Journal of Structural Biology 213 (2021) 107800

Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbi

Graphical Structural Biology Reviews

Markov state modeling of membrane transport proteins

Matthew C. Chan [a], Diwakar Shukla [a,b,c,d,e,f,g,h,*]

[a] Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, United States
[b] Center for Biophysics and Quantitative Biology, University of Illinois at Urbana-Champaign, Urbana, IL, United States
[c] Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL, United States
[d] Center for Digital Agriculture, University of Illinois at Urbana-Champaign, Urbana, IL, United States
[e] NIH Center for Macromolecular Modeling and Bioinformatics, University of Illinois at Urbana-Champaign, Urbana, IL, United States
[f] Cancer Center at Illinois, University of Illinois at Urbana-Champaign, Urbana, IL, United States
[g] National Center for Supercomputing Applications, University of Illinois, Urbana, IL, United States
[h] Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL, United States

ARTICLE INFO

Keywords:
Membrane transporter proteins
Adaptive sampling
Markov state model
Molecular dynamics simulations

ABSTRACT

The flux of ions and molecules in and out of the cell is vital for maintaining the basis of various biological processes. The permeation of substrates across the cellular membrane is mediated through the function of specialized integral membrane proteins commonly known as membrane transporters. These proteins undergo a series of structural rearrangements that allow a primary substrate binding site to be accessed from either side of the membrane at a given time. Structural insights provided by experimentally resolved structures of membrane transporters have aided in the biophysical characterization of these important molecular drug targets. However, characterizing the transitions between conformational states remains challenging to achieve both experimentally and computationally. Though molecular dynamics simulations are a powerful approach to provide atomistic resolution of protein dynamics, a recurring challenge is its ability to efficiently obtain relevant timescales of large conformational transitions as exhibited in transporters. One approach to overcome this difficulty is to adaptively guide the simulation to favor exploration of the conformational landscape, otherwise known as adaptive sampling. Furthermore, such sampling is greatly benefited by the statistical analysis of Markov state models. Historically, the use of Markov state models has been effective in quantifying slow dynamics or long timescale behaviors such as protein folding. Here, we review recent implementations of adaptive sampling and Markov state models to not only address current limitations of molecular dynamics simulations, but to also highlight how Markov state modeling can be applied to investigate the structure–function mechanisms of large, complex membrane transporters.

1. Introduction

Membrane transport proteins, or transporters, are integral membrane proteins that regulate the movement of ions and small molecules across the cellular membrane. Despite adopting diverse structural topologies, transporters operate via a universal paradigm known as the alternating access mechanism (Jardetzky, 1966). In this model, the transport of substrates across the membrane involves a series of structural rearrangements that enable access of the primary binding site from either side of the membrane at a given time. As such, the transporter adopts characteristic conformational states: an outward-facing (OF) state in which the pathway to the binding site is opened from the extracellular space; an occluded (OC) intermediate state in which the binding site is closed from both sides of the membrane; and an inward-facing (IF) state in which the binding site is accessible from the intracellular space (Fig. 1). Biophysical characterization of various transporters further supports a mechanism of alternative access among this diverse class of proteins. Structural biology techniques, such as X-ray crystallography and cryogenic electron microscopy, have historically paved the way in establishing the structural basis of transporter function. Despite these advancements, an experimentally obtained structure provides only a static conformation of the transporter, a glimpse into its intrinsic, dynamic nature. Therefore, a single structure cannot provide the mechanistic basis of conformational transitions and substrate transport, and as such, the understanding of transporter dynamics is limited.

* Corresponding author at: Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL, United States.
E-mail address: diwakar@illinois.edu (D. Shukla).

https://doi.org/10.1016/j.jsb.2021.107800
Received 9 August 2021; Received in revised form 20 September 2021; Accepted 20 September 2021
Available online 29 September 2021

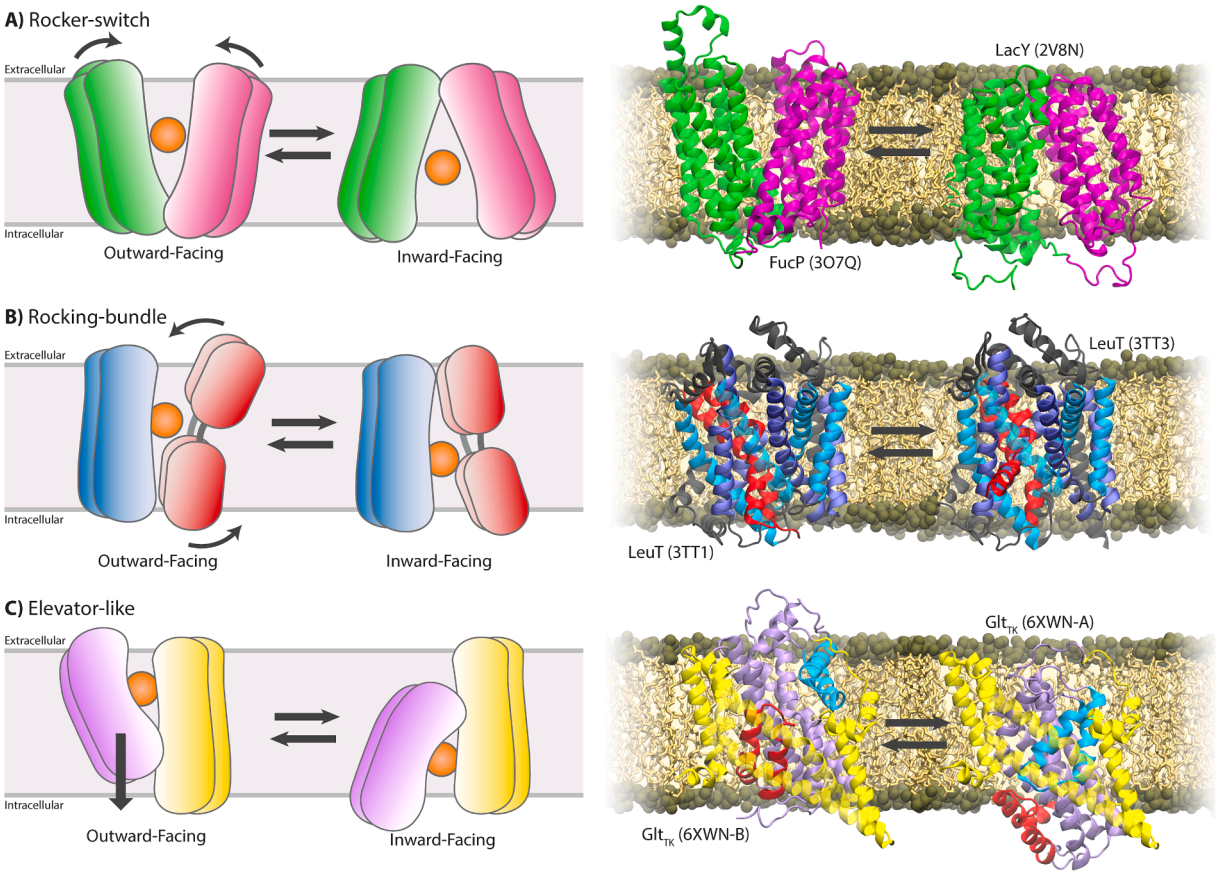

Fig. 1. Alternating access mechanisms of membrane transporters. Membrane transporters undergo a series of conformational transitions that allow the primary substrate binding site to be accessed from either the extracellular or intracellular space. (A) Conformational transitions in the rocker-switch scheme are enabled by two domains (colored green and pink) that cooperatively open and close the transporter. The rocker-switch mechanism is identified in the major-facilitator superfamily (MFS) of transporters as shown in sugar transporters FucP (PDB 3O7Q) and LacY (PDB 2V8N). The two rocker domains are shown as a cartoon representation, colored in green and pink. The lipid bilayer is shown in yellow sticks, with the phospholipid head shown in spheres. (B) In the rocking-bundle mechanism, access to the substrate binding site is mediate by gating helices (red) that pivot towards and away the stationary scaffold domain (blue). The rocking-bundle mechanism is commonly observed in transporters that adopt the LeuT fold (PDB 3TT1,3TT3). The gating helices are colored in red cartoons, the two-fold symmetric scaffold domain is colored in cyan and dark blue. (C) Transporters that adopt an elevator-like mechanism involve a rigid transport domain (purple) that translates across a stationary scaffold domain (yellow), thereby transporting the substrate across the membrane. A monomeric unit of the glutamate transporter homolog Glt_{Tk} (PDB 6XWN) is shown in cartoon representation. The transport domain is colored in purple, with characteristic hairpins that surround the substrate binding site colored in red and cyan. The scaffold domain is colored in yellow.

2. Investigating protein dynamics with molecular simulations

The use of molecular dynamics (MD) simulations as a biophysical method to probe protein dynamics has been effective in addressing this gap of knowledge. MD simulations model the motion of atoms by numerically solving Newton's equations of motion. The interactions that act upon atoms (bonds, angles, torsions, electrostatics, van der Waals, etc.) are represented by force field models that are parameterized based on experimental and/or quantum mechanical calculations (Tian et al., 2019). Each iteration of a MD timestep, δt, first involves calculating all the forces acting upon all atoms at time t. Newton's equations of motion are then integrated to obtain positions at $t + \delta t$. This updates the simulation with the new positions. The simulations are then extended by progressing the time to $t + \delta t$ until sufficient sampling has been observed for a desired process. To further investigate systems of larger size, coarse-grained simulations, which represents groups of atoms as "beads", can be employed to sample longer timescales. In the past decades, improvements in algorithms and computational hardware, most notably graphical processing units (GPUs), have greatly increased the performance and accuracy of MD simulations (Phillips et al., 2020), further solidifying it as an attractive method to characterize protein structure and function at atomistic resolution.

The timescales associated with large structural rearrangements and substrate transport may occur up to hundreds of nanoseconds (10^{-9} seconds) to microseconds (10^{-6} seconds) and beyond. Thus, obtaining sufficient sampling of such long timescale events remains a reoccurring challenge among MD simulations. For reference, a typical integration timestep of all-atom MD simulations is of the order of femtoseconds (10^{-12} seconds). In its purest sense, MD simulations model the equilibrium behavior of the simulated system, and thus the sampling will favor processes with low free energy barriers. When an energetic barrier is too high, the system will remain trapped in a local energy minimum and may not yield insightful observables. Such energetic barriers decrease the probability of observing certain conformational states and their related transition paths which occur either *in vitro* or *in vivo* but cannot be observed directly in experiments because of resolution differences between simulation and experiments. Common approaches to enhance the conformational sampling is to conduct non-equilibrium simulations in which an external force or energy bias is introduced (Yang et al., 2019). Examples include using accelerated MD and metadynamics to enable the system to overcome higher free energy barriers, string method path optimization to identify an transition path between states, and steered MD, targeted MD, and umbrella sampling to restrict the simulation along specified reaction coordinates to directly sample

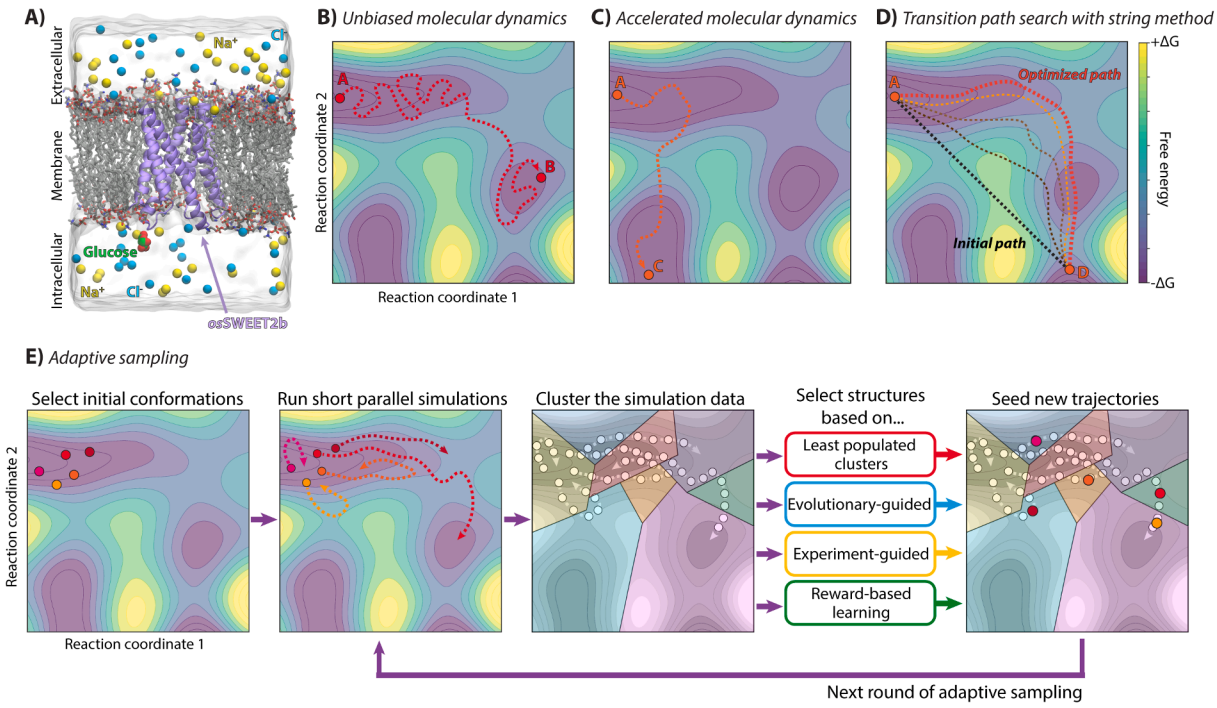

Fig. 2. Molecular dynamics simulations. (A) A fully atomistic molecular dynamics system of a sugar transporter *os*SWEET2b (represented in purple cartoons) embedded in a phospholipid bilayer membrane (gray sticks). The system is solvated with water molecules and sodium and chloride ions, depicted as yellow and cyan spheres, respectively. The substrate for *os*SWEET2b, glucose, is also added to the system and shown as green and red spheres. (B-E) Examples of molecular dynamics techniques to explore a simple conformational landscape. (B) In unbiased molecular dynamics, no external energy or force is supplemented to the system. As such, the simulation initiated from state A is restricted to exploring pathways of lower free energy barriers. Furthermore, a majority of the simulation is spent trapped in a energetic minimum before reaching to state B. (C) Accelerated molecular dynamics is an enhanced sampling method that introduces a potential energy bias when the system falls under an designated threshold. As a result, the free energy barrier of transitions is decreased, increasing the probability of observing the pathway from state A to state C. (D) If a particular pathway between two states (A and D) is desired, the transition path must first be identified before sampling. Shown here is an example of four iterations of utilizing the string method optimize an initial pathway, shown as a black dashed line, and obtain a minimal free energy path between the two states (orange dashed line). (E) The adaptive sampling methodology involves the simulations to be iteratively guided to explore the landscape. In this example, five initial starting structures for simulations are selected. The simulations are conducted in parallel and often of relatively short length to increase efficiency. Next, the simulation trajectories are clustered based on a metric or reaction coordinate of interest. The starting structures for the subsequent round of simulations are selected. For the example shown, the simulation data were grouped into eight clusters, and five structures were chosen based on least populated clusters. Other selection criteria for obtaining starting structures included evolutionary-coupling guided (Shamsi et al., 2017), experimental-guided (Zhao and Shukla, 2018), or reinforcement or reward-based learning (Shamsi et al., 2018; Zimmerman and Bowman, 2015). The process is repeat until the sampling is deemed converged or sufficient. In all, adaptive sampling leverages the utility of distributive simulations to provides an unbiased approach in exploring multiple transition pathways and increasing the sampling of rare events.

relevant dynamics (Yang et al., 2019). However, with many biased techniques, the appropriate reaction coordinates in which to sample along may not be intuitive or involve several degrees of freedom. Thus sampling along inappropriate reaction coordinates may affect the system to produce unphysical conformations and dynamics.

Adaptive sampling is an alternative unbiased sampling technique in which the simulations are guided towards relevant conformations or observables (Fig. 2E) (Hruska et al., 2018). The adaptive sampling scheme is an iterative approach in which multiple, short, unbiased simulations are conducted. These simulations may be initiated from different starting structures or initial velocities to increase diversity among trajectories. The ensemble of simulations are then clustered into states based on the reaction coordinates of interest. Next, structures for the subsequent iteration of adaptive sampling are seeded based on various criteria. For example, least-counts based adaptive sampling selects structures from clusters with the lowest populations, thereby increasing the probability of exploring undiscovered regions of the conformational landscape. Alternatively, the adaptive sampling can be guided through evolutionary (Shamsi et al., 2017) or experimental constraints (Zhao and Shukla, 2018) to maximize the efficiency of sampling relevant conformations. Lastly, adaptive sampling may further optimize the conformational exploration by utilizing reward-based algorithms to systematically select poorly sampled states along a designated reaction coordinate (Shamsi et al., 2018; Zimmerman and Bowman, 2015). Once the structures are selected for the next iteration, the procedure is repeated until the sampling is deemed sufficient or converges. Compared to traditional "long-MD" simulation, adaptive sampling allows for robust combination of exploration and exploitation of the landscape to sufficiently sample rare transition states and identify alternate conformational pathways (Zimmerman et al., 2018).

3. Markov state model: A statistical approach to characterize protein dynamics

With the increased performance of algorithms and computational hardware over the past decades, the collection of MD data has become more accessible to researchers. However, large datasets, such as ones obtained from MD simulations, present inherent difficulties in analyzing and obtaining relevant insights. Markov state models (MSMs) provide a powerful approach to efficiently process and quantify large MD trajectory datasets. Having historically been applied to study protein folding (Voelz et al., 2010), MSMs may also be implemented to model long timescale behaviors such as catalytic or transport cycles, substrate binding pathways, protein–protein association, allosteric modulation, and intrinsically disordered protein dynamics. An MSM is a statistical framework that expresses the kinetics of the system as transition

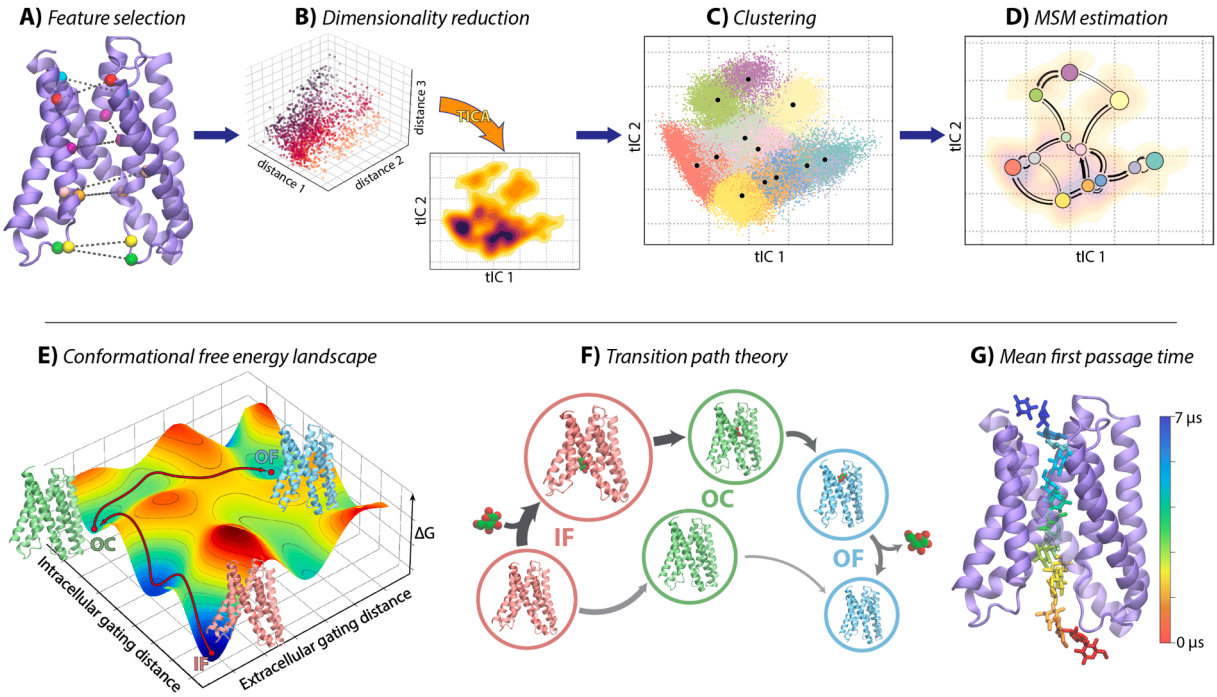

Fig. 3. Construction of the Markov state model (MSM). An MSM is a kinetic model that represents a system as a series of discretized states that are each related by a transition probability. (A-D) Given sufficient sampling of the system, an MSM is constructed through a sequence of statistical modeling. A simplified example of the MSM workflow is presented. (A) First, a selection of features that represent the dynamics of the system are chosen. Example features include contact distance pairs, as shown, or torsion angles of the protein backbone. (B) The high dimensional feature space is then transformed into a reduced dimensional space to increase efficiency and identify key featurization components. A popular decomposition method is time-lagged component analysis (tICA) which constructs linear combinations of the features to identify the reactions coordinates of the slow timescale processes. Shown here, a three dimensional feature space of distances is reduced by tICA and projected onto a two dimensional tICA space defined by the first two time-lagged components (tICs). (C) The simulation trajectories are then discretized in the tICA space using a clustering algorithm such as *k*-means. As such, structures of similar conformational are grouped in the same cluster. The frames of the simulations are also assigned based on their designated cluster. As an example shown here, the tICA space has been clustered in 12 states and are colored independently. (D) For the calculated Markovian lag time, the transitions between clusters are counted and converted to transition probabilities. The transition probabilities are calculated for all transitions to and from every state, including self-transitions. The transitions probabilities can be visualized as a graph, where the nodes represents each MSM state, the size of the nodes represents the self-transition probability, and the thickness of the directed edges are the transition probability from two states. (E-G) Further analysis of the MSM provides powerful thermodynamic and kinetic observables of the dynamics of the simulated system. (E) Trajectories, typically obtained from adaptive sampling, may be reweighted by the equilibrium probability to obtain a MSM-weighted conformational free energy landscape. (F) Kinetic analysis such as transition path theory examines the probability flux from an initial to an end state. An example of the transition fluxes for *os*SWEET2b when initialized from the inward-facing (IF) state and ending at the outward-facing (OF) state is shown. The relative flux is indicated by arrow thickness, signifying that conformational transitions are promoted by the binding of the substrate, glucose (green spheres). (G) Rates of substrate transport, as showed in sticks and colored based on simulated time, or conformations changes can be measured by calculating the mean first passage time of the MSM.

probabilities between states (Husic and Pande, 2018). MSMs utilize the Markovian property, in which the system becomes "memoryless" and the transitions to a state do not depend on its past transitions, only the current state itself. This method greatly complements adaptive sampling as multiple rounds of adaptive sampling can be unified into a representative model.

The construction of the MSM involves a series of statistical calculations (Fig. 3A) (Husic and Pande, 2018; Wehmeyer et al., 2019). The initial step is to represent the Cartesian coordinates of the MD trajectory as a collection of metrics that best describes the processes of the system. Such feature metrics may include contact distance pairs, dihedral angles, and root-mean-square deviation (RMSD). MSMs are highly dependent on the selected features, and for complex or elusive dynamics, distinguishing such relevant metrics may not always be intuitive. Machine learning algorithms offer an alternative approach to systematically identify relevant features of the system. For example, a genetic algorithm searches among all possible contact distances to obtain the most optimal features for MSM construction (Chen et al., 2018). Ward et al. developed self-supervised autoencoders to automatically identify features that distinguish multiple datasets (Ward et al., 2021). Another effective approach, named VAMPnets, decomposes the coordinate space with the use of variational encoders (Mardt et al., 2018). In many cases, a set of features consist of many degrees of freedom; therefore, reducing the dimensionality, not only improves computational efficiency, but also amplifies principle features that describe the overall dynamics of the system. A popular decomposition method is time-lagged independent component analysis (tICA) in which linear combinations of the features are constructed to maximize the autocorrelation between features thus yielding reaction coordinates that best represent slow dynamic processes (Pérez-Hernández et al., 2013).

The transformed dataset is subsequently clustered into kinetically relevant microstates using a clustering algorithm such as *k*-means. The number of microstates influences the resolution of the system's dynamics: increasing number of microstates discretizes the feature space into finer-grained clusters, while decreasing number of microstates trends to a coarse-grained resolution. Finally, the MSM is estimated in which the number of transitions between microstates at an interval of the Markovian lag time τ is counted from the trajectories. The value of τ is determined when the relaxation timescales of the system are approximately constant with respect to τ. The normalized counts to and from each microstate are collectively known as the probability transition matrix which is of size n x n, where n is the number of microstates. Typically, to enforce detailed balance among discretized states, a maximum likelihood estimate is used to calculate the counts and

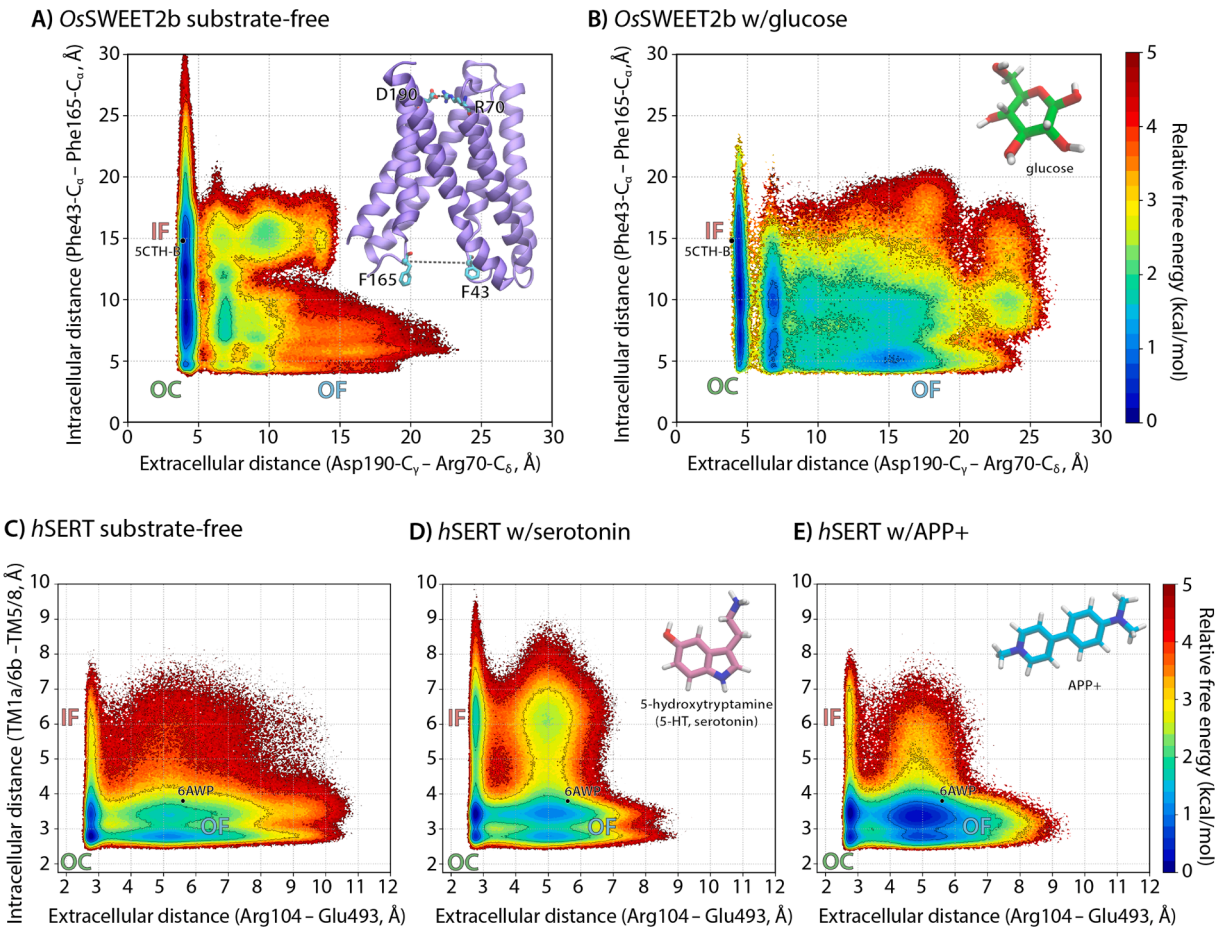

Fig. 4. Substrate-induced changes of the conformational free energy landscape. The Markov state model-weighted free energy landscapes of two transporters, sugar transporter *Os*SWEET2b and neurotransmitter reuptake transporter *h*SERT, are recreated from (Selvam et al., 2019; Chan et al., 2020; Young et al., 2021) and projected on the coordinates defined by the extracellular and intracellular gating distances. Color represents the relative free energy as estimated by the MSM-weighted simulation data. (A) Conformational transitions of the sugar transporter from *Oryza sativa*, *Os*SWEET2b in the absence of the substrate. The inward-facing (IF) and occluded (OC) states of *Os*SWEET2b are relative stable with relative free energy values of <1 kcal/mol. However, transitions to the outward-facing state are restricted with high free energy barriers of >4 kcal/mol. (B) As *Os*SWEET2b transports the substrate glucose, the transition free energy barriers decrease, most notably the transitions between occluded to outward-facing conformations. (C) Similarly, in simulations of the human serotonin transporter (*h*SERT), transitions to the inward-facing state are of relatively higher free energy as compared to outward-facing and occluded transitions. (D) In the presence of the substrate serotonin, the inward-facing state is further stabilized to promote conformational transitions and substrate transport. (E) When *h*SERT transports an unphysiological substrate such as APP+, a fluorescent neurotransmitter analog, the outward-facing and occluded states are further stabilized. Furthermore, the increased rigidity of APP+ restricts transitions to the inward-facing state, ultimately decreasing the rate of transport.

transition probabilities (Bowman et al., 2009). After validating the MSM to ensure proper Markovian behavior, the model is a host for several thermodynamic and kinetic analyses, including free energy surface projection along defined reaction coordinates, flux analysis with transition path theory, and kinetic measurements by calculating mean first passage times (Fig. 3B).

4. Applications of Markov state models on membrane transporters

The synergistic combination of efficient conformational exploration by adaptive sampling followed with robust statistical analysis from MSMs presents an effective method to computationally investigate membrane transporters. Recent applications of Markov state modeling on transporters have characterized the conformational heterogeneity and how substrates induced modulations in the conformational free energy landscapes (Selvam et al., 2018; Selvam et al., 2019; Chan et al., 2020; Young et al., 2021). An example is illustrated with the sugar transporter from *Oryza sativa*, *Os*SWEET2b (Fig. 4A,B) (Selvam et al., 2019). Simulations show, in the absence of the substrate, *Os*SWEET2b freely transitions from inward-facing to occluded conformations. However, the formation of the outward-facing state is greatly destabilized with free energy barriers of ~4 kcal/mol (Fig. 4A). In the presence of the substrate, glucose, the outward-facing state is stabilized to ~1–2 kcal/mol, thus promoting the complete transitions from inward-facing to outward-facing (Fig. 4B). Similar observations of how the substrate promotes the conformational transitions of transporters are seen in simulations of the human serotonin transporter, *h*SERT (Fig. 4C,D) (Chan et al., 2020). While transitions from outward-facing to occluded remain relatively low in free energy with and without the substrate serotonin, the most significant influence of the substrate is observed in the stabilization of the inward-facing state to promote substrate transport (Fig. 4D). When transporting the fluorescent neurotransmitter analog, APP+, the outward-facing and occluded states are further stabilized; however *h*SERT adopts a conformational landscape that closely resembles to an inhibitor-bound transporter with restricted transitions to the inward-facing state (Fig. 4E) (Young et al., 2021). Taken together, the Markov state modeling of these transporters show how the substrate facilitates the necessary structural rearrangements by decreasing the free energy barriers of transitions. Furthermore, the conformational free

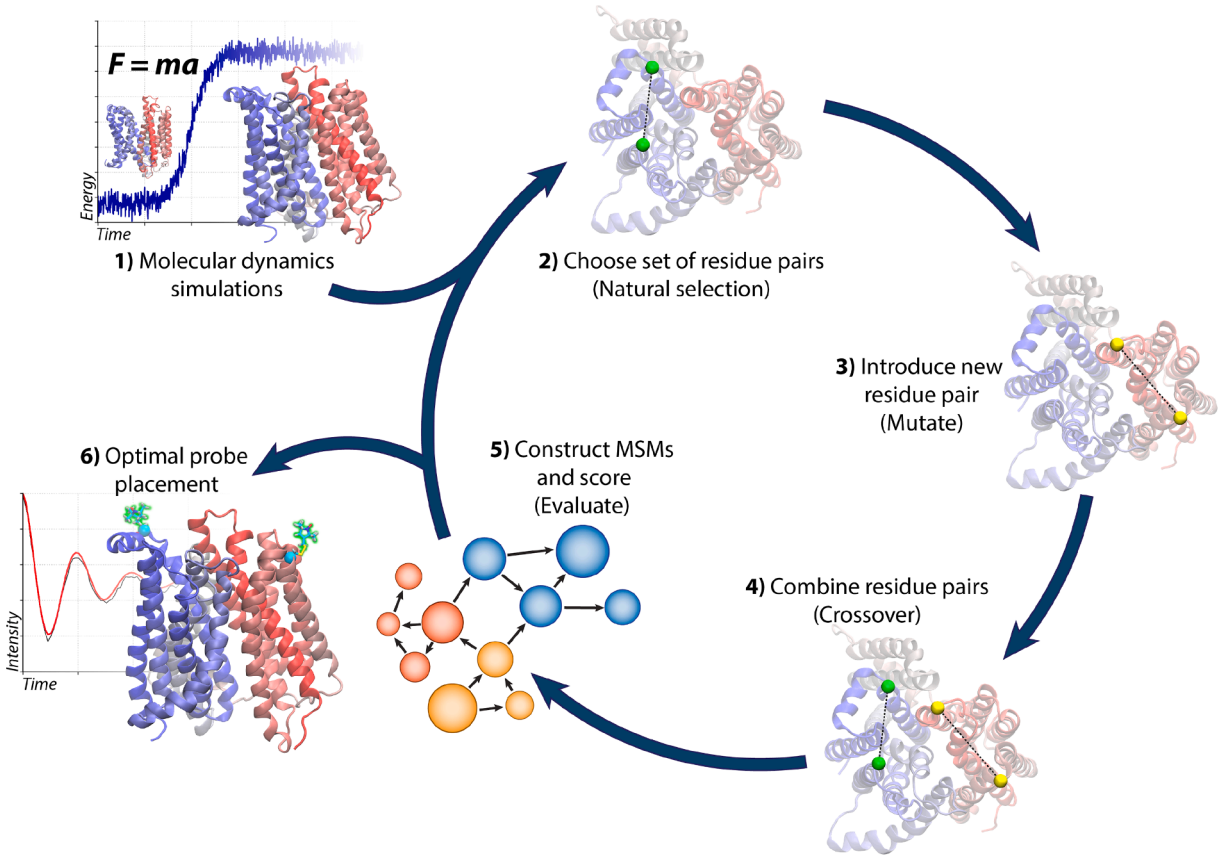

Fig. 5. Markov state model (MSM)-guided experimental design for optimal placement of spin labelling probes. An advantage of the MSM is its ability to characterize long timescale behavior, such as conformational transitions observed in electron paramagnetic resonance spectroscopy. Shown here is a genetic algorithm implementation to utilize MSM to objectively identify key probe positions for distance measurement (Mittal and Shukla, 2017). (1) MD simulations are first conducted to explore the conformational heterogeneity of the transporter. (2) A set of residue pairs are randomly chosen from all possible pair combinations. This step is analogous to natural selection in population genetics. Only residue pairs that can be experimentally measured will be considered. For example, residues buried in the transporter core or transmembrane-spanning regions will not selected. (3) In some sets, a residue pair will be "mutated" in which that pair will be replaced with a new residue pair. (4) "Crossover" allows sets to combine residue pairs to further increase genetic diversity. (5) The sets of residue pairs are used to construct a MSM. Each MSM is quantitatively scored with the generalized matrix Rayleigh quotient (GMRQ), which measures the MSM ability to capture long-timescales processes (Husic and Pande, 2018). For the next iteration, the population of distances that obtained the highest GMRQ score "survives" as the next generation and progresses through the subsequent iteration of the genetic algorithm. (6) Upon convergence of the GMRQ score, the genetic algorithm stops and the distance pairs that obtained the highest GMRQ score are suggested for optimal probe placement.

energy landscapes highlight how evolution has optimized its selectivity to efficiently transport its endogenous substrate.

The kinetic and thermodynamic insights obtained from MSM provides a powerful means to characterize slow conformational transitions, and in doing so presents opportunities to complement biochemical and biophysical experiments. Electron paramagnetic resonance techniques utilize site-directed spin probes to obtain distance measurements upon structural rearrangement of the protein. As such, the placement of the spin probes directly influences the observed distance measurements; however, often times, the placement the spin probes is not intuitive. Using a genetic algorithm, MSMs can be leveraged to systemically obtain probe placements that best captures characteristic conformational changes of proteins (Fig. 5) (Mittal and Shukla, 2017). Classically implemented to model population genetics, the genetic algorithm incorporates the existing principles of mutation, crossover, and selection to achieve the set of features that best describe the dynamics of the system. First, a population of sets of randomly selected distance pairs is initialized. A subset of the population is then "mutated" in which certain residue pairs are randomly replaced by other residue pairs. In the "crossover" step, residue pairs among sets are swapped with one another to further increase genetic diversity. The sets of distances are used to construct a respective MSM and evaluated based on how successful the model identifies the slowest timescales of the system. To do this, the generalized matrix Rayleigh quotient (GMRQ) for each MSM is calculated (Husic and Pande, 2018). The GMRQ is a variational cross-validated quantity that is equal to the sum of eigenvalues of the transition probability matrix. Models that fail to identify the slow dynamics will overfit and result in a low GMRQ, whereas GMRQ values that approach the upper limit signify MSMs that capture long timescale processes. As such, MSMs with the top GMRQ scores are retained for the next iteration of the genetic algorithm and its associated distance pairs serve as the next generation of the population. The process is then repeated until convergence of the highest GMRQ score is obtained. In doing so, identifying optimal probe placements for various spectroscopy techniques not only aids in design of biophysical experiments to study transporter function, but may serve as additional validation of the MSM (Selvam et al., 2018).

Understanding the fundamentals of substrate transport is essential for characterizing all cellular processes. Structural determination techniques provide invaluable atomic details of the structural architecture of transporters, and more so, the recent emergence of deep learning algorithms has not only impressively accelerated accurate protein structure predictions (Jumper et al., 2021), but also expand the applicability of Markov state models (Konovalov et al., 2021). As computational efficiency continues to increase over time, the atomistic resolution of protein dynamics obtained from molecular dynamics simulations and

Markov state models provides unparalleled insights to complement biophysical experiments, furthering our understanding of membrane transporter structure and function.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

M.C.C. wishes to thank Jiming Chen, Austin T. Weigle, and Tinnie J. Louie for critical evaluation and discussion of this review. D.S. acknowledge support from the NSF MCB 18–45606.

References

Bowman, G.R., Beauchamp, K.A., Boxer, G., Pande, V.S., 2009. Progress and challenges in the automated construction of markov state models for full protein systems. J. Chem. Phys. 131 (12), 124101 https://doi.org/10.1063/1.3216567.

M. Chan, B. Selvam, H. Young, E. Procko, D. Shukla, The substrate import mechanism of the human serotonin transporter doi:10.26434/chemrxiv.9922301.v2. URL:https://doi.org/10.26434/chemrxiv.9922301.v2.

Chen, Q., Feng, J., Mittal, S., Shukla, D., 2018. Automatic feature selection in markov state models using genetic algorithm. J. Comput. Sci. Educ. 9 (2), 14–22. https://doi.org/10.22369/issn.2153-4136/9/2/2.

Hruska, E., Abella, J.R., Nüske, F., Kavraki, L.E., Clementi, C., 2018. Quantitative comparison of adaptive sampling methods for protein dynamics. J. Chem. Phys. 149 (24), 244119 https://doi.org/10.1063/1.5053582.

Husic, B.E., Pande, V.S., 2018. Markov state models: From an art to a science. J. Am. Chem. Soc. 140 (7), 2386–2396. https://doi.org/10.1021/jacs.7b12191.

Jardetzky, O., 1966. Simple allosteric model for membrane pumps. Nature 211 (5052), 969–970. https://doi.org/10.1038/211969a0.

J. Jumper, R. Evans, A. Pritzel, T. Green, M. Figurnov, O. Ronneberger, K. Tunyasuvunakool, R. Bates, A. Žídek, A. Potapenko, A. Bridgland, C. Meyer, S.A.A. Kohl, A.J. Ballard, A. Cowie, B. Romera-Paredes, S. Nikolov, R. Jain, J. Adler, T. Back, S. Petersen, D. Reiman, E. Clancy, M. Zielinski, M. Steinegger, M. Pacholska, T. Berghammer, S. Bodenstein, D. Silver, O. Vinyals, A.W. Senior, K. Kavukcuoglu, P. Kohli, D. Hassabis, Highly accurate protein structure prediction with AlphaFold, Nature doi:10.1038/s41586-021-03819-2. URL:https://doi.org/10.1038/s41586-021-03819-2.

K.A. Konovalov, I.C. Unarta, S. Cao, E.C. Goonetilleke, X. Huang, Markov state models to study the functional dynamics of proteins in the wake of machine learning, JACS Au doi:10.1021/jacsau.1c00254. URL:https://doi.org/10.1021/jacsau.1c00254.

A. Mardt, L. Pasquali, H. Wu, F. Noé, VAMPnets for deep learning of molecular kinetics, Nature Communications 9 (1). doi:10.1038/s41467-017-02388-1. URL:https://doi.org/10.1038/s41467-017-02388-1.

Mittal, S., Shukla, D., 2017. Predicting optimal DEER label positions to study protein conformational heterogeneity. J. Phys. Chem. B 121 (42), 9761–9770. https://doi.org/10.1021/acs.jpcb.7b04785.

Pérez-Hernández, G., Paul, F., Giorgino, T., Fabritiis, G.D., Noé, F., 2013. Identification of slow molecular order parameters for markov model construction. J. Chem. Phys. 139 (1), 015102 https://doi.org/10.1063/1.4811489.

Phillips, J.C., Hardy, D.J., Maia, J.D.C., Stone, J.E., Ribeiro, J.V., Bernardi, R.C., Buch, R., Fiorin, G., Hénin, J., Jiang, W., McGreevy, R., Melo, M.C.R., Radak, B.K., Skeel, R.D., Singharoy, A., Wang, Y., Roux, B., Aksimentiev, A., Luthey-Schulten, Z., Kalé, L.V., Schulten, K., Chipot, C., Tajkhorshid, E., 2020. Scalable molecular dynamics on CPU and GPU architectures with NAMD. J. Chem. Phys. 153 (4), 044130 https://doi.org/10.1063/5.0014475.

Selvam, B., Mittal, S., Shukla, D., 2018. Free energy landscape of the complete transport cycle in a key bacterial transporter. ACS Central Science 4 (9), 1146–1154. https://doi.org/10.1021/acscentsci.8b00330.

Selvam, B., Yu, Y.-C., Chen, L.-Q., Shukla, D., 2019. Molecular basis of the glucose transport mechanism in plants. ACS Central Science 5 (6), 1085–1096. https://doi.org/10.1021/acscentsci.9b00252.

Shamsi, Z., Moffett, A.S., Shukla, D., 2017. Enhanced unbiased sampling of protein dynamics using evolutionary coupling information. Sci. Rep. 7 (1) https://doi.org/10.1038/s41598-017-12874-7.

Shamsi, Z., Cheng, K.J., Shukla, D., 2018. Reinforcement learning based adaptive sampling: REAPing rewards by exploring protein conformational landscapes. J. Phys. Chem. B 122 (35), 8386–8395. https://doi.org/10.1021/acs.jpcb.8b06521.

Tian, C., Kasavajhala, K., Belfon, K.A.A., Raguette, L., Huang, H., Migues, A.N., Bickel, J., Wang, Y., Pincay, J., Wu, Q., Simmerling, C., 2019. ff19sb: Amino-acid-specific protein backbone parameters trained against quantum mechanics energy surfaces in solution. J. Chem. Theory Comput. 16 (1), 528–552. https://doi.org/10.1021/acs.jctc.9b00591.

Voelz, V.A., Bowman, G.R., Beauchamp, K., Pande, V.S., 2010. Molecular simulation of ab initio protein folding for a millisecond folder NTL9(1–39). J. Am. Chem. Soc. 132 (5), 1526–1528. https://doi.org/10.1021/ja9090353.

M.D. Ward, M.I. Zimmerman, A. Meller, M. Chung, S.J. Swamidass, G.R. Bowman, Deep learning the structural determinants of protein biochemical properties by comparing structural ensembles with DiffNets, Nature Communications 12 (1). doi:10.1038/s41467-021-23246-1. URL:https://doi.org/10.1038/s41467-021-23246-1.

Wehmeyer, C., Scherer, M.K., Hempel, T., Husic, B.E., Olsson, S., Noé, F., 2019. Introduction to markov state modeling with the PyEMMA software [article v1.0]. J. Comput. Mol. Sci. 1 (1) https://doi.org/10.33011/livecoms.1.1.5965.

Yang, Y.I., Shao, Q., Zhang, J., Yang, L., Gao, Y.Q., 2019. Enhanced sampling in molecular dynamics. J. Chem. Phys. 151 (7), 070902 https://doi.org/10.1063/1.5109531.

H.J. Young, M. Chan, B. Selvam, S.K. Szymanski, D. Shukla, E. Procko, Deep mutagenesis of a transporter for uptake of a non-native substrate identifies conformationally dynamic regions doi:10.1101/2021.04.19.440442. URL:https://doi.org/10.1101/2021.04.19.440442.

Zhao, C., Shukla, D., 2018. SAXS-guided enhanced unbiased sampling for structure determination of proteins and complexes. Sci. Rep. 8 (1) https://doi.org/10.1038/s41598-018-36090-z.

Zimmerman, M.I., Bowman, G.R., 2015. FAST conformational searches by balancing exploration/exploitation trade-offs. J. Chem. Theory Comput. 11 (12), 5747–5757. https://doi.org/10.1021/acs.jctc.5b00737.

Zimmerman, M.I., Porter, J.R., Sun, X., Silva, R.R., Bowman, G.R., 2018. Choice of adaptive sampling strategy impacts state discovery, transition probabilities, and the apparent mechanism of conformational changes. J. Chem. Theory Comput. 14 (11), 5459–5475. https://doi.org/10.1021/acs.jctc.8b00500.