Release from the Dopamine Transporter

Received: 03 October 2016

Accepted: 30 November 2016

Published: 06 January 2017

Asghar M. Razavi¹, George Khelashvili¹ & Harel Weinstein^{1,2}

The dopamine transporter (DAT) belongs to the neurotransmitter:sodium symporter (NSS) family of membrane proteins that are responsible for reuptake of neurotransmitters from the synaptic cleft to terminate a neuronal signal and enable subsequent neurotransmitter release from the presynaptic neuron. The release of one sodium ion from the crystallographically determined sodium binding site Na2 had been identified as an initial step in the transport cycle which prepares the transporter for substrate translocation by stabilizing an inward-open conformation. We have constructed Markov State Models (MSMs) from extensive molecular dynamics simulations of human DAT (hDAT) to explore the mechanism of this sodium release. Our results quantify the release process triggered by hydration of the Na2 site that occurs concomitantly with a conformational transition from an outward-facing to an inward-facing state of the transporter. The kinetics of the release process are computed from the MSM, and transition path theory is used to identify the most probable sodium release pathways. An intermediate state is discovered on the sodium release pathway, and the results reveal the importance of various modes of interaction of the N-terminus of hDAT in controlling the pathways of release.

The dopamine transporter (DAT) is a transmembrane (TM) secondary transporter protein belonging to the Neurotransmitter:Sodium Symporter (NSS) family, which also includes the closely related serotonin (SERT) and norepinephrine (NET) transporters¹. The NSS are responsible for clearance of released neurotransmitters (e.g. dopamine (DA), by DAT) from the synaptic cleft through their translocation back into the presynaptic nerve terminals. The function of NSS transporters in neuronal signaling implicate them in the mechanisms of action of abused psychostimulants, such as cocaine and amphetamine^{2,3}, and in various psychiatric and neurological disorders including drug addiction, schizophrenia, and Parkinson's disease³. These essential neurophysiological roles have made NSS transporters primary targets for antidepressant medications.

Breakthrough crystallographic determinations of members of the NSS family includes the bacterial homolog LeuT^{4,5}, and most recently the structures of *Drosophila* DAT (dDAT)^{6–8} and SERT⁹. Structures of these transporters have been determined in complex with various substrates and ligands in the primary (S1) and secondary (S2) binding sites and with sodium ions bound in sites Na1 and Na2. They have offered an essential molecular context for the investigation of the mechanism of uphill neurotransmitter reuptake transport enabled by the coupling with the transmembrane Na⁺ gradient^{1,10}. Our current understanding of functional mechanisms in NSS has been further shaped by the large body of structure-function studies of LeuT and other related bacterial transporters, carried out both experimentally^{11–17} and computationally^{11–14,16,18–23}, and more recently by findings from molecular dynamics (MD) simulations of DAT^{24–29} and SERT^{30,31} constructs. Together, these studies have suggested for the transport cycle in NSS proteins an allosteric process that is consistent with the alternating access mechanism³² in which concerted dynamic rearrangements on the extracellular (EC) and intracellular (IC) sides of the transporter result from ion- and ligand-specific conformational rearrangements. These motions allow the transporter to transition between conformational states in which the substrate binding site is either occluded from the protein environment and does not allow water penetration either from the EC or the IC end, or alternatively exposed to the EC vestibule (in an outward-open configuration) and the IC vestibule (in an inward-open one).

One of the key mechanistic components of the early steps in the transport cycle in these transporters, which emerged from these studies, relates to the binding/release of the Na⁺ ion from the Na2 site. Indeed, studies of LeuT and other related bacterial transporters^{16,22,33} have proposed that an occupied Na2 site is essential for

¹Department of Physiology and Biophysics, Weill Cornell Medical College of Cornell University, New York, NY 10065, USA. ²Institute for Computational Biomedicine, Weill Medical College of Cornell University, New York, NY 10065, USA. Correspondence and requests for materials should be addressed to H.W. (email: haw2002@med.cornell.edu)

LIPID	Extracellular	Intracellular	Total
Cholesterol	29	24	53
POPC	125	26	151
POPE	0	92	92
POPS	0	20	20
PI(4,5)P ₂	0	18	18
SM	12	0	12
Number of lipids	166	180	346

Table 1. Lipid composition* of the model membrane used in this study. *The lipid abbreviations are as follows: POPC – 1-palmitoyl-2-oleyl-sn-glycero-3-phosphocholine, POPE – 1-palmitoyl-2-oleyl-sn-glycero-3-phosphoethanolamine, POPS – 1-palmitoyl-2-oleyl-sn-glycero-3-phosphoserine, PI(4,5)P₂ – phosphatidylinositol-4,5-bisphosphate, SM-sphingomyelin.

maintaining the transporter in the inward-closed state and that inward-opening and the subsequent substrate release process are energetically more favorable once the Na⁺ ion has left the Na₂ site. Consistent with this model, MD simulations of DAT²⁸ and SERT³¹ have displayed spontaneous outward-open to inward-open transition upon destabilization and inward release of the sodium from the Na₂ site. From analysis of microsecond-long MD trajectories²⁸ we discovered that the release of the sodium from the Na₂ site (referred to hereafter as Na⁺/Na₂) in the human DAT (hDAT) can be allosterically triggered by interactions of the N-terminus region of the transporter (residues 1–59) with various other regions of the transporter, and that some of these interactions are electrostatically driven and supported by the highly charged (−4e at neutral pH) PI(4,5)P₂ (phosphatidylinositol 4,5-bisphosphate) lipids in the membrane^{28,34,35}.

One important aspect of the transport cycle that has not been resolved relates to the kinetics of Na⁺ ion dissociation from the Na₂ site, which is essential for building kinetic models that could describe quantitatively the allosteric mechanisms involved in function of the NSS³⁶. Here we describe results quantifying the rates of Na⁺ ion release from the Na₂ site of hDAT from the analysis of the trajectories of extensive equilibrium atomistic MD simulations (~50 μs) using a Markov State Model (MSM)³⁷. We find that starting from the transporter in the occluded state, the average time for Na⁺ ion to diffuse from the Na₂ site to the IC solution is ~800 ns after the initial destabilization of the functionally important IC gates^{26,28,38} that maintain hDAT in the inward-closed state. From this analysis we also identify the existence of a meta-stable intermediate state in the Na⁺/Na₂ release process, in which the Na⁺ ion that has left the Na₂ site is temporarily retained within the IC vestibule of hDAT by strong interactions with residue E428. These results, presented here in the context of detailed molecular structure, provide novel insights into kinetic aspects of allosteric mechanisms in NSS transporters that had previously not been attainable from computation³⁶.

Methods

System preparation. Full-length model of hDAT. The molecular model of full-length hDAT (human dopamine transporter) used in this study is the same as described and investigated earlier²⁸, built from the outward-facing X-ray structure of dDAT (drosophila dopamine transporter), PDB ID code: 4M48, and including the substrate dopamine (DA) in the S1 site, as well as the two Na⁺ ions and a Cl⁻ ion in the respective ion binding sites. The complete molecular model includes the full length N- and C-termini in the conformations described in ref. 34 and is attached to the transmembrane domain as described before²⁸ so that the positioning of the C-terminus closely resembled that in the dDAT X-ray structure, and the N-terminus does not contact any residue in the TM bundle in this initial model (see ref. 28 for additional details).

The hDAT-membrane system. The complete hDAT model was immersed in a pre-equilibrated compositionally asymmetric bilayer membrane (see Table 1 for final lipid content, after the protein insertion), designed to resemble a neuronal cell plasma membrane²². This hDAT-membrane system was solvated in a 150 mM K⁺Cl⁻ TIP3P water solution with ions added to neutrality, resulting in a final atom count of ~150,000.

Molecular Dynamics Simulations. All-atom molecular dynamics (MD) simulations were initiated with the multi-step equilibration protocol established previously²⁸ and carried out with NAMD software version 2.10³⁹. During this stage, the backbone of the protein was first fixed and then harmonically restrained. The solvent was initially prevented from entering the lipid-water interface. The constraints on the protein backbone were released gradually in three steps of 1 ns each, changing the force constants from 1, to 0.5, and 0.1 kcal/(mol Å²), respectively. This step was followed by relatively short (~20 ns) unbiased MD simulations performed with a 2 fs integration time-step and under the NPT ensemble (at T = 310 K), using the Particle-Mesh-Ewald (PME) method for electrostatics⁴⁰ and the Nose-Hoover Langevin piston⁴¹ to control the target 1 atm pressure, with Langevin piston period and decay parameters set to 100 fs and 50 fs, respectively.

After this equilibration phase, the velocities of all the atoms in the system were reset (at T = 310 K using random number seed) and 50 independent ~1 μs long unbiased MD simulations were carried out using the ACEMD software⁴², resulting in a cumulative MD simulation time of ~50 μs. The simulations with ACEMD implemented the PME method for electrostatic calculations, and were carried out according to the protocol developed at AcCellera⁴² with 4 fs integration time-step and the standard mass repartitioning procedure for hydrogen atoms

implemented in ACEMD. The simulations were conducted under the NVT ensemble (at $T = 310\text{ K}$), using the Langevin Thermostat with Langevin Damping Factor set to 0.1 ps^{-1} .

The latest all-atom CHARMM36 force fields for proteins⁴³, lipids^{44,45}, and ions⁴⁶ were used throughout. As recommended⁴⁵, the total electrostatic charge on individual PI(4,5)P₂ lipids was kept at $-4e$ by protonating 50% of PI(4,5)P₂ molecules at the “4” site and 50% at the “5” site.

Definition of extracellular and intracellular vestibules in hDAT. To facilitate the description of large structural changes that result in water penetration into various regions of the molecule, we defined a series of local water-fillable cavities. The templates for these definitions are the extracellular (EC) and intracellular (IC) vestibules in hDAT, which had been defined before²⁸ and characterized by counting water molecules. Briefly, a water molecule is considered to be in the EC vestibule if (a)-its oxygen atom is within 26 \AA of the C_β atom of F326, but not within 5 \AA of lipid atoms, and (b)-if its z-coordinate is larger than the C_β atom of F326 by no more than 23 \AA (the z-axis is perpendicular to the membrane with its positive direction toward the EC side). Pari passu, an EC cavity is defined as a region of the EC vestibule in which water can penetrate with the oxygen atom positioned within 10 \AA of the center of mass of the ligand dopamine. On the intracellular side, the algorithm assumes that a water molecule belongs to the IC vestibule if its oxygen atom is within 15 \AA of the dopamine center of mass but not within 5 \AA of any lipid, and if its z-directional distance to the C_β atom of D436 does not exceed 15.5 \AA . The IC vestibule is further divided into two volumes: IC channel and IC cavity. The IC channel is defined as the part of the IC vestibule that is within 12 \AA of the backbone of residue T269, whereas the IC cavity is the part of the IC vestibule that lies outside the IC channel region.

Dimensionality reduction using the “time-structure based independent component analysis” (tICA) approach. To construct a Markov model it is necessary first to reduce the dimensionality of the system in order to both remove the redundant information stored in the atomic coordinates, and provide a framework for accurate clustering of conformations (see below). For a dynamic system with multiple kinetic modes, a promising approach for accurate clustering of conformations is to separate them based on their kinetic similarity. This can be achieved by projecting the conformations of the system on its slowest reaction coordinates⁴⁷. A metric called “time-structure based independent component analysis” (tICA) was shown recently to identify the slowest reaction coordinates of a system^{47–50}.

Briefly, tICA is based on constructing a time-lagged covariance matrix (TLCM): $C_{TL}(\tau) = \langle X(t)X^T(t+\tau) \rangle$ and the usual covariance matrix $C = \langle X(t)X^T(t) \rangle$, where $X(t)$ is the data vector at time t , τ is the lag-time, and the symbol $\langle \dots \rangle$ denotes the time average. The slowest reaction coordinates are then identified by solving the generalized eigenvalue problem: $C_{TL}V = CV\Lambda$, where Λ and V are the eigenvalue and eigenvector matrices, respectively. The eigenvectors corresponding to the largest eigenvalues identify the slowest reaction coordinates. These reaction coordinates are not unique and depend on the choice of data vector X . The number of elements in tICA eigenvectors is the same as the number of parameters comprising the data vector X . Thus, there is a one-to-one relation between parameters in X and elements in tICA eigenvectors. Moreover, the relative magnitude of each element in tICA eigenvectors reflects the importance of the corresponding parameter in forming a particular tICA reaction coordinate. A common practice in defining tICA parameters is to use C_α or C_β distances among all residues. While this approach is unbiased and previous studies have shown that it provides excellent tICA reaction coordinates in protein folding studies^{48,51,52}, in our case this general approach couldn't provide useful tICA reaction coordinates along which the simulation data could be discretized, because the Na⁺/Na₂ release process itself is accompanied by only a limited change in secondary and tertiary structure of the protein. Indeed, projection of simulation data on such reaction coordinates yielded Gaussian distributions (data not shown) indicating that they do not contain information regarding Na⁺/Na₂ release dynamics. The specific choice of the parameters for constructing tICA reaction coordinates in hDAT is described in the Results section.

Construction of the Markov Model. Markov State Models (MSMs) have proven to be powerful tools for describing the kinetics of equilibrium processes in protein folding^{53,54}, and in studies of functional mechanisms of various membrane proteins such as GPCRs, or activation pathways of kinases^{55,56}. MSM validation and implementation is well documented and described in several excellent reviews^{57–59}. Briefly, the MSM analysis of molecular dynamics is based on the idea that global conformational changes of a system can be modeled as a Markov chain, given that the transitions between different conformations are sampled at long enough time intervals so that each transition is Markovian, i.e. a transition from one state to another state is independent of the previous transitions of states. Owing to their Markovian property, the MSMs can yield information about long-timescale events from a combination of information from short time-scale events⁶⁰.

To construct a MSM from a combination of MD trajectories, the conformational space from all trajectories is reduced by a transformation to a space defined by only a few reaction coordinates of interest. The reduced conformational space is then discretized into multiple (usually on the order of hundreds or thousands) zones (*microstates*) using automated clustering algorithms. When the microstates are defined, each conformation in the MD trajectory is assigned to the closest microstate, which produces the so-called *assignment space*. A *transition count matrix* is constructed in the next step by counting the number of transitions among all microstates. To satisfy detailed balance and local equilibrium, the transition count matrix is symmetrized using its transpose matrix or a maximum likelihood approach, as described in ref. 61. Finally, the probabilities of transitions among all microstates are calculated by normalizing the transition count matrix, which yields the *transition probability matrix* (TPM). To ensure that the behavior is Markovian, multiple TPMs are constructed for different time interval between transitions (referred to as “lag times”), and the relaxation timescales of the system are extracted by using the relation:

$$\tau_i = \frac{\tau'}{\ln \lambda_i} \quad (1)$$

where τ' is the lag-time used for building the TPM, λ_i is the i^{th} eigenvalue of the transition probability matrix and the resulting τ_i is the so-called *implied timescale* corresponding to the i^{th} relaxation mode of the system. If the transition probability matrix is Markovian, then τ_i will be independent of τ' and this is the basis for the test of “Markovian behavior” for a particular choice of parameters (a specific illustration of the test is given in the Results section).

To extract information from the Markovian TPM that contains all the information about the thermodynamics and kinetics of the system, the TPM is decomposed into its eigenvalues and eigenvectors. The eigenvector corresponding to the largest eigenvalue (that is, equaling one) represents the equilibrium population of the system. The remaining eigenvectors contain kinetic information of the system and the one corresponding to the second largest eigenvalue represents the slowest relaxation dynamics of the system. In turn, the eigenvector corresponding to the third largest eigenvalue represents the second slowest dynamics of the system, and so on.

Selection of parameters for Markov model construction. The quality of a Markov model depends on a number of parameters including the number of tICA eigenvectors used in the discretization of the conformational space, the number of microstates used in the representation of the space, and the clustering algorithm used for grouping conformations into microstates^{57,61}. The discretization of the conformational space required for the construction of the MSM can be projected on the problem of interest by the protocol of dimensionality reduction. The number of eigenvectors used impacts the quality of MSM, as more tICA eigenvectors provide discretized states of higher resolution, but also reduce the statistical significance of the states due to finite sampling errors⁶².

To find the best combination of parameters for constructing the MSM, we used a recently introduced MSM scoring method called “generalized matrix Rayleigh quotient” (GMRQ)⁶³ that is implemented in the MSMBuilder software package⁶¹. Briefly, this method utilizes all the simulation data to obtain the tICA space; the subsequent *assignment space* is obtained by discretization of the tICA space as described above. In the next step of the GMRQ method, the *assignment space* is divided into two sets: a training set and a test set. A TPM is then constructed for each set at a lag time that produces Markovian behavior (here we used 48 ns, see Results) and is decomposed into its eigenvectors and eigenvalues. The GMRQ algorithm then tests how well eigenvectors of the training set diagonalize the TPM of the test set, producing a score defined as:

$$R = \text{Tr}(\mathbf{V}^T \mathbf{C} \mathbf{V} (\mathbf{V}^T \mathbf{S} \mathbf{V})^{-1}) \quad (2)$$

where Tr denotes trace of the inner matrix product in which \mathbf{V} contains the first n eigenvectors of the training set TPM; \mathbf{S} is the diagonal matrix composed of equilibrium populations of the test set (note that the equilibrium population is the first eigenvector of the TPM); and \mathbf{C} is the inner matrix product of \mathbf{S} and the TPM of the test set (\mathbf{T}):

$$\mathbf{C} = \mathbf{S} \mathbf{T} \quad (3)$$

The best MSM parameters are those that yield the highest performance (highest GMRQ score) on the *test* set.

Transition path theory (TPT) analysis. Besides providing thermodynamic and kinetic information for a system, Markov models can also be utilized to obtain the most probable pathways for a kinetic process, such as the most probable pathways for the release of Na^+/Na_2 we examine here. Transition path theory (TPT) is an elegant approach to obtain and quantify these pathways. Briefly, TPT is based on first constructing a *flux* matrix for all states of the system. A definition of the elements of the flux matrix defined in an analogous manner to the definition in ref. 64 is:

$$J_{ij} = \pi_i (1 - P_{\text{target},i}) T_{ij} P_{\text{target},j} \quad (4)$$

where π_i is the equilibrium population of state i , $P_{\text{target},i}$ is the P_{target} for state i and T_{ij} is the transition probability from state i to state j . $P_{\text{target},i}$ is the probability of visiting the target state(s) (in our case the states with Na^+/Na_2 released) before the other states (in our case the states with Na^+/Na_2 bound), and takes values between 0 and 1.

The most probable pathways are obtained by finding those with the highest flux between the states in the flux matrix by using a graph theory algorithm, and we employed here the Dijkstra algorithm⁶⁵ implemented in the MSMbuilder software⁶¹, which finds the top path. Subsequent paths are obtained by first removing the top path from the flux matrix and then repeating the algorithm. This iterative approach was used here to obtain the first 7 paths, which contribute to >60 percent of the total flux (see the Result section).

We note that whereas the MSM is built on the microstate level (with hundreds to thousands of states) in order to provide accurate timescales for the kinetics of a system, the TPT analysis, is usually performed on a *macrostate* level (on the order of a few tens of states) to facilitate the visualization of the most probable pathways. To bridge these levels, microstates are grouped together into a macrostate based on their similarity in specific aspects relevant to the information sought from the MSM analysis. Here, we used the Robust Perron Cluster Analysis (PCCA⁺)⁶⁶ algorithm, which groups microstates based on their kinetic similarity.

Results

Molecular mechanism of Na^+ release from the Na_2 site of hDAT. PIP_2 -mediated N-terminus/IL4 interactions destabilize intracellular gates in hDAT. Our previous computational studies²⁸ have demonstrated the mechanistic involvement of PIP_2 -mediated interactions between the N-terminus domain (N-term) and the intracellular loop 4 (IL4) of hDAT in the inward opening of the transporter and the concomitant release of the Na^+/Na_2 .

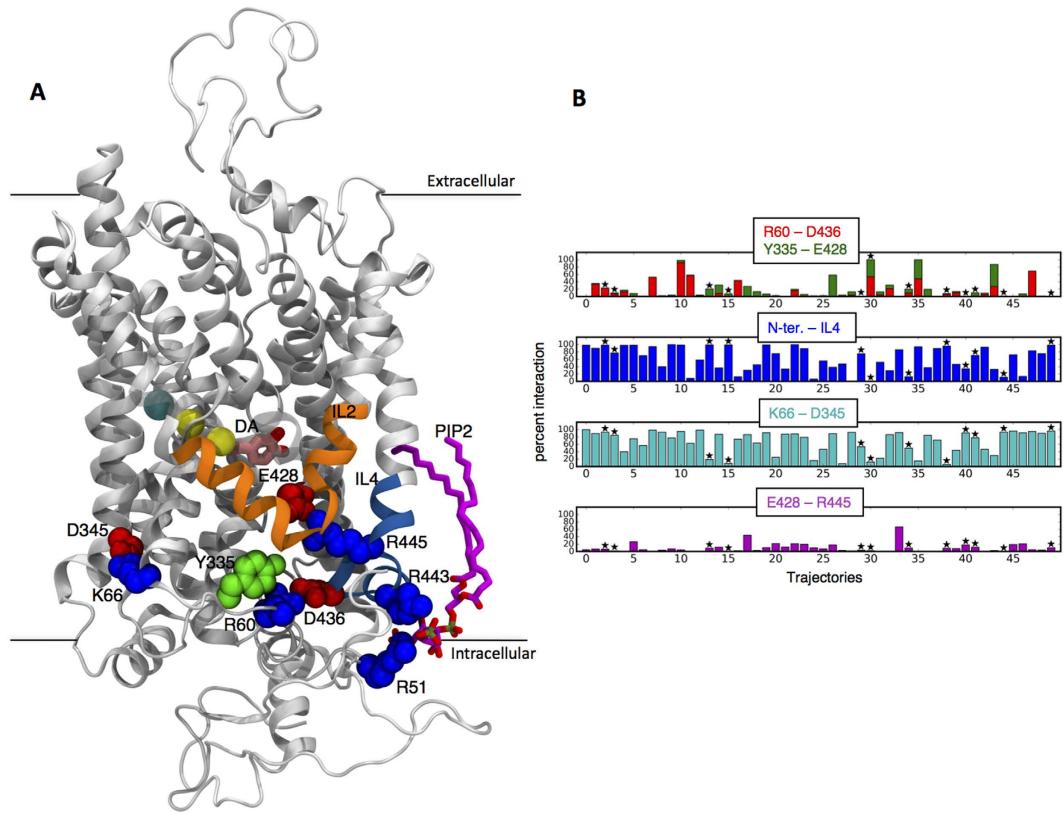


Figure 1. Interactions of the N-terminus with the intracellular end of hDAT in the process of Na^+/Na_2 release. (A) Full atomistic representation of hDAT. Residues mentioned in the analysis are shown in space-filling representation, labeled and color-coded based on their polarity, with positively charged residues in blue, polar residues in green, and the negatively charged ones in red. Na^+ and Cl^- ions are shown as yellow and cyan spheres, respectively. Dopamine (DA) is shown in pink licorice. IL4 (in blue cartoon), and IL2 (in orange cartoon) are also highlighted. PIP_2 (magenta) mediates interaction between IL4 and the N-terminus (here, R51). (B) The frequencies of interactions between key residues in all 50 trajectories (expressed as percent), averaged over the entire simulation time. Stars indicate the trajectories in which Na^+/Na_2 is released. The N-ter.-IL4 panel represents PIP_2 mediated interactions (within 6 Å) between various Arg and Lys residues from the N-terminus (residue 1 to 59) with Arg443 from IL4 (defined as the interaction of the same PIP_2 with both Arg443 and any Arg or Lys residue of the N-terminus).

Here, we found the same mode of N-term/IL4 association recapitulated during the initial stages of the MD simulations we analyzed (Figure S1). Consistent with our previous results²⁸, we found that the N-term/IL4 interactions resulted in hDAT conformations in which the functional intracellular gates formed by residues R60, Y335, E428, R445, and D436 (Fig. 1A)^{26,33}, were destabilized. To expand the exploration of the conformational space, we initiated from this state of the hDAT an ensemble of 50 new, ~1 μs-long, statistically independent atomistic MD simulations. The analysis revealed that while in more than half of the trajectories the N-term could disengage from IL4 and initiate other modes of interaction with the intracellular region of hDAT, the interactions defining the intracellular gates remained nevertheless broken in the majority of the simulations (Fig. 1B). Importantly, in this ensemble of 50 simulations we observed multiple instances of spontaneous inward release of Na^+/Na_2 , the Na^+ ion from the Na_2 site.

The isomerization of hDAT from an outward-facing to an inward-facing configuration with concomitant release of Na^+/Na_2 is determined by the extent of solvation of functional sites. Spontaneous inward release of the Na^+/Na_2 was observed in 12 of the 50 ~1 μs-long MD trajectories (Figure S2). The fastest observed release occurred after only 80 ns of MD simulation (not including the initial equilibration phase), while the slowest observed release took place after 900 ns (Figure S2). Analysis of the trajectories in which Na^+ leaves the Na_2 site reveals that the process is largely determined by the extent of water penetration from the intracellular region into the functional sites of the transporter. Indeed, increase in the distance of Na^+/Na_2 from the stable Na^+ ion in the Na_1 site is accompanied by an increase in the number of water molecules surrounding Na^+/Na_2 . This is evidenced by the correlation between the water coordination number of Na^+/Na_2 , and the distance of the Na^+/Na_2 from the sodium at site Na_1 calculated for each of the 12 trajectories in which Na^+/Na_2 is released to the intracellular environment, as shown in Table S1.

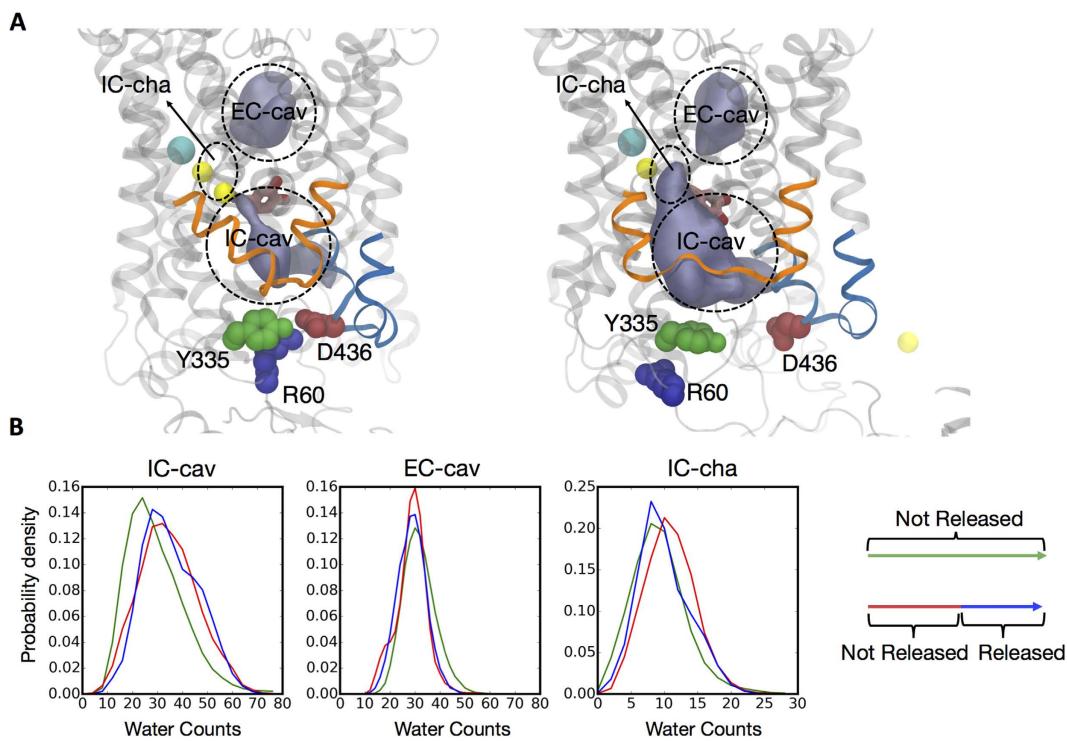


Figure 2. Solvation of the IC and EC cavities (IC-cav; EC-cav respectively). (A) Representation of the changes in water counts in EC and IC, and the breaking of the interactions among residues R60, D436, and Y335 at the intracellular end of hDAT. Note that when the interactions are broken (right panel), the water-occupied volume (in grey surface representation) is much increased both in the IC-cav and in the region including the Na⁺ site, identified as the IC channel (IC-cha). The color code for structural elements is the same as in Fig. 1A. (B) Water distribution in the intracellular cavity (IC-cav), the extracellular cavity (EC-cav), and the intracellular channel (IC-cha). Blue curves represent all the conformations in the 12 separate trajectories in which Na⁺/Na₂ has been released to the intracellular environment. Red curves represent all the conformations in which Na⁺ is still in the Na₂ site in these 12/50 trajectories from which Na⁺/Na₂ is eventually released. Green curves represent all the conformations where sodium is in the Na₂ site in the 38/50 trajectories in which Na⁺/Na₂ is not released. Confidence intervals at 68% and 95% were calculated for the EC-cav distribution that has normal distribution properties (Table S2).

Concomitant with this movement of Na⁺/Na₂, we observed an increase in the water count inside the intracellular (IC) vestibule (Fig. 2A), which is enabled by the open IC bridges between R60-D436, Y335-E428, and E428-R445 (see Fig. 1B). The accumulation is highest in the trajectories in which Na⁺/Na₂ is released, as shown in Fig. 2 (see IC-cav). Notably, the disruption of the K66-D345 salt bridge (e.g., see replica 30 in Fig. 1B) allows water influx through another region, located between TM1a and IL3.

In extracellular (EC) vestibule the number of water molecules is seen to decrease somewhat as Na⁺/Na₂ leaves (left shift of the distribution plots in Fig. 2, EC-cav), in tandem with the increase in the solvation of the IC-cav, which points to the transition of hDAT from an outward-open/inward-closed configuration to a more outward-closed/inward-open one that accompanies the release of Na⁺/Na₂. The confidence intervals calculated for the water distribution in the EC-cav for the left shift to lower numbers of water molecules observed in EC-cav water distributions for the trajectories in which the Na⁺/Na₂ is released, compared to the ones in which Na⁺/Na₂ is not released, are presented in Table S2. This finding is consistent mechanistically with the observation in the dynamics of extracellular gates that the main extracellular gate (R85 and D476³³) is more stable in trajectories in which Na⁺/Na₂ is released compared to the other trajectories (Figure S3). The water count also increases in a somewhat deeper local region that includes the Na₂ site, termed IC-cha for IC channel (Fig. 1B) (see Methods).

Notably, there is a decrease in the number of water molecules near the Na₂ site after the release of the Na⁺/Na₂ ion (Fig. 2B, IC-cha, compare red and blue traces), which relates to a penetration of K⁺ ions from the IC region towards the Na₂ site near residue D79. The calculated probability of finding K⁺ in the Na₂ site is shown in Fig. 3. This finding is intriguing in light of a known role for K⁺ ions in functional mechanism of the related SERT transporter⁶⁷, where transitions from the inward-facing to the outward-facing state are K⁺ dependent (see Discussion).

Kinetics of the Na⁺/Na₂ release. *The energy landscape for the release process of the Na⁺/Na₂ specified with the tICA method.* The extensive conformational sampling of the system (and the Na⁺/Na₂ dynamics in particular) was subjected to dimensionality reduction with the tICA method by projecting all the trajectory snapshots on tICA reaction coordinates as described in Methods. The construction of the tICA space required a definition of parameters as detailed in Methods. To obtain the specific set of parameters for constructing the tICA

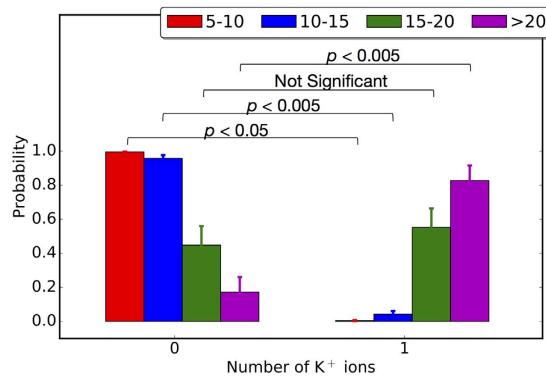


Figure 3. Probability of finding K⁺ ions in the binding site. The probability of finding K⁺ ions within 7 Å of D79 is compared for different states of the system in all 50 trajectories. The states are defined by the distance between the two sodiums (color code on top is in Å), and K⁺ is seen to penetrate the Na2 site only after the Na⁺/Na2 has cleared the hydration channel (i.e., Na2–Na1 distances > 10 Å). The error bars are calculated using the bootstrap method with randomly selecting between 26 to 38 trajectories repeated for 100 times. *p*-values are calculated using the student T-test for the mean of variables. The *p*-values indicate the significance of probability differences between corresponding states in the presence (1 ion) and absence (0 ion) of potassium.

Parameters reflecting Na ⁺ motion	Parameters related to water penetration
Na ⁺ /Na2 to Na ⁺ /Na1 distance	R60–Y335 distance
Na ⁺ /Na2 to E428 distance	R60–E446 distance
Na ⁺ /Na2 to D421 distance	R60–E428 distance
Na ⁺ /Na2 to D79 distance	R60–D436 distance
	Y335–E428 distance
	D436–R445 distance
	E428–R445 distance
	Na ⁺ /Na2 water coordination number

Table 2. Parameters used for dimensionality reduction and tICA construction.

reaction coordinates for the release of Na⁺/Na2 (see Table 2 and Fig. 4) we used the observations of conformational changes described above, and the structural and dynamic inferences from the Na⁺/Na2 release process. The chosen parameters reflect the changes in (1) the Na⁺ ion motion as it exits through the intracellular vestibule; (2) water penetration into the IC vestibule as described by the distance among residues that form the intracellular gate (cf. Fig. 1); and (3) the water coordination number of the Na⁺/Na2 as a parameter for measuring amount of water in the binding site (cf. Fig. 2). These parameters are listed in Table 2 and represented in Fig. 4 in the context of the transporter structure.

It has been shown previously that tICA reaction coordinates are not critically dependent on the lag time used for the construction of the time-lagged covariance matrix, TLCM, described in the Methods⁶⁸. We chose a lag time of 16 ns for the TLCM calculation, and mapped all simulation data on all 12 tICA eigenvectors. Analysis of the contributions from these eigenvectors showed that only the first three tICA eigenvectors produced distributions of the simulation data that were different from normal distribution (Figure S4), indicating that the first 3 eigenvectors are sufficient for discretization. Furthermore, statistical analysis of Markov models (see section below and Figure S6) showed that the quality of a Markov Model built from the first two tICA reaction coordinates is higher, and is thus preferable to the one built from the first three tICA reaction coordinates. Therefore we chose to use the first two tICA reaction coordinates for all subsequent analyses. The components of the tICA eigenvectors show that the first tICA reaction coordinate focuses on the dynamics of Na⁺/Na2, while the second describes dynamics of the N-terminus (Figure S5). Inspection of conformations that belong to different areas of the tICA space revealed that the tICA space spanned by the first two eigenvectors can be divided into the three major regions shown in Fig. 4 which represent, respectively, states with Na⁺ bound to Na2 site, an intermediate state, and the states in which the Na⁺/Na2 had been released.

Markov state based kinetic model for the release of Na⁺/Na2. After dimensionality reduction with the tICA method, the next steps in building the MSM include the discretization of the tICA space using a clustering algorithm, and then finding the best number of microstates for MSM construction. As described in Methods section, we used a recently developed MSM scoring method called generalized matrix Rayleigh quotient (GMRQ)⁶³, and took the highest GMRQ score to define a set of parameters that result in the best MSM. The GMRQ scores were obtained from the described protocol by randomly selecting half of the simulation trajectories (25 trajectories: 6 in which Na⁺/Na2 releases and 19 in which it doesn't release) as the test set, and using the other half to build the MSM. The GMRQ scores for all MSMs are shown in Figure S6. We found that the best MSM is obtained by

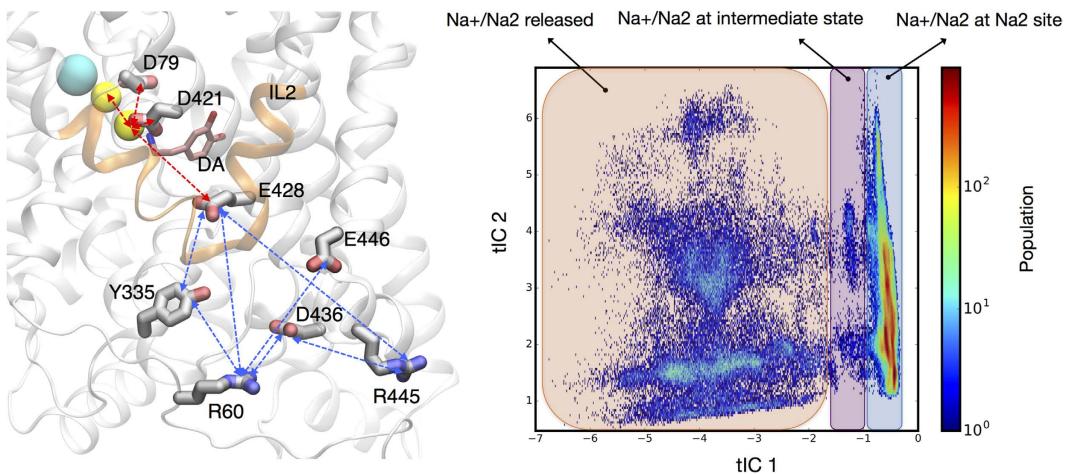


Figure 4. Dimensionality reduction of the entire sampled configurational space. The left panel depicts the two sets of parameters (distances) used for tICA construction (see also Table 2), in the structural context of the transporter. The distance parameters describing the intracellular gate dynamics are indicated with blue arrows; the ones used to describe the Na^+/Na_2 motion are shown with red arrows. Dopamine (DA) is shown in pink licorice, IL2 in orange cartoon, sodium ions as yellow spheres, and the chloride ion as a cyan sphere. The right panel depicts the two-dimensional tICA space. The color map represents the population of states (see color code on the right). The three major regions of this population of states are identified by their colored background: one in which the Na^+/Na_2 is still bound in the Na_2 site (light blue background, on the right); the intermediate state (purple background); and configurations after the Na^+/Na_2 was released (tan color background).

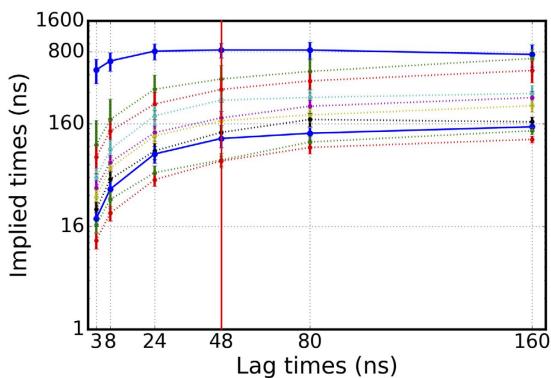


Figure 5. Implied-timescales for the Na^+/Na_2 release kinetics. The Markovian behavior of the system improves as MSM lag-time increases. The 48 ns lag-time (highlighted with red line) was used for constructing transition probability matrix and subsequent analysis. The 1st and 8th relaxation modes are highlighted in bold blue curves. Error bars are calculated using the bootstrap method by using 10 subsamples each generated by randomly choosing 50 trajectories with replacement method yielding between 28 to 36 unique trajectories for each subsample.

using the k -means clustering algorithm to obtain about 100 microstates on the first two tICA eigenvectors (see discussion in Figure S6 caption).

Using the best set of parameters identified with the GMRQ analysis, we obtained multiple transition probability matrices (TPMs) in order to construct the *implied time-scales plot* (see Methods). We checked the Markovian behavior of the system by plotting the implied timescales corresponding to the first 10 relaxation modes of each TPM, against multiple lag times (Fig. 5). The final MSM was chosen from the plot by identifying the position at which the curves saturate; a lag-time of ~ 50 ns is seen to ensure Markovian behavior with minimal loss of data.

As shown in Fig. 5, the transition probability matrix constructed at 48 ns lag time fulfills the Markovian behavior criterion and this transition probability matrix was chosen for the subsequent analyses. The eigenvector corresponding to the first relaxation mode of the selected transition probability matrix is depicted in Fig. 6A. Negative and positive values correspond, respectively, to states in which Na^+/Na_2 was released, and states in which Na^+/Na_2 was still bound (Fig. 6B). Since the population flows from states with positive values to states with negative values, we conclude that this relaxation mode captures the overall dynamics of the Na^+/Na_2 release process. The Markov model predicts a timescale of about 800 ns for this relaxation mode (Figs 5 and 6). This predicted timescale should be considered as the average time for Na^+/Na_2 release *after* the IC gates are broken, since the

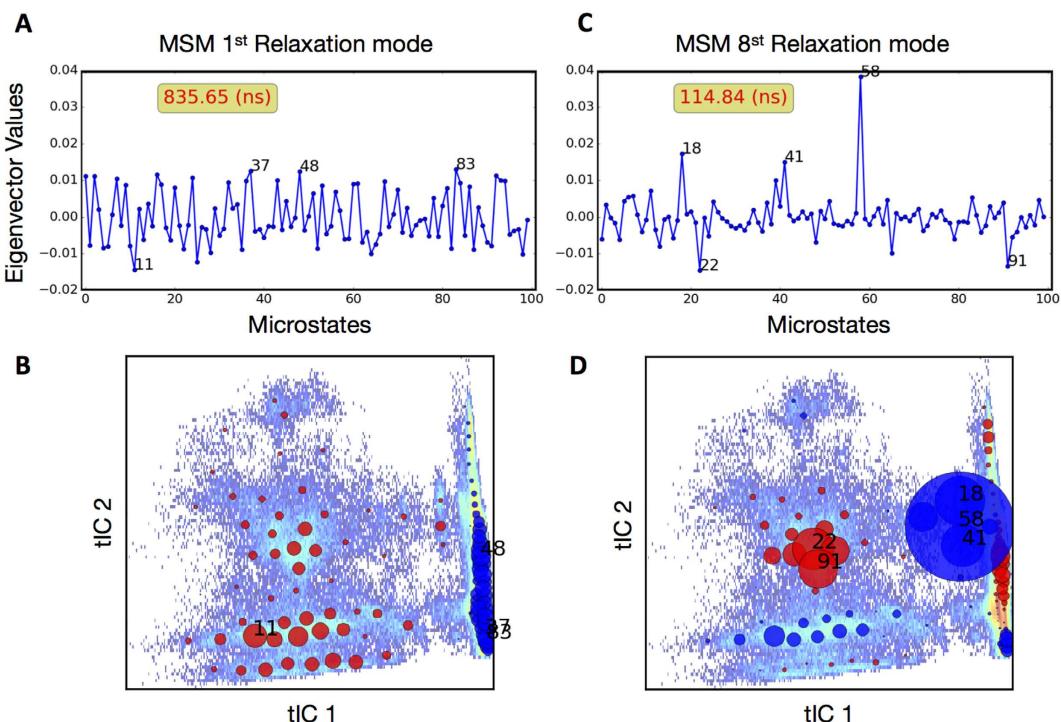


Figure 6. The first and eighth MSM relaxation modes. (A) Shows the contribution of each microstate to the first relaxation mode (eigenvector values). The microstates with the largest contributions are numbered. (B) The location of each microstate contributing to the first MSM relaxation mode is shown in the space of the first two tICA eigenvectors. Red dots indicate negative values, and blue dots indicate positive values, with the diameters indicating relative contributions. Note that all blue (positive) dots belong to the “ $\text{Na}^+/\text{Na}2$ bound” tICA space (cf. Fig. 4), whereas all red (negative) dots belong to the “ $\text{Na}^+/\text{Na}2$ released” domain in tICA space. Panels C and D show the same data as A and B, for the eighth MSM relaxation mode. This relaxation mode corresponds to the intermediate state (see text).

ensemble simulations covering the conformational space were initiated from the hDAT conformation in which the IC gates were destabilized due to PIP_2 -mediated N-term/IL4 interactions²⁸.

Identification of an intermediate state on the release pathway of $\text{Na}^+/\text{Na}2$. For trajectories in which $\text{Na}^+/\text{Na}2$ is released, we calculated the free energy profile of the system based on populations observed along the 1st tICA reaction coordinate (Fig. 7). The profile identifies three major basins, and analysis of conformations belonging to each free energy basin shows that the one with the lowest free energy corresponds to a state where Na^+ is bound in the Na2 site. The second lowest free energy basin corresponds to a state in which $\text{Na}^+/\text{Na}2$ was released, and the third, middle basin, corresponds to states in which $\text{Na}^+/\text{Na}2$ has left the Na2 site, is coordinated by E428, and therefore has not yet been released to the intracellular solution (Fig. 7). Tracking the $\text{Na}^+/\text{Na}2$ release pathway along the tICA landscape reveals that in some of the release trajectories the metastable intermediate state is more populated than in others in which this state is relatively short-lived (Fig. 8 and Figure S7). The microstates corresponding to the intermediate state (microstates 18, 41, and 58) dominate the relaxation mode that is captured by the 8th MSM eigenvector (Fig. 6C, Figures S8 and S9), which identifies the eighth eigenvector as the most relevant relaxation mode corresponding to this intermediate state; the kinetics of the intermediate state is thus quantified as being on the order of 100 ns (Fig. 6C and D).

Detailed analysis of the trajectories suggests that the intermediate state becomes accessible as a result of disruption of contacts between residues E428 and R445. In Figure S10 the plot of the R445-E428 distance against tIC 1 shows that this interaction is fully broken only in the intermediate state. This establishes the E428/R445 pair as part of the functional IC network (see Fig. 1) that stabilizes hDAT in an inward-closed state as observed in the simulations (see above). Breaking the *ionic lock* between these residues upon PIP_2 -mediated interactions between the N-terminus and IL4, frees E428 to coordinate the $\text{Na}^+/\text{Na}2$ and form the intermediate state.

Sodium release pathways. The transition path theory (TPT)^{64,69} described in the Methods section was used to examine the sodium release pathways and to identify relevant conformational changes associated with each pathway. The analysis is implemented in the MSMBuilder software package⁶¹ and provides information about the most probable pathways and the overall flux associated with each pathway (see Methods). For TPT analysis, the 100 microstates were grouped into 15 *macrostates* (Fig. 9) based on their kinetic similarity determined with the Robust Perron Cluster Analysis (PCCA^+) algorithm⁶⁶. Briefly, the PCCA^+ algorithm lumps microstates into macrostates assuming that microstates with the same signs (positive or negative) in the MSM eigenvectors, will have similar kinetics⁷⁰.

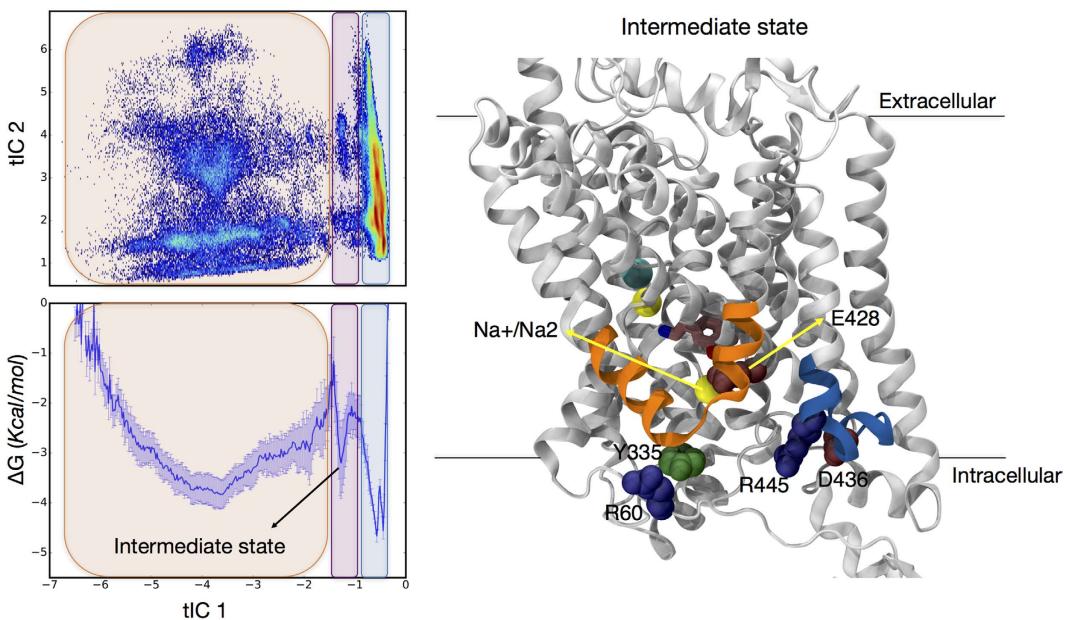


Figure 7. Characteristics of an intermediate state. In the intermediate state the Na^+/Na_2 has left the binding site, is interacting with E428, but has not yet been released to the intracellular solution. Panels on the left show the two-dimensional tICA space (upper) and the corresponding free energy profile (lower panel) obtained by taking the natural log of populations based on tICA 1st eigenvector. The right panel shows a conformation corresponding to the intermediate state. The R445–E428 interaction is broken and Na^+/Na_2 is stabilized by E428. Color code is the same as in Fig. 1A. Error bars for the free energy values are calculated using the bootstrap method on blocks of frames with 180 ± 70 ns time range that are randomly selected from all the trajectories in which Na^+/Na_2 is released to the intracellular environment.

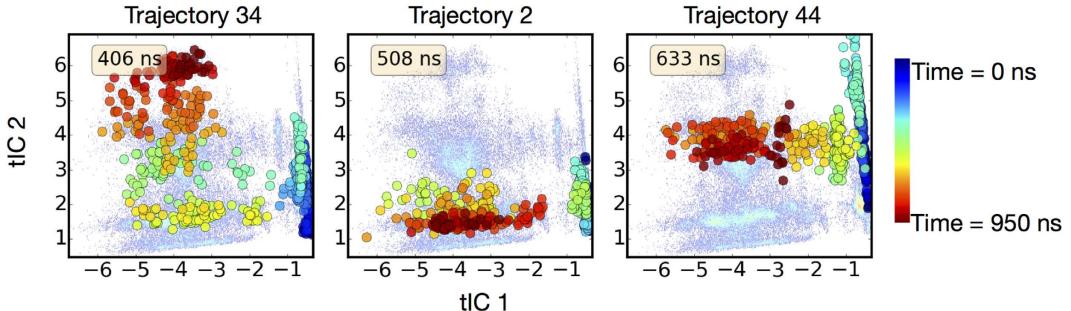


Figure 8. Na^+/Na_2 release pathway on tICA space. Tracking the Na^+/Na_2 release pathway in three representative trajectories (identified on top of each panel) in tICA space. The light blue background shows the tICA landscape obtained from all 50 trajectories. The large colored dots indicate the position of the conformation in tICA space as function of time in the trajectory. The color code of the large dots indicates the time in the trajectory (see color bar). Specifically, the time sequence starts as dark blue color and ends with dark red. Thus, blue dots represent initial states and dark red dots represent final states in each trajectory. The insert in each panel indicates the approximate time where Na^+/Na_2 is released to the intracellular environment. Note that in some trajectories (#44) the metastable intermediate state is well-populated (as seen from the number of yellow and light orange dots in the region of the transition state), while in others (#2) it is not visited. For similar plots for all other trajectories in which Na^+/Na_2 is released, see Figure S7.

The full transition flux matrix obtained from the model is shown in Figure S12 and error estimates for fluxes are given in Table S3. The top 7 Na^+/Na_2 release pathways identified by the TPT analysis are shown in Fig. 9 (see also Table S4); they contribute >60% of the total Na^+/Na_2 release flux. These pathways could be further clustered into three major sodium release pathways categorized as follows:

The most probable pathway is highlighted in the green box on tICA space in Fig. 9. This pathway is created when the interaction pairs R60-D436 and E428-R445 are replaced by the R60-E428 salt bridge, resulting in the opening of IC gates that allow water penetration as far as the Na_2 site. The Na^+/Na_2 is released with a short dwell time in the intermediate state because E428 is engaged in the interaction with R60 and cannot stabilize the Na^+ ion.

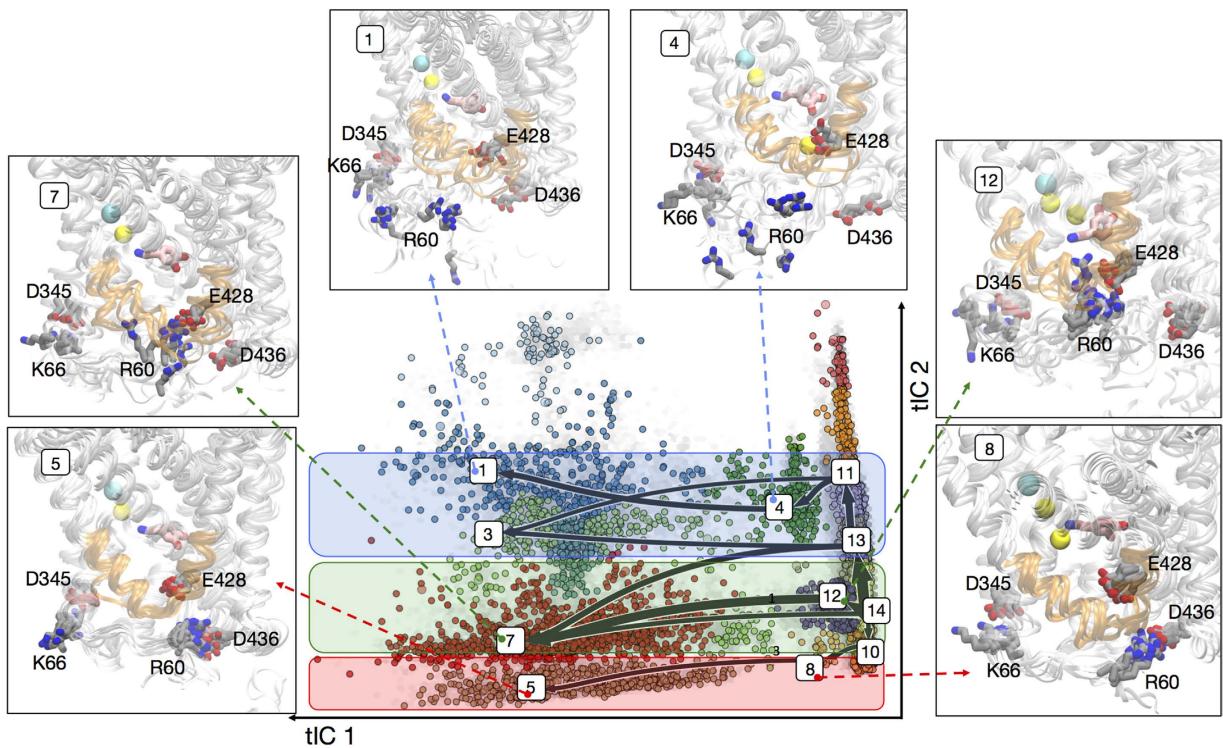


Figure 9. Sodium release pathways. The tICA space (light gray background) is divided into 15 (color coded) macrostates using the PCCA⁺ algorithm, and the most probable release pathways determined with TPT (see text), are indicated on the tICA landscape by arrows that connect macrostates indicated by their numbers. The arrows point from Na⁺-not-released macrostates (on the right) to Na⁺-released macrostates (on the left). The thickness of the arrows indicates the relative magnitude of the flux (i.e., likelihood of release) of the pathway. Flux values are given in Table S3. The tICA spaces corresponding to three distinct pathways (see text) are highlighted in red, green, and blue boxes. Structural representations for key macrostates (numbered) are shown in the structure panels. Residues R60, D436, K66, D345, and E428 are rendered in licorice, IL2 in orange cartoon, dopamine in pink licorice, and the sodium and chloride ions in yellow and cyan spheres, respectively.

The second major pathway (blue box in Fig. 9) is formed when R60 disengages from D436, and interacts through PIP2 lipids with Lys residues (K257, K260, K264) from intracellular loop 2 (IL2) (Figure S11). These interactions lead to partial unfolding of IL2, a structural feature that has been suggested to be involved in the substrate release mechanism in the related LeuT-fold bacterial MhsT transporter³³ (see Discussion). In this pathway, Na⁺ is stabilized in the intermediate state through interactions with the available E428.

In the third major pathway (red box in Fig. 9) R60 is engaged with D436, hence the intracellular gate is not fully open. In this pathway the Na2 site is hydrated through a different channel compared to the other two, as the water molecules penetrating the Na2 site reach it through an opening between TM1a and IL3 that results from the breaking of the K66-D345 interaction (cf. Fig. 1B).

It is noteworthy that the TPT analysis of the Markov Models identifies several pathways of release of the Na⁺/Na2 which engage the same set of structural elements but somewhat different configurations and kinetics. These differences depend on specific variations in the mode of interaction of the N-terminus with the intracellular regions of the transporter.

Discussion

The role of the Na⁺/Na2 in functional mechanisms of mammalian NSS and bacterial homologs, has been addressed in a variety of studies^{16,22,28,30,31}. The binding of the Na⁺ ion in the Na2 site was shown in such studies to contribute to the stability of an inward-closed conformation of these transporters, so that release of the Na⁺/Na2 ion is a likely step in the initiation of the inward transport cycle of substrate and sodium. Indeed, we described recently²⁸ how the release of the Na⁺/Na2 ion in hDAT is coupled to a spontaneous inward-opening of the transporter in a process triggered by PIP₂-mediated interactions between the N-terminus and IL4 domains. In particular, these PIP₂-mediated interactions can disrupt the intracellular network of interactions^{24,26} among residues R60, D436, E428, Y335, and R445 that stabilize hDAT in an inward-closed state. Starting from the state in which the intracellular network of interactions is weakened, our Markov model predicts that once the intracellular gates in hDAT are broken, the spontaneous release step of Na⁺/Na2 in the 12 out of 50 replicas of ~1 μs simulations, proceeds relatively fast (timescale of about 800 ns). Indeed, inspection of all 50 trajectories indicates that in longer simulation times the Na⁺/Na2 would have been released in all of them, judging from the accumulation of elements in the dynamic pattern associated with release.

Of the three major release pathways for the $\text{Na}^+/\text{Na}2$ revealed by the TPT analysis, one is seen to go through a metastable intermediate state in which $\text{Na}^+/\text{Na}2$ is coordinated by E428 before its escape (blue box in Fig. 9). We found the average dwell-time of $\text{Na}^+/\text{Na}2$ in this intermediate state to be ~ 100 ns and showed that the formation of this state is related to disruption of the E428-R445 ionic bond when the N-terminus interacts with intracellular loop 4 (“N-term/IL4 interactions” in Fig. 1B). The E428 side chain can thus participate in both the stabilization of the intracellular gates, and in the coordination of the Na^+ ion released from the Na2 site. This finding from our analysis is consonant with an experimentally identified functional role of residue E428 in hDAT⁷¹: the E428Q substitution in hDAT expressed in HEK293 cells was found to impair the binding of the inhibitor CFT, suggesting that this mutation favors an inward-open state in hDAT over the outward-open conformation to which CFT binds. Our computational findings are consistent with these results, as the E428Q substitution will weaken the interaction with R445 and thereby influence the outward-open to inward-open transition equilibrium. This is substantiated as well by available structural data on the bacterial homolog LeuT⁴ in which the Gln that occupies the position aligning with E428 in hDAT, does not interact with the R445-equivalent residue in LeuT (R375).

It is noteworthy that the E428-dependent intermediate state in the $\text{Na}^+/\text{Na}2$ release pathway we discovered, seems to enable K^+ ions from the intracellular environment to partition into the Na2 site. The role of K^+ ions in the function of the related SERT transporter is well-established⁶⁷, and we have recently described significant effects that the presence of a K^+ ion in the Na2 site can have on the dynamics of the extracellular end of the cognate LeuT and its transport cycle(see refs 72 and 73). While it is tempting to speculate that the ability of K^+ ions to enter the intracellular region of DAT could facilitate $\text{Na}^+/\text{Na}2$ release from the intermediate state, a more complete evaluation of the impact that K^+ ions may have on various elements of NSS mechanisms will require future considerations based on quantitative free energy-based computations.

The other major pathway for the $\text{Na}^+/\text{Na}2$ release, enabled when R60 breaks from D436 and engages with Lys residues in the intracellular loop 2 (IL2) through PIP₂-mediated interactions (Figure S11), is also intriguing as it appears to involve partial unfolding of the IL2 segment. Such a structural perturbation was seen recently in the X-ray structure of related bacterial transporter MhsT³³, and has been suggested to be a required mechanistic feature of the NSS transporters to allow release of the $\text{Na}^+/\text{Na}2$. Thus, to our knowledge, we provide here the first direct supporting evidence for a dynamic mode of involvement of the IL2 loop unfolding in the $\text{Na}^+/\text{Na}2$ release process.

The third major pathway for the $\text{Na}^+/\text{Na}2$ release is created by water molecules entering the IC vestibule through the region between TM1a and IL3 as a result of the loss of K66-D345 interaction. The involvement of D345 in $\text{Na}^+/\text{Na}2$ release mechanisms identified in this pathway is interesting in light of experimental evidence strongly implicating this residue in the transport mechanisms of hDAT⁷¹. Specifically, the D345N mutant of hDAT expressed in HEK293 cells has a significantly impaired maximal velocity for dopamine uptake (V_{\max}) and a reduced maximal capacity of CFT binding (B_{\max}). Furthermore, these studies concluded that D345N hDAT had a reduced turnover rate, which, in turn, was shown to be largely determined by the rate of reorientation of the unloaded inward-facing transporter. Unlike the wild type Asp at position 345, the Asn would not engage in the same strong interaction with K66 and therefore, according to our computational results, would increase the probability for opening the $\text{Na}^+/\text{Na}2$ release pathway. Since the release concomitantly widens the intracellular vestibule, the impairment of the K66-D345 interaction is then predicted to affect the kinetics of inward-open to inward-close transition, consistent with the conclusions from the cited experiments.

The detailed structural rearrangements and membrane-supported interactions underlying the kinetic model for $\text{Na}^+/\text{Na}2$ release after the intracellular gates are broken due to the PIP₂-mediated N-term/IL4 association, represent steps in an allosteric mechanism of a key mechanistic stage in the transport cycle of NSS transporters^{36,74}. We showed that this allosteric mechanism could follow distinct transition paths in which the same set of structural elements are engaged, but with somewhat different configurations of interactions. Thus, the different modes in which the N-terminus can interact with the intracellular region of the transporter, and variations in the manner in which E428 can stabilize different salt bridges and thereby modulate the mode of $\text{Na}^+/\text{Na}2$ release, can result in different kinetic pathways even for the relatively simple step of $\text{Na}^+/\text{Na}2$ release in the transport cycle of hDAT. Since these pathways were shown in our analysis to correspond to different energy basins, it is clearly desirable to develop a comprehensive kinetic model of the entire process, including the intracellular gate-breaking step, and to base the mechanistic analysis of the transition from the outward-facing configuration to the inward-facing configuration of the transporter on similar considerations of the energetics. Further studies are now ongoing to collect such information.

References

1. Broer, S. & Gether, U. The solute carrier 6 family of transporters. *Br J Pharmacol* **167**, 256–278 (2012).
2. Sulzer, D., Sonders, M. S., Poulsen, N. W. & Galli, A. Mechanisms of neurotransmitter release by amphetamines: a review. *Prog Neurobiol* **75**, 406–433 (2005).
3. Sora, I. *et al.* Monoamine transporter as a target molecule for psychostimulants. *Int Rev Neurobiol* **85**, 29–33 (2009).
4. Yamashita, A., Singh, S. K., Kawate, T., Jin, Y. & Gouaux, E. Crystal structure of a bacterial homologue of Na^+/Cl^- -dependent neurotransmitter transporters. *Nature* **437**, 215–223 (2005).
5. Penmatsa, A. & Gouaux, E. How LeuT shapes our understanding of the mechanisms of sodium-coupled neurotransmitter transporters. *J Physiol* **592**, 863–869 (2014).
6. Penmatsa, A., Wang, K. H. & Gouaux, E. X-ray structures of Drosophila dopamine transporter in complex with nisoxetine and reboxetine. *Nature structural & molecular biology* **22**, 506–508 (2015).
7. Wang, K. H., Penmatsa, A. & Gouaux, E. Neurotransmitter and psychostimulant recognition by the dopamine transporter. *Nature* **521**, 322–327 (2015).
8. Penmatsa, A., Wang, K. H. & Gouaux, E. X-ray structure of dopamine transporter elucidates antidepressant mechanism. *Nature* **503**, 85–90 (2013).
9. Coleman, J. A., Green, E. M. & Gouaux, E. X-ray structures and mechanism of the human serotonin transporter. *Nature* **532**, 334–339 (2016).
10. Penmatsa, A. & Gouaux, E. How LeuT shapes our understanding of the mechanisms of sodium-coupled neurotransmitter transporters. *J Physiol* **592**, 863–869 (2013).

11. Kazmier, K. *et al.* Conformational dynamics of ligand-dependent alternating access in LeuT. *Nature structural & molecular biology* **21**, 472–479 (2014).
12. Zhao, Y., Terry, D., Shi, L., Weinstein, H., Blanchard, S. C. & Javitch, J. A. Single-molecule dynamics of gating in a neurotransmitter transporter homologue. *Nature* **465**, 188–193 (2010).
13. Zhao, Y. *et al.* Substrate-modulated gating dynamics in a Na⁺-coupled neurotransmitter transporter homologue. *Nature* **474**, 109–113 (2011).
14. Claxton, D. P. *et al.* Ion/substrate-dependent conformational dynamics of a bacterial homolog of neurotransmitter: sodium symporters. *Nature structural & molecular biology* **17**, 822–829 (2010).
15. Kazmier, K., Sharma, S., Islam, S. M., Roux, B. & McHaourab, H. S. Conformational cycle and ion-coupling mechanism of the Na⁺/hydantoin transporter Mhp1. *Proc Natl Acad Sci USA* **111**, 14752–14757 (2014).
16. Shi, L., Quick, M., Zhao, Y., Weinstein, H. & Javitch, J. A. The mechanism of a neurotransmitter:sodium symporter–inward release of Na⁺ and substrate is triggered by substrate in a second binding site. *Mol Cell* **30**, 667–677 (2008).
17. Quick, M., Shi, L., Zehnpfennig, B., Weinstein, H. & Javitch, J. A. Experimental conditions can obscure the second high-affinity site in LeuT. *Nature structural & molecular biology* **19**, 207–211 (2012).
18. Zhao, C., Stolzenberg, S., Gracia, L., Weinstein, H., Noskov, S. & Shi, L. Ion-controlled conformational dynamics in the outward-open transition from an occluded state of LeuT. *Biophys J* **103**, 878–888 (2012).
19. Quick, M., Winther, A. M., Shi, L., Nissen, P., Weinstein, H. & Javitch, J. A. Binding of an octylglucoside detergent molecule in the second substrate (S2) site of LeuT establishes an inhibitor-bound conformation. *Proc Natl Acad Sci USA* **106**, 5563–5568 (2009).
20. Cheng, M. H. & Bahar, I. Complete mapping of substrate translocation highlights the role of LeuT N-terminal segment in regulating transport cycle. *PLoS Comput Biol* **10**, e1003879 (2014).
21. Shaikh, S. A. & Tajkhorshid, E. Modeling and dynamics of the inward-facing state of a Na⁺/Cl⁻ dependent neurotransmitter transporter homologue. *PLoS Comput Biol* **6**, e1000905 (2010).
22. Kiessling, V., Wan, C. & Tamm, L. K. Domain coupling in asymmetric lipid bilayers. *Biochimica Et Biophysica Acta-Biomembranes* **1788**, 64–71 (2009).
23. LeVine, M. V., Khelashvili, G., Shi, L., Quick, M., Javitch, J. A. & Weinstein, H. Role of Annular Lipids in the Functional Properties of Leucine Transporter LeuT Proteomicelles. *Biochemistry* **55**, 850–859 (2016).
24. Shan, J., Javitch, J. A., Shi, L. & Weinstein, H. The substrate-driven transition to an inward-facing conformation in the functional mechanism of the dopamine transporter. *PLoS One* **6**, e16350 (2011).
25. Guptaroy, B. *et al.* A juxtamembrane mutation in the N terminus of the dopamine transporter induces preference for an inward-facing conformation. *Mol Pharmacol* **75**, 514–524 (2009).
26. Kniazeff, J., Shi, L., Loland, C. J., Javitch, J. A., Weinstein, H. & Gether, U. An intracellular interaction network regulates conformational transitions in the dopamine transporter. *J Biol Chem* **283**, 17691–17701 (2008).
27. Stockner, T. *et al.* Mutational analysis of the high-affinity zinc binding site validates a refined human dopamine transporter homology model. *PLoS Comput Biol* **9**, e1002909 (2013).
28. Khelashvili, G. *et al.* Spontaneous inward opening of the dopamine transporter is triggered by PIP2-regulated dynamics of the N-terminus. *ACS Chem Neurosci* **6**, 1825–1837 (2015).
29. Cheng, M. H. & Bahar, I. Molecular Mechanism of Dopamine Transport by Human Dopamine Transporter. *Structure* **23**, 2171–2181 (2015).
30. Borre, L., Andreassen, T. F., Shi, L., Weinstein, H. & Gether, U. The second sodium site in the dopamine transporter controls cation permeation and is regulated by chloride. *J Biol Chem* **289**, 25764–25773 (2014).
31. Koldso, H., Autzen, H. E., Grouleff, J. & Schiott, B. Ligand induced conformational changes of the human serotonin transporter revealed by molecular dynamics simulations. *PLoS One* **8**, e63635 (2013).
32. Jardetzky, O. Simple allosteric model for membrane pumps. *Nature* **211**, 969–970 (1966).
33. Malinauskaitė, L. *et al.* A mechanism for intracellular release of Na⁺ by neurotransmitter/sodium symporters. *Nature structural & molecular biology* **21**, 1006–1012 (2014).
34. Khelashvili, G., Doktorova, M., Sahai, M. A., Johner, N., Shi, L. & Weinstein, H. Computational modeling of the N-terminus of the human dopamine transporter and its interaction with PIP2-containing membranes. *Proteins* **83**, 952–969 (2015).
35. Hamilton, P. J. *et al.* PIP2 regulates psychostimulant behaviors through its interaction with a membrane protein. *Nat Chem Biol* **10**, 582–589 (2014).
36. LeVine, M. V., Cuendet, M. A., Khelashvili, G. & Weinstein, H. Allosteric Mechanisms of Molecular Machines at the Membrane: Transport by Sodium-Coupled Symporters. *Chem Rev* **116**, 6552–6587 (2016).
37. Chodera, J. D., Singhal, N., Pande, V. S., Dill, K. A. & Swope, W. C. Automatic discovery of metastable states for the construction of Markov models of macromolecular conformational dynamics. *J Chem Phys* **126**, 155101 (2007).
38. Loland, C. J., Granas, C., Javitch, J. A. & Gether, U. Identification of intracellular residues in the dopamine transporter critical for regulation of transporter conformation and cocaine binding. *J Biol Chem* **279**, 3228–3238 (2004).
39. Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *Journal of Computational Chemistry* **26**, 1781–1802 (2005).
40. Essmann, U., Perera, L., Berkowitz, M. L., Darden, T., Lee, H. & Pedersen, L. G. A Smooth Particle Mesh Ewald Method. *Journal of Chemical Physics* **103**, 8577–8593 (1995).
41. Evans, D. J. & Holian, B. L. The Nose-Hoover Thermostat. *Journal of Chemical Physics* **83**, 4069–4074 (1985).
42. Harvey, M., Giupponi, G. & De Fabritiis, G. ACEMD: Accelerated molecular dynamics simulations in the microseconds timescale. *Journal of Chemical Theory and Computation* **5**, 1632–1639 (2009).
43. Best, R. B. *et al.* Optimization of the additive CHARMM all-atom protein force field targeting improved sampling of the backbone phi, psi and side-chain chi(1) and chi(2) dihedral angles. *J Chem Theory Comput* **8**, 3257–3273 (2012).
44. Klauda, J. B. *et al.* Update of the CHARMM All-Atom Additive Force Field for Lipids: Validation on Six Lipid Types. *Journal of Physical Chemistry B* **114**, 7830–7843 (2010).
45. Wu, E. L. *et al.* CHARMM-GUI Membrane Builder toward realistic biological membrane simulations. *J Comput Chem* **35**, 1997–2004 (2014).
46. Venable, R. M., Luo, Y., Gawrisch, K., Roux, B. & Pastor, R. W. Simulations of anionic lipid membranes: development of interaction-specific ion parameters and validation using NMR data. *The journal of physical chemistry B* **117**, 10183–10192 (2013).
47. Perez-Hernandez, G., Paul, F., Giorgino, T., De Fabritiis, G. & Noe, F. Identification of slow molecular order parameters for Markov model construction. *J Chem Phys* **139**, 015102 (2013).
48. Schwantes, C. R. & Pande, V. S. Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. *J Chem Theory Comput* **9**, 2000–2009 (2013).
49. Molgedey, L. & Schuster, H. G. Separation of a mixture of independent signals using time delayed correlations. *Phys Rev Lett* **72**, 3634–3637 (1994).
50. Naritomi, Y. & Fuchigami, S. Slow dynamics in protein fluctuations revealed by time-structure based independent component analysis: the case of domain motions. *J Chem Phys* **134**, 065101 (2011).
51. Zhou, G. & Voelz, V. A. Using Kinetic Network Models To Probe Non-Native Salt-Bridge Effects on alpha-Helix Folding. *The journal of physical chemistry B* **120**, 926–935 (2016).
52. Razavi, A. M., Wuest, W. M. & Voelz, V. A. Computational screening and selection of cyclic peptide hairpin mimetics by molecular simulation and kinetic network models. *J Chem Inf Model* **54**, 1425–1432 (2014).

53. Voelz, V. A., Bowman, G. R., Beauchamp, K. & Pande, V. S. Molecular Simulation of ab Initio Protein Folding for a Millisecond Folder NTL9. *Journal of the American Chemical Society* **132**, 1526–1528 (2009).
54. Beauchamp, K. A., McGibbon, R., Lin, Y. S. & Pande, V. S. Simple few-state models reveal hidden complexity in protein folding. *Proc Natl Acad Sci USA* **109**, 17807–17813 (2012).
55. Kohlhoff, K. J. et al. Cloud-based simulations on Google Exacycle reveal ligand modulation of GPCR activation pathways. *Nat Chem* **6**, 15–21 (2014).
56. Shukla, D., Meng, Y., Roux, B. & Pande, V. S. Activation pathway of Src kinase reveals intermediate states as targets for drug design. *Nat Commun* **5**, 3397 (2014).
57. Prinz, J. H. et al. Markov models of molecular kinetics: generation and validation. *J Chem Phys* **134**, 174105 (2011).
58. Noe, F. & Fischer, S. Transition networks for modeling the kinetics of conformational change in macromolecules. *Curr Opin Struct Biol* **18**, 154–162 (2008).
59. Pande, V. S., Beauchamp, K. & Bowman, G. R. Everything you wanted to know about Markov State Models but were afraid to ask. *Methods* **52**, 99–105 (2010).
60. Chodera, J. D., Swope, W. C., Pitera, J. W. & Dill, K. A. Long-Time Protein Folding Dynamics from Short-Time Molecular Dynamics Simulations. *Multiscale Modeling & Simulation* **5**, 1214–1226 (2006).
61. Beauchamp, K. A., Bowman, G. R., Lane, T. J., Maibaum, L., Haque, I. S. & Pande, V. S. MSMBuilder2: Modeling Conformational Dynamics at the Picosecond to Millisecond Scale. *J Chem Theory Comput* **7**, 3412–3419 (2011).
62. Lyman, E. & Zuckerman, D. M. On the structural convergence of biomolecular simulations by determination of the effective sample size. *The journal of physical chemistry B* **111**, 12876–12882 (2007).
63. McGibbon, R. T. & Pande, V. S. Variational cross-validation of slow dynamical modes in molecular kinetics. *J Chem Phys* **142**, 124105 (2015).
64. Berezhkovskii, A., Hummer, G. & Szabo, A. Reactive flux and folding pathways in network models of coarse-grained protein dynamics. *J Chem Phys* **130**, 205102 (2009).
65. Dijkstra, E. W. A note on two problems in connexion with graphs. *Numerische mathematik* **1**, 269–271 (1959).
66. Deuflhard, P. & Weber, M. Robust Perron cluster analysis in conformation dynamics. *Linear algebra and its applications* **398**, 161–184 (2005).
67. Torres, G. E., Gainetdinov, R. R. & Caron, M. G. Plasma membrane monoamine transporters: structure, regulation and function. *Nat Rev Neurosci* **4**, 13–25 (2003).
68. Razavi, A. M. & Voelz, V. A. Kinetic network models of tryptophan mutations in beta-hairpins reveal the importance of non-native interactions. *J Chem Theory Comput* **11**, 2801–2812 (2015).
69. Vanden-Eijnden E, E. W. Towards a Theory of Transition Paths. *Journal of Statistical Physics* **123**, 503–523 (2006).
70. Chodera, J. D. & Noé, F. Markov state models of biomolecular conformational dynamics. *Current opinion in structural biology* **25**, 135–144 (2014).
71. Chen, N., Vaughan, R. A. & Reith, M. E. The role of conserved tryptophan and acidic residues in the human dopamine transporter as characterized by site-directed mutagenesis. *Journal of neurochemistry* **77**, 1116–1127 (2001).
72. Khelashvili, G. et al. Conformational dynamics on the extracellular side of LeuT controlled by Na⁺ and K⁺ ions and the protonation state of E290. *J Biol Chem* **291**, 19786–19799 (2016).
73. Billésbølle, C. B. et al. Transition metal ion FRET uncovers K⁺ regulation of a neurotransmitter/sodium symporter. *Nat Commun* **7**, 12755 doi: 10.1038/ncomms12755 (2016).
74. Stolzenberg, S., Michino, M., LeVine, M. V., Weinstein, H. & Shi, L. Computational approaches to detect allosteric pathways in transmembrane molecular machines. *Biochim Biophys Acta* **1858**, 1652–1662 (2016).

Acknowledgements

The authors are grateful to Drs. Michel A Cuendet and Michael V. LeVine for critical reading of the manuscript and excellent suggestions. The work was supported by NIH Grants P01 DA012408 and U54 GM087519. The following computational resources are gratefully acknowledged: resources of the Oak Ridge Leadership Computing Facility (ALCC allocation BIP109) at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725; an allocation at the National Energy Research Scientific Computing Center (NERSC, repository m1710) supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231; and the computational resources of the David A. Cofrin Center for Biomedical Information in the HRH Prince Alwaleed Bin Talal Bin Abdulaziz Alsaud Institute for Computational Biomedicine at Weill Cornell Medical College.

Author Contributions

All authors contributed to conceiving the project and experimental design, A.M.R. carried out the MSM construction, G.K. carried out the MD simulations, all authors participated in the interpretation of results and in writing the manuscript.

Additional Information

Supplementary information accompanies this paper at <http://www.nature.com/srep>

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Razavi, A. M. et al. A Markov State-based Quantitative Kinetic Model of Sodium Release from the Dopamine Transporter. *Sci. Rep.* **6**, 40076; doi: 10.1038/srep40076 (2016).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>

A Markov State-based Quantitative Kinetic Model of Sodium Release from the Dopamine Transporter

Asghar M. Razavi¹, George Khelashvili¹, Harel Weinstein^{1,2*}

¹Department of Physiology and Biophysics, Weill Cornell Medical College of Cornell University, New York, NY, 10065, USA. ²Institute for Computational Biomedicine, Weill Medical College of Cornell University, New York, NY 10065, USA. *Corresponding author: (email: haw2002@med.cornell.edu)

Supplementary Information

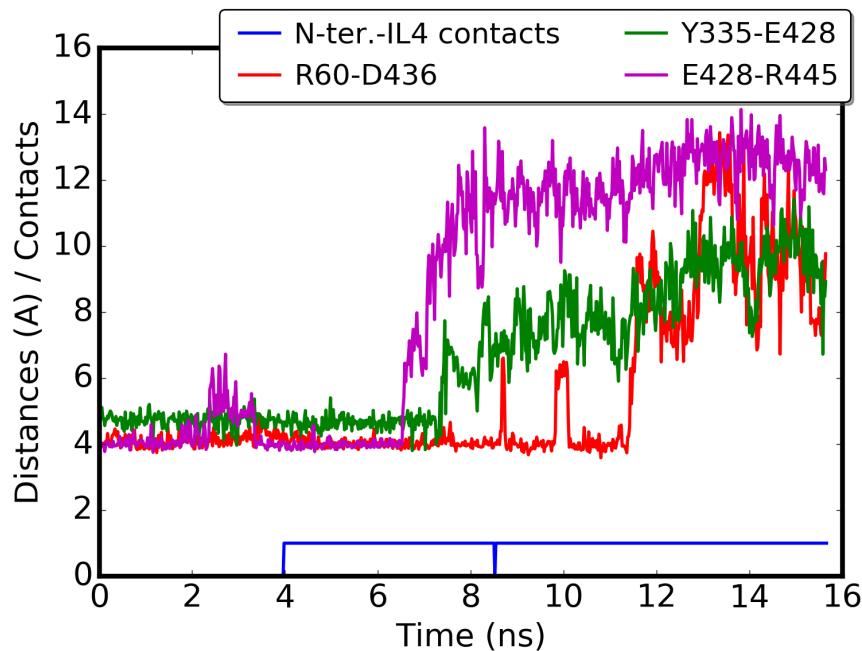


Figure S1. Intracellular gate opening during the equilibration phase. As the PIP₂ mediated interaction of the N-terminus (R51 here) and IL4 (R443) is established, the network of interactions at the intracellular gate (R60–D436, Y335–E428, and R445–E428) is destabilized.

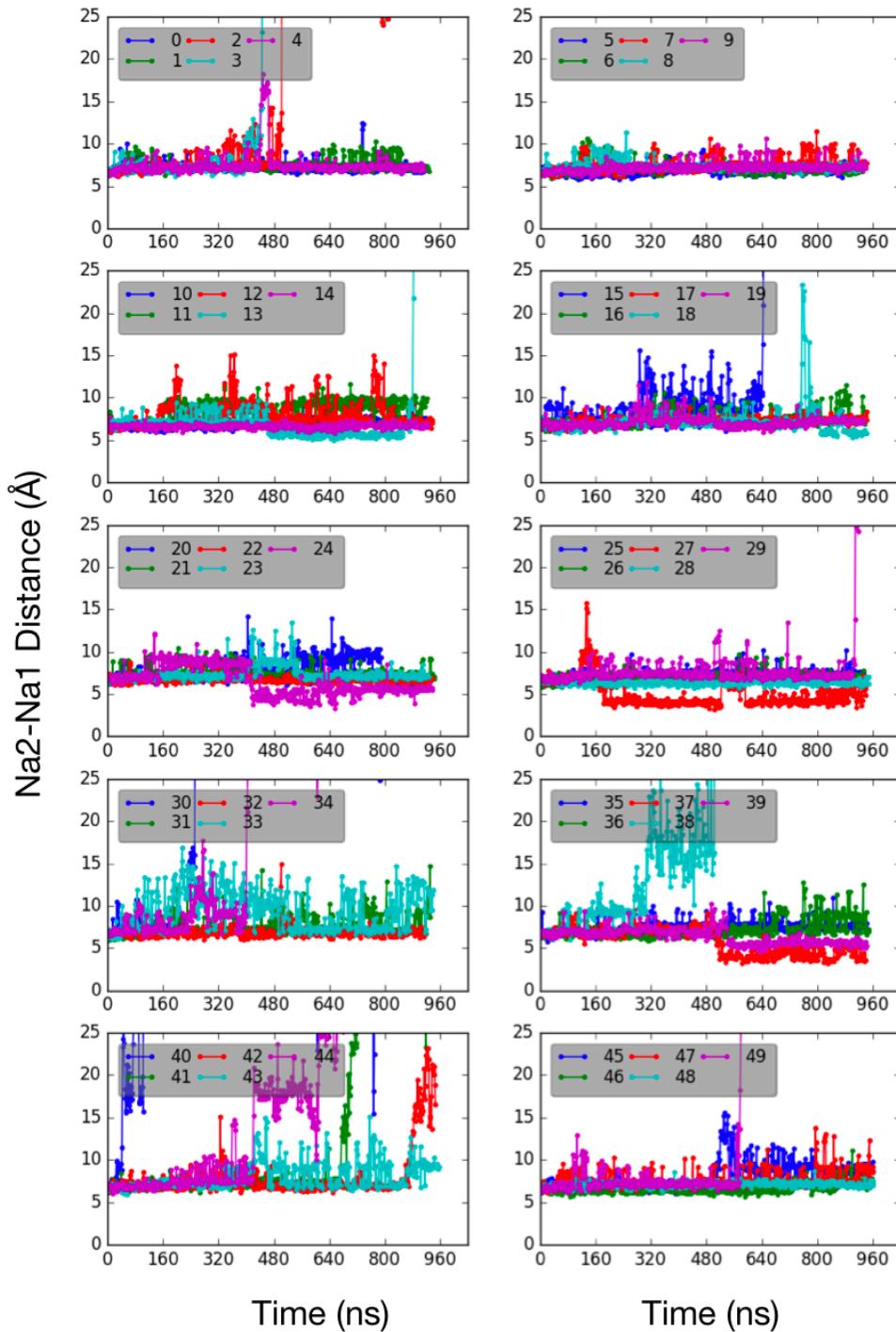


Figure S2. Time evolution of the Na1-Na2 distance in all 50 MD trajectories. Na⁺/Na₂ release into intracellular solution (distance beyond 25 Å) was observed in 12 trajectories.

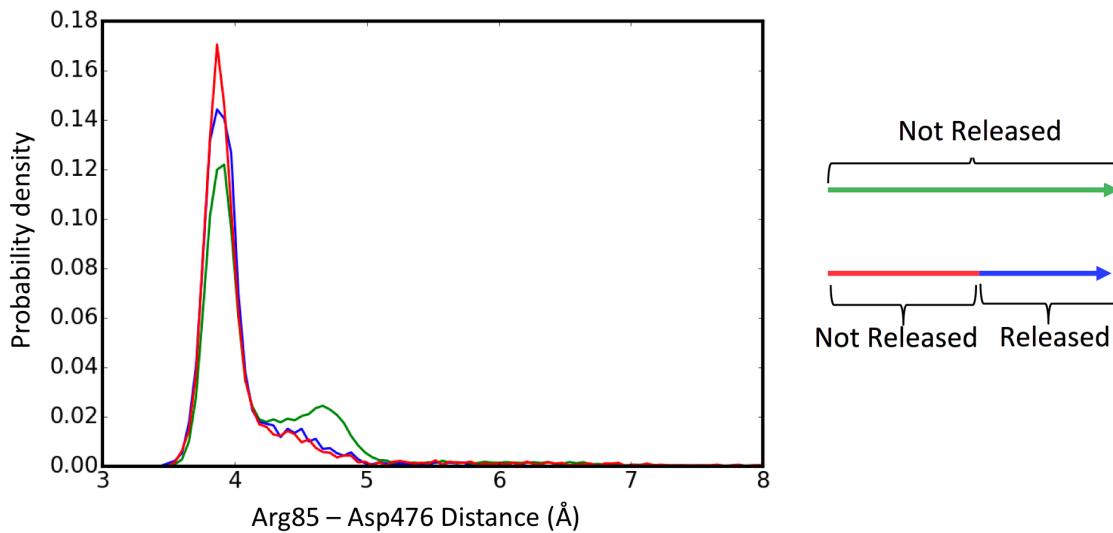


Figure S3. Extracellular gate dynamics represented by the Arg85 and Asp476 distance measured between terminal carbon atoms in both Arg and Asp. In trajectories that Na^+/Na_2 releases to intracellular (red and blue curves) the extracellular gate is more closed than trajectories that Na^+/Na_2 doesn't release to intracellular (green curve). For a detailed description of classifying trajectories in red, blue, and green curves see the caption of Figure 2 in the main text.

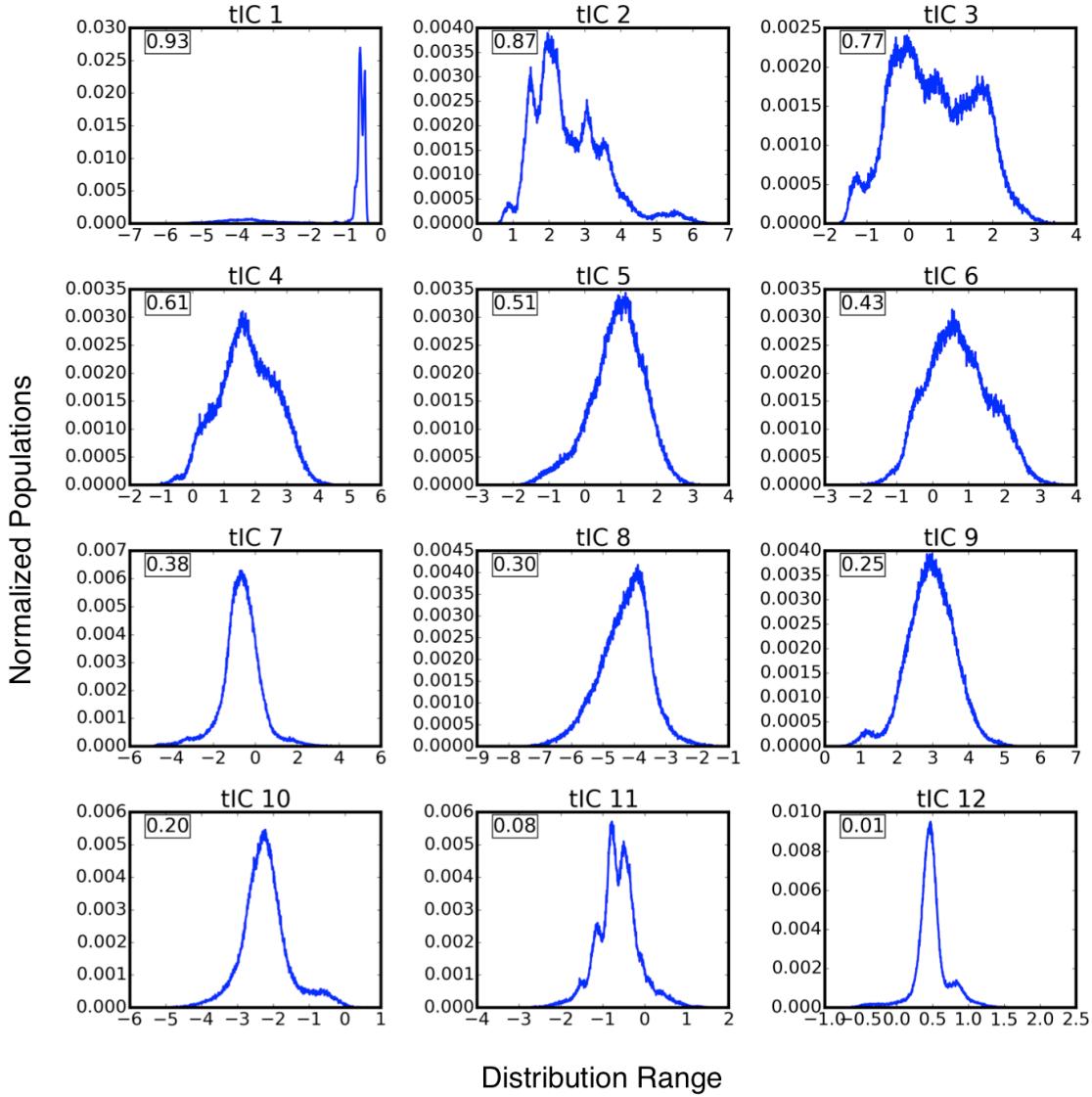


Figure S4. Distributions of simulation data projected on all 12 tICA eigenvectors. Only the first three eigenvectors produce a clear non-Gaussian distribution, and thus can be used for conformational space discretization. The insets show the corresponding eigenvalue for each eigenvector.

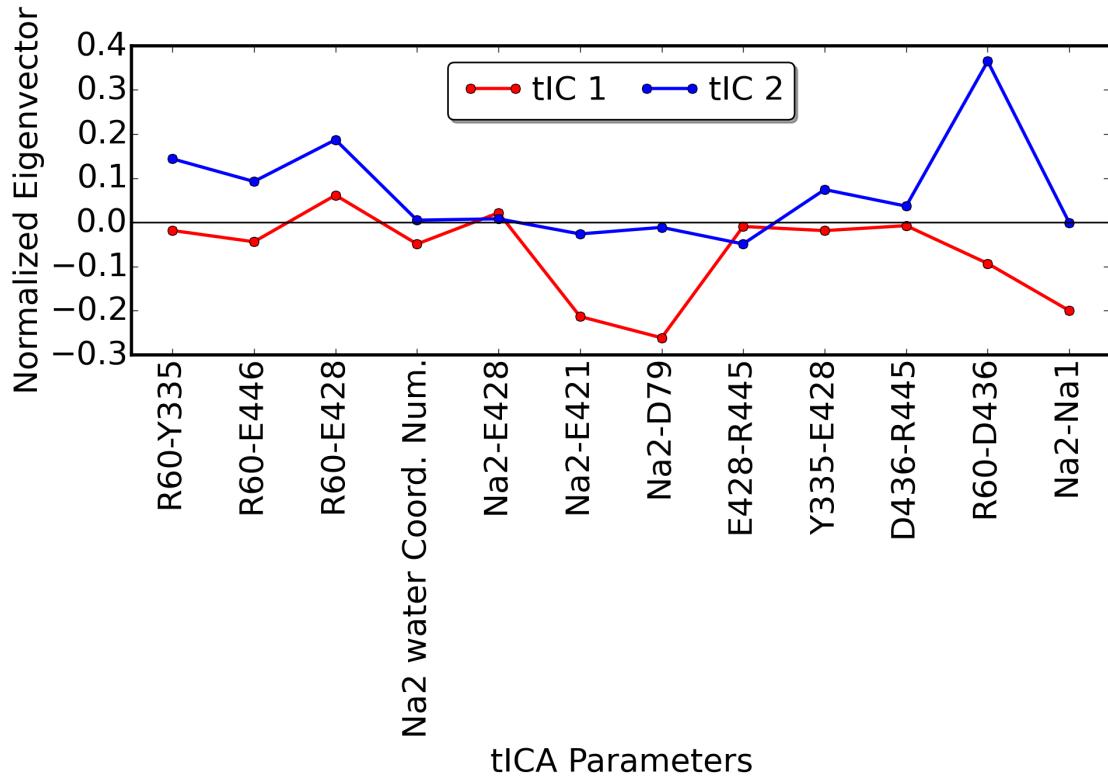


Figure S5. Contribution of tICA parameters to the first two eigenvectors. Parameters reflecting Na^+ motion contribute more to tIC 1; parameters reflecting water penetration have higher contributions in tIC 2.

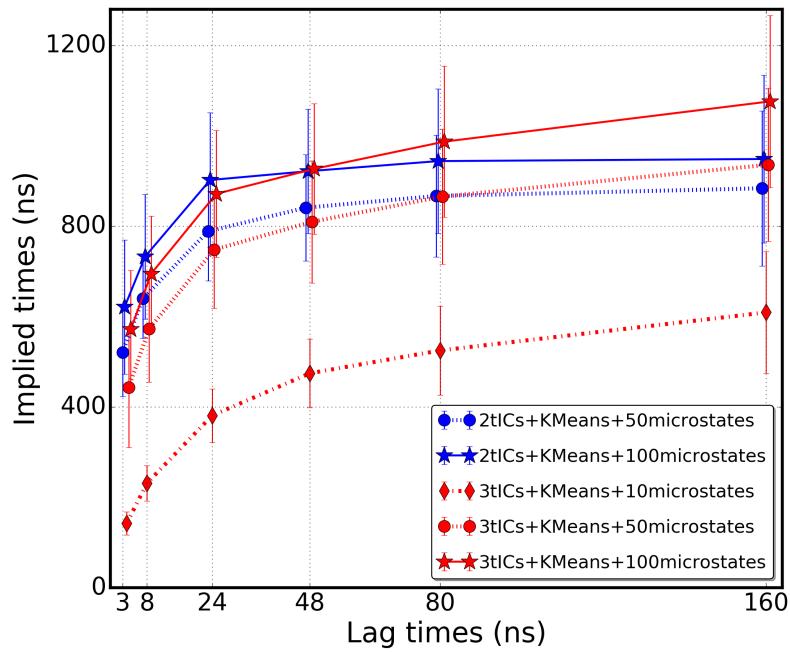
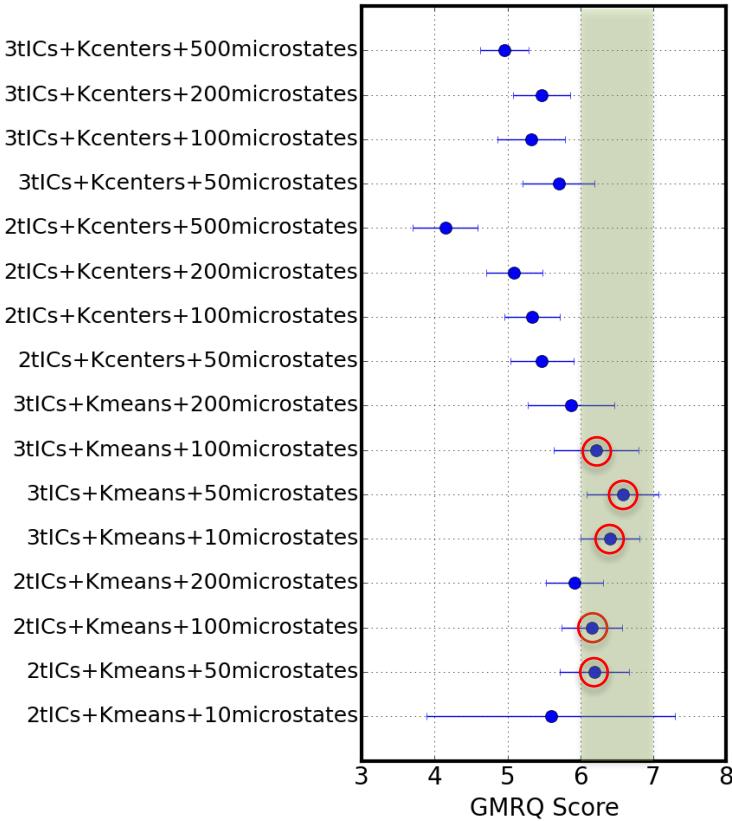


Figure S6. Top panel: GMRQ score for MSM parameters. The highlighted shaded area indicates the (higher) scores of models that are statistically superior to others. For each set of parameters (identified by a name describing the number of tICA

eigenvectors + the clustering algorithm + the number of microstates), the mean and error bar are the result of 200 GMRQ scores calculated by randomly dividing the data in two sets: a training set and a test set (see Methods). The first 10 MSM eigenvectors obtained with a lag time of 48 ns are used (as described in Methods) for all models. The k -means clustering algorithm is consistently performing better than the k -centers algorithm in this task. The models with highlighted red circles were subjected to further analysis based on implied time-scales plots. *Lower panel:* The first implied timescale calculated for each of the models highlighted with red circles in the top panel. The MSMs with 2tICs (blue curves) act Markovian shortly after about 24 ns lag time. On the other hand, MSMs with 3tIC s (red curves) don't satisfy Markovian criterion even after 80 ns. Hence, the MSMs with 2tICs are considered superior to MSMs with 3tICs. Since MSMs always underestimate relaxation timescales (due to discretization of the conformational landscape) the model with the slower implied time (2tICs+KMeans+100microstae) is preferable to the other one (2tICs+KMeans+50microstae). This model was chosen for all subsequent analyses presented in the main text. Error bars are calculated using a bootstrap method with randomly selecting 50 trajectories and calculating implied timescales and repeating the process 1000 times for each model. As the lag time increases, the error bars also increase due to the increase in finite sampling errors.

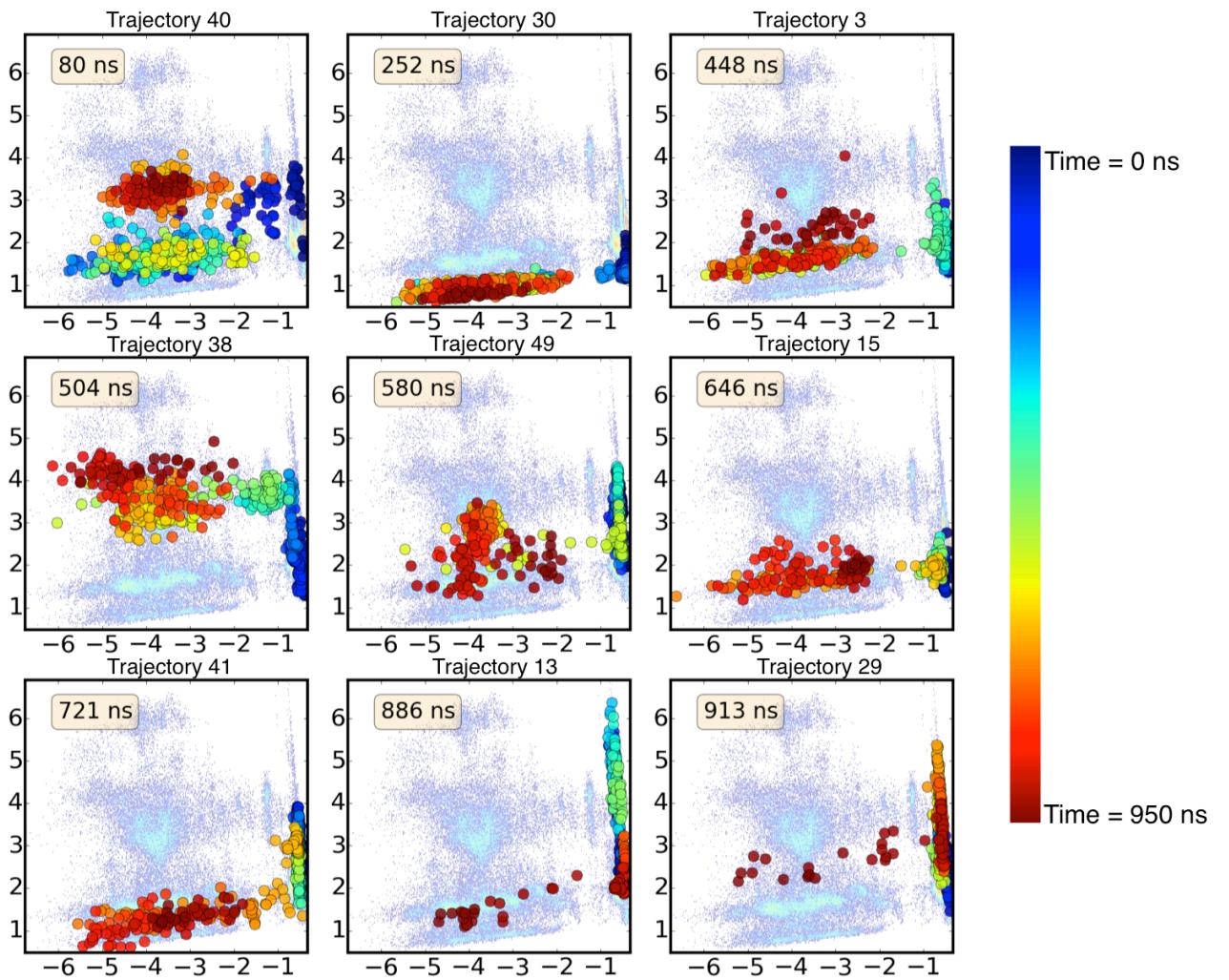


Figure S7. Time evolution of the trajectories where Na^+/Na_2 is released at different times (trajectory lengths). For each release time to the intracellular environment, (shown as insert), the colored dots indicate the time evolution (from blue to red) for the trajectories that released at that time, mapped on the first two tICA coordinates (cf. Fig 8 in main text).

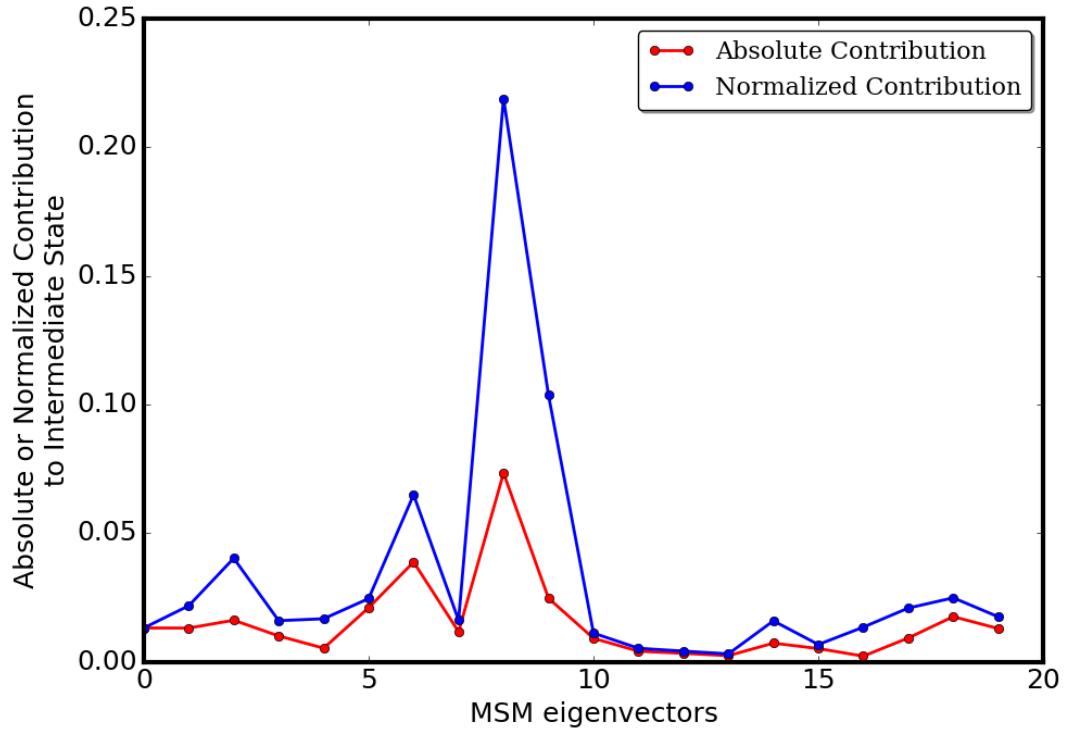


Figure S8. Identification of the MSM relaxation mode for the intermediate state. The total contribution of each eigenvector to the intermediate state is shown as both absolute ($\sum_i \text{abs}(v_i)$) and normalized ($\sum_i \frac{\text{abs}(v_i)}{\sum_{\text{all}} v}$) values; v_i are the eigenvectors corresponding to intermediate state microstates.

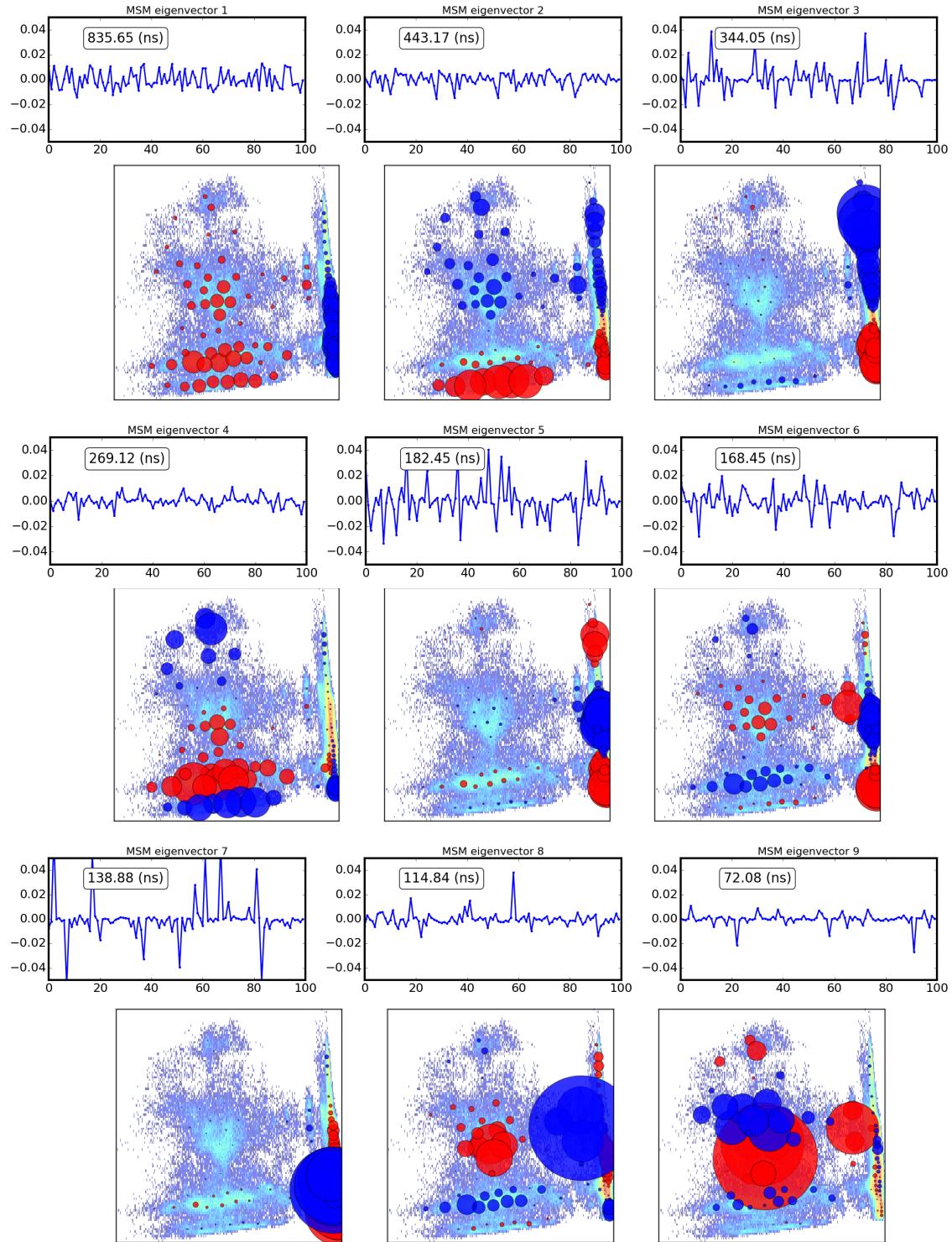


Figure S9. Markov State Model eigenvectors and corresponding relaxation modes on tICA landscapes (cf. Figure 6 in main text). Colored circles identify the center of the microstate into which the trajectory snapshots are clustered. Red and blue colors represent negative and positive eigenvectors, respectively, and the size of the

circles is proportional to eigenvectors value. The relaxation times are also shown. The MSM 8th eigenvector contributes most to the intermediate state.

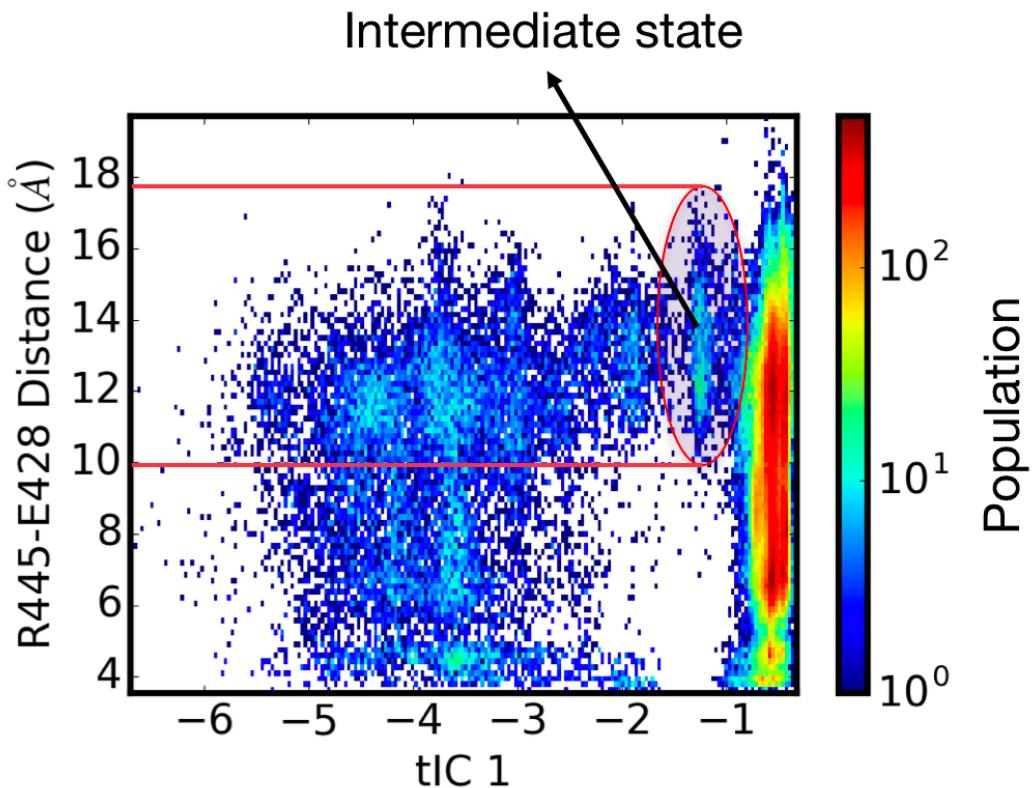


Figure S10. The population weighted 2D histogram of R445-E428 distance against tIC 1. The oval highlights the intermediate state and lines show R445-E428 distance range in \AA .

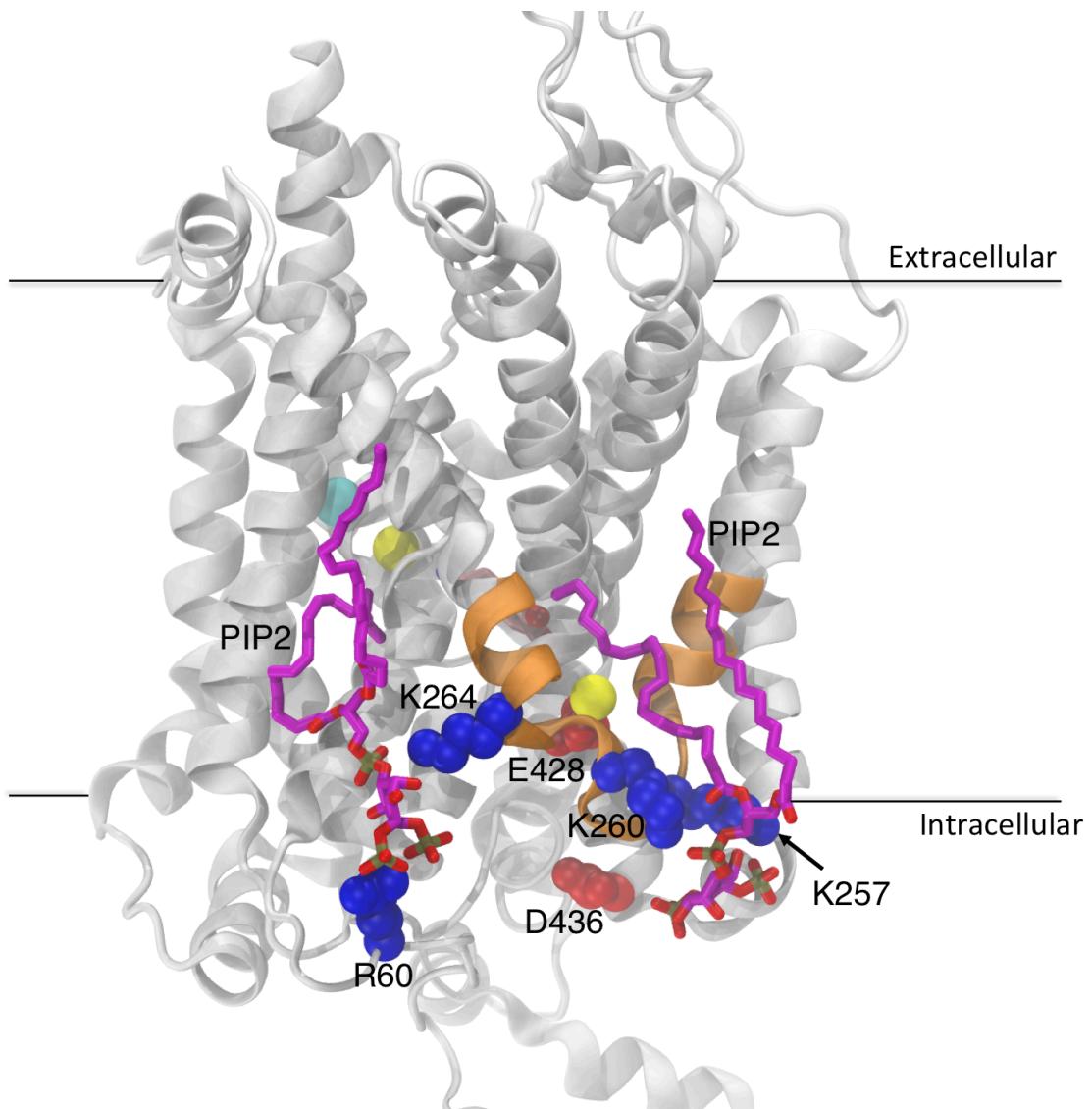


Figure S11. PIP₂ mediated N-terminus interactions with IL2. R60 from the N-terminus is interacting with K264 from IL2 (orange cartoon) through a PIP₂ lipid (in magenta). Other Lys residues from IL2 (K257 and K260) are interacting with PIP₂ as well. These interactions result in partial unfolding of IL2. Na⁺/Na₂ is interacting with E428. Color code is the same as in Figure 1A.

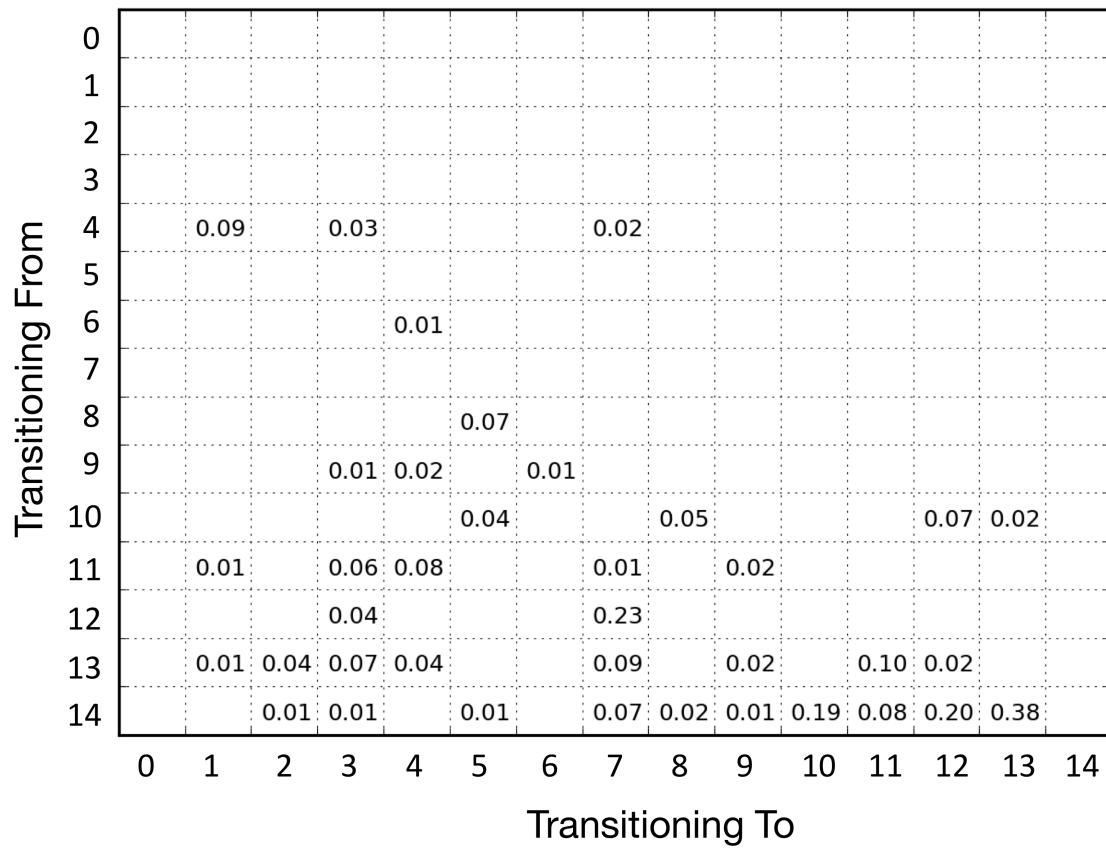


Figure S12. The transition flux matrix presents the flux values for transitions from macrostates numbered on the vertical axis to those identified on the horizontal axis. Only transition values of 0.01 and higher are shown in the matrix. Error estimates for these transitions are given in Table S3.

Table S1. Correlation values for water coordination number of Na+/Na2 and distance between Na+/Na2 and sodium at Na1 site. The correlation numbers are calculated by considering all frames up to the time when Na+/Na2 is released to the intracellular environment.

Trajectory number	Absolute Na+/Na2 release time (ns)	Correlation between Na2–Na1 distance and water coordination number of Na+/Na2
2	504	0.76
3	448	0.70
13	886	0.56
15	646	0.38
29	913	0.41
30	252	0.74
34	406	0.60
38	504	0.75
40	80	0.78
41	721	0.50
44	633	0.62
49	580	0.44
Average correlation for the 12 trajectories: 0.60		

Table S2. Confidence levels (at 68% and 95%) for water distribution in the extracellular cavity, obtained by assuming normal distributions.

Trajectories		68 percent confidence interval	95 percent confidence interval
Na+/Na2 is Not Released		25.7 – 38.7	19.4 – 45.0
Na+/Na2 is Released	before the release	22.9 – 35.2	17.0 – 41.2
	after the release	23.5 – 35.0	17.9 – 40.6

Table S3. Absolute values and standard deviations of fluxes calculated for all transitions (only fluxes >0.01 are listed). Standard deviations are calculated using the bootstrap method with 10 subsamples selected randomly from 40 individual trajectories out of the total 50 trajectories.

Macrostate Transition	Flux
4 --> 1	0.094 ± 0.067
11 --> 1	0.013 ± 0.006
13 --> 1	0.010 ± 0.008
13 --> 2	0.043 ± 0.040
14 --> 2	0.014 ± 0.012
4 --> 3	0.035 ± 0.023
9 --> 3	0.013 ± 0.010
11 --> 3	0.056 ± 0.026
12 --> 3	0.036 ± 0.014
13 --> 3	0.074 ± 0.034
14 --> 3	0.011 ± 0.008
6 --> 4	0.014 ± 0.006
9 --> 4	0.016 ± 0.008
11 --> 4	0.079 ± 0.068
13 --> 4	0.041 ± 0.018
8 --> 5	0.066 ± 0.072
10 --> 5	0.037 ± 0.047
14 --> 5	0.014 ± 0.012
9 --> 6	0.011 ± 0.005
4 --> 7	0.019 ± 0.011
11 --> 7	0.013 ± 0.005
12 --> 7	0.232 ± 0.094
13 --> 7	0.094 ± 0.023
14 --> 7	0.074 ± 0.043
10 --> 8	0.051 ± 0.053
14 --> 8	0.025 ± 0.012
11 --> 9	0.017 ± 0.011
13 --> 9	0.016 ± 0.008
14 --> 9	0.012 ± 0.007
14 --> 10	0.189 ± 0.069
13 --> 11	0.097 ± 0.055
14 --> 11	0.082 ± 0.018
10 --> 12	0.068 ± 0.039
13 --> 12	0.023 ± 0.022
14 --> 12	0.197 ± 0.034
10 --> 13	0.022 ± 0.014
14 --> 13	0.377 ± 0.046

Table S4. MSM-predicted sodium release pathways and fluxes.

Pathway	Normalized Flux	Accumulated Flux
[14 12 7]	0.20	0.20
[14 13 7]	0.10	0.30
[14 13 11 4 1]	0.08	0.38
[14 7]	0.07	0.45
[14 13 3]	0.07	0.52
[14 11 3]	0.06	0.58
[14 10 8 5]	0.05	0.63