12-2022

Comparative Analysis of Conformational Transition Pathways in Homologous Proteins

Dylan Sebastien Ogden
University of Arkansas, Fayetteville

Comparative Analysis of Conformational Transition Pathways in Homologous Proteins

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy in Cell and Molecular Biology

by

Dylan Ogden
University of Arkansas
Bachelor of Science in Chemistry, 2017

December 2022
University of Arkansas

This dissertation is approved for recommendation to the Graduate Council

Mahmoud Moradi, Ph.D.
Dissertation Director:

Colin Heyes, Ph.D.
Committee member

Suresh Thallapuranam, Ph.D.
Committee member

Yong Wang, Ph.D.
Committee member

Abstract

Molecular dynamics (MD) simulations are routinely used to study the dynamics of proteins. However, conventional MD limited to the sampling of local conformational changes as the functionally important conformational transitions of proteins often extend beyond the timescales of the simulations employed, for example, membrane transport proteins. We have determined the combination of multiple MD based techniques that allows for a rigorous characterization of energetics and kinetics of large-scale conformational changes in membrane proteins. The methodology is based on biased, nonequilibrium collective variable based simulations including nonequilibrium pulling, string method with swarms of trajectories, bias-exchange umbrella sampling, and rate estimation techniques. Severe acute respiratory syndrome (SARS) coronaviruses 1 and 2 (SARS-CoV-1 and SARS-CoV-2) spike proteins undergo conformational changes in the receptor binding domain (RBD) in order to bind to the host cell angiotensin converting enzyme 2 (ACE2). However, the mechanistic details that describe the large-scale conformational changes associated with spike protein activation or deactivation are still somewhat unknown. We have employed a simulation strategy that we have shown to be successful in characterizing the confromational changes in membrane proteins and have elucidated the transition pathways of protein deactivation for the SARS-CoV-1 (CoV-1) and SARS-CoV-2 (CoV-2) prefusion spike proteins. The resulting transition pathways were shown to be unique to both CoV-1 and CoV-2 . A number of key residues that form various interdomain salt-bridges have been identified. We have also constructed the free energy profiles along the transition pathways for both CoV-1 and CoV-2. The CoV-2 spike protein must overcome larger free energy barriers to undergo conformational changes towards protein activation or deactivation, when compared to CoV-1. We have performed a variety of biased and unbiased MD simulations on the bacterial proton-coupled oligopeptide transporter GkPOT, glucose transporter 1 (GluT1), and glycerol-3-phosphate transporter (GlpT) to compare the similarities and differences of the conformational dynamics of three different classes of MFS transporters. Here we have simulated the apo protein in an explicit membrane environment. Our results suggest a very similar conformational transition involving interbundle salt-bridge formation/disruption coupled with the

orientation changes of transmembrane (TM) helices, specifically H1/H7 and H5/H11, resulting in an alternation in the accessibility of water at the cyto- and periplasmic gates.

Acknowledgement

Firstly, I would like to thank my research supervisor, Dr. Mahmoud Moradi, for his guidance and support over the course of my PhD program. He has been a great mentor to me during my time as a graduate student. I could not be any more grateful for the opportunity he has given me, it has truly been an honor. I have had the opportunity to be a part of something that is really making a difference and I will cherish the memories I have made over the course of the last five years. I would also like to thank the committee members, Dr. Suresh Kumar Thallapuranam, Dr. Colin Heyes, and Dr. Yong Wang, for their willingness to help and provide support over the course of my program.

I also extend my thanks to my colleagues from the Biomolecular Simulations Group (Vivek Kumar, Adithya Polasa, Ugochi Isu, James Losey, Mortaza Molayousefi, Matthew Brownd and Shadi Badiee). We all have grown so close over these last few years and I will miss them dearly as we all move forward.

Finally, I would like to thank my family and friends for their love and support during my tenure here at the University of Arkansas. They have stood by me through thick and thin, and without them I wouldn't be the man who I am today.

Table of contents

List of published/submitted papers

1. Chapter 1 – **Ogden, D.**, Moradi, M. (2021).Molecular Dynamics–Based Thermodynamic and Kinetic Characterization of Membrane Protein Conformational Transitions. In: Schmidt-Krey, I., Gumbart, J.C. (eds) Structure and Function of Membrane Proteins. Methods in Molecular Biology, vol 2302. Humana, New York, NY. https://doi.org/10.1007/978-1-0716-1394-8_16

2. Chapter 2 **Ogden, D.**, Moradi, M. (2022). Atomic-level characterization of the conformational transition pathways in SARS-CoV-1 and SARS-CoV-2 spike proteins. bioRxiv 2022.11.29.518406; https://doi.org/10.1101/2022.11.29.518406

3. Chapter 3 – **Ogden, D.**, Immadisetty, K., Moradi, M (2019). Conformational Transition Pathways in Major Facilitator Superfamily Transporters. bioRxiv 708289; https://doi.org/10.1-101/708289

Introduction

With the every growing advances in modern supercomputing technology, the continuing improvement of all-atom force fields, and the increasing number of available structures of proteins, the molecular dynamics (MD) simulation technique [1, 2, 3] has emerged as a prominent computational method for determining the structural dynamics of proteins and biomolecules. MD is a technique that is widely used to study the local fluctuations of proteins in functional states that have been determined by X-ray crystallography, cryogenic electron microscopy (cryoEM), or homology modeling. A major hindrance that tends to arise in most MD studies is the timescale gap between local fluctuations and large-scale conformational changes, which often doesn't allow for the study of functionally relevant events such as the state transitions of proteins or complex binding interactions.

The conformational dynamics of proteins can be modeled as a protein diffusing along the protein conformational landscape [4, 5, 6], with the landscape being composed of various basins, with various minima, and saddle points that represent the stable and transition states. A viral protein that transitions between protein activation and inactivation will have a number of closely related free energy minima around the free energy basins describing these two functional states, another example would be a membrane transport protein whose conformational landscape is composed of a number of free energy basins, with a number of free energy minima, that are often associated with the states of outward-facing (OF), inward-facing (IF), or protein occlusion (Occ).

Although, studying the known functional states of proteins will provide valuable insight into various local fluctuations, due to sampling around the free energy basins of the functional state, much more insight can be gained by exploring the conformational landscape in order to determine the various states that connect known functional states. In order to explore or sample along the conformational landscape and observe functionally relevant events, such as the transitioning or "jumping" between major free energy basins, often a very rare phenomenon, one needs to employ biased and/or nonequilibrium MD simulations to induce protein conformational changes, and overcome the free energy barriers that hinder the protein from transitioning to inaccessible functional

states. Many enhanced sampling methods are used routinely for studying functionally relevant transitions; however, utilizing these methods can be quite difficult when studying complex systems. Here we outline a number of methods to be used in finding and characterizing protein conformational changes along specific transition pathways; through the use of path-finding algorithms and enhanced sampling techniques. These techniques include: nonequilibrium pulling simulations (such as targeted MD, steered MD, and similar methods), path optimization algorithms (such as string method with swarms of trajectories [7, 8]), along-the-path free-energy calculations (such as bias-exchange umbrella sampling [8]), and transition rate estimation methods.

This work focuses on comparing the conformational transition pathways at the molecular level for homologous proteins, which is currently only possible using enhanced sampling techniques. We have specifically chosen proteins from two different protein classes; the first including the Severe acute respiratory syndrome (SARS) coronaviruses 1 and 2 (SARS-CoV-1 and SARS-CoV-2), which mediates viral entry into host cell receptors [9, 10, 11, 12], and the second class of proteins includes a bacterial proton-coupled oligopeptide transporter, namely GkPOT [13, 14], the human glucose transporter 1 (GluT1) [15, 16], and the bacterial glycerol-3-phosphate transporter (GlpT) [17], all of which are important members from the Major Facilitator Family (MFS) of membrane transport proteins. Structural similarities of proteins, such as the ones chosen here, suggest that these proteins may undergo similar conformational changes during their inter-conversions between either known or unknown functional states. By employing an accurate and effective enhanced sampling scheme, the similarities and differences between homologous proteins, undergoing similar conformational changes, can be better understood.

SARS-CoV-1 and SARS-CoV-1, first class of proteins in this study, must undergo large-scale confromational changes in the receptor binding domain (RBD) in order to bind with angiotensin-converting enzyme (ACE2) [9, 10, 11, 12]. The RBD transitions from the inactive state to the active state to facilitate binding of the receptor binding motif (RBM) with ACE2 [18]. Recent cryo-EM studies have proven successful in resolving structures of both spike proteins in the inactive state, active unbound state, and active ACE2-bound state [19, 20, 21, 22, 23]. However, the cryo-

EM studies were successful in capturing static pictures of specific protein conformations, it is somewhat unclear whether all relevant conformational states were captured and resolved. A recent single-molecule fluorescence resonance energy transfer (smFRET) study was able to capture an alternative inactive conformation for the SARS-CoV-2 spike protein [24], this conformation was deemed to be inconsistent with the current resolved cryo-EM structures. It is important to decipher the differential conformational landscapes of the SARS-CoV-1 and SARS-CoV-2 spike proteins in terms of both important functional states and their dynamics.

By employing a rigorous computational study, we were successful in facilitating conformational transition pathway characterization and along-the-path free energy calculations for both the SARS-CoV-1 and SARS-CoV-2 spike proteins. The molecular details observed in this work will shed light on two differing transition pathways associated with the deactivation of the active protomer. We also observe multiple local interactions along the pathways in which the active RBD of each protein forms key saltbridges along the pathways with domains adjacent to the active RBD, suggesting that these interactions may be directly involved in the activation/deactivation process and provide insights into complex multi-step state transitions specific to SARS-CoV-1 and SARS-CoV-2. Furthermore, we observe differing dynamics of the inactive protomers, specifically in the RBDs. Finally, the determined free energy profiles for both proteins have been constructed and further our results from our previous study [25]. Protein activation or deactivation in SARS-CoV-2 is hindered by a higher free energy barriers than SARS-CoV-1.

MFS transporters, second class of proteins, can undergo protein function in a number of different ways, some of which include active and passive transport. Active transport has two variations, either symport or antiport [17, 26, 27] and passive transport can be described as uniport. Despite these distinct functions, MFS transporters all share a common overall mode of function known as the "alternating access" mechanism, which is shared with other membrane transporters [28, 27, 26]. According to this mechanism the protein alternates between an inward- (IF) and an outward-facing (OF) conformation while never leaving the binding site of the protein expose to either side of the membrane at the same time.

By employing a similar enhanced sampling approach to the SARS-CoV-1 and SARS-CoV-2 study, we sought out to elucidate the paths that connect IF to OF in the MFS transport proteins. We performed steered molecular dynamics (SMD) [29] to generate the transition pathway for GkPOT, which is otherwise unlikely to undergo conformational changes due to a strong interdomain salt-bridge locking protein in the IF conformation. Admittedly, the transition between the IF and OF conformation may be sampled using unbiased MD for certain proteins. Our simulations were successful in capturing the transition between protein Occ and the OF state for GluT1. The conformational transition pathways of GlpT have already been reported previously for the *apo* and phosphate-bound proteins [30]. The data reported for the*apo* protein was used in making a thorough comparison between the classes of the MFS transporters. We aim to provide strong justification in describing how similar and different each of these proteins undergoes both global and local conformational changes.

References

[1] T. Hansson, C. Oostenbrink, and W. F. van Gunsteren, "Molecular dynamics simulations," *Curr. Opin. Struct. Biol.*, vol. 12, no. 2, pp. 190–196, 2002.

[2] M. Karplus and J. A. McCammon, "Molecular dynamics simulations of biomolecules," *Nat. Struct. Biol.*, vol. 265, pp. 654–652, 2002.

[3] M. Karplus and J. Kuriyan, "Molecular dynamics and protein function," *Proc. Natl. Acad. Sci. USA*, vol. 102, pp. 6679–6685, 2005.

[4] L. Maragliano, A. Fischer, E. Vanden-Eijnden, and G. Ciccotti, "String method in collective variables: minimum free energy paths and isocommittor surfaces," *J. Chem. Phys.*, vol. 125, p. 24106, 2006.

[5] W. E and E. Vanden-Eijnden, "Transition-path theory and path-finding algorithms for the study of rare events," *Annu. Rev. Phys. Chem.*, vol. 61, p. 391, 2010.

[6] M. E. Johnson and G. Hummer, "Characterization of a dynamic string method for the construction of transition pathways in molecular reactions," *J. Phys. Chem. B*, vol. 116, 2012.

[7] A. C. Pan, D. Sezer, and B. Roux, "Finding Transition Pathways Using the String Method with Swarms of Trajectories," *The Journal of Physical Chemistry B*, vol. 112, pp. 3432–3440, mar 2008.

[8] M. Moradi, G. Enkavi, and E. Tajkhorshid, "Atomic-level characterization of transport cycle thermodynamics in the glycerol-3-phosphate:phosphate antiporter," *Nature Communications*, vol. 6, p. 8393, sep 2015.

[9] M. A. Tortorici and D. Veesler, "Structural insights into coronavirus entry," in *Advances in Virus Research*, vol. 105, pp. 93–116, Academic Press Inc., 2019.

[10] A. C. Walls, X. Xiong, Y. J. Park, M. A. Tortorici, J. Snijder, J. Quispe, E. Cameroni, R. Gopal, M. Dai, A. Lanzavecchia, M. Zambon, F. A. Rey, D. Corti, and D. Veesler, "Unexpected Receptor Functional Mimicry Elucidates Activation of Coronavirus Fusion," *Cell*, vol. 176, no. 5, pp. 1026–1039.e15, 2019.

[11] X. Ou, Y. Liu, X. Lei, P. Li, D. Mi, L. Ren, L. Guo, R. Guo, T. Chen, J. Hu, Z. Xiang, Z. Mu, X. Chen, J. Chen, K. Hu, Q. Jin, J. Wang, and Z. Qian, "Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV," *Nature Communications*, vol. 11, no. 1, 2020.

[12] W. Tai, L. He, X. Zhang, J. Pu, D. Voronin, S. Jiang, Y. Zhou, and L. Du, "Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine," *Cellular and Molecular Immunology*, vol. 17, no. 6, pp. 613–620, 2020.

[13] N. Solcan, J. Kwok, P. W. Fowler, A. D. Cameron, D. Drew, S. Iwata, and S. Newstead, "Alternating access mechanism in the POT family of oligopeptide transporters," *EMBO J.*, vol. 31, no. 16, pp. 3411–3421, 2012.

[14] F. Guettou, E. M. Quistgaard, L. Tresaugues, P. Moberg, , C. Jorgerschold, A. J. Jong, P. Nordlund, and C. Loew, "Structural insights into substrate recognition in proton-dependent oligopeptide transporters.," *EMBO Rep.*, vol. 14, pp. 804–810, 2013.

[15] B. Thorens and M. Mueckler, "Glucose transporters in the 21st century," *Am. J. Physiol. – Endoc. M.*, vol. 298, no. 2, pp. E141–E145, 2010.

[16] D. Deng, C. Xu, P. Sun, J. Wu, C. Yan, M. Hu, and N. Yan, "Crystal structure of the human glucose transporter GLUT1," *Nature*, vol. 510, no. 7503, p. 121, 2014.

[17] M. G. Lawrence, L. Lindahl, and J. M. Zengel, "Effects on translation pausing of alterations in protein and RNA components of the ribosome exit tunnel," *J. Bacteriol.*, vol. 190, pp. 5862–5869, 2008.

[18] R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, Y. Bi, X. Ma, F. Zhan, L. Wang, T. Hu, H. Zhou, Z. Hu, W. Zhou, L. Zhao, J. Chen, Y. Meng, J. Wang, Y. Lin, J. Yuan, Z. Xie, J. Ma, W. J. Liu, D. Wang, W. Xu, E. C. Holmes, G. F. Gao, G. Wu, W. Chen, W. Shi, and W. Tan, "Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding," *The Lancet*, vol. 395, no. 10224, pp. 565–574, 2020.

[19] D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C. L. Hsieh, O. Abiona, B. S. Graham, and J. S. McLellan, "Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation," *Science*, vol. 367, no. 6483, pp. 1260–1263, 2020.

[20] A. C. Walls, Y.-J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, and D. Veesler, "Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein," *Cell*, vol. 181, no. 2, pp. 281 – 292.e6, 2020.

[21] Y. Yuan, D. Cao, Y. Zhang, J. Ma, J. Qi, Q. Wang, G. Lu, Y. Wu, J. Yan, Y. Shi, X. Zhang, and G. F. Gao, "Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains," *Nature Communications*, vol. 8, no. 1, p. 15092, 2017.

[22] C. Xu, Y. Wang, C. Liu, C. Zhang, W. Han, X. Hong, Y. Wang, Q. Hong, S. Wang, Q. Zhao, Y. Wang, Y. Yang, K. Chen, W. Zheng, L. Kong, F. Wang, Q. Zuo, Z. Huang, and Y. Cong, "Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM," *Science Advances*, vol. 7, no. 1, p. eabe5575, 2021.

[23] W. Song, M. Gui, X. Wang, and Y. Xiang, "Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2," *PLoS Pathogens*, vol. 14, no. 8, p. e1007236, 2018.

[24] M. Lu, P. D. Uchil, W. Li, D. Zheng, D. S. Terry, J. Gorman, W. Shi, B. Zhang, T. Zhou, S. Ding, R. Gasser, J. Prévost, G. Beaudoin-Bussières, S. P. Anand, A. Laumaea, J. R. Grover, L. Liu, D. D. Ho, J. R. Mascola, A. Finzi, P. D. Kwong, S. C. Blanchard, and W. Mothes, "Real-Time Conformational Dynamics of SARS-CoV-2 Spikes on Virus Particles," *Cell Host and Microbe*, vol. 28, no. 6, pp. 880–891.e8, 2020.

[25] V. Govind Kumar, D. S. Ogden, U. H. Isu, A. Polasa, J. Losey, and M. Moradi, "Prefusion spike protein conformational changes are slower in sars-cov-2 than in sars-cov-1," *Journal of Biological Chemistry*, vol. 298, p. 101814, 2022.

[26] I. S. Wood and P. Trayhurn, "Glucose transporters (GLUT and SGLT): expanded families of sugar transport proteins," *British Journal of Nutrition*, vol. 89, no. 1, pp. 3–9, 2003.

[27] N. Yan, "Structural advances for the major facilitator superfamily (MFS) transporters," *Trends in Biochemical Sciences*, vol. 38, no. 3, pp. 151–159, 2013.

[28] V. S. Reddy, M. A. Shlykov, R. Castillo, E. I. Sun, and M. H. Saier, "The major facilitator superfamily (MFS) revisited," *FEBS J.*, vol. 279, pp. 2022–2035, June 2012.

[29] B. Isralewitz, J. Stone, and K. Schulten, "Timeline: an interactive raster plot to identify events in molecular dynamics trajectories," 2011. Submitted.

[30] M. Moradi, G. Enkavi, and E. Tajkhorshid, "Atomic-level characterization of transport cycle thermodynamics in the glycerol-3-phosphate: phosphate antiporter," *Nature Communications*, vol. 6, 2015.

Molecular Dynamics Based Thermodynamic and Kinetic Characterization of Membrane Protein Conformational Transitions

Dylan Ogden[1], and Mahmoud Moradi[1] *

[1]Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, AR 72701

Abstract

Molecular dynamics (MD) simulations are routinely used to study structural dynamics of membrane proteins. However, conventional MD is often unable to sample functionally important conformational transitions of membrane proteins such as those involved in active membrane transport or channel activation process. Here we describe a combination of multiple MD based techniques that allows for a rigorous characterization of energetics and kinetics of large-scale conformational changes in membrane proteins. The methodology is based on biased, nonequilibrium collective variable based simulations including nonequilibrium pulling, string method with swarms of trajectories, bias-exchange umbrella sampling, and rate estimation techniques.

Introduction

With advances in supercomputing technology, continued improvement of all-atom force fields, and increasing number of available structures of membrane proteins, molecular dynamics (MD) simulation technique [1, 2, 3] has emerged as a prominent computational method for determining the structural dynamics of membrane proteins in their membrane environment. MD is a technique that is routinely used to study the local fluctuations of membrane proteins around given functional states, often determined by X-ray crystallography, cryogenic electron microscopy (cryoEM), or homology modeling. The timescale gap between local fluctuations and large-scale conformational changes, however, has hindered the use of MD to study the functionally important conformational changes such as those involved in the state transition of transporters or activation of channels and receptors.

Conformational dynamics of proteins can be modeled as a diffusion in the protein conforma-

tional landscape [4, 5, 6]. The conformational free energy landscape has various basins and saddle points that represent the stable and transition states, respectively. A membrane protein, whether it is a channel, transporter, receptor, etc, is associated with many free energy minima, most of which are clustered around a few major free energy basins, representing the functional states of protein. For instance, a channel may have various closely related free energy minima around a large free energy basin that represents its active state, and one or a few free energy basins that represent its inactive state(s). Similarly, a membrane transporter has at least three free energy basins: one for the outward-facing (OF) state, one for the inward-facing (IF) state, and one for the occluded (Occ) state. Although the conformational landscape of a protein is generally very vast, most of this landscape is associated with large free energies and can be ignored. This is due to the presence of intra- and intermolecular forces that restrict the movement of atoms and molecular domains. These forces allow fluctuations around free energy basins and "rarely" allow the system to jump between these free energy basins by crossing the free energy barriers.

MD practitioners often use the available structures to study the local fluctuations of proteins around given free energy basins employing MD simulations of tens to hundreds of nanoseconds, and more recently up to several microseconds. Jumping between the major free energy basins, however, rarely happens. In order to induce such jumps, one often needs to employ biased and/or nonequilibrium MD simulations. Many enhanced sampling techniques have been developed that can be employed to facilitate the sampling of rate events [7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18]. A handful of these methods are used routinely for the study of functionally relevant transitions; however, applying many of these methods to complex biological systems such as membrane proteins is challenging. Here we outline a number of methods to be used in finding and characterizing the membrane protein conformational transition pathways through the use of path-finding algorithms and enhanced sampling techniques. The specific techniques include: nonequilibrium pulling simulations (such as targeted MD, steered MD, and similar methods), path optimization algorithms (such as string method with swarms of trajectories [19, 20]), along-the-path free-energy calculations (such as bias-exchange umbrella sampling [20]), and transition rate estimation meth-

ods. The following section outlines the theory behind some of the techniques employed in this protocol.

Theory

Dimensionality reduction is a necessary part of computational studies of protein dynamics due to the large number of degrees of freedom in the atomic coordinate space of proteins. Collective variable (colvar) based enhanced sampling techniques such as umbrella sampling (US) [7, 8] and its nonequilibrium counterparts [9, 11, 10, 21, 13, 22] effectively work within a reduced space. Collective variables can be defined intuitively to describe the slow degrees of freedom associated with functionally important protein conformational changes [23], e.g., an interdomain molecular distance as in steered MD (SMD) [11] or the RMSD from a reference structure as in targeted MD[9]. Several collective variable suites/modules [24, 25, 26, 27] have recently been developed to allow for the system-specific design of collective variables such as path colvars [28, 29] and orientation colvars [26, 30].

An ideal collective variable could represent the "reaction coordinate" [31] as often described in transition state theory. However, even if a one-dimensional reaction coordinate exists, it is not known *a priori*. This has led to the development of several path-finding algorithms, which implicitly or explicitly approximate the reaction coordinate by the arc-length of a curve in the multi-dimensional space of atomic coordinates [32, 33] or colvars [4, 19]. Many of the colvar based enhanced sampling techniques implicitly or explicitly use a diffusion model to describe the effective dynamics in the colvar space [4, 5]. We have recently developed a Riemannian diffusion model for protein conformational dynamics that provides a robust framework for conformational free energy calculation methods and path-finding algorithms [34]. Unlike their Euclidean counterparts, the Riemannian potential of mean force (PMF) and minimum free energy path (MFEP) are *invariant under coordinate transformations* [34]. However, the protocol discussed here can be used with or without the Riemannian treatment of the colvar space.

Suppose that the dynamics of a high-dimensional atomic system ($\mathbf{x}$) can be simplified as an

effective dynamics in a reduced but generally multidimensional colvar space, ζ. The effective dynamics can be described by a Brownian motion in the ζ space with an effective potential energy $G(\zeta)$, that is the PMF of the atomic system in the ζ space:

$$G(\zeta_0) = -k_B T \log \left\langle \delta(\zeta(\mathbf{x}) - \zeta_0) \right\rangle, \tag{1}$$

where $\left\langle . \right\rangle$ is an ensemble average, k_B and T are the Boltzmann constant and temperature, respectively, and δ is the Dirac delta function.

One may sample the regions around a given point ζ_0 in the colvar space by adding a biasing term to the potential of the atomic system such as $U_0(\zeta) = \frac{k}{2}(\zeta - \zeta_0)^2$ in which k is the force constant. The free energy of the biased system (or the perturbed free energy) at point ζ_0 is:

$$F(\zeta) = -k_B T \log \int \mathbf{d}\zeta \, \exp\big(-\beta(G(\zeta) + U_0(\zeta))\big), \tag{2}$$

where $\beta = (k_B T)^{-1}$. For large force constants, the PMF can be approximated using the perturbed free energy $F(\zeta)$. Otherwise, other methods can be used to estimate the PMF as briefly described below. If the biasing center is different in different simulations (or windows/images), we have: $U_i(\zeta) = \frac{k}{2}(\zeta - \zeta_i)^2$, where i is the window/image index.

Methods such as US rely on calculating the relative free energy of different points by biasing the system around those points in different simulations. Alternatively, the biasing center could change by time, e.g., replacing ζ_i by $\eta(t)$. Examples of such simulations are SMD and targeted MD. Here, we refer to such methods as nonequilibrium pulling simulations, which may use any colvar(s) with any schedule of change (i.e., $\eta(t)$ may or many not be linear in time). The biasing potential is described by: $U(\zeta, t) = \frac{k}{2}(\zeta - \eta(t))^2$. The accumulated nonequilibrium work at any given time is: $W(t) = \int_0^t dt' \, \frac{\partial}{\partial t'} U(\zeta, t')$. Nonequilibrium work can be used to estimate the perturbed free energy using the Jarzynski relation [35] or other nonequilibrium work relations [36, 37]. In this protocol, we use pulling simulations only to generate initial pathways for other simulation protocols. The nonequilibrium pulling simulations may use multidimensional colvars; however, a

specific 1D pathway needs to be selected in order to perform the simulations.

Ideally, one may use a 1D collective variable for defining the effective dynamics as well as the biasing protocol. In practice, however, this may only be possible for extremely simple systems. A practical solution to this problem is to keep the collective variable space multidimensional, while sampling only around a particular pathway, represented by a 1D curve $\zeta(\xi)$, parametrized by ξ. The choice of the pathway is obviously crucial here and determines the relevance of the free energy results to the transition of interest. Now $\xi(\mathbf{x})$ can be treated as a 1D colvar defined as a function of atomic coordinates $\mathbf{x}$, and $G(\xi)$ is the PMF associated with ξ,

$$\exp(-\beta G(\xi)) = \left\langle \delta(\xi(\mathbf{x}^t) - \xi) \right\rangle. \tag{3}$$

Assuming ξ dynamics can be effectively described by a diffusive model, we have,

$$d\xi = (-\beta D(\xi)\frac{d}{d\xi}G(\xi) + \frac{d}{d\xi}D(\xi))dt + \sqrt{2D(\xi)}dB. \tag{4}$$

in which $D(\xi)$ is a position-dependent diffusion constant, and $B(t)$ is a Wiener process such that $\left\langle B(t) \right\rangle = 0$ and $\left\langle B^2(t) \right\rangle = t$. Fokker-Planck (or Smoluchowsky) equation associated with this process is:

$$\frac{\partial}{\partial t}p(\xi,t|\xi_0,0) = \frac{\partial}{\partial \xi}(D(\xi)\exp(-\beta G(\xi))\frac{\partial}{\partial \xi}(\exp(\beta G(r))p(\xi,t|\xi_0,0))), \tag{5}$$

in which $p(\xi,t|\xi_0,t)$ is the likelihood of finding the system at ξ after time t, given it was at ξ_0 at time 0. If two major free energy minima exist at points A and B, with no other basins outside the region spanning from A to B, the mean-first-passage time (MFPT) from A to B ($\bar{\tau}_{FP}$) can be estimated using the following relation[38]:

$$\bar{\tau}_{FP} = \int_{\xi_A}^{\xi_B} d\xi \, \frac{\int_{\xi_A}^{\xi} d\xi' \exp(-\beta G(\xi'))}{D(\xi)\exp(-\beta G(\xi))}. \tag{6}$$

The aim of the protocols described here is to find the MFEP in a multidimensional colvar space,

representing the most probable transition path between two free energy basins associated with two functional states of a protein. We start by generating an approximate path using nonequilibrium pulling simulations [23] (path generation), followed by path optimization in the multidimensional colvar space using string method [19, 20], followed by along-the-path free energy calculations using bias-exchange US [20], and finally followed by estimating the transition rate between the two states using the estimated free energies and diffusion constants.

String Method with Swarms of Trajectories (SMwST)

The SMwST algorithm [19, 20] starts from an initial string, defined by N points/images $\{\mathbf{x}_i\}$, where i is any integer from 1 to N. Colvar ζ primarily defines the biasing potential, which is $U_i(\zeta) = \frac{k}{2}(\zeta - \zeta_i)^2$ for M copies of image i. The initial values for the image centers are determined from the initial string: $\zeta_i = \zeta(\mathbf{x}_i)$. The SMwST algorithm consists of three iterative steps as follows. (Step I) Restraining: Each system is restrained for τ_R (restraining time) using the harmonic potential described above centered at the current image ζ_i. (Step II) Drifting: The simulations are continued after being released from restraints for τ_D (drifiting time). (Step III) Reparametrization: The new center for each image i is determined by averaging over all observed $\zeta(\mathbf{x})$ values of M systems associated with image i at time $\tau_R + \tau_D$ and using a linear interpolation algorithm to keep the image centers equidistant. By iterating over these steps, the string will converge to the zero-drift path, around which the string centers oscillate (upon convergence). The zero-drift path is an approximation of the MFEP [6, 34].

Bias Exchange Umbrella Sampling (BEUS)

Once the MFEP (parametrized by ξ) is known, $F(\xi)$ can be estimated using a generalization of US [39], termed BEUS [20]. Similar to the SMwST method, ξ is discretized and N umbrella windows/images are defined with biasing potentials $U_i(\zeta^t) = \frac{k}{2}(\zeta^t - \zeta_i)^2$ for $i = 1, \ldots, N$. This scheme can be thought of as a 1D US along the reaction coordinate ξ with an additional restraint on the (shortest) distance from the $\zeta(\xi)$ curve. Perturbed free energies $F_i = F(\zeta(i))$ can be estimated

(up to an additive constant) by self-consistently solving the equations [40, 41, 42]:

$$e^{-\beta F_i} = \sum_t \frac{e^{-\beta U_i(\zeta^t)}}{\sum_j T_j e^{-\beta(U_j(\zeta^t)-F_j)}}, \tag{7}$$

in which $\sum_t$ sums over all collected samples (irrespective of which replica or image they belong to) and T_j is the number of samples collected for image j.

With appropriate reweighting, PMF can be reconstructed in any arbitrary collective variable space, given sufficient sampling in that space. w^t, the unnormalized weight of configuration $\mathbf{x}^t$ can be estimated via [41]:

$$w^t = (\sum_i T_i e^{-\beta(U_i(\zeta^t)-F_i)})^{-1} \tag{8}$$

in which $\{F_i\}$ are estimated via Equation (7). Alternatively [41], one may estimate $\{w^t\}$ and $\{F_i\}$ by iteratively solving Equation (8) and:

$$e^{-\beta F_i} = \sum_t w^t e^{-\beta U_i(\zeta^t)}. \tag{9}$$

The PMF in terms of $\eta(\mathbf{x})$, an arbitrary collective variable, is estimated (up to an additive constant) as:

$$G(\eta) = -\beta^{-1} \log(\sum_t w^t \mathbf{K}(\eta(\mathbf{x}^t) - \eta)) \tag{10}$$

in which $\mathbf{K}$ is a kernel function. The above estimator is not accurate if the sampling in $\eta(\mathbf{x})$ is not converged which is the case if $\eta(\mathbf{x})$ has a slow dynamics and is not strongly correlated with ζ. For the special case of $\eta = \zeta$, the perturbed free energies $\{F_i\}$ can be used directly to estimate the PMF only within the stiff-spring approximation.

Finally, for averaging an arbitrary quantity $A(\mathbf{x})$ along the pathway $\zeta(\xi)$, one may use the weighted average $\bar{A}(\xi) = \sum_t w^t A(\mathbf{x}^t)\delta(\zeta^t - \zeta(\xi))$. However, the unweighted estimator $\bar{A}_i = \langle A(\mathbf{x}^t)\rangle_i$ is often more efficient. $\bar{\sigma}_i^2 = \langle A^2(\mathbf{x}^t) - \bar{A}^2\rangle_i / g$ provides an estimate for the variance, given $g = 1 + 2\tau_{ac}^A/\tau_{lag}$ is the statistical inefficiency in which τ_{ac}^A is the autocorrelation time associated with quantity A and τ_{lag} is the lag time between the data points used in the analysis [23].

Transition Rate Estimation

Discretizing Relation (5) results in [43]:

$$\mathbf{P}(\delta t) = (\mathbf{1} + \mathbf{R}\delta t)\mathbf{P}(0), \tag{11}$$

where $\mathbf{P}(t)$ is a vector with elements $P_i = p(\xi_i, t|\xi_0, 0)$, δt is a small time step, and $\mathbf{R}$ is a tridiagonal matrix with elements $R_{ii} = -R_{ii+1} - R_{ii-1}$ and:

$$R_{ii\pm1} = \delta\xi^2 D(\xi_{ii\pm\frac{1}{2}}) \exp(-\beta(G(\xi_i) - G(\xi_{i\pm\frac{1}{2}}))), \tag{12}$$

where $\delta\xi = \xi_{i+1} = \xi_i$ for any i. More generally, for any lag time Δt and any time t, we have:

$$\mathbf{P}(t + \Delta t) = \exp(\mathbf{R}\Delta t)\mathbf{P}(t), \tag{13}$$

which implies that the likelihood of finding a system at bin j at time $t + \Delta t$, given it was at bin i at time t, is proportional to $\exp(\mathbf{R}\Delta t)_{ij}$. Therefore, assuming neither $G(\xi)$ nor $D(\xi)$ is known, one may find both, as in Ref. [43], by maximizing the likelihood $L = \prod_\alpha \exp(\mathbf{R}\Delta t)_{i_\alpha j_\alpha}$ ($\prod_\alpha$ runs over all observations of trajectories starting at the bin i_α at a given time t and being found at the bin j_α at time $t + \delta t$). Assuming $G(\xi)$ is known, one may find $D(\xi)$ using a similar maximum likelihood approach [44]. For any given $D(\xi)$, $\mathbf{R}$ can be evaluated, resulting in the log-likelihood,

$$l = \sum_\alpha \log(\exp(\mathbf{R}\Delta t)_{i_\alpha j_\alpha}), \tag{14}$$

which can be maximized using a Metropolis Monte Carlo algorithm. We first estimate the factors $\exp(-\beta(G(\xi_i) - G(\xi_{i\pm\frac{1}{2}})))$ in $R_{ii\pm1}$, where $G(\xi_i)$ is determined for $i = 1, 2, \ldots, N$ from the BEUS simulations and $G(\xi_{i+\frac{1}{2}})$ is estimated by interpolation. An arbitrary series $D_{i+\frac{1}{2}}$, $i = 1, \ldots, N-1$ can be used as an initial guess for $D(\xi_{i+\frac{1}{2}})$. $R_{ii\pm1}$ and R_{ii} values are then calculated to estimate the log-likelihood l. For a faster convergence, one may start with the estimates of $\mathbf{R}$ associated with

the $\Delta t \to 0$ limit of Relation (12) (i.e., Relation (11)) to maximize the log-likelihood in (14). It is easy to show that the following values for $R_{i,i\pm1}$ maximize the log-likelihood in (14) at the $\Delta t \to 0$ limit:

$$R_{ii\pm1} = \frac{1}{\Delta t} \frac{N_{ii\pm1} + N_{i\pm1\,i}}{N_{ii} \exp(-\beta(G(\xi_{i\pm1}) - G(\xi_i))) + N_{i\pm1\,i\pm1}}, \tag{15}$$

in which $N_{i\,j}$ is the number of observed jumps from bin i to j with lag time Δt. Diagonal values of $\mathbf{R}$ can be also estimated using $R_{ii} = -R_{ii+1} - R_{ii-1}$, while the other elements are zero. For an arbitrary lag time Δt, the log-likelihood in Relation (14) can be evaluated using the values of $\mathbf{N}$ matrix as:

$$l = \sum_{\alpha} N_{i\,j} \log(\exp(\mathbf{R}\delta t)_{i\,j}). \tag{16}$$

Starting from the $\Delta t \to 0$ limit of $\mathbf{R}$, one can use a Metropolis Monte Carlo algorithm to maximize the log-likelihood l in Relation (16). $D_{i+\frac{1}{2}}$ is then can be estimated using:

$$D_{i+\frac{1}{2}} = \delta\xi^2 R_{ii+1} \exp(\beta(G(\xi_i) - G(\xi_{i\pm\frac{1}{2}}))). \tag{17}$$

$D(\xi_i)$ can be estimated by interpolation $(D_{i-\frac{1}{2}} + D_{i-\frac{1}{2}})/2$. Finally, the MFPT ($\bar{\tau}_{FP}$) can be estimated numerically using:

$$\bar{\tau}_{FP} = \sum_{i=1}^{N} \frac{\sum_{j=1}^{i}, \exp(-\beta G(\xi_j))}{D(\xi_i) \exp(-\beta G(\xi_i))}. \tag{18}$$

Methods

Initial Preparation

1. Begin by preparing a membrane-embedded, water-solvated model of protein using one of its available structures. The suggested protocol here may use information from multiple structures in the next steps, but only one initial model needs to be prepared for all MD simulations (*see* **Notes 1 and 2**).

2. Before employing any biased or nonequilibrium simulations, it is important to run an equi-

librium, unbiased simulation of the protein as in a conventional MD simulation. The next steps of the protocol will suffer particularly in terms of convergence if they are initiated from unequilibrated structures.

3. The length of the initial equilibration simulation can vary on how quickly a stable conformation can be reached. This is typically examined by monitoring the root-mean-square deviation (RMSD) of the protein (*see* **Note 3**).

4. The last snapshot of the equilibrium MD simulation can be used as the initial conformation for the pulling simulations. If longer simulations have been performed, multiple snapshots may be used to examine the reproducibility as long as the selected structures resides in the equilibrated region.

Path Generation: Nonequilibrium Pulling Simulations

1. The choice of colvars should be specific to the protein of interest. A set of colvars used successfully to induce the transition of interest in one protein may not be applicable to another protein. Some common examples include the RMSD with respect to a target structure (as in targeted MD) and the distance between the mass centers of two specific molecules or molecular domains (as in SMD). The orientation based colvars have particularly proven to be very effective in describing the orientations of transmembrane helices or helical bundles of transmembrane proteins and are highly recommended as an alternative to RMSD and distance.

2. When determining what colvar to use, it is important to be familiar with the proposed conformational transition of the protein, in particular the target state. It is always useful to have a model of the target state even if the model is not complete or not accurate. The target model will not be used to run an actual MD simulation so it can typically only contain the C_α atoms of the protein or the domains that will be steered (e.g., transmembrane helices).

3. Once a target model is available, one may run a targeted MD simulation with the target

model to generate a transition path. The targeted MD based transition pathways are not often reliable since they typically generate pathways that are not close to the MFEP. However, they could provide a reference to compare other protocols that are based on other colvars. Multistemp targeted MD simulations also provide an alternative method for generating initial pathways, if intermediate target models are available (*see* **Note 5**).

4. The number of colvars used is completely dependent on how simple or complex the transition pathway may be. Multiple colvars may be used in a single nonequilibrium pulling protocol; however, a specific schedule need to be provided for changing the center of bias for each colvar. In other words, the system is steered along a specific path in the colvar space, if multiple colvars are used.

5. The choice of force constant is also dependent on both the colvar type—the unit of force constant is that of energy divided by the square of the colvar unit—and the particular transition of interest—the barriers that need to be crossed as the system is driven along the predetermined pathway is particularly important. The force constant should be large enough to induce the conformational transition of interest. If the force constant is too large, however, the simulation could become unstable, and the molecular system could undergo deformation or distortion. One may need to start with an educated guess and perform a short simulation to determine whether or not the colvars change according to their schedule. If not, the force constant can be increased. If there is very little deviation from the schedule, the force constant can be lowered to allow some deviation without introducing a delay in the schedule that increases significantly over time (*see* **Note 6**).

6. The simulation time is also dependent on the choice of the colvars, the desired transition, and even the force constant chosen. It is reasonable to start with relatively short simulations (a few nanoseconds) to roughly determine the quality of the protocol and fine tune the parameters; however, the final simulation that will seed the next step of our protocol (i.e., SMwST) should be long enough (at least 100 ns) to allow for relaxation of orthogonal degrees of

freedom not involved in the colvars used at least to some extent.

7. Since pulling simulations are relatively inexpensive, it is advantageous to repeat and try many different protocols to identify one or more that may lead to a reasonable transition pathway for the given protein of interest (*see* **Fig. 1**).

8. A reasonable protocol must satisfy the following four criteria:

   (a) The desired transition must occur (*see* **Note 7**). To monitor this, one may use various measures depending on the particular transition of interest. For instance, for the activation of a channel, one may monitor the number of water molecules within the transmembrane region of protein (*see* **Fig. 1**) or measure the pore radius using programs such as HOLE [45].

   (b) The protocol must not introduce undesired structural distorsions such as major secondary structural changes (unless it is part of the transition mechanism).

   (c) The protocol should not require large amounts of work (e.g., over 500 kcal/mol). Large work indicates that the generated transition pathway is too far from equilibrium and it may not easily relax to a converged pathway in the next step (i.e., SMwST).

   (d) Finally, it is important to follow up the pulling simulations with an equilibrium simulation to release the system of all restraints and to ensure that the system will maintain a relatively stable conformation following the transition. This is to be done by slowly releasing all restraints and then allowing the system to converge towards a new equilibrium state. A protocol may satisfy all three criteria above but the final conformation may not maintain the desired functional state (e.g., the channel may not stay open). In this case, the protocol is not considered a successful protocol (*see* **Note 8**).

Path Optimization: String Method with Swarms of Trajectories

1. After performing the pulling simulations and the post equilibrium simulations, an initial set of conformations along the generated transition pathway can be extracted from the nonequi-

librium trajectories to initiate the SMwST algorithm (*see* **Note 9**).

2. The number of conformations extracted as snapshots from the trajectory files will be the number of images used in the algorithm. The number of images to be used in the algorithm may depend on the complexity of the pathway that is to be refined (*see* **Note 10**).

3. SMwST may be performed as a series of simulations, as originally implemented or it may be run in parallel as a single job as currently implemented in NAMD (using a TCL script) and Amber (using the NFE suite). In the parallel version, each image is represented by a number of independent copies that are first restrained to stay around the current image center (the restraining step) and then released (drifting stage). The total number of replicas is determined by the number of copies of each image multiplied by the number of images (*see* **Note 11**).

4. The collective variables to be used in the string method simulations may or may not be similar to those used in the pulling simulations. Obviously, there needs to be more than one colvar to have a meaningful path optimization. There is no particular limitation on the number of colvars used as long as the colvar space represents a smooth space (*see* **Note 12**).

5. The force constants to be used for the restraining step may be at least as large as those used in previous pulling simulations if the same colvars are used. However, it is recommended to use larger force constants at this stage to ensure that the restrained conformations reach and stay around the desired image center during short restraining simulations. Distortion is unlikely in these short simulations and thus larger force constants can be employed (*see* **Note 13**).

6. The number of iterations to be employed in the string method simulations will be dependent on the initial pathway generated. The closer to the MFEP the initial pathway is, the faster the SMwST simulations converge. It is thus important to ensure a reasonable pulling protocol is used to generate the initial pathway before employing computationally costly SMwST

simulations.

7. The convergence of the SMwST can be monitored using measures such as string RMSD from a reference structure such as initial string or final string. The string RMSD between two strings (of N images) is defined as the root mean square distance of the individual colvars. For instance, if n colvars are used, the string RMSD between strings i and j would be $\sqrt{\frac{1}{nN}\sum_{k=1}^{N}\sum_{l=1}^{n} d(\xi_{k,l,i},\xi_{k,l,j})^2}$, where $d(a,b)$ is the distance between colvars a and b, and $\xi_{k,l,i}$ is the l'th colvar of the k'th image of string i (*see* **Fig. 2**).

Free Energy Calculations: Bias-Exchange Umbrella Sampling

1. The converged SMwST string provides an approximation for the MFEP. To estimate the free energy along this pathway, the US simulations can be carried out using the converged SMwST image centers as the center of US windows with the same colvars used for the SMwST simulations and the last snapshots of SMwST simulations as the initial conformations of the US simulations.

2. The BEUS scheme is recommended over the conventional US scheme since it allows for the diffusion of the individual replicas along the pathway through the exchange procedure of the BEUS scheme. The faster and more reliable convergence is expected as a result.

3. Rather than using one copy per image/window, one may use multiple copies of each image in the BEUS scheme. This is particularly convenient since the SMwST algorithm already provides us with multiple copies of each image if the fully parallel version is used (*see* **Note 14**).

4. BEUS uses a similar restraining bias as that used by the SMwST simulations, but unlike the restraining used in SMwST, the force constant cannot be too large. The restraining is used to keep the protein from drifting away from the current image centers as in the SMwST method; however, an additional feature of BEUS is that the biasing potentials are also used to determine the exchange criteria. Therefore, the force constant cannot be too high, otherwise,

there will not be adequate exchange between the neighboring windows. The force constant can be selected such that all replicas have 10 to 50 % of successful exchange at every attempt (*see* **Note 15**).

5. By employing a nonparametric reweighting scheme as discussed above in Section 2, the perturbed free energies can be estimated to reconstruct the free energy profile (*see* **Fig. 3**) and various other ensemble averages along the MFEP. The perturbed free energies are only a good estimate of the PMF along the path, if the force constant is large enough (*see* **Note 16**).

6. An alternative method to estimate the PMF is to use the weighting factors (w^t's) to construct a PMF along a given 1D collective variable such as the first principal component obtained from the principal component analysis of C_α atoms of proteins. The latter method has an advantage over the former method in that it does not require the large force constant condition.

Transition Rate Estimation

1. Assuming that we have identified the MFEP relatively accurately in a relevant colvar space such that the effect dynamics of the system can be assumed to be diffusive along the identified MFEP, we can use a 1D diffusion model to describe the effective kinetics of the system as described in Section 2 above.

2. To determine the transition rate, the free energy profile (or the PMF) along the MFEP is needed, which is already obtained from BEUS simulations. In addition, the the position-dependent diffusion constant along the MFEP is also needed to accurately estimate the transition rate using Relation (6). The diffusion constant estimation can be carried out by estimating inter-image transition rates measured using unbiased simulations along the MFEP.

3. Multiple copies of conformations per image/window will be extracted from BEUS trajectories to initiate these unbiased simulations. If multiple copies of images were used in BEUS

simulations as recommended, one may simply use the last snapshots of all BEUS trajectories as the starting point for these unbiased simulations.

4. Although these are unbiased simulations, it is recommended to collect the colvar values during the simulations to monitor jumps between the images. These are the same colvars used in the BEUS simulations and the colvar values can be used to first assign an image to each sampled conformation at any given time and then count the number of transitions between different images. One may then build an empirical transition matrix based on these counts. The empirical transition matrix will be dependent on the lag time used to collect the data (or to count the jumps). It is recommended to use multiple lag times to determine the behavior of the estimated transition rates as a function of the lag time (*see* **Note 17**).

5. Once an empirical transition rate is constructed for a given lag time, a Metropolis Monte Carlo algorithm will be used as described in Section 2, to estimate diffusion constants $D(\xi_i)$ from the empirical transition matrix and the BEUS estimate free energies $G(\xi_i)$ and eventually estimate the MFPT and the overall transition rate (*see* **Fig. 4** *and* **Note 18**).

Notes

Initial Preparation

1. In the case of membrane transporters, a typical crystal structure, cryo-EM structure, or homology model will be in an inward-facing (IF), outward-facing (OF), or occluded (Occ) state. The choice of the starting point is often based on the quality and reliability of the structure. E.g., a homology model is less reliable than a crystal structure; a mutant or engineered crystal structure is less reliable than a wild-type one.

2. The choice of lipid composition, salt type and concentration, protonation states of titratable amino acids, force field parameters, temperature, box size, and other MD simulation parameters is determined at this stage. Care must to be taken in making these choices as any changes

of these parameters in the next steps may complicate the interpretability of the results.

3. Although it is common to monitor the RMSD of the protein with respect to initial frame (or preferably initial model that usually represents the known (e.g., crystal) structure), it is recommended to also monior the RMSD with respect to the last frame. If the RMSD with respect to the last frame stays small for a long enough period, it is a much stronger evidence for the stability of the final conformation than if the RMSD with respect to the initial frame stays constant.

Nonequilibrium Pulling

4. The orientation based colvars as implemented in NAMD and LAMPS (also currently available for Amber as a patch) are based on the orientation quaternion formalism that measure the rotation of a semi-rigid-body molecule or molecular domain with respect to the same molecule or molecular domain in a reference conformation. The colvar may fully describe all three rotational degrees of freedom (that is in the form of an unit orientation quaternion) or only describe the angle of rotation (orientation angle) or the angle of rotation with respect to a specific axis (tilt angle or spin angle). A 1D orientation based colvar such as an orinetation, spin, or tilt angle is typically easier to use particularly at this stage of protcol. In SMwST/BEUS simulations, the orientation colvar that contains all three degrees of freedom is more appropriate to use.

5. Another method to employ the pulling simulations is the definition of individual initial, final, and hypothetical intermediate state along the transition pathway. This will be carried out in many individual targeted simulations until the final state has been reached.Targeting of the intermediates allows to target local minima which would otherwise not be sampled by employing a single target. The intermediate may be based on available crystal structures (e.g., the Occ state of a transporter) or they may be generated using other modeling techniques (e.g., coarse-graining or isotropic network models).

6. Although a short simulation could be quite informative with regards to the choice of the force constant, one needs to also examine whether the force constant is large enough for the other stages of the transition. Due to the presence of many metastable states and barriers, it is quite possible for the system to get trapped in one of the metastable states along the pathway and may never reach the final desired state. Increasing the force constant may help overcoming such issues; however, often times changing the choice of the colvars or their schedule is more effective.

7. The first step to monitor whether the desired transition has occurred is to monitor the colvars and how they follow their imposed schedule. However, it is important to note that the final targeted value may not be always reached, which admittedly, such is the case for TMD simulations. As long as the final conformation has the desired functional features (e.g., it is an open channel), the desired transition is considered to have been induced.

8. The post-pulling equilibrium MD simulations must be long enough to allow for the relaxation of the system. If the system relaxes to a conformation that is at the desired functional state but significantly differs from the initial target used for pulling simulations, one may use the equilibrated conformation as a new target to generate an alternative transition path.

Path Optimization

9. The extracted snapshots maybe extracted from the nonequilibrium pulling simulations in an equitemporal manner. In addition, one may include a few snapshots from the post-pulling equilibrium simulation trajectory as well. This often helps speeding up the convergence of the SMwST simulations.

10. If the number of images is small, the reparametrization step may introduce large changes to the image centers that cannot be easily achieved during the restraining step. Using too many images, on the other hand, introduces unnecessary curvatures. The ideal number of images

is highly dependent on the nature and number of the colvars. A typical number of images would be between 50 and 200.

11. Depending on the computational resources available, one may use a hybrid version of serial and parallel SMwST. This is again implemented in both NAMD (smwst script) and Amber (NFE suite). In this version, one chooses to use less number of copies (or only one copy) per image, but sample each copy more than once before averaging the drift and updating the image centers. One may choose to use one copy per image and 20 samples per copy, or 20 copies per image and only one sample per copy, or anything in between, e.g., 5 copies per image and 4 samples per copy. The recommended number of copies $\times$ number of samples is at least 20.

12. One may even use the atomic coordinate space (of select atoms, e.g., C_α's). The orientation based colvars, however, provide a smoother space and are expected to provide a faster and more reliable convergence.

13. It is important to monitor the progress of the SMwST simulations, particularly in the first few iterations, to ensure the appropriateness of the parameters chosen. For instance, one should check whether all copies of each image end up around the desired image center during the restraining stage before they are released. For instance, one may plot both the image centers and the actual colvar values of the sampled conformations in different 2D colvar spaces to ensure the sampled colvar values are closely distributed around each image center at the end of each restraining stage.

Free Energy Calculations

14. Using multiple copies of an image/window in a BEUS simulation, given the availability of supercomputers that allow executing large jobs, not only allows for faster sampling but more importantly also allows for a better uncertainty estimation. The multiple copies of images

are effectively independent simulations and provide uncorrelated data points for unbiased uncertainty estimation.

15. Having adequate sampling overlap between the neighboring windows is important in US simulations. Similarly, having adequate exchange rate is important in BEUS simulations. Care must be taken in choosing the number of images and the force constants to ensure each image is close enough to its neighboring image, allowing some exchange between the neighboring replicas. This can be remedied by either lowering the force constant between neighboring images that may be experiencing lower exchange rates or by adding more images allowing for more evenly spaced neighboring images and promoting a much greater exchange rate between neighboring images. The above parameters can be optimized iteratively using short runs with the goal of achieving similar rates of exchange between neighboring replicas.

16. Within the stiff-spring approximation, i.e., the large force constant assumption, the perturbed free energy and the PMF are equal. However, the first-order correction of the approximation can be estimated *a posteriori* from the estimated perturbed free energy $F(\xi)$ as: $\frac{1}{2\beta k}(\beta(\frac{d}{d\xi}F(\xi))^2 - \frac{d^2}{d\xi^2}F(\xi))$.

Transition Rate Estimation

17. If the lag time is too short, the data will be too correlated and the diffusion constants and transition rates will be overestimated. If the lag time is too long, given the limited simulation time, few transitions will be observed and the estimated diffusion constants and transition rates will be associated with large errors. Using multiple lag times allows for identifying the optimum lag time.

18. The simulation time is again system dependent; however, since the estimated free energies already provide relative inter-image transition rates $R_{ii+1}/R_{i+1i} = \exp(-\beta(G(\xi_i) - G(\xi_{i+1})))$, the only information needed to fully construct the transition rate matrix is the downhill inter-

image transition rates. Without the BEUS free energy estimates, both uphill and downhill inter-image transition rates need to be estimated, which requires considerably more time.

Figure Captions

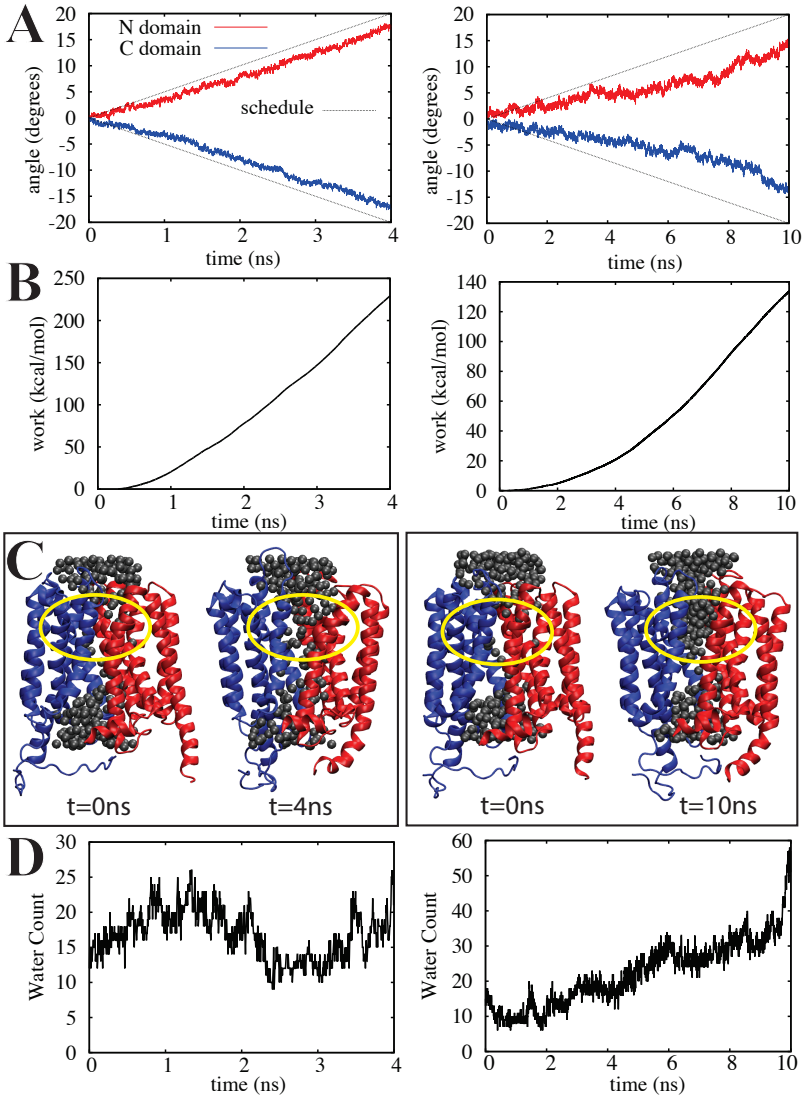

Figure 1: Comparing two nonequilibrium pulling protocols to induce an IF→OF transition in membrane transporter GlpT [20]. Both protocols induce rotational changes on the N- and C-bundle domains of GlpT using spin angles (left) or orientations (right) of the two domains. Both protocols use a force constant of 3 kcal/mol.deg^2. **A** Time series of spin (left) and orientation (right) angles. The black line shows the schedule of the time-dependent center of harmonic potentials. **B** Time series of nonequilibrium work. **C** Snapshots of protein at the beginning and end of simulations. The water molecules within the transmembrane region are shown. **D** The number of water molecules in the periplasmic side of the transmembrane region as a function of time.

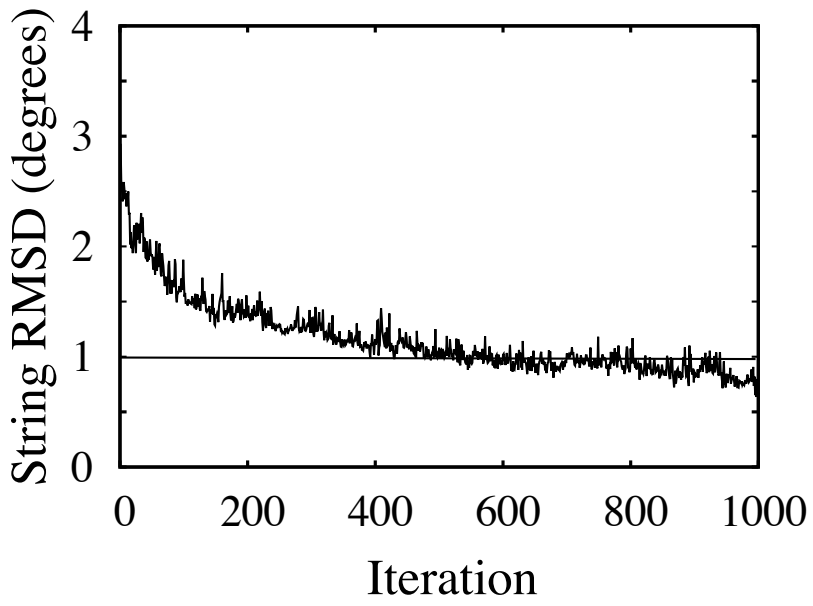

Figure 2: Monitoring the convergence of the SMwST algorithm. String RMSD at each iteration with respect to the last string from SMwST simulations of If→OF GlpT transition using 12 transmembrane orientations. Comparing the horizontal line with measured RMSD string shows that after about 400 iterations, the SMwST does not significantly change.

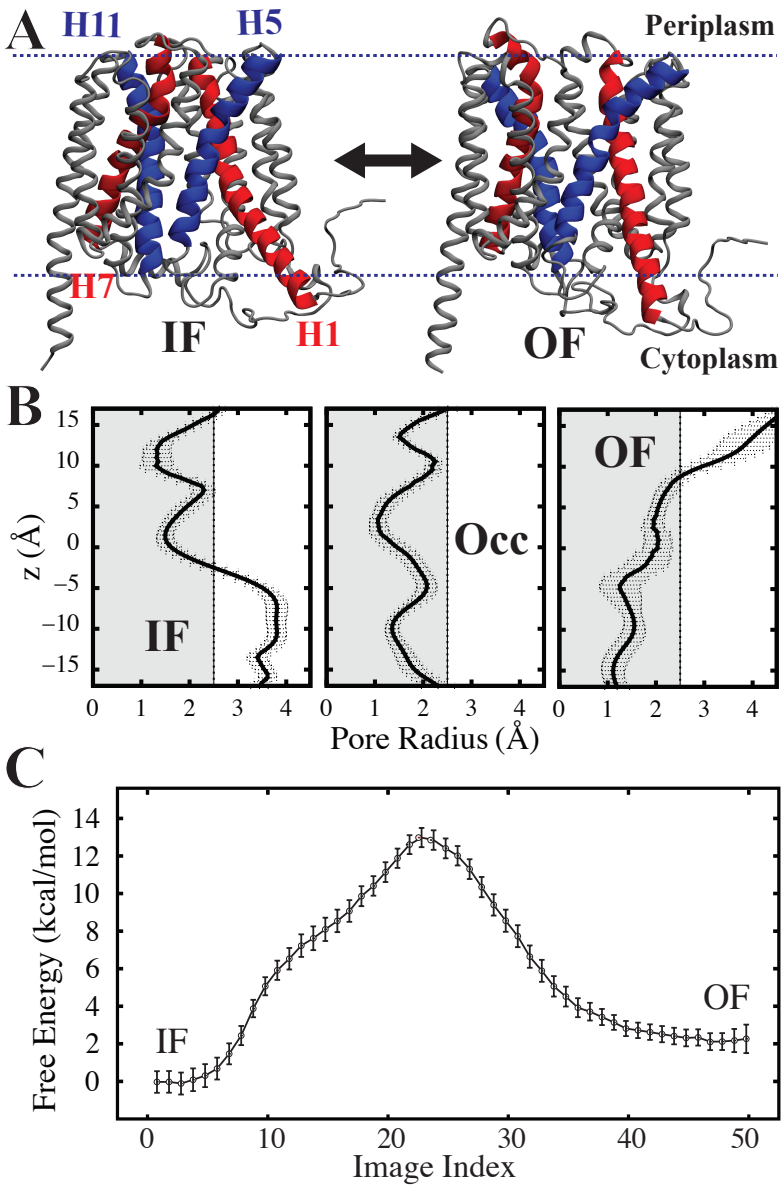

Figure 3: BEUS simulation results of IF→OF GlpT transition. **A** Snapshots of the first ($i = 1$) and last ($i = 50$) images, representing the IF and OF state, respectively. **B** The pore radius along the pore based on the snapshots of the first (IF) and last (OF) as well as an intermediate (Occ) image. The latter represents an occluded state. **C** The perturbed free energies estimated from the BEUS simulations [20].

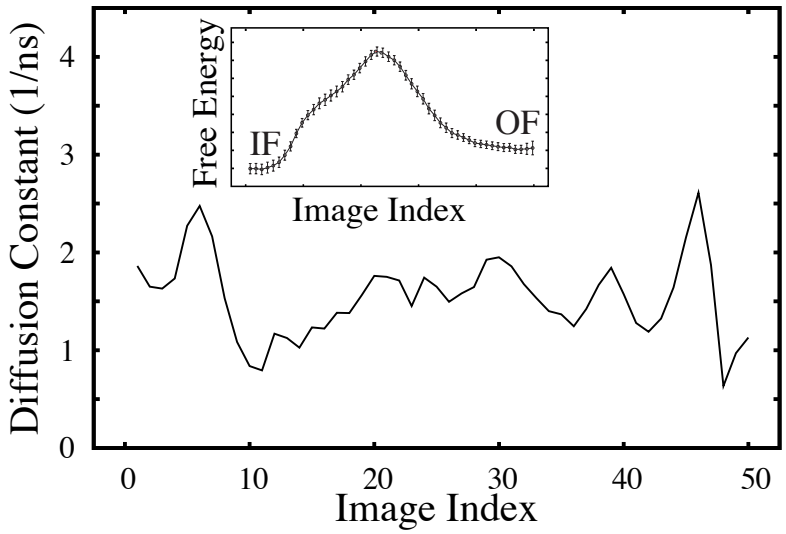

Figure 4: The diffusion constant as a function of image index estimated based on GlpT BEUS simulations (for free energies) and follow-up equilibrium simulations using a lag time of 0.5 ns, which was determined to be the optimum lag time. The MFPT estimated based on these calculations for the IF→OF GlpT transition is approximately 6 seconds.

References

[1] T. Hansson, C. Oostenbrink, and W. F. van Gunsteren, "Molecular dynamics simulations," *Curr. Opin. Struct. Biol.*, vol. 12, no. 2, pp. 190–196, 2002.

[2] M. Karplus and J. A. McCammon, "Molecular dynamics simulations of biomolecules," *Nat. Struct. Biol.*, vol. 265, pp. 654–652, 2002.

[3] M. Karplus and J. Kuriyan, "Molecular dynamics and protein function," *Proc. Natl. Acad. Sci. USA*, vol. 102, pp. 6679–6685, 2005.

[4] L. Maragliano, A. Fischer, E. Vanden-Eijnden, and G. Ciccotti, "String method in collective variables: minimum free energy paths and isocommittor surfaces," *J. Chem. Phys.*, vol. 125, p. 24106, 2006.

[5] W. E and E. Vanden-Eijnden, "Transition-path theory and path-finding algorithms for the study of rare events," *Annu. Rev. Phys. Chem.*, vol. 61, p. 391, 2010.

[6] M. E. Johnson and G. Hummer, "Characterization of a dynamic string method for the construction of transition pathways in molecular reactions," *J. Phys. Chem. B*, vol. 116, 2012.

[7] G. M. Torrie and J. P. Valleau, "Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling," *J. Comp. Phys.*, vol. 23, pp. 187–199, 1977.

[8] S. H. Northrup, M. R. Pear, C. Y. Lee, J. A. McCammon, and M. Karplus, "Dynamical theory of activated processes in globular proteins," *Proc. Natl. Acad. Sci. USA*, vol. 79, pp. 4035–4039, 1982.

[9] J. Schlitter, M. Engels, P. Krüger, E. Jacoby, and A. Wollmer, "Targeted molecular dynamics simulation of conformational change — application to the T-R transition in insulin," *Mol. Sim.*, vol. 10, no. 2–6, pp. 291–308, 1993.

[10] T. Huber, A. E. Torda, and W. F. van Gunsteren, "Local elevation: A method for improving the searching properties of molecular dynamics simulation," *J. Comput.-Aided Mol. Des.*, vol. 8, p. 695, 1994.

[11] S. Izrailev, S. Stepaniants, M. Balsera, Y. Oono, and K. Schulten, "Molecular dynamics study of unbinding of the avidin-biotin complex," *Biophys. J.*, vol. 72, pp. 1568–1581, 1997.

[12] Y. Sugita and Y. Okamoto, "Replica-exchange molecular dynamics method for protein folding," *Chem. Phys. Lett.*, vol. 314, pp. 141–151, 1999.

[13] A. Laio and M. Parrinello, "Escaping free energy minima," *Proc. Natl. Acad. Sci. USA*, vol. 99, no. 20, pp. 12562–12566, 2002.

[14] E. Darve, D. Rodríguez-Gómez, and A. Pohorille, "Adaptive biasing force method for scalar and vector free energy calculations," *J. Chem. Phys.*, vol. 128, p. 144120, 2008.

[15] C. F. Abrams and E. Vanden-Eijnden, "Large-scale conformational sampling of proteins using temperature-accelerated molecular dynamics," *Proc. Natl. Acad. Sci. USA*, vol. 107, no. 11, pp. 4961–4966, 2010.

[16] C. Templeton, S.-H. Chen, A. Fathizadeh, and R. Elber, "Rock climbing: A local-global algorithm to compute minimum energy and minimum free energy pathways," *J. Chem. Phys.*, vol. 147, no. 15, p. 152718, 2017.

[17] L. T. Chong, A. S. Saglam, and D. M. Zuckerman, "Path-sampling strategies for simulating rare events in biomolecular systems," *Curr. Opin. Struct. Biol.*, vol. 43, pp. 88–94, 2017.

[18] A. Laio, A. Z. Panagiotopoulos, and D. M. Zuckerman, "Preface: Special topic on enhanced sampling for molecular systems," *J. Chem. Phys.*, vol. 149, no. 7, p. 072001, 2018.

[19] A. C. Pan, D. Sezer, and B. Roux, "Finding Transition Pathways Using the String Method with Swarms of Trajectories," *The Journal of Physical Chemistry B*, vol. 112, pp. 3432–3440, mar 2008.

[20] M. Moradi, G. Enkavi, and E. Tajkhorshid, "Atomic-level characterization of transport cycle thermodynamics in the glycerol-3-phosphate:phosphate antiporter," *Nature Communications*, vol. 6, p. 8393, sep 2015.

[21] G. Hummer and I. G. Kevrekidis, "Coarse molecular dynamics of a peptide fragment: Free energy, kinetics, and long-time dynamics computations," *J. Chem. Phys.*, vol. 118, p. 10762, 2003.

[22] M. Moradi and E. Tajkhorshid, "Driven metadynamics: Reconstructing equilibrium free energies from driven adaptive-bias simulations," *J. Phys. Chem. Lett.*, vol. 4, no. 11, pp. 1882–1887, 2013.

[23] M. Moradi and E. Tajkhorshid, "Computational recipe for efficient description of large-scale conformational changes in biomolecular systems," *J. Chem. Theory Comp.*, vol. 10, no. 7, pp. 2866–2880, 2014.

[24] M. Bonomi, D. Branduardi, G. Bussi, C. Camilloni, D. Provasi, P. Raiteri, D. Donadio, F. Marinelli, F. Pietrucci, R. Broglia, and M. Parrinello, "PLUMED: a portable plugin for free energy calculations with molecular dynamics," *Comput. Phys. Commun.*, vol. 180, p. 1961, 2009.

[25] V. Babin, V. Karpusenka, M. Moradi, C. Roland, and C. Sagui, "Adaptively biased molecular dynamics: An umbrella sampling method with a time-dependent potential," *Int. J. Quantum Chem.*, vol. 109, no. 15, pp. 3666–3678, 2009.

[26] G. Fiorin, M. L. Klein, and J. Hénin, "Using collective variables to drive molecular dynamics simulations," *Mol. Phys.*, vol. 111, p. 3345, 2013.

[27] H. Sidky, Y. J. Colón, J. Helfferich, B. J. Sikora, C. Bezik, W. Chu, F. Giberti, A. Z. Guo, X. Jiang, J. Lequieu, J. Li, J. Moller, M. J. Quevillon, M. Rahimi, H. Ramezani-Dakhel, V. S. Rathee, D. R. Reid, E. Sevgen, V. Thapar, M. A. Webb, J. K. Whitmer, and J. J. de Pablo, "Ssages: Software suite for advanced general ensemble simulations," *J. Chem. Phys.*, vol. 148, no. 4, p. 044104, 2018.

[28] D. Branduardi, F. L. Gervasio, and M. Parrinello, "From a to b in free energy space," *J. Chem. Phys.*, vol. 126, no. 5, p. 054103, 2007.

[29] A. Berteotti, A. Cavalli, D. Branduardi, F. L. Gervasio, M. Recanatini, and M. Parrinello, "Protein conformational transitions: The closure mechanism of a kinase explored by atomistic simulations," *J. Am. Chem. Soc.*, vol. 131, no. 1, pp. 244–250, 2009.

[30] M. Moradi and E. Tajkhorshid, "Mechanistic picture for conformational transition of a membrane transporter at atomic resolution.," *Proc. Natl. Acad. Sci. USA*, vol. 110, no. 47, pp. 18916–18921, 2013.

[31] F. Legoll and T. Leliévre, "Effective dynamics using conditional expectations," *Nonlinearity*, vol. 23, no. 9, p. 2131, 2010.

[32] R. Czerminski and R. Elber, "Reaction path study of conformational transitions and helix formation in a tetrapeptide," *Proc. Natl. Acad. Sci. USA*, vol. 86, p. 6963—6967, September 1989.

[33] G. Mills and H. Jónsson, "Quantum and thermal effects in h_2 dissociative adsorption: Evaluation of free energy barriers in multidimensional quantum systems," *Phys. Rev. Lett.*, vol. 72, pp. 1124–1127, 1994.

[34] A. Fakharzadeh and M. Moradi, "Effective Riemannian diffusion model for conformational dynamics of biomolecular systems," *J. Phys. Chem. Lett.*, vol. 7, pp. 4980–4987, 2016.

[35] C. Jarzynski, "Nonequilibrium equality for free energy differences," *Phys. Rev. Lett.*, vol. 78, pp. 2690–2693, 1997.

[36] G. E. Crooks, "Path-ensemble averages in systems driven far from equilibrium," *Phys. Rev. E*, vol. 61, pp. 2361–2366, 2000.

[37] G. Hummer and A. Szabo, "Free energy reconstruction from nonequilibrium single-molecule pulling experiments," *Proc. Natl. Acad. Sci. USA*, vol. 98, pp. 3658–3661, 2001.

[38] S. Lifson and J. L. Jackson, "On the self-diffusion of ions in a polyelectrolyte solution," *The Journal Of Chemical Physics*, vol. 36, no. 9, pp. 2410–2414, 1962.

[39] G. M. Torrie and J. P. Valleau, "Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling," *J. Comput. Phys.*, vol. 23, p. 187, 1977.

[40] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg, "The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method," *J. Comp. Chem.*, vol. 13, pp. 1011–1021, 1992.

[41] C. Bartels, "Analyzing biased monte carlo and molecular dynamics simulations," *Chem. Phys. Lett.*, vol. 331, p. 446, 2000.

[42] M. R. Shirts and J. D. Chodera, "Statistically optimal analysis of samples from multiple equilibrium states," *J. Chem. Phys.*, vol. 129, p. 124105, 2008.

[43] G. Hummer, "Position-dependent diffusion coefficients and free energies from bayesian analysis of equilibrium and replica molecular dynamics simulations," *New Journal of Physics*, vol. 7, p. 34, 2005.

[44] A. Singharoy, C. Chipot, M. Moradi, and K. Schulten, "Chemomechanical coupling in hexameric protein–protein interfaces harness energy within V-type ATPases," *J. Am. Chem. Soc.*, vol. 139, pp. 293—-310, 2017.

[45] O. S. Smart, J. G. Neduvelil, X. Wang, B. A. Wallace, and M. S. P. Sansom, "HOLE: A program for the analysis of the pore dimensions of ion channel structural models," *J. Mol. Graphics*, vol. 14, pp. 354–360, 1996.

Atomic-level Characterization of the Conformational Transition Pathways in SARS-CoV-1 and SARS-CoV-2 Spike Proteins

Dylan Ogden[1], and Mahmoud Moradi[1] *

[1]Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, AR 72701

Abstract

Severe acute respiratory syndrome (SARS) coronaviruses 1 and 2 (SARS-CoV-1 and SARS-CoV-2) derive transmissibility from spike protein activation in the receptor binding domain (RBD) and binding to the host cell angiotensin converting enzyme 2 (ACE2). However, the mechanistic details that describe the large-scale conformational changes associated with spike protein activation or deactivation are still somewhat unknown. Here, we have employed an extensive set of nonequilibrium all-atom molecular dynamics (MD) simulations, utilizing a novel protocol, for the SARS-CoV-1 (CoV-1) and SARS-CoV-2 (CoV-2) prefusion spike proteins in order to characterize the conformational pathways associated with the active-to-inactive transition. Our results indicate that both CoV-1 and CoV-2 spike proteins undergo conformational transitions along pathways unique to each protein. We have identified a number of key residues that form various inter-domain salt-bridges, suggesting a multi-stage conformational change along the pathways. We have also constructed the free energy profiles along the transition pathways for both CoV-1 and CoV-2 spike proteins. The CoV-2 spike protein must overcome larger free energy barriers to undergo conformational changes towards protein activation or deactivation, when compared to CoV-1.

Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) mediates viral entry into host cell receptors [1, 2, 3, 4] and is the major cause for the coronavirus disease pandemic in 2019 (COVID-19) [5, 6, 7, 8, 9]. SARS-CoV-2 is by far the most effective coronavirus, among SARS coronavirus 1 (SARS-CoV-1) and Middle East respiratory syndrome (MERS) [10], in terms of being easily transmissible to millions of people around the world [11]. Genomic analysis of SARS-related

coronaviruses has indicated the coronavirus spike protein is the most variable region of the entire genome [5]. This particular protein is involved in host cell recognition, by binding the human angiotensin-converting enzyme 2 (ACE2) receptor, and viral entry by mediating membrane fusion. Interestingly, the CoV-2 spike protein has been shown to have higher affinity for human ACE2 than the CoV-1 spike protein [10].

SARS-CoV-2 must undergo large-scale confromational changes in the receptor binding domain (RBD) in order to interact and bind with ACE2 [1, 2, 3, 4]. The RBD transitions from the inactive state ("down") to the active state ("up") to facilitate binding of the receptor binding motif (RBM) with ACE2 [6]. Recent Cryo-EM studies have successfully resolved structures of both spike proteins in the inactive state, active unbound state, and active ACE2-bound state [12, 10, 13, 14, 15]. However, cryo-EM and X-ray crystallography studies capture static pictures of specific protein conformations and do not provide detailed information on the dynamic behavior associated with the major protein conformational changes [16, 17, 18]. In addition, given the substantial differences in the experimental and physiological conditions, it is not clear whether all relevant conformational states are captured using techniques such as cryo-EM. For instance, a recent single-molecule fluorescence resonance energy transfer (smFRET) study has captured an alternative inactive conformation for the CoV-2 spike protein [19] that is not consistent with those obtained from cryo-EM. It is important to decipher the differential conformational landscapes of the CoV-1 and CoV-2 spike proteins in terms of both important functional states and their dynamics. Understanding the activation mechanism of the coronavirus spike proteins and its potential variant-dependent nature is key to the development of efficient coronavirus vaccines and therapeutics with long-term efficacy.

All-atom molecular dynamics (MD) simulations can provide dynamic pictures of biomolecular processes at a detailed molecular level. Often times MD simulations are often unable functionally relevant events, such as those involved in spike protein activation, tend to extend beyond the the time scales of most MD simulations. The scope of MD simulation studies of large protein complexes is most often limited to the local conformational changes of individual states starting from

available crystal structures or homology models [20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30]. An efficient computational framework for describing large-scale conformational changes using a combination of several distinct enhanced sampling techniques, without compromising atomic details, has been previously developed [31, 32, 33, 34] and employed in this work.

In order to generate functionally relevant state transitions between protein activation and inactivation in both SARS-CoV-1 and SARS-CoV-2, we have employed Steered Molecular Dynamics (SMD) to generate initial transition pathways and followed by String Method with Swarms of Trajectories (SMwST) for further transition path characterization and refinement in high-dimensional collective variable spaces. Finally, Bias Exchange Umbrella Sampling (BEUS) is employed to effectively sample along the transition pathways using optimized reaction coordinates in order to determine the free energy associated with the conformational changes.

By employing a novel enhanced sampling scheme, we have conducted an extensive computational study to characterize the transition pathways for both SARS-CoV-1 and CoV-2 spike proteins. Our results provide details indicating two very different transition pathways associated with the deactivation of the active protomer in both systems. We also observe multiple local interactions along the pathways in which the active RBD of each protein forms key salt-bridges along the pathways with an adjacent protomer's (promters B and C) RBD, N-terminal domain (NTD), and S2 domains, suggesting the forming and breaking of these salt-bridges may be required in the activation/inactivation process and provide insights into complex multi-step state transitions specific to CoV-1 and CoV-2. Furthermore, we observe differing dynamics of the inactive protomer RBDs for both CoV-1 and 2 along the transition pathways. Finally, the determined free energy profiles for both proteins have been constructed and further our results from our previous study [35]; protein activation or deactivation in CoV-2 is hindered by a higher free energy barriers than CoV-1.

Methods

All-atom equilibrium MD simulations.

We have used all-atom equilibrium MD simulations to characterize the active-to-inactive conformational transition pathways of the two spike proteins SARS-CoV-2 and SARS-CoV-1. Our simulations were derived from the two cryo-EM structures of the SARS-CoV-2 spike protein (PDB entry:6VYB) [10] and SARS-CoV-1 spike protein (PDB entry:5X5B) [13], both containing one active promoter. The protein was solvated in a box of TIP3P waters, and 0.15 M NaCl (in addition to the counterions used to neutralize the protein) using CHARMM-GUI [36, 37]. The box size for the CoV-2 active model was 198 x 198 x 198 $Å^3$ with 850000 atoms. The box size for the CoV-1 active model was 193 x 193 x 193 $Å^3$ with 750000 atoms.

All simulations were conducted using the NAMD 2.13/14 [38] simulation package with systems parameterized with the CHARMM36m all-atom additive force field [39]. As described in previous work [35], each system was energy-minimized for 10,000 steps using the conjugate gradient algorithm [40]. Both systems were relaxed using restrained MD simulations in a step-wise manner using the standard CHARMM-GUI protocol [36, 37] (“relaxation step”). Backbone and sidechain restraints were used for 10 ns with a force constant of 1 kcal/mol.$Å^2$ and 0.5 kcal/mol.$Å^2$ respectively (”restraining step”). The systems were then equilibrated with no bias for another 10 ns (”equilibration step”). The initial relaxation was performed in an NVT ensemble while the rest of the simulations were performed in an NPT ensemble. Simulations were carried out using a 2-fs time step at 310 K using a Langevin integrator with a damping coefficient of $\gamma = 0.5$ $ps^{-}1$. The pressure was maintained at 1 atm using the Nose-Hoover Langevin piston method [40, 41]. The smoothed cutoff distance for non-bonded interactions was set at 10 to 12 Å and long-range electrostatic interactions were computed with the particle mesh Ewald (PME) method [42].

Steered Molecular Dynamics (SMD)

To induce activation/inactivation of a protomer initially in the inactive/active conformation, we defined collective variables based on the the Cα RMSD of each protomer for both the CoV-1 and CoV-2 systems. Reference coordinates were taken from the corresponding active/inactive structure for both CoV-1 and CoV-2 protomers. The atoms chosen were based on the total number of resolved and modeled residues in the CoV-2 structures. Structural analysis of CoV-1 and CoV-2 was employed to ensure that equivalent Cα atoms were steered in all simulation sets. 1037 atoms were steered for any given protomer and the following residue range was used: 27 to 239, 244 to 315, 322 to 662, 673 to 809, and 831 to 1104. These atoms span the entire protomer, starting from the NTD and ending approximately at the C-terminus of the S2 region. A force constant of 250 $kcal/mol/\text{Å}^2$ was used for SMD simulations involving a single protomer. Altogether 40 simulations were conducted (2 directions x 2 proteins x 10 replicas) for 100 ns (400 ns total) of SMD. One pathway was chosen for each protein whose resulting pathway required the least amount of nonequilibrum work (active-to-inactive transition pathway) [35]. Each of the two pathways were reduced to 100 equidistant images (conformations or windows) to represent the initial transition pathway.

String Method with Swarms of Trajectories (SMwST)

The resulting generated pathways are far from equilibrium and require path optimization. Path refinement through the use of an automatic path finding algorithm known as string method with swarms of trajectories (SMwST) [33] was employed. SMwST has proven successful in generating reliable transition pathways in the past [33, 34, 43], while taking into account the nonlinearity of multidimensional collective variables. For the first set of SMwST, we employed 200 iterations for SARS-CoV-2 and 125 iterations for SARS-CoV-1 using the atomic coordinates of the Cα, with a force constant of 10 kcal/mol.Å^2, atoms of the active protomer as the collective variable to characterize the transition pathway from active to inactive, which encompassed 1037 atoms in both systems, the same as SMD. Both systems utilized 2000 replicas in total (i.e. 100 images x 20

copies/image), each iteration of SMwST included 10 ps of restraining followed by 1 ps of restraint release. The total simulation times were 4.4 μs for SARS-CoV-2 and 2.75 μs for SARS-CoV-1.

A second set of SMwST was used to further characterize the transition pathways using collective variables to more accurately describe the conformational changes. In our approach, we employed more system-specific collective variables such as the "orientation quaternion" [44, 45, 46, 31]. Orientation quaternions have been proven to be efficient collective variables at inducing a rotational change of a given molecular domain or can be utilized more simply by restraining a domain's orientation. For each system we employed four collective variables, the orientation quaternion of the three domains: RBD, NTD, and S2. Additionally, the distance (distance vector) between the RBD and S2 domain (active protomer only) was used as the fourth collective variable. The force constants used were 200 kcal/mol.Å^2 for distance and 500,000 kcal/mol.rad^2 for the orientation quaternion collective variables. Furthermore, 200 iterations of SMwST were employed for CoV-2 and 120 iterations for CoV-1. Each iteration of SMwST included 10 ps of restraining and 2 ps of restraint release. Similar to the first set, each system contained 2000 replicas (i.e. 100 images x 20 copies/image). The final simulation times were 4.4 μs for CoV-2 and 2.88 μs for CoV-1. In total the simulation time for CoV-2 is 8.8 μs and 5.63 μs for CoV-1, and in grand total: 14.43 μs for all SMwST simulations.

Bias Exchange Umbrella Sampling (BEUS)

Once optimized pathways have been generated from SMwST, the pathways were be further characterized by calculating the free energy along the transition paths. The novel Bias Exchange Umbrella Sampling (BEUS) scheme utilized in this work, allows for very accurate sampling along a transition pathway best represented by the optimal transition pathways generated from the SMwST simulations [31, 32, 33, 34]. The two converged SMwST simulations were reduced to 100 replicas, 1 replica extracted from each of the 100 images. We performed 25 ns of BEUS, a replica exchange was attempted every 2 ps, in the reparameterized collective variable space described above, to provide enough sampled dynamics for accurate construction of active-to-inactive free energy profiles.

The force constants used for each of the collective variables was 2000 kcal/mol.rad^2 for the orientation quaternion restraints (both systems) and for the distance restraints 2 kcal/mol.Å^2 for CoV-2 and 0.5 kcal/mol.Å^2 for CoV-1. The exchange rates for 100 replicas in both CoV-1 and CoV-2 were as follows: 64% for CoV-1 and 63% for CoV-2.

Theoretical Framework

The SMwST algorithm [47, 33] starts from an initial string, defined by N points/images $\{\mathbf{x}_i\}$, where i is any integer from 1 to N. Colvar ζ primarily defines the biasing potential, which is $U_i(\zeta) = \frac{k}{2}(\zeta - \zeta_i)^2$ for M copies of image i. The initial values for the image centers are determined from the initial string: $\zeta_i = \zeta(\mathbf{x}_i)$. The SMwST algorithm consists of three iterative steps as follows. (Step I) Restraining: Each system is restrained for τ_R (restraining time) using the harmonic potential described above centered at the current image ζ_i. (Step II) Drifting: The simulations are continued after being released from restraints for τ_D (drifiting time). (Step III) Reparametrization: The new center for each image i is determined by averaging over all observed $\zeta(\mathbf{x})$ values of M systems associated with image i at time $\tau_R + \tau_D$ and using a linear interpolation algorithm to keep the image centers equidistant. By iterating over these steps, the string will converge to the zero-drift path, around which the string centers oscillate (upon convergence). The zero-drift path is an approximation of the MFEP [48, 49].

Once the MFEP (parametrized by ξ) is known, $F(\xi)$ can be estimated using a generalization of US [50], termed BEUS [33]. Similar to the SMwST method, ξ is discretized and N umbrella windows/images are defined with biasing potentials $U_i(\zeta^t) = \frac{k}{2}(\zeta^t - \zeta_i)^2$ for $i = 1, \ldots, N$. This scheme can be thought of as a 1D US along the reaction coordinate ξ with an additional restraint on the (shortest) distance from the $\zeta(\xi)$ curve. Perturbed free energies $F_i = F(\zeta(i))$ can be estimated (up to an additive constant) by self-consistently solving the equations [51, 52, 53]:

$$e^{-\beta F_i} = \sum_t \frac{e^{-\beta U_i(\zeta^t)}}{\sum_j T_j e^{-\beta(U_j(\zeta^t) - F_j)}}, \tag{19}$$

in which $\sum_t$ sums over all collected samples (irrespective of which replica or image they belong to) and T_j is the number of samples collected for image j.

With appropriate reweighting, PMF can be reconstructed in any arbitrary collective variable space, given sufficient sampling in that space. w^t, the unnormalized weight of configuration $\mathbf{x}^t$ can be estimated via [52]:

$$w^t = (\sum_i T_i \mathrm{e}^{-\beta(U_i(\zeta^t)-F_i)})^{-1} \tag{20}$$

in which $\{F_i\}$ are estimated via Equation (19). Alternatively [52], one may estimate $\{w^t\}$ and $\{F_i\}$ by iteratively solving Equation (20) and:

$$\mathrm{e}^{-\beta F_i} = \sum_t w^t \mathrm{e}^{-\beta U_i(\zeta^t)}. \tag{21}$$

The PMF in terms of $\eta(\mathbf{x})$, an arbitrary collective variable, is estimated (up to an additive constant) as:

$$G(\eta) = -\beta^{-1} \log(\sum_t w^t \mathbf{K}(\eta(\mathbf{x}^t) - \eta)) \tag{22}$$

in which $\mathbf{K}$ is a kernel function. The above estimator is not accurate if the sampling in $\eta(\mathbf{x})$ is not converged which is the case if $\eta(\mathbf{x})$ has a slow dynamics and is not strongly correlated with ζ. For the special case of $\eta = \zeta$, the perturbed free energies $\{F_i\}$ can be used directly to estimate the PMF only within the stiff-spring approximation.

Finally, for averaging an arbitrary quantity $A(\mathbf{x})$ along the pathway $\zeta(\xi)$, one may use the weighted average $\bar{A}(\xi) = \sum_t w^t A(\mathbf{x}^t)\delta(\zeta^t - \zeta(\xi))$. However, the unweighted estimator $\bar{A}_i = \left\langle A(\mathbf{x}^t) \right\rangle_i$ is often more efficient. $\bar{\sigma}_i^2 = \left\langle A^2(\mathbf{x}^t) - \bar{A}^2 \right\rangle_i / g$ provides an estimate for the variance, given $g = 1 + 2\tau_{ac}^A/\tau_{lag}$ is the statistical inefficiency in which τ_{ac}^A is the autocorrelation time associated with quantity A and τ_{lag} is the lag time between the data points used in the analysis [32].

Results and Discussion

We recently performed a combination of all-atom microsecond-level and SMD simulations of the non-glycosylated spike proteins SARS-CoV-1 and the original SARS-CoV-2 strain [35]. In all but one of the systems simulated, no large-scale conformational change occurred within the timescales of the unbiased simulations. A simulation that was initiated from the active conformation for CoV-1 transitioned from active to a "pseudo-inactive" and provided details on a number local interactions between the CoV-1 spike NTD and RBD, some of which are unique to CoV-1, and were directly involved in the deactivation of the spike protein. Furthermore, the SMD simulations showed that it is relatively difficult for the CoV-2 spike protein to undergo a large-scale conformational transition between active and inactive states, when compared to the CoV-1 spike protein.

To further the work previously employed, we aimed to more accurately characterize the conformational transition of the active protomer in both CoV-1 and CoV-2 using a computational protocol that has proven to be reliable at accurately describing large-scale conformational changes [31, 32, 33, 34]. Using the pathways that were generated from the previous sets of SMD simulations [35], we chose one pathway for both CoV-1 and CoV-2 to be used as the initial defining pathway to be further characterized by the SMwST simulations. These particular pathways were derived from the sets of simulations in the direction of active-to-inactive. In order to determine the appropriate pathways to be utilized, we employed a "work-based" analysis when choosing them, (i.e. the pathways that required the least amount of nonequilibrum work) under the assumption that the pathways generated were energetically favored. Currently, the pathway that connects the active and inactive states in both proteins is not known but the cryo-EM structures provide details on the two end states (active and inactive) [13, 10].

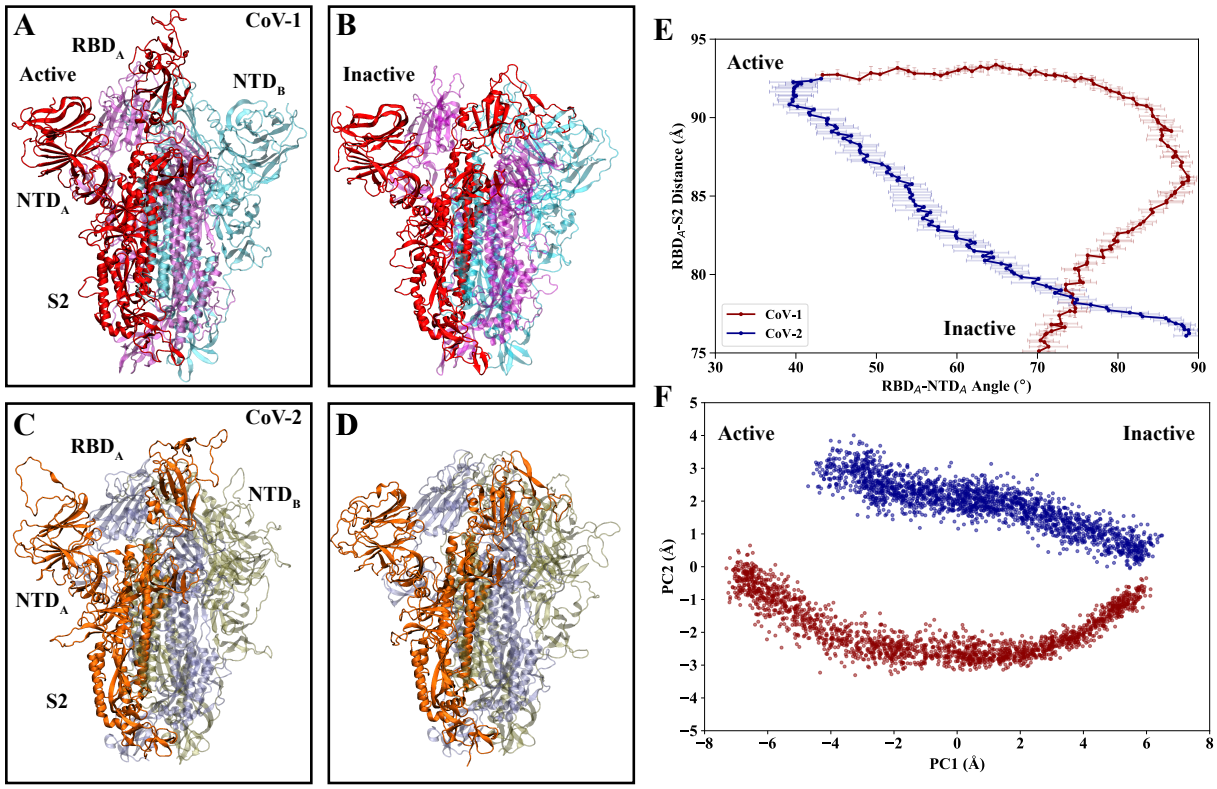

Figure 1: **Global conformational changes of the CoV-1 and CoV-2 spike proteins (A, B)** Side view of active and inactive states for CoV-1. The active protomer, A, is colored in red, protomer B is colored in cyan, and protomer C is colored in magenta Protomers B and C are represented transparent. **(C, D)** Protomer A is colored in orange, protomer B is colored in tan, and protomer C is colored in blue. Protomers B and C are represented transparent. **(E)** RBD-S2 center of mass distance vs. RBD$_A$-NTD$_A$ interdomain angle, each domain is represented by the roll axis (longest principle axis). Each point is the average value of the 20 copies for each image and the error bars are defined by the standard deviation of the 20 copies for each image. **(F)** The first two principle components of the protomer A Cα atoms.

The work that is presented here is primarily derived from an extensive set of SMwST simulations (data reported is generated from the last iteration of the final set of string method simulations) that employed string method using different definitions for the collective variables used in biasing the active protomer. The first set focused on biasing the atomic coordinates of the of Cα atoms that encompass the active protomer. Atomic coordinate [43] biasing has shown to be a useful application for SMwST simulations when characterizing transition pathways. The second set employed distance and orientation quaternion based collective variables in order to describe the conformational changes of the active protomer deactivation more appropriately. Supplementary figures 1-4

show the convergence of the two sets of string method simulations with collective variables that provide insight into the global conformational changes of the CoV-1 and CoV-2 spike protein's active protomer along the transition pathways.

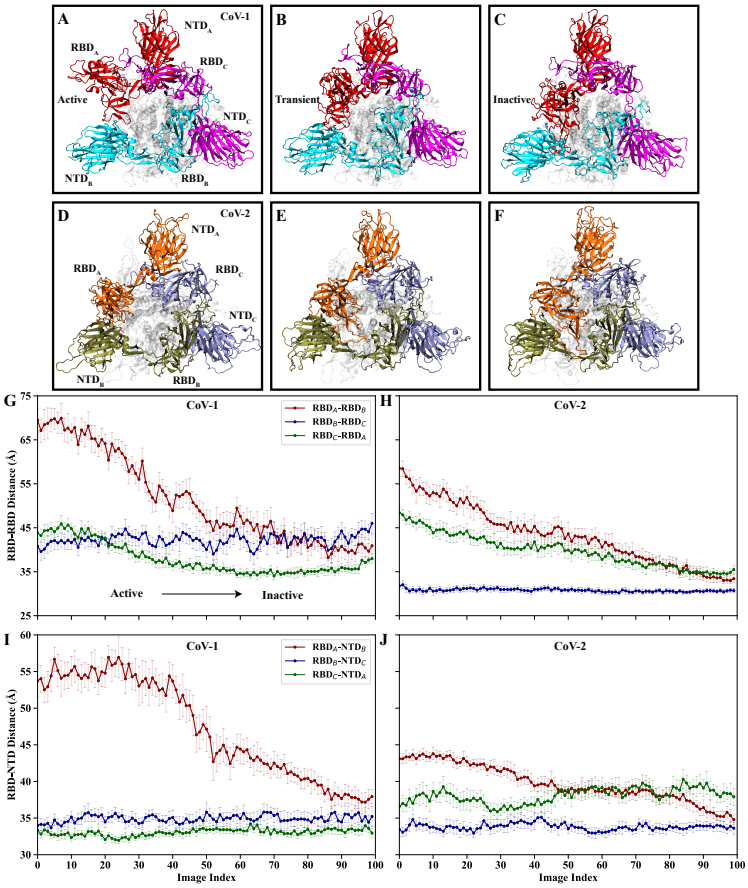

Figure 2: **Receptor binding domains show a differential dynamic behavior. (A-C)** Top view of active, transient, and inactive states for CoV-1. (D-F) Top view of active, transient, and inactive states for CoV-2. **(G, H)** RBD-RBD center of mass distances with respect to the image index. **(I, J)** RBD-NTD center of mass distance of each RBD and its adjacent NTD respect to the image index. Each point is the average value of the 20 copies for each image and the error bars are defined by the standard deviation of the 20 copies for each image.

Global protein conformational changes

With the transition pathways characterized, the global protein conformational changes along the deactivation pathways can be described. Figure 1 A-D, emphasizes the similarities and differences between two proteins when in the active and inactive conformations, with CoV-1 adopting a more "active" conformation (Fig.1A) as the RBD appears to be pointed straight up, while the RBD in

CoV-2 is angled downward in the direction toward NTD_B (Fig.1C). The inactive conformations (Fig.1B,D) show a noticeable difference in the RBD orientations with the RBD in CoV-1 being slightly tilted downwards and away from the adjacent RBDs and the RBD in CoV-2 tilted slightly upwards and towards the adjacent RBDs. Figure 1E shows the coupling between same protomer domains as the conformational changes is occurring along the deactivation pathway. The RBD in CoV-1 rotates away from the NTD before any change in the RBD-S2 distance occurs, the RBD-S2 distance remains constant at $\sim$93 Å until the angle between the two domains approaches closer to 90° and then the trend between the two calculations becomes quite linear. Unlike CoV-1, the transition of the active protomer is coupled linearly between the two RBD-S2 distance the RBD-NTD angle. The deactivation of the RBD in both proteins follows two different paths, the RBD in CoV-1 rotates farther from the NTD before transitioning closer to the S2 domain, while the RBD deactivation in CoV-2 is essentially negatively correlated between the two definitions. Furthermore, we have also monitored the degree of the ACE2 accessibility (Supplementary Fig.3), using a definition derived by Peng et al [54]. The relative degree of the ACE2 accessibility for both CoV-1 and CoV-2 is in line with the thresholds of 52.2°-84.8°, with ACE2 being completely inaccessible for images (windows) $>$50.

To further track the conformational changes of the CoV-1 and CoV-2 spike proteins, more insight was gained by projecting the Cα atoms of the active protomer, taken from the last frame of the SMwST simulation trajectories, onto its principle components. Figure 1F further illustrates how the two protein active protomers take two different pathways during active-to-inactive transition. PC1 and PC2 account for $\sim$83% of the variance with PC1 accounting for $\sim$63% of the variance. PC1 roughly describes the motions of the RBD rotating downward towards the inactive state, with slight deviations in both the NTD and S2 domains, and PC2 describes a rotational change of the RBD in the XY plane. Following along the pathways from active-to-inactive, both CoV-1 and 2 exhibit paths similar to those provided in Figure 1E. CoV-2 transitions linearly and CoV-1 undergoes a non-linear path in the PC space.

Active and inactive RBD conformations

The top views of the CoV-1 and CoV-2 spike proteins (Fig.2A-F) further exhibit the differences between the two ends states (Fig.1A-D) and provide insight into two example transient conformations (image 50) along the transition pathways. The RBD, when in the active conformation, CoV-1 is oriented equidistant between NTD_A and NTD_B, while RBD_C remains in close proximity of RBD_A (Fig.2A). The active RBD in CoV-2 is positioned closer to the adjacent protomer NTD_B (Fig.2D). The transient conformation for both CoV-1 and CoV-2 spike proteins (Fig2.B,E) provides insight into the differing orientations of RBD activation halfway through the transition pathway. Further inspection of the top views (Fig.2A-F) provide insight into how the inactive RBDs (protomers B and C) adopt different orientations as the active RBD transitions towards the inactive state. All three RBDs in CoV-2 begin to pack tighter together (Fig.2E-F) as the active RBD transitions towards being inactive. In CoV-1 the RBD's move away and position closer towards the each adjacent protomer NTD (Fig.2B,C), this is extenuated in (Fig.2C) as there is a gap between all three RBDs and much of the S2 domain is visible below (represented colorless and transparent). This is not the case for CoV-2 as all three RBDs are positioned close enough that the S2 domain is significantly less visible underneath.

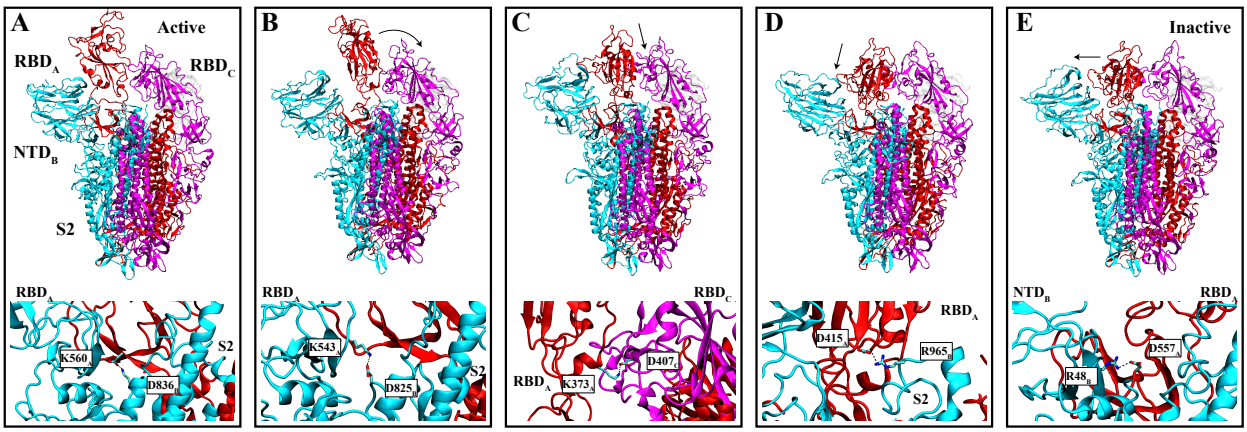

Figure 3: **Multi-step transition pathway of CoV-1 coupled through key salt-bridges (A-E)** Tracking the conformational changes of the CoV-1 spike protein along the transition pathway. Arrows point in the direction the active protomer RBD transitions along the pathway. Key salt-bridges are indicated below the protein snapshots that are formed along the transition pathway.

Figure 2 G-J tracks the inter-domain center of mass distances between the three RBD's and

the center of mass distances between each RBD and its adjacent protomer's NTD. The trends in the RBD-RBD inter-domain distances provide further indication that the receptor binding domains tend to remain in close proximity, amongst all three, in CoV-2 (Fig.2H) and the RBDs in CoV-1 (Fig.2G) transition closer to the RBD that is adjacent or "points to" when either transitioning towards inactive or residing in the inactive conformation. Similar to the trend that is observed in Figure 1E, the RBD-NTD distance remains constant across the first 40 images and then becomes linear as it is moving closer to NTD_B. The active RBD in CoV-1 moves closer to other neighboring domains (Fig.3) before transitioning closer towards NTD_B. The RBDs in CoV-1 tend to adopt conformations that remain pointing towards the adjacent RBD but lie closer in the proximity to the adjacent protomer's NTD (Fig.2A-C,I) unlike the RBDs in CoV-2, which remain in equal proximity to each of the adjacent domains (Fig.2D-F,J).

Coupling between global and local conformational changes

Figures 3,4 illustrate the coupling between global conformational changes of the active protomer and local conformational changes through the forming and breaking of key salt-bridge interactions between the active RBD and neighboring domains along the transition pathways. These events are used to characterize the transitions of the active protomers into a number of key events along the pathways. Supplementary Figures 2,3 provide the list of all key salt-bridge interactions, with respect to the image index, for both the CoV-1 and CoV-2 spike proteins.

The RBD in CoV-1 undergoes a rotational change along its Z axis (Fig.3A) and forms a salt-bridge between D560 in the linker region between RBD_A and the $S2_A$ and K836 of the S2 domain of protomer B. The formation of salt-bridge is quickly followed by the formation of a salt-bridge between K543 of RBD_A and D825 of $S2_B$ (Fig.3B). RBD_A continues deactivating and transitions closer to RBD_B (Fig.3C) and forms two salt-bridges. The first being, forms very briefly and breaks, K373 of RBD_A and D407 of RBD_C. The second salt-bridge forms between (Supporting Fig.2) D312 of RBD_A and K447 of RBD_C. The RBD of protomer A is following by a transition downwards towards the S2 domain and transitions in the direction towards protomer B (Fig.3D,E). A

salt-bridge is formed between K373 of RBD_A and E730 of $S2_B$ as the RBD continues transitioning towards inactive. The stabilization of the inactive state is required by the formation of three salt-bridges. The first, a brief forming transient salt-bridge, between RBD_A and the NTD and S2 domains of protomer B. These salt-bridges include D572 of the linker region between RBD_A and the $S2_A$ and R829 of $S2_B$, D415 of RBD_A and R965 $S2_B$, and D557 of RBD_A and R48 of NTD_B.

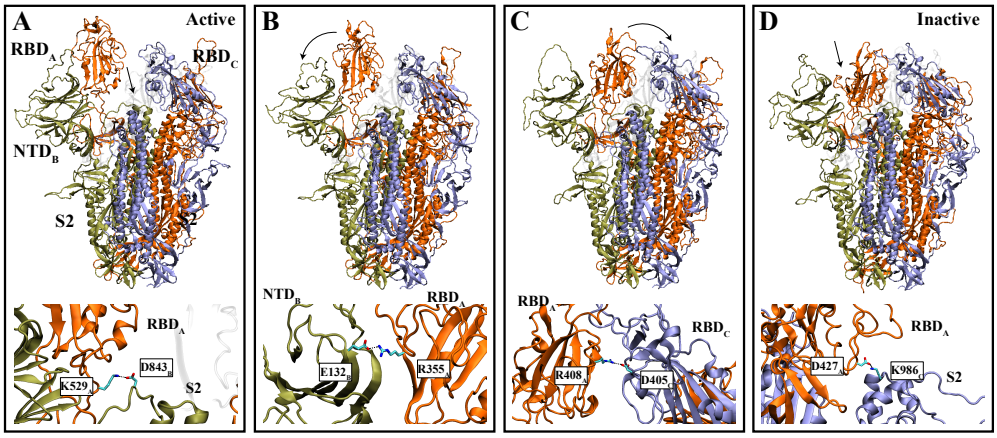

Figure 4: **Multi-step transition pathway of CoV-2 coupled through key salt-bridges (A-D)** Tracking the conformational changes of the CoV-2 spike protein along the transition pathway. Arrows point in the direction the active protomer RBD transitions along the pathway. Key salt-bridges are indicated below the protein snapshots that are formed along the transition pathway.

The deactivation of RBD_A in CoV-2 begins by transitioning downwards and forms a salt-bridge between K529 of RBD_A and D843 of $S2_B$ (Fig.4A). The formation of this salt-bridge is quickly followed by the formation of a salt-bridge between K528 of RBD_A and D839 of $S2_B$ (Supporting Fig.3). RBD_A transitions closer towards the adjacent NTD of protomer B (Fig.4 B) and forms a salt-bridge between R355 of RBD_A and E132 of NTD_B. Next, RBD_A transitions towards RBD_C and forms a salt-bridge between R408 of RBD_A and D405 of RBD_C. Finally, RBD_A continues transitioning towards and inactive and is stabilized by three salt-bridges, two between the RBD and NTD of protomer B, and third forming with the S2 of protomer C. The three formed salt-bridges include E484 of RBD_A and K528 of RBD_B, K462 of RBD_A and D198 NTD_B, and D427 of RBD_A and K986 of $S2_C$.

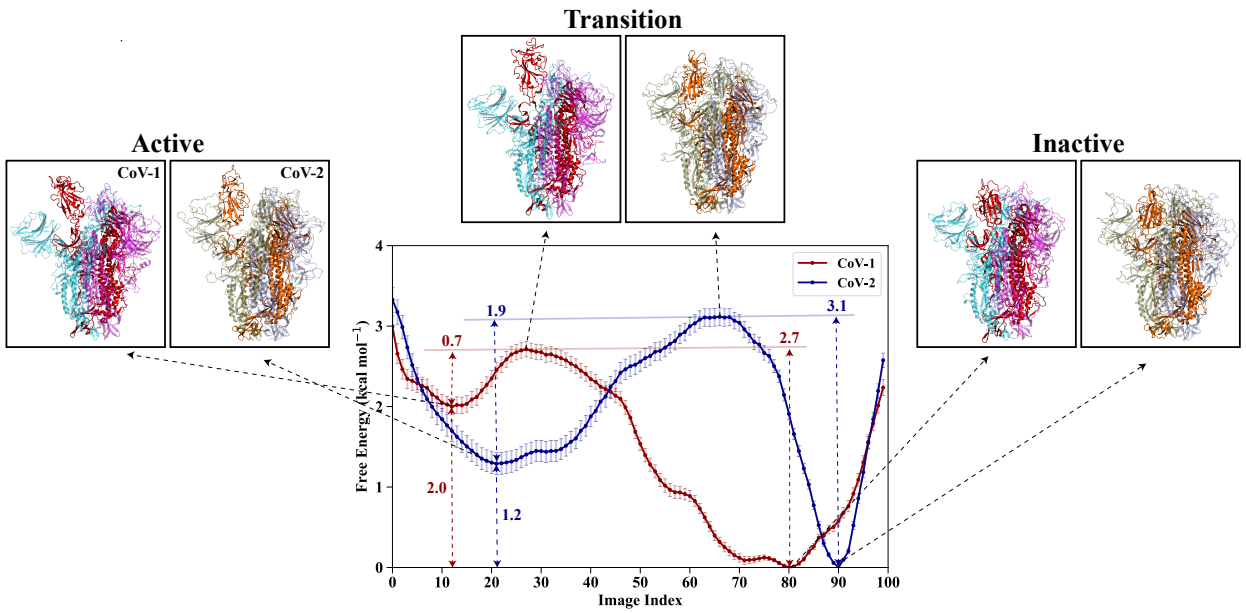

Figure 5: **Free energy profiles for the CoV-1 and CoV-2 spike proteins A** Free energy profile along the transition pathways. Each image (window) represents a particular conformation for CoV-1 and CoV-2 on the discretized pathways composed of 100 images each. The error bars represent the standard deviation based on Bayesian block bootstrapping. Specific conformations of the spike proteins are utilized to represent the two minima (local and global) and the transition states. Protomers B and C are represented transparent in both proteins.

Transition path free energy calculations

We have computed the free energy profiles along the transition pathways, derived from the second set of the SMwST simulations, for both CoV-1 and CoV-2 spike proteins and utilized the same reaction coordinates described by the same collective variable space as indicated in the methods section. The free energy profiles were computed by excluding the first 5 ns, equilibrium phase, and binned the last 20 ns into 4-5 ns bins for bootstrap re-sampling. The constructed free energy profiles (Fig.5) with respect to the image index (includes snapshots of the spike proteins at the associated image index, representing the minima and transition states), indicates a higher free energy barrier for Cov-2, with height 3.1 kcal.mol^{-1}. The profile also contains a broad free energy well located in the region of active conformations, composed of images 5-40 and local minima with a free energy of 1.2 kcal.mol^{-1}. The global minimum resides with an inactive conformation at image 90. CoV-1 also has well formed, local minima, along the active conformations, images 2-20, with a free energy of 2.0 kcal.mol^{-1}. The well is followed by a lower free energy barrier that CoV-2

at 2.7 kcal.mol^{-1}. The global minimum for CoV-1 is located at image 80. Interestingly, a broad free energy well is present in CoV-1, similar to CoV-2, which resides in conformations associated with being in the inactive conformation. The presence of these two very broad free energy wells are described by a higher contribution of conformational entropy, and thus, the proteins can easily occupy or transition between any one of the states that encompass these free energy wells. These results are in line with our previous work [35] with CoV-1 exhibiting a free energy profile that indicates a low free energy barrier when transitioning towards the inactive conformation and CoV-2 is hindered by higher free energy barriers when either transitioning towards active or inactive.

Conclusion

By employing a rigorous computational study, we have demonstrated that the SARS-CoV-1 and SARS-CoV-2 spike proteins exhibit differential dynamic behavior by characterizing the large-scale conformational changes associated with the active-to-inactive transition pathways using biased molecular dynamics simulations. More specifically, the transition pathways were first determined through the use of SMD, further refined by employing two sets of SMwST, and effectively sampled the transition pathways using BEUS to determine the associated free energy barriers along the pathways. The amount of detail provided here will allow for the testing of various experiments and even the trapping of specific intermediate states along the transition pathways.

Several experiments could be performed in order to test the results presented in our computational study. For instance, the importance certain residues play in forming key salt-bridge interactions between the active protomer RBD and adjacent domains, for both CoV-1 and CoV-2, could be investigated via site-directed mutagenesis. This could shed light on some additional insights on the conformational dynamics of the CoV-1 and CoV-2 spike proteins. Additionally, smFRET experiments could be used to investigate potential, RBD-RBD, RBD-NTD, and RBD-S2 interactions by measuring the distance between fluorophores attached at specific points in each domain. Disulfide cross-linking experiments could also be used to investigate residues in the NTD, RBD, and S2 that potentially interact with each other.

MD simulations provide atomic-level elucidation of the dynamic behavior of proteins [18, 17] and other biomolecules. Here, we have performed simulations using the non-glycosylated spike proteins of CoV-1 and CoV-2. A recent study has shown that glycosylation of the spike proteins might play an important role in the conformational dynamics of the RBD [55, 56]. However there exists the difficulty in determining whether conformational changes occur as a result of the intrinsic protein dynamics or the differential glycosylation patterns of the CoV-1 and CoV-2 spike proteins imposed by modeling. Admittedly, it is advantageous to employ the same protocol described in this work using glycoslated models, as much is still unknown about the large-scale confromational changes that connect to the states between the active and inactive proteins.

Our results for non-glycosylated SARS-CoV-2 spike protein somewhat resembles that of the glycosylated SARS-CoV-2 spike protein, recently reported by Pang et al. [57] (higher and broader free energy well for the open state) but the free energy values are scaled down by a factor of $\sim$3 due to the absence of glycan chains [57]. Similar reduction in the free energies was observed when the glycans were removed; however, the free energy difference between the open and closed states also disappears in the absence of the glycan chains [57]. We believe since we have used transition path finding (string method); our results are perhaps more reliable than those obtained from direct BEUS simulations without any path optimization. Another possibility to explain the difference is that free energy values are generally collective variable dependent and we have not used the same collective variables as Pang et al. [57].

As discussed previously, our study primarily sheds light on the conformational dynamics of the SARS-CoV-1 and SARS-CoV-2 spike proteins. While differences in the dynamic behavior of these spike proteins almost certainly contribute to differences in transmissibility and infectivity, factors such as spike protein glycosylation and the behavior of other viral proteins also need to be considered in order to provide a more complete hypothesis. Additional experimental and computational studies are thus needed to fully investigate the differential infectivity and transmissibility of SARS-CoV-1 and SARS-CoV-2. Our simulations provide valuable insight into the dynamic behavior of the CoV-1 and CoV-2 spike proteins when transitioning between active-to-inactive.

Employing the simulation protocol presented here, in the reverse direction: inactive-to-active, is required to determine the reversibility of the characterized transition pathways (i.e. global and local conformational changes and location of free energy barriers). An improved understanding of the conformational changes regarding activation or inactivation of the spike proteins, is critical to the effective development of novel therapeutics and vaccines using a structure-based drug design framework.

References

[1] M. A. Tortorici and D. Veesler, "Structural insights into coronavirus entry," in *Advances in Virus Research*, vol. 105, pp. 93–116, Academic Press Inc., 2019.

[2] A. C. Walls, X. Xiong, Y. J. Park, M. A. Tortorici, J. Snijder, J. Quispe, E. Cameroni, R. Gopal, M. Dai, A. Lanzavecchia, M. Zambon, F. A. Rey, D. Corti, and D. Veesler, "Unexpected Receptor Functional Mimicry Elucidates Activation of Coronavirus Fusion," *Cell*, vol. 176, no. 5, pp. 1026–1039.e15, 2019.

[3] X. Ou, Y. Liu, X. Lei, P. Li, D. Mi, L. Ren, L. Guo, R. Guo, T. Chen, J. Hu, Z. Xiang, Z. Mu, X. Chen, J. Chen, K. Hu, Q. Jin, J. Wang, and Z. Qian, "Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV," *Nature Communications*, vol. 11, no. 1, 2020.

[4] W. Tai, L. He, X. Zhang, J. Pu, D. Voronin, S. Jiang, Y. Zhou, and L. Du, "Characterization of the receptor-binding domain (RBD) of 2019 novel coronavirus: implication for development of RBD protein as a viral attachment inhibitor and vaccine," *Cellular and Molecular Immunology*, vol. 17, no. 6, pp. 613–620, 2020.

[5] P. Zhou, X. L. Yang, X. G. Wang, B. Hu, L. Zhang, W. Zhang, H. R. Si, Y. Zhu, B. Li, C. L. Huang, H. D. Chen, J. Chen, Y. Luo, H. Guo, R. D. Jiang, M. Q. Liu, Y. Chen, X. R. Shen, X. Wang, X. S. Zheng, K. Zhao, Q. J. Chen, F. Deng, L. L. Liu, B. Yan, F. X. Zhan, Y. Y. Wang, G. F. Xiao, and Z. L. Shi, "A pneumonia outbreak associated with a new coronavirus of probable bat origin," *Nature*, vol. 579, no. 7798, pp. 270–273, 2020.

[6] R. Lu, X. Zhao, J. Li, P. Niu, B. Yang, H. Wu, W. Wang, H. Song, B. Huang, N. Zhu, Y. Bi, X. Ma, F. Zhan, L. Wang, T. Hu, H. Zhou, Z. Hu, W. Zhou, L. Zhao, J. Chen, Y. Meng, J. Wang, Y. Lin, J. Yuan, Z. Xie, J. Ma, W. J. Liu, D. Wang, W. Xu, E. C. Holmes, G. F. Gao, G. Wu, W. Chen, W. Shi, and W. Tan, "Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding," *The Lancet*, vol. 395, no. 10224, pp. 565–574, 2020.

[7] J. F. W. Chan, S. Yuan, K. H. Kok, K. K. W. To, H. Chu, J. Yang, F. Xing, J. Liu, C. C. Y. Yip, R. W. S. Poon, H. W. Tsoi, S. K. F. Lo, K. H. Chan, V. K. M. Poon, W. M. Chan, J. D. Ip, J. P. Cai, V. C. C. Cheng, H. Chen, C. K. M. Hui, and K. Y. Yuen, "A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster," *The Lancet*, vol. 395, no. 10223, pp. 514–523, 2020.

[8] C. Huang, Y. Wang, X. Li, L. Ren, J. Zhao, Y. Hu, L. Zhang, G. Fan, J. Xu, X. Gu, Z. Cheng, T. Yu, J. Xia, Y. Wei, W. Wu, X. Xie, W. Yin, H. Li, M. Liu, Y. Xiao, H. Gao, L. Guo, J. Xie, G. Wang, R. Jiang, Z. Gao, Q. Jin, J. Wang, and B. Cao, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *The Lancet*, vol. 395, no. 10223, pp. 497–506, 2020.

[9] N. Zhu, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G. F. Gao, and W. Tan, "A novel coronavirus from patients with pneumonia in China, 2019," *New England Journal of Medicine*, vol. 382, no. 8, pp. 727–733, 2020.

[10] A. C. Walls, Y.-J. Park, M. A. Tortorici, A. Wall, A. T. McGuire, and D. Veesler, "Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein," *Cell*, vol. 181, no. 2, pp. 281 – 292.e6, 2020.

[11] E. Petersen, M. Koopmans, U. Go, D. H. Hamer, N. Petrosillo, F. Castelli, M. Storgaard, S. Al Khalili, and L. Simonsen, "Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics," *The Lancet Infectious Diseases*, vol. 20, pp. e238–e244, 2020.

[12] D. Wrapp, N. Wang, K. S. Corbett, J. A. Goldsmith, C. L. Hsieh, O. Abiona, B. S. Graham, and J. S. McLellan, "Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation," *Science*, vol. 367, no. 6483, pp. 1260–1263, 2020.

[13] Y. Yuan, D. Cao, Y. Zhang, J. Ma, J. Qi, Q. Wang, G. Lu, Y. Wu, J. Yan, Y. Shi, X. Zhang, and G. F. Gao, "Cryo-EM structures of MERS-CoV and SARS-CoV spike glycoproteins reveal the dynamic receptor binding domains," *Nature Communications*, vol. 8, no. 1, p. 15092, 2017.

[14] C. Xu, Y. Wang, C. Liu, C. Zhang, W. Han, X. Hong, Y. Wang, Q. Hong, S. Wang, Q. Zhao, Y. Wang, Y. Yang, K. Chen, W. Zheng, L. Kong, F. Wang, Q. Zuo, Z. Huang, and Y. Cong, "Conformational dynamics of SARS-CoV-2 trimeric spike glycoprotein in complex with receptor ACE2 revealed by cryo-EM," *Science Advances*, vol. 7, no. 1, p. eabe5575, 2021.

[15] W. Song, M. Gui, X. Wang, and Y. Xiang, "Cryo-EM structure of the SARS coronavirus spike glycoprotein in complex with its host cell receptor ACE2," *PLoS Pathogens*, vol. 14, no. 8, p. e1007236, 2018.

[16] A. Punjani and D. J. Fleet, "3D Variability Analysis: Resolving continuous flexibility and discrete heterogeneity from single particle cryo-EM. https://www.biorxiv.org/content/early/2021/01/13/2020.04.08.032466," *bioRxiv*, 2021.

[17] M. Torrens-Fontanals, T. M. Stepniewski, D. Aranda-García, A. Morales-Pastor, B. Medel-Lacruz, and J. Selent, "How do molecular dynamics data complement static structural data of GPCRs," *International Journal of Molecular Sciences*, vol. 21, no. 16, p. 5933, 2020.

[18] A. Srivastava, T. Nagai, A. Srivastava, O. Miyashita, and F. Tama, "Role of computational methods in going beyond x-ray crystallography to explore protein structure and dynamics," *International Journal of Molecular Sciences*, vol. 19, no. 11, p. 3401, 2018.

[19] M. Lu, P. D. Uchil, W. Li, D. Zheng, D. S. Terry, J. Gorman, W. Shi, B. Zhang, T. Zhou, S. Ding, R. Gasser, J. Prévost, G. Beaudoin-Bussières, S. P. Anand, A. Laumaea, J. R. Grover, L. Liu, D. D. Ho, J. R. Mascola, A. Finzi, P. D. Kwong, S. C. Blanchard, and W. Mothes, "Real-Time Conformational Dynamics of SARS-CoV-2 Spikes on Virus Particles," *Cell Host and Microbe*, vol. 28, no. 6, pp. 880–891.e8, 2020.

[20] J. Li, P.-C. Wen, M. Moradi, and E. Tajkhorshid, "Mechanistic view of membrane transporters through computational microscopy of molecular dynamics," *Curr. Opin. Struct. Biol.*, vol. 31, pp. 96–105, 2015. invited.

[21] J. Gumbart, M. C. Wiener, and E. Tajkhorshid, "Mechanics of force propagation in TonB-dependent outer membrane transport," vol. 93, pp. 496–504, 2007.

[22] P.-C. Wen and E. Tajkhorshid, "Dimer opening of the nucleotide binding domains of ABC transporters after ATP hydrolysis," *Biophys. J.*, vol. 95, no. 11, pp. 5100–5110, 2008.

[23] Z. Huang and E. Tajkhorshid, "Dynamics of the extracellular gate and ion-substrate coupling in the glutamate transporter," vol. 95, pp. 2292–2300, 2008.

[24] C. J. Law, G. Enkavi, D.-N. Wang, and E. Tajkhorshid, "Structural basis of substrate selectivity in the glycerol-3-phosphate:phosphate antiporter GlpT," vol. 97, pp. 1346–1353, 2009.

[25] J. Gumbart, M. C. Wiener, and E. Tajkhorshid, "Coupling of calcium and substrate binding through loop alignment in the outer membrane transporter BtuB," vol. 393, pp. 1129–1142, 2009.

[26] Z. Huang and E. Tajkhorshid, "Identification of the third Na^+ site and the sequence of extracellular binding events in the glutamate transporter," vol. 99, pp. 1416–1425, 2010.

[27] G. Enkavi and E. Tajkhorshid, "Simulation of spontaneous substrate binding revealing the binding pathway and mechanism and initial conformational response of GlpT," vol. 49, pp. 1105–1114, 2010.

[28] P.-C. Wen and E. Tajkhorshid, "Conformational coupling of the nucleotide-binding and the transmembrane domains in the maltose ABC transporter," vol. 101, no. 3, pp. 680–690, 2011.

[29] P.-C. Wen, B. Verhalen, S. Wilkens, H. S. Mchaourab, and E. Tajkhorshid, "On the origin of large flexibility of P-glycoprotein in the inward-facing state," vol. 288, no. 26, pp. 19211–19220, 2013.

[30] S. A. Shaikh, J. Li, G. Enkavi, P. Wen, Z. Huang, and E. Tajkhorshid, "Visualizing functional motions of membrane transporters with molecular dynamics simulations," vol. 52, no. 4, pp. 569–587, 2013.

[31] M. Moradi and E. Tajkhorshid, "Mechanistic picture for conformational transition of a membrane transporter at atomic resolution.," vol. 110, no. 47, pp. 18916–18921, 2013.

[32] M. Moradi and E. Tajkhorshid, "Computational recipe for efficient description of large-scale conformational changes in biomolecular systems," vol. 10, no. 7, pp. 2866–2880, 2014.

[33] M. Moradi, G. Enkavi, and E. Tajkhorshid, "Atomic-level characterization of transport cycle thermodynamics in the glycerol-3-phosphate:phosphate transporter," *Nat. Commun.*, vol. 6, p. 8393, 2015.

[34] A. Singharoy, C. Chipot, M. Moradi, and K. Schulten, "Chemomechanical coupling in hexameric protein–protein interfaces harness energy within V-type ATPases," vol. 139, pp. 293—-310, 2017.

[35] V. Govind Kumar, D. S. Ogden, U. H. Isu, A. Polasa, J. Losey, and M. Moradi, "Prefusion spike protein conformational changes are slower in sars-cov-2 than in sars-cov-1," *Journal of Biological Chemistry*, vol. 298, p. 101814, 2022.

[36] S. Jo, T. Kim, V. G. Iyer, and W. Im, "CHARMM-GUI: A web-based graphical user interface for CHARMM," *J. Comp. Chem.*, vol. 29, pp. 1859–1865, 2008.

[37] J. Lee, X. Cheng, J. M. Swails, M. S. Yeom, P. K. Eastman, J. A. Lemkul, S. Wei, J. Buckner, J. C. Jeong, Y. Qi, S. Jo, V. S. Pande, D. A. Case, I. Charles L. Brooks, J. Alexander D. MacKerell, J. B. Klauda, and W. Im, "CHARMM-GUI input generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM simulations using the CHARMM36 additive force field," vol. 12, no. 1, pp. 405–413, 2016.

[38] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, "Scalable molecular dynamics with NAMD," vol. 26, pp. 1781–1802, 2005.

[39] J. Huang, S. Rauscher, G. Nawrocki, T. Ran, M. Feig, B. L. de Groot, H. Grubmüller, and A. D. MacKerell Jr, "CHARMM36m: an improved force field for folded and intrinsically disordered proteins," *Nature methods*, vol. 14, pp. 71–73, jan 2017.

[40] J. K. Reid, "On the method of conjugate gradients for the solution of large sparse systems of linear equations," in *Large Sparse Sets of Linear Equations* (J. K. Reid, ed.), pp. 231–254, London: Academic Press, 1971.

[41] G. J. Martyna, D. J. Tobias, and M. L. Klein, "Constant pressure molecular dynamics algorithms," *J. Chem. Phys.*, vol. 101, pp. 4177–4189, 1994.

[42] T. Darden, D. York, and L. Pedersen, "Particle mesh Ewald: An $N\cdot\log(N)$ method for Ewald sums in large systems," *J. Chem. Phys.*, vol. 98, pp. 10089–10092, 1993.

[43] H. Chen, D. Ogden, S. Pant, W. Cai, E. Tajkhorshid, M. Moradi, B. Roux, and C. Chipot, "A Companion Guide to the String Method with Swarms of Trajectories: Characterization, Performance, and Pitfalls," *Journal of Chemical Theory and Computation*, vol. 18, pp. 1406–1422, mar 2022.

[44] E. A. Coutsias, C. Seok, and K. A. Dill, "Using quaternions to calculate RMSD," *J. Chem. Phys.*, vol. 25, no. 15, pp. 1849–1857, 2004.

[45] B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *J. Opt. Soc. Am.Ã*, vol. 4, no. 4, pp. 629–642, 1987.

[46] G. Fiorin, M. L. Klein, and J. Hénin, "Using collective variables to drive molecular dynamics simulations," *Mol. Phys.*, vol. 111, no. 22–23, pp. 3345–3362, 2013.

[47] J. Pan, S. Tristram-Nagle, N. Kucerka, and J. F. Nagle, "Temperature dependence of structure, bending rigidity, and bilayer interactions of dioleoylphosphatidylcholine bilayers," *Biophys. J.*, vol. 94, pp. 117–124, 2008.

[48] M. E. Johnson and G. Hummer, "Characterization of a dynamic string method for the construction of transition pathways in molecular reactions," *J. Phys. Chem. B*, vol. 116, 2012.

[49] A. Fakharzadeh and M. Moradi, "Effective Riemannian diffusion model for conformational dynamics of biomolecular systems," vol. 7, pp. 4980–4987, 2016.

[50] G. M. Torrie and J. P. Valleau, "Nonphysical sampling distributions in monte carlo free-energy estimation: Umbrella sampling," *J. Comput. Phys.*, vol. 23, p. 187, 1977.

[51] S. Kumar, D. Bouzida, R. H. Swendsen, P. A. Kollman, and J. M. Rosenberg, "The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method," *J. Comp. Chem.*, vol. 13, pp. 1011–1021, 1992.

[52] C. Bartels, "Analyzing biased monte carlo and molecular dynamics simulations," *Chem. Phys. Lett.*, vol. 331, p. 446, 2000.

[53] M. R. Shirts and J. D. Chodera, "Statistically optimal analysis of samples from multiple equilibrium states," *J. Chem. Phys.*, vol. 129, p. 124105, 2008.

[54] Y. S. X. W. K. M. Y. Y. X. Z. Z. X. Cheng Peng, Zhengdan Zhu and W. Zhu, "Computational insights into the conformational accessibility and binding strength of sars-cov-2 spike protein to human angiotensin-converting enzyme 2," *The Journal of Physical Chemistry Letters*, vol. 11, pp. 10482–10488, 2020.

[55] L. Casalino, Z. Gaieb, J. A. Goldsmith, C. K. Hjorth, A. C. Dommer, A. M. Harbison, C. A. Fogarty, E. P. Barros, B. C. Taylor, J. S. McLellan, E. Fadda, and R. E. Amaro, "Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein," *ACS Central Science*, vol. 6, no. 10, pp. 1722–1734, 2020.

[56] T. Sztain, S.-H. Ahn, A. T. Bogetti, L. Casalino, J. A. Goldsmith, R. S. McCool, F. L. Kearns, J. Andrew McCammon, J. S. McLellan, and R. E. Amaro, "A glycan gate controls opening of the SARS-CoV-2 spike protein. https://www.biorxiv.org/content/10.1101/2021.02.15.431212v1," *bioRxiv*, 2021.

[57] D. L. L. A. P. Yui Tik Pang, Atanu Acharya and J. C. Gumbart, "Sars-cov-2 spike opening dynamics and energetics reveal the individual roles of glycans and their collective impact," *Communications Biology*, vol. 5, 2022.

Conformational Transition Pathways in Major Facilitator Superfamily Transporters

Dylan Ogden[1], Kalyan Immadisetty[1], and Mahmoud Moradi[1] *

[1]Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, AR 72701

Abstract

Major facilitator superfamily (MFS) of transporters consists of three classes of membrane transporters: symporters, uniporters, and antiporters. Despite such diverse functions, MFS transporters are believed to undergo similar conformational changes within their distinct transport cycles. While the similarities between conformational changes are noteworthy, the differences are also important since they could potentially explain the distinct functions of symporters, uniporters, and antiporters of MFS superfamily. We have performed a variety of equilibrium, non-equilibrium, biased, and unbiased all-atom molecular dynamics (MD) simulations of bacterial proton-coupled oligopeptide transporter GkPOT, glucose transporter 1 (GluT1), and glycerol-3-phosphate transporter (GlpT) to compare the similarities and differences of the conformational dynamics of three different classes of transporters. Here we have simulated the *apo* protein in an explicit membrane environment. Our results suggest a very similar conformational transition involving interbundle salt-bridge formation/disruption coupled with the orientation changes of transmembrane (TM) helices, specifically H1/H7 and H5/H11, resulting in an alternation in the accessibility of water at the cyto- and periplasmic gates

Introduction

Membrane transporters facilitate the exchange of materials across lipid bilayers. Among these transporters, the major facilitator superfamily (MFS) is the largest family of secondary membrane transporters, comprised by more than 10,000 members [1, 2]. Examples of MFS transporters that have been studied structurally include the lactose permease (LacY) from *Escherichia coli* (a sugerporter) [3, 4, 5], the glycerol-3-phosphate transporter (GlpT) [6], xylose transporter(XylE) [7], and the multidrug transporter EmrD [8].

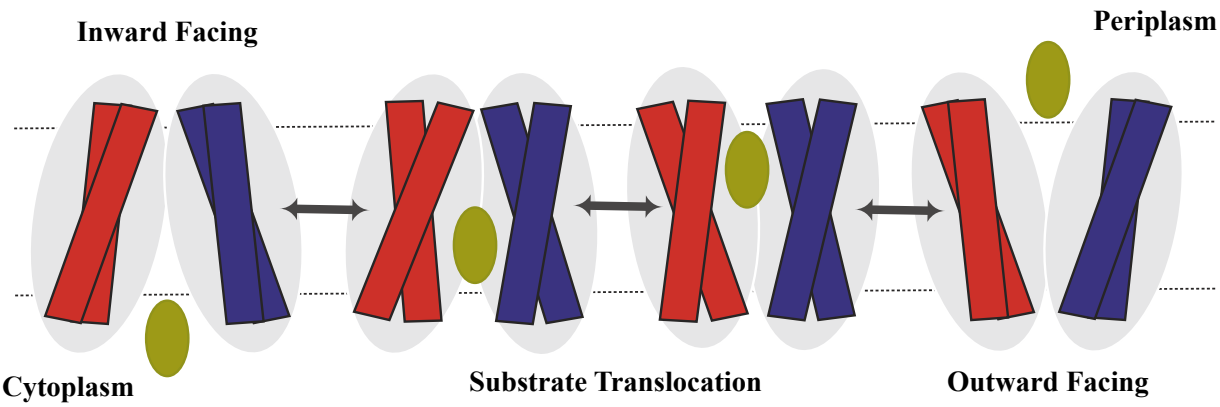

Figure 1: Schematic representation of the MFS facillitative diffusion of a substrate across cell membrane.

All MFS transporters share a common 12 TM helical structure, consisting of an N- and a C-bundle domain, displaying a twofold pseudo-symmetry [9]. MFS transporters function in a number of different ways including active and passive transport. The latter function is uniport (transport of only one molecule in one direction) and the former is either symport or antiport [6, 10, 9] (transport of two distinct molecules in the same direction and transport of two distinct molecules in opposite directions). Despite these distinct functions, MFS transporters all share a common overall mode of function known as the "alternating access" mechanism, which is shared with other membrane transporters [1, 9, 10]. According to this mechanism, the binding site is never exposed to both sides of membrane at the same time; instead, the protein alternates between an inward- (IF) and an outward-facing (OF) conformation. A more specific model of alternating access in MFS transporters has been proposed, which is known as the "rocker-switch" mechanism [6]. In order for transport to ensue, the N- and C-bundle domains undergo concerted conformational changes, resembling a rocker-switch-like mode of movement that exposes the binding site to the two sides of membrane alternatively and provides a mechanism for substrate translocation (Figure 1). In addition to the IF and OF states, the transport mechanism also involves occluded intermediates, where the binding site is exposed to neither side of the membrane.

Structural similarities of the MFS transporters suggest that these transporters may undergo similar conformational changes during their inter-conversions between the IF and OF states; although

the coupling of these conformational changes to different binding/unbinding events is likely to be quite different in various symporters, uniporters, and antiporters. In order to study the similarities and differences of the structural changes in three different classes of MFS transporters, here we present a comparative view of IF-OF conformational changes of three proteins from three different classes of MFS transporters, including a bacterial proton-coupled oligopeptide transporter, namely GkPOT [11, 12], the human glucose transporter 1 (GluT1) [13, 14], and the bacterial glycerol-3-phosphate transporter (GlpT) [6]. We specifically limit the discussion to the *apo* state of these proteins; however, we are aware that a full picture can only be provided if the full transport cycle is simulated. This includes the binding, unbinding, and translocation of substrates and their cotransported species. The study of *apo* protein IF-OF transition is the first step in characterizing the structural changes of MFS transporters within their transport cycle, which is the subject of the current study.

GluT1, a uniporter belonging to the sugar porter subfamily of MFS transporters, mediates the uptake of glucose by passively transporting it along its concentration gradient [13]. In GluT1, the N and C domains are connected by an intracellular helical bundle (ICH) which comprises four short α-helices. The ICH domain is observed in the structures of other sugar porter members, such as XylE, but not in other MFS transporters, suggesting that the ICH may be a unique feature of the sugar porter subfamily. The ICH in sugar porter members acts as "latch" to secure closure of the intracellular gate in the OF conformation [14]. Although GluT1 is a passive transporter, recent structural studies suggest that the transport mechanism follows an alternating access mode of function similar to active membrane transporters [14, 15, 16].

Proton-coupled oligopeptide transporters (POTs) are among the symporter members of the MFS superfamily. POTs uptake small peptides and peptide like molecules to the cell across the cell membrane using the inward directed proton electrochemical gradient as the source of energy for their active symport function [17, 18]. A very important feature among POTs is substrate promiscuity [19] attributed to the binding site accommodating a range of peptides and peptide like molecules in multiple orientations [20]. Mammalian POTs are yet to be crystallized; how-

ever, several bacterial POT members have recently been crystallized including $PepT_{so}$ [21, 20], $PepT_{st}$ [11, 22], $PepT_{so2}$ [12], and GkPOT [18]. Particularly, the POT transporter from *Geobacillus kaustophilus* bacterium (GkPOT) has been crystallized in a lipidic environment to a remarkably high resolution of 1.9 Å [18].

GlpT is an antiporter member of the MFS superfamily that uptakes glycerol-3-phosphate using the outward directed inorganic phosphate concentration gradient as the source of energy for its active antiport activity [6, 23]. The binding of inorganic phosphate facilitates the transition from the IF-OF states, and the replacement of glycerol-3-phosphate reverts the protein back to the IF conformation [24, 25].

Molecular dynamics simulations have been a useful tool in the extensive study of proteins and membrane transporters such as GkPOT [26], GluT1 [27], and GlpT [28]. Unfortunately, due to the short timescales of typical unbiased molecular dynamics simulations, this technique is often incapable of capturing large-scale conformational changes. To be able to sample such conformations, enhanced sampling techniques are typically sought after, due to their ability to capture longer timescales. These include simulations such as steered molecular dynamics (SMD) [29] or targeted molecular dynamics (TMD) [30]. Admittedly, the transition between the IF and OF conformation may be sampled using unbiased MD for certain proteins. Recent work has shown a complete transition of GluT1 in 1-1.5 μs [27]. For slower IF-OF transitions, we have recently proposed a computational recipe that allows for reconstructing the IF-OF transition of various timescales [31, 32, 28]. We have recently improved this protocol by employing a Riemannian description of conformational landscape of protein [33]. The conformational transition pathways of GlpT has already been reported previously for the *apo* and phosphate-bound proteins [28]. Here we compare the IF-OF conformational transition pathway of *apo* GkPOT and GluT1 to that of GlpT and discuss the similarities and differences of these pathways in detail.

Theoretical Methods

Characterizing structural transitions of membrane transporters without compromising the detailed chemical description of these systems and their environments is computationally challenging mainly due to the prohibitively long timescales involved in such processes. Here we have employed all-atom MD to study the IF-OF conformational transitions of MFS transporters in the *apo* state. The GlpT simulations were previously reported [28] and they are only included here for comparison. We were able to capture the GluT1 IF→OF transition using unbiased equilibrium MD; however, the GkPOT IF→OF transition was not captured using such simulations [26]. We therefore used the Riemannian [33] version of the protocols developed for enhanced sampling simulations of GlpT [28] to capture the IF→OF transition in *apo* GkPOT.

In each case, we used a membrane-embedded model of the wild-types *apo* protein (GkPOT [26], GluT1, and GlpT [28]) in its respective IF state based on the available crystal structure of protein (PDB: 4IKV [18], 4PYP [14], and 1PW4 [25], for GkPOT, GluT1, and GlpT, respectively). GkPOT and GlpT were placed in the 1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphoethanolamine (POPE) lipids and all the crystal structure waters were removed. The protocols for generating these models have been previously reported for GlpT [28] and GkPOT [26] in detail. For GluT1, we used CHARMM-GUI [34, 35] for building the simulation system. The THR 45 mutated residue in GluT1 crystal stricture (PDB: 4PYP) [14] was substituted by the wild-type amino acid and GluT1 was placed in a lipid bilayer of 1-Palmitoyl-2-oleoyl-sn-glycero-3-phosphocholine (POPC), solvated with TIP3P waters [36], ionized with 0.15 M KCl, and placed in a box of $\approx 106 \times 105 \times 104$ Å^3.

We used CHARMM36 all-atom additive force field to describe all molecules [37, 38]. All MD simulations were performed with NAMD 2.11 [39]. Prior to equilibrium runs, each system was energy minimized for 10,000 steps using conjugate gradient algorithm [40] and relaxed using a $\sim$1 ns multi-step restraining procedure (CHARMM-GUI's default procedure for membrane proteins [34]). This initial relaxation was performed in an NVT ensemble while all production runs were performed in an NPT ensemble. Simulations were carried out using a 2 fs timestep at 310 K using a Langevin integrator with a damping coefficient of $\gamma = 0.5\,\mathrm{ps}^{-1}$. The pressure was

maintained at 1 atm using the Nosé-Hoover Langevin piston method [41, 42]. The smoothed cutoff distance for non-bonded interactions was set to $10-12$ Å, and long-range electrostatic interactions were computed with the particle mesh Ewald (PME) method [43] and periodic boundary conditions were used in all simulations. We performed two sets (replicas) of equilibrium MD simulations for GluT1. Each replica was equilibrated for 700 ns. The simulation times employed were long enough to capture the IF→OF transition in the *apo* GluT1.

For GkPOT, the starting point for the simulations presented here was the equilibrated model from the unbiased simulations reported previously for the *apo* GkPOT [26]. We then performed SMD simulations using various biasing protocols, described in detail in Supporting Information. Here we only present the results of the most successful protocol, which uses a biasing potential based on the orientation of TM helices H1, H2, H7, and H8 to induce the IF→OF transition of *apo* GkPOT. The simulation time for this SMD protocol was 100 ns. During this simulation, a harmonic restraint was imposed on the orientation quaternions of the aforementioned helices with a varying harmonic center starting from the orientation quaternions of the initial conformation of the SMD simulation and changing towards those of a target structure that was built based on a homology model of MFS protein LacY, whose crystal structure is in the OF state (PDB: 3O7Q) [44]. The biasing potential was $\frac{1}{2}k\Omega^2$, where $k = 10,000\, kcal/(mol\mathring{A}^2)$ is the harmonic constant and Ω is the geodesic distance between the current orientation quaternion and a nonlinear interpolation of the initial and target orientation quaternions based on the Riemannian geometry. The outcome of the SMD simulation was equilibrated again for approximately 200 ns to determine the stability of the generated OF state of GkPOT. Prior to this follow-up equilibrium simulation, the final harmonic restraint on orientation quaternions were released gradually within a 15-ns SMD simulation with a fixed center an a variable force constant from 10,000 to 0 $kcal/(mol\mathring{A}^2)$.

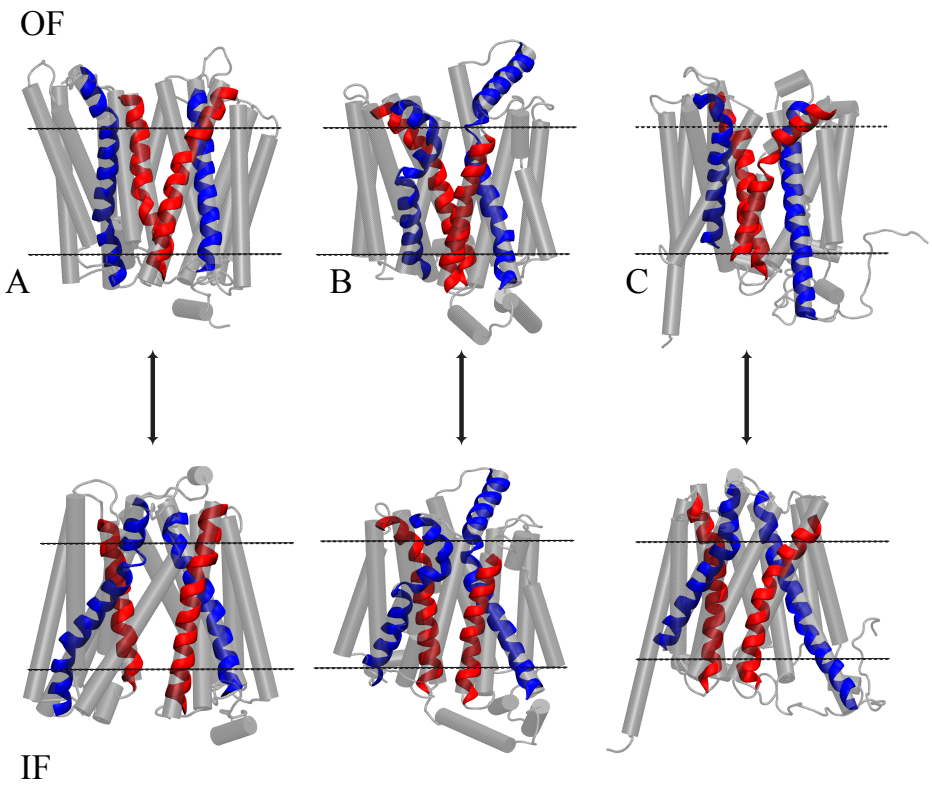

Figure 2: MFS trasnporters GkPOT (A), GluT1 (B), and GlpT (C) in the OF (top) and IF (bottom) conformations. Helices 1,7 are colored in blue and helices 5,11 are colored in red. The remaining helices are colored grey to extenuate the changes in the helices that play dominant roles in the cytoplasmic and periplasmic/extracelluar and intracellular gating.

Results and Discussion

In a previously reported [26] study of GkPOT, 400ns simulations in various *apo* and substrate-bound conditions have revealed the shortcomings of unbiased MD for the study of global conformational changes in this protein. Therefore, the equilibrated structure of *apo* GkPOT was used here to initiate nonequilibrium pulling SMD simulations to capture the IF→OF conformational transition as described in Theoretical Methods. A number of TM helix biasing protocols were conducted to steer the conformational change to the unknown OF conformation. The combination of H1, H2, H7 and H8 provided the most successful attempt, evaluated by the stability of the resulting OF conformation; i.e., once equilibrated, the resulting conformation from the SMD simulation should stay in an OF conformation with an open periplasmic gate and a closed cytoplasmic gate. Similar to GkPOT, the simulations of GlpT were previously conducted using a similar bias-

ing protocol, which only involved the TM helices H1 and H7 [28]. This protocol was trialed in our GkPOT studies (see Supporting Information); however, the resulting OF conformation was not stable enough as compared to that generated using the H1/H2/H7/H8 protocol.

Two sets of equilibrium simulations for the *apo* GluT1 were initiated from the IF conformation. Both sets undergo a conformational change to the OF conformation at different time scales. Within the first 200 ns for Set 1 and within the last 300 ns for Set 2. Note that GluT1 has been crystallized [14] under conditions that are different from those used in this simulation. Specifically, N45T/E329Q mutant GluT1 has been used for crystallography, which is different from the E329Q mutant GluT1 in our simulations. The fast transition is likely to be attributed to the differences between the wild-type and N45T/E329Q mutant GluT1 proteins. However, here we assume that the transition observed in our simulations more or less reveal the important features of the IF-OF transition of *apo* GluT1.

Global Protein Conformational Dynamics

Our previous simulations of GlpT revealed an important role for TM helices H1, H5, H7, H11 and the conformational changes among them by defining interhelical angle pairings between H1 and H7 ($\langle$H1,H7$\rangle$) and between H5 and H11 ($\langle$H5,H11$\rangle$) [28]. The former is directly involved in periplasmic/extracellular gating and the latter is involved in cytoplasmic/intracellular gating. Figure 2 provides insight into the IF and OF conformations of the respective proteins and helices involved in the cyto-/intra- and periplasmic/extracellular gating, the TM helix pairings are colored blue (H1, H7) and red (H5, H11). These proteins share a very similar topology (12 transmembrane structure and similar openings at the two gates). GluT1 (Fig.2B), unlike the two proteins GkPOT and GlpT, contains the additional domain, ICH, at the intracellular gate.

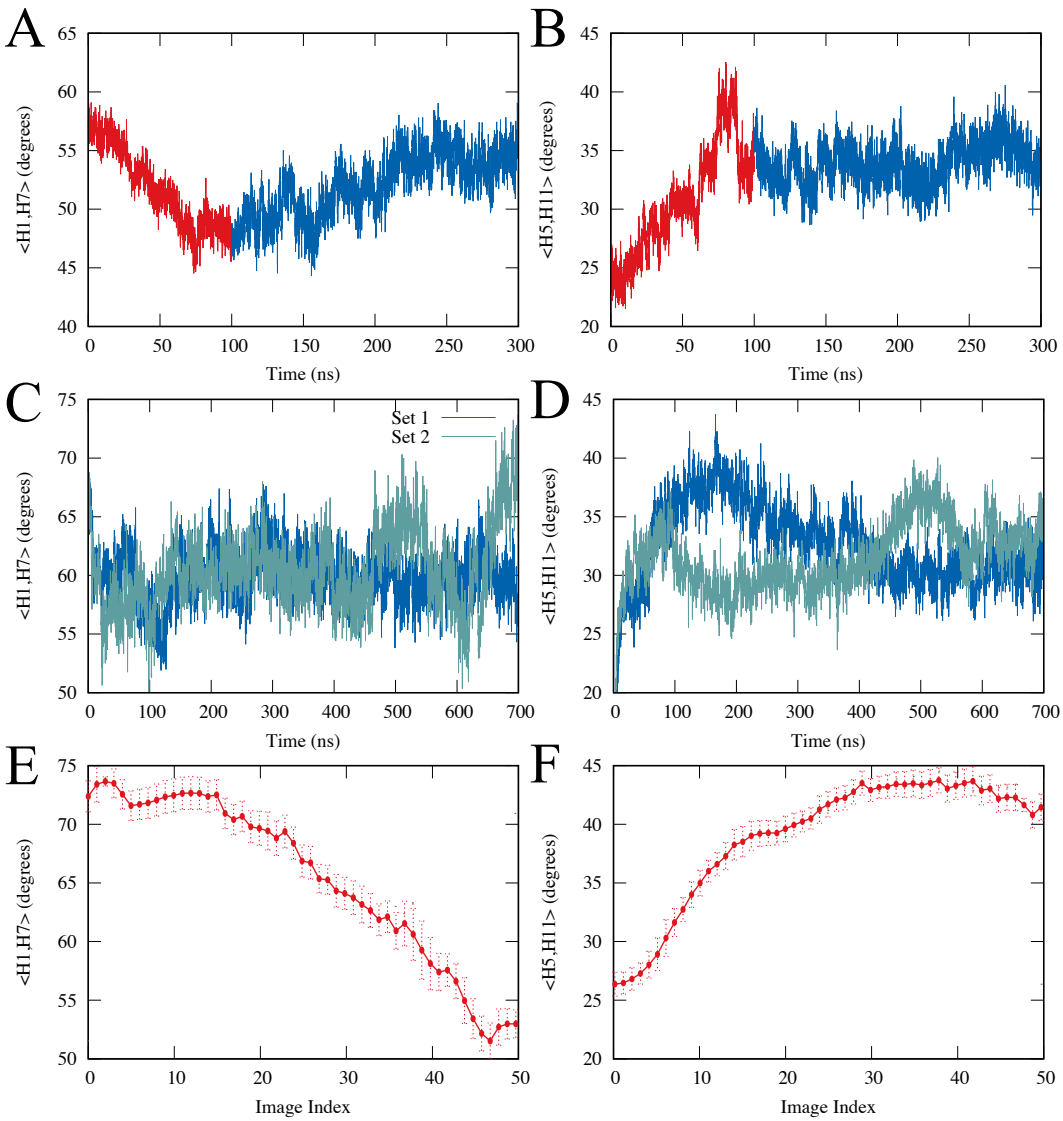

Figure 3: Changes in interhelical angles $\langle$H1,H7$\rangle$ and $\langle$H5,H11$\rangle$ in GkPOT (A,B), GluT1 (C,D), and GlpT (E,F) simulations. GkPOT time series include SMD (red) and follow-up release (blue) simulations. GluT1 time series is based on unbiased equilibrium simulations. GlpT values are based on biase-exchange umbrella sampling simulations, where images 0 and 50 correspond to IF and OF states, respectively [28].

Figure 3 describes the changes associated with these interhelical angles for GkPOT, GluT1, and GlpT. The nature of the three sets of simulations are different; however, one can determine the trend of the changes in these interhelical angles in all three sets of simulations, with the SMD and follow-up release simulations described in Theoretical Methods, GkPOT transitions to a stable OF conformation. We have shown both the SMD and the post equilibration for GkPOT. Both Sets 1

and 2 of GluT1 undergo conformational changes at very quick timescales. Both proteins transition from the IF conformation to an OCC conformation and then to the OF conformation within the 100 ns in Set 1 (Fig. 3C) and at ∼400 ns for Set 2. The transitions are observed by changes in the interhelical angles for 1, 7 and 5, 11, as indicated in Figure 3 C-D.

GluT1 and GlpT contain the pronounced rotational changes in the interhelical angle between H1 and H7. Set 2 in GluT1, during the conformational transition, $\langle$H1,H7$\rangle$ decreases from 68° to 52° and followed by a rotational increase to ∼74°. GkPOT undergoes a similar conformational change to GlpT along the interhelical angles. We note that during the post SMD equilibration, the $\langle$H1,H7$\rangle$ angle increases again to some extent in GkPOT simulations. This is potentially due to the fact that the change imposed on the $\langle$H1,H7$\rangle$ angle based on the LacY structure (see Theoretical Methods) is more than the actual change in $\langle$H1,H7$\rangle$ during the IF-OF transition. The outcome after the equilibration, is more closely similar to that of GluT1, which is less than 10 degrees. On the other hand, the $\langle$H1,H7$\rangle$ change in GlpT is about 20 degrees. For $\langle$H5,H11$\rangle$, GluT1 is the transporter with the highest change. All three proteins, however, show a significant correlation between the changes in $\langle$H1,H7$\rangle$ and $\langle$H5,H11$\rangle$, in line with the rocker-switch mechanism.

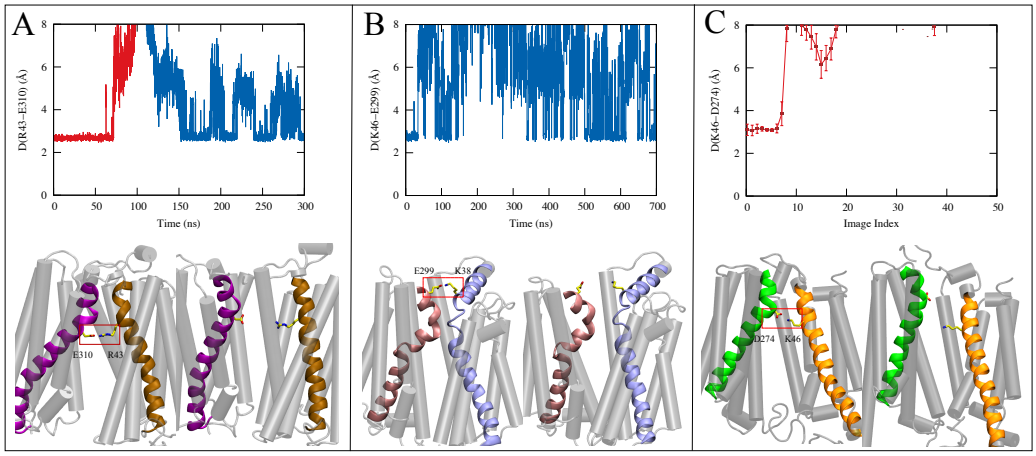

Figure 4: Minimum donor-acceptor distance measurements between the key inter-bundle salt-bridge forming amino acids (A) R43 and E310 in GkPOT, (B) K38 and E299 in GluT1 (Set 1), and (C) K46 and D274 in GlpT.

Local Conformational Changes

As GkPOT transitions toward the OF state, an important inter-bundle salt bridge between R43 in H1 and E310 in H7 becomes disrupted, and appears to break after 75 ns of the SMD pulling (Fig. 4A). The salt bridge forms again after the protein reaches the equilibrated OF conformation, although the strength of this salt bridge is extremely weakened. The salt bridge forms and breaks multiple times during the equilibrated portion of the simualtion.

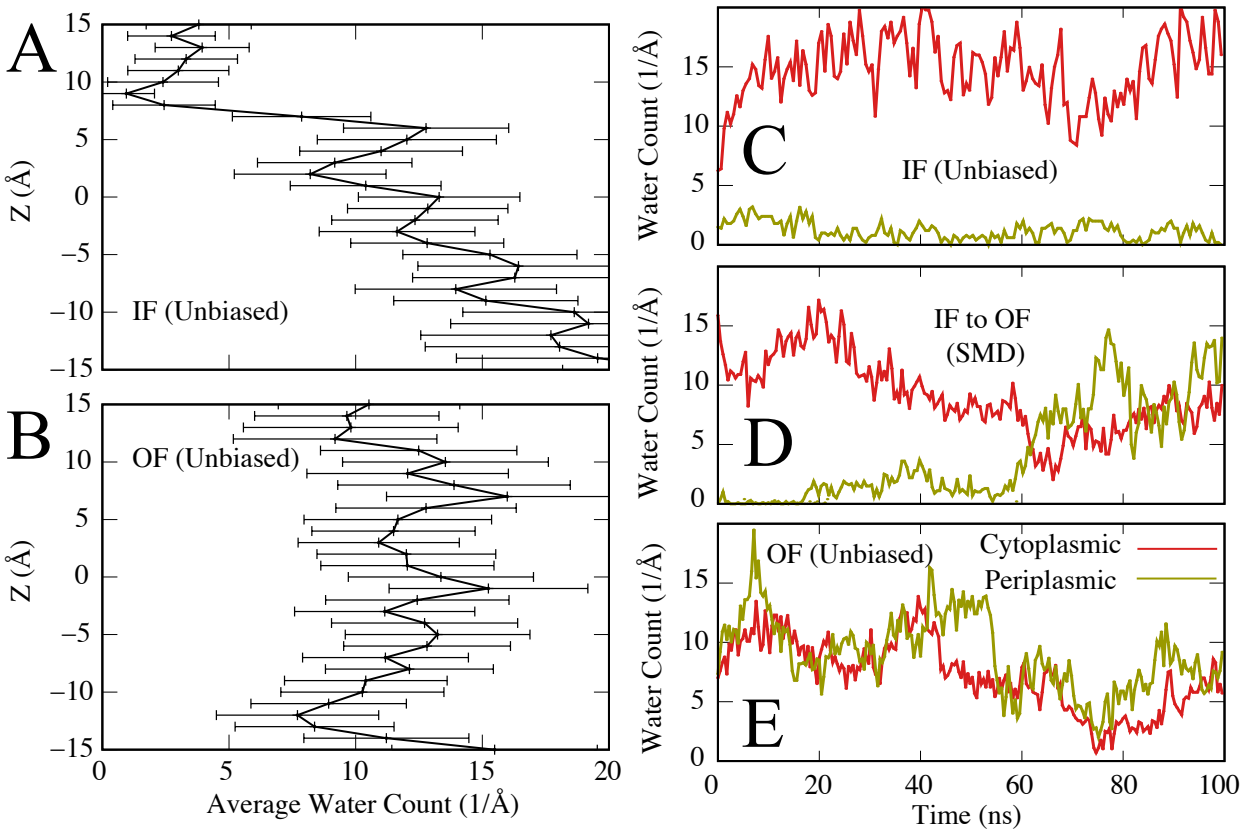

Figure 5: GkPOT water accessibility along the pore (A,B) and at peri- (bronze) and cytoplasmic (red) gates during different stages of equilibrium and nonequilibrium simulations: (C) 100 ns of equilibrium simulation of GkPOT in the IF state prior to SMD (D) 100 ns of SMD as described before for GkPOT, and (E) First 100 ns of follow-up unbiased MD after transitioning to the OF state. The water profiles and the error bars in (A) and (B) are generated from the same simulations used for panels (C) and (E).

The GluT1 extracelluar gating is directly influenced by the interactions between TM helices H1 and H7. This is consistent with both GkPOT and GlpT (Figure 3). However, unlike GkPOT

and GlpT, the interactions between H1 and H7 take place on the extracellular side of the protein as opposed to being in transmembrane region (Fig. 4B). The salt bridges observed in the TM region of both GkPOT and GlpT are necessary for protein stability and play a key role in substrate translocation (Fig. 4).

A salt bridge between Lys 38 in H1 and Glu 299 in H7 is observed in the very beginning of the GluT1 simulation (Fig. 4 and Supplementary), as it remains in the IF. While in the occluded conformation the salt bridge tries to reform. Additional salt-bridge forming residues have been observed and appear to be directly involved in allowing GluT1 to move into the OF conformation, the ICH provides the cytoplasmic gating by moving into a position to interact with both H5 and H11. By forming two very strong salt bridges with H5 and H11 (Glu 243 and Arg 153, Glu 247 and Arg 400) (Supplementary Information), this allows for the ICH to remain locked into position. From this stability we observe the interhelical angles between H5-H11 to exhibit the same angular change of about 15-20 degrees during the transitions between IF and OF conformations.

GlpT, on the other had, has the strongest correlation between the local events (i.e., salt bridge formation/disruption) and global conformational changes (interhelical angle changes). The interbundle salt bridges seem to play an important role in the transport mechanism of MFS transporters; however, the exact location and the exact function of these salt bridges make it quite difficult to draw any conclusion on this issue. The interbundle salt bridges seem to stabilize the IF conformation in all three proteins and the IF-OF transition seems to require the disruption of this salt bridge. The OF state, however, may or may not be associated with the presence or absence of the salt bridge.

Water Accessibility at Cytoplasmic and Peripliasmic Gates

Figure 5 depicts the change in water count along the protein pore and the average water count at the cytoplasmic and periplasmic gates of GkPOT from the different stages of the equilibrium and nonquilibrum simulations. Fig. 5 (A-B) shows the comparison of the water along the protein pore when in the IF and OF conformations both before and after the nonequilibrium simulation. The

comparison between the two states shows a much broader degree of the water at the cytoplasmic side of the protein, this attributed to GkPOT remaining IF the entirety of the unbiased simulation. While the OF conformation from the post unbiased simulation shows a water profile that has water both at the cytoplasmic and periplasmic sides. We attribute this profile to the overall function of GkPOT and the ability to accommodate different substrates in different binding orientations, which would account for a slightly "leaky" bottleneck at both gates. This observation is furthered by panels (C-E). The average water at the periplasmic gate trends above the average water found at the cytoplasmic gate during the nonequilibrium and continues to remain greater during the post nonequilibrium simulation. Admittedly, (E) only represents the first 100ns of the simulation, however (B) is the water profile from all 200ns of unbiased equilibration.

GluT1 (Fig. 6) exhibits similar water profiles to that of GkPOT. GluT1 transitions very quickly to an OCC conformation within the first few nanseconds of the simulations and is indicated in the water profiles of Fig. 6 (A-B). Water profiles from different stages of the equilibrium simulations showing the increase in water found at the extracellular gate are indicated in C-D. It can be seen that the average water along the pore in the GluT1 when in the OF conformation is very comparable to that of the OF conformation of GkPOT. The average water counts at the intra- and extracellular gates are also shown as the full 700ns time series and also shown at different time intervals after the GluT1 has transitioned to the OF conformation Fig.6 (E-H). There is clear separation in the amount of water present at the extracellular gate in both simulation sets (E-H). The extracellular gate appears to close briefly for Set 2 (Fig.6 H) but reverts back as the protein is trying to stabilize in the OF conformation at time 550 ns. The water profiles and time series of the water counts at both the peri-/extra- and cytoplasmic/intracellular gates furthers the alternating-access mechanism in both GkPOT and GluT1. Although we do not observe GluT1 transition back towards the IF conformtation in the two sets of simulations, the trends indicated in Figures 5 and 6 shows that proteins do not open at either side of the membrane at the same time and suggests the conformational changes of the *apo* proteins follow ideal gating modes.

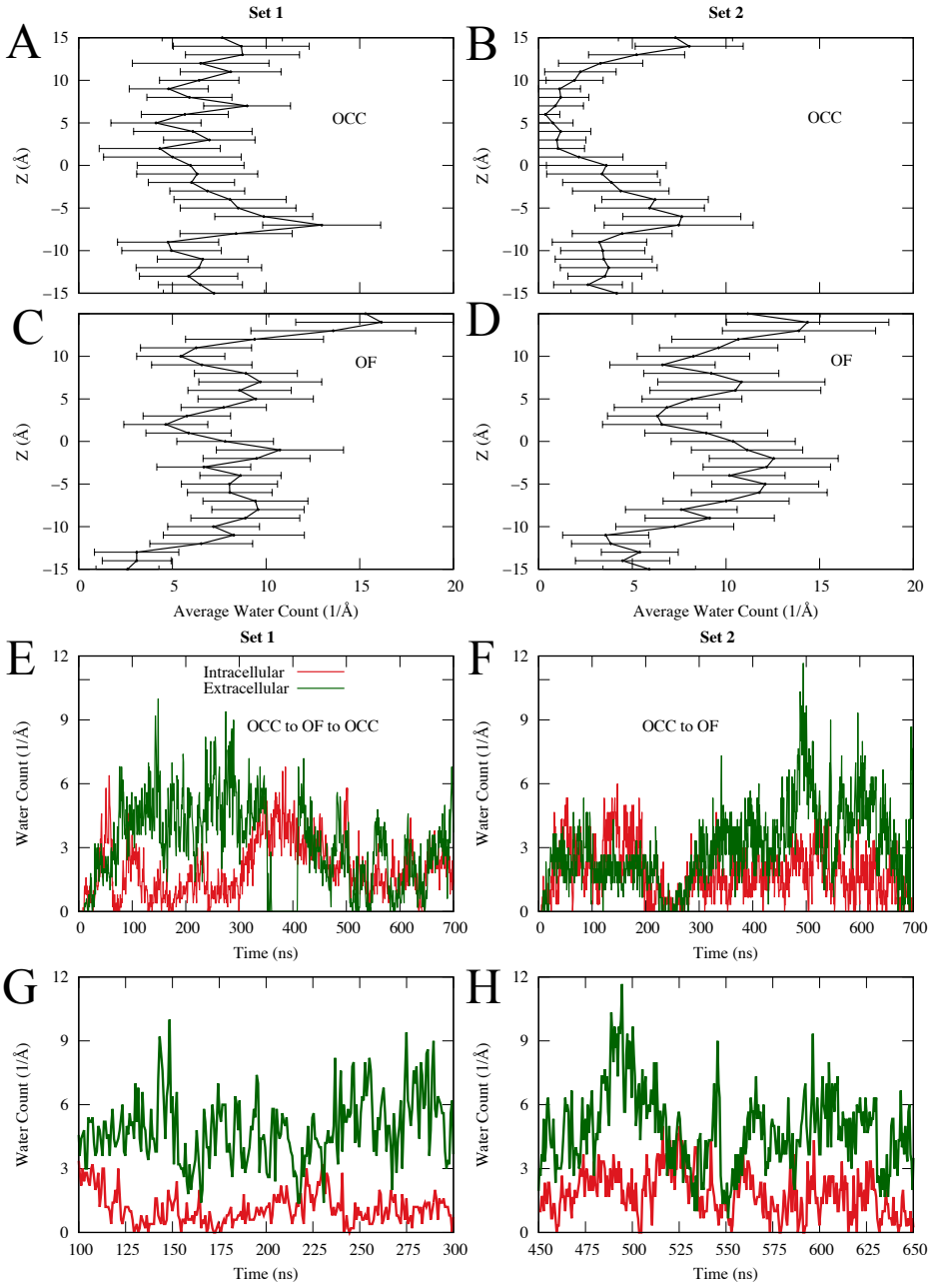

Figure 6: GluT1 water accessibility of the two simulation sets along the pore (A-D) and at the extra- (green) and intracellular (red) gates during different stages of the equilibrium simulations (E-H): (A, B) 100 ns of equilibrium simulation of GluT1 in the OCC state prior to transitioning to OF. (C, D) Average water over 20 ns of simulation after transitioning to OF. (E, F) Entire time series of Water Count at the extra- and intracellular gates for both sets. (G, H) 200 ns of the OF water count at the associated time intervals for a given set.

Conclusions

We have employed equilibrium and nonequilibrium all-atom MD simulations of GkPOT, GluT1, and GlpT. Through this study we have demonstrated a thorough comparison of the conformational dynamics of the three different classes of MFS transporters; i.e., symporters, uniporters, and antiporters, with emphasis placed on the simulation of the *apo* protein.

The three systems (classes) of MFS transporters examined are quite different in terms of the timescale of the IF-OF transition; however, we observe various similarities between the three systems including the involvement of specific TM helices that undergo rotational changes along the transition pathways. These include, the H1/H7 helices which are involved in periplasmic/extracellular gating and the H5/H11 helices, involved in cytoplasmic/intracellular gating. Additionally, these three systems contain an interbundle salt bridge that forms between the H1 and H7. The formation of these interbundle salt bridge stabilizes these three systems of MFS transporters in the IF state. Admittedly, there are notable differences between these three systems including: (1) the smaller change of $\langle$H1,H7$\rangle$ angle in GkPOT as compared to GluT1 and GlpT, (2) the larger change of $\langle$H5,H11$\rangle$ angle in GluT1 as compared to GkPOT and GlpT, and (3) the formation of the interbundle salt bridge in the OF state for the GkPOT protein as compared to GluT1 and GlpT.

The presented study provides a detailed picture of the similarities and differences associated with the IF$\leftrightarrow$OF conformational transition of three transporters from three distinct classes of MFS superfamily. If combined with similar simulations in the presence of physiologically relevant substrates and cotransported species, the reported simulations could provide a full description of transport cycles of MFS transporters.

References

[1] V. S. Reddy, M. A. Shlykov, R. Castillo, E. I. Sun, and M. H. Saier, "The major facilitator superfamily (MFS) revisited," *FEBS J.*, vol. 279, pp. 2022–2035, June 2012.

[2] S. S. Pao, I. T. Paulsen, and M. H. Saier, "Major Facilitator Superfamily," *Microbiology and Molecular Biology Reviews*, vol. 62, pp. 1 LP – 34, mar 1998.

[3] J. Abramson, I. Smirnova, V. Kasho, G. Verner, H. R. Kaback, and S. Iwata, "Structure and Mechanism of the Lactose Permease of Escherichia coli," *Science*, vol. 301, no. 5633, pp. 610–615, 2003.

[4] L. Guan, Y. Hu, and H. R. Kaback, "Aromatic stacking in the sugar binding site of lactose permease," *Biochemistry*, vol. 42, pp. 1377–1382, 2003.

[5] L. Guan and H. Kaback, "Lessons from lactose permease.," *Annu. Rev. Biophys. Biomol. Struct.*, vol. 35, pp. 67–91, 2006.

[6] M. G. Lawrence, L. Lindahl, and J. M. Zengel, "Effects on translation pausing of alterations in protein and RNA components of the ribosome exit tunnel," *J. Bacteriol.*, vol. 190, pp. 5862–5869, 2008.

[7] E. M. Quistgaard, C. Löw, P. Moberg, L. Trésaugues, and P. Nordlund, "Structural basis for substrate transport in GLUT-homology family of monosaccharide transporters," *Nat. Struct. Mol. Biol.*, vol. 20, no. 6, pp. 766–768, 2013.

[8] Y. Yin, M. Ø. Jensen, E. Tajkhorshid, and K. Schulten, "Sugar binding and protein conformational changes in lactose permease," vol. 91, pp. 3972–3985, 2006.

[9] N. Yan, "Structural advances for the major facilitator superfamily (MFS) transporters," *Trends in Biochemical Sciences*, vol. 38, no. 3, pp. 151–159, 2013.

[10] I. S. Wood and P. Trayhurn, "Glucose transporters (GLUT and SGLT): expanded families of sugar transport proteins," *British Journal of Nutrition*, vol. 89, no. 1, pp. 3–9, 2003.

[11] N. Solcan, J. Kwok, P. W. Fowler, A. D. Cameron, D. Drew, S. Iwata, and S. Newstead, "Alternating access mechanism in the POT family of oligopeptide transporters," *EMBO J.*, vol. 31, no. 16, pp. 3411–3421, 2012.

[12] F. Guettou, E. M. Quistgaard, L. Tresaugues, P. Moberg, , C. Jorgerschold, A. J. Jong, P. Nordlund, and C. Loew, "Structural insights into substrate recognition in proton-dependent oligopeptide transporters.," vol. 14, pp. 804–810, 2013.

[13] B. Thorens and M. Mueckler, "Glucose transporters in the 21st century," *Am. J. Physiol. – Endoc. M.*, vol. 298, no. 2, pp. E141–E145, 2010.

[14] D. Deng, C. Xu, P. Sun, J. Wu, C. Yan, M. Hu, and N. Yan, "Crystal structure of the human glucose transporter GLUT1," *Nature*, vol. 510, no. 7503, p. 121, 2014.

[15] D. Deng, P. Sun, C. Yan, M. Ke, X. Jiang, L. Xiong, W. Ren, K. Hirata, M. Yamamoto, S. Fan, *et al.*, "Molecular basis of ligand recognition and transport by glucose transporters," *Nature*, vol. 526, no. 7573, p. 391, 2015.

[16] N. Yan, "A glimpse of membrane transport through structures—advances in the structural biology of the GLUT glucose transporters," *Journal of molecular biology*, vol. 429, no. 17, pp. 2710–2725, 2017.

[17] N. Yan, "Structural advances for the major facilitator superfamily (MFS) transporters," *Trends Biochem. Sci.*, vol. 38, no. 3, pp. 151–159, 2013.

[18] S. Doki, H. E. Kato, N. Solcan, M. Iwaki, M. Koyama, M. Hattori, N. Iwase, T. Tsukazaki, Y. Sugita, H. Kandori, S. Newstead, R. Ishitani, and O. Nureki, "Structural basis for dynamic mechanism of proton-coupled symport by the peptide transporter POT," *Proc. Natl. Acad. Sci. USA*, vol. 110, pp. 11343–11348, July 2013.

[19] Q. Yang, Y. Ma, Y. Zhao, Z. She, L. Wang, J. Li, C. Wang, and Y. Deng, "Accelerated drug release and clearance of PEGylated epirubicin liposomes following repeated injections: a new challenge for sequential low-dose chemotherapy," *International Journal of Nanomedicine*, vol. 8, pp. 1257 – 1268, 2013.

[20] J. A. Lyons, J. L. Parker, N. Solcan, A. Brinth, D. Li, S. T. Shah, M. Caffrey, and S. Newstead, "Structural basis for polyspecificity in the POT family of proton-coupled oligopeptide transporters.," vol. 15, pp. 886 – 893, 2014.

[21] S. Newstead, D. Drew, A. D. Cameron, V. L. Postis, X. Xia, P. W. Fowler, J. C. Ingram, E. P. Carpenter, M. S. Sansom, M. J. McPherson, S. A. Baldwin, and S. Iwata, "Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2.," *EMBO J.*, vol. 30, pp. 417–426, Jan. 2011.

[22] E. M. Quistgaard, M. Martinez Molledo, and C. Löw, "Structure determination of a major facilitator peptide transporter: Inward facing PepTSt from streptococcus thermophilus crystallized in space group p3121," *PLOS ONE*, vol. 12, pp. 1–20, 03 2017.

[23] C. M. Elvin, C. M. Hardy, and H. Rosenberg, "Pi exchange mediated by the GlpT-dependent sn-glycerol-3-phosphate transport system in Escherichia coli.," *Journal of Bacteriology*, vol. 161, pp. 1054 LP – 1058, mar 1985.

[24] C. J. Law, J. Almqvist, A. Bernstein, R. M. Goetz, Y. Huang, C. Soudant, A. Laaksonen, S. Hovmöller, and D.-N. Wang, "Salt-bridge Dynamics Control Substrate-induced Conformational Change in the Membrane Transporter GlpT," *Journal of Molecular Biology*, vol. 378, no. 4, pp. 828–839, 2008.

[25] Y. Huang, M. J. Lemieux, J. Song, M. Auer, and D.-N. Wang, "Structure and Mechanism of the Glycerol-3-Phosphate Transporter from Escherichia coli," *Science*, vol. 301, pp. 616 LP – 620, aug 2003.

[26] K. Immadisetty, J. Hettige, and M. Moradi, "What Can and Cannot Be Learned from Molecular Dynamics Simulations of Bacterial Proton-Coupled Oligopeptide Transporter GkPOT?," *The Journal of Physical Chemistry B*, vol. 121, pp. 3644–3656, apr 2017.

[27] T. Galochkina, M. Ng Fuk Chong, L. Challali, S. Abbar, and C. Etchebest, "New insights into GluT1 mechanics during glucose transfer," *Scientific Reports*, vol. 9, no. 1, p. 998, 2019.

[28] M. Moradi, G. Enkavi, and E. Tajkhorshid, "Atomic-level characterization of transport cycle thermodynamics in the glycerol-3-phosphate:phosphate antiporter," *Nature Communications*, vol. 6, p. 8393, sep 2015.

[29] B. Isralewitz, J. Stone, and K. Schulten, "Timeline: an interactive raster plot to identify events in molecular dynamics trajectories," 2011. Submitted.

[30] T. Schlick, B. Li, and W. Olson, "The influence of salt on the structure and energetics of supercoiled DNA," *Biophys. J.*, vol. 67, pp. 2146–2166, 1994.

[31] M. Moradi and E. Tajkhorshid, "Mechanistic picture for conformational transition of a membrane transporter at atomic resolution.," vol. 110, no. 47, pp. 18916–18921, 2013.

[32] M. Moradi and E. Tajkhorshid, "Computational Recipe for Efficient Description of Large-Scale Conformational Changes in Biomolecular Systems," *Journal of Chemical Theory and Computation*, vol. 10, pp. 2866–2880, jul 2014.

[33] A. Fakharzadeh and M. Moradi, "Effective Riemannian diffusion model for conformational dynamics of biomolecular systems," vol. 7, pp. 4980–4987, 2016.

[34] S. Jo, T. Kim, and W. Im, "Automated builder and database of protein/membrane complexes for molecular dynamics simulations," *PLoS One*, vol. 2, p. e880, 2007.

[35] J. Lee, X. Cheng, J. M. Swails, M. S. Yeom, P. K. Eastman, J. A. Lemkul, S. Wei, J. Buckner, J. C. Jeong, Y. Qi, S. Jo, V. S. Pande, D. A. Case, I. Charles L. Brooks, J. Alexander D. MacKerell, J. B. Klauda, and W. Im, "Charmm-gui input generator for namd, gromacs, amber, openmm, and charmm/openmm simulations using the charmm36 additive force field," vol. 12, no. 1, pp. 405–413, 2016.

[36] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, "Comparison of simple potential functions for simulating liquid water," *J. Chem. Phys.*, vol. 79, no. 2, pp. 926–935, 1983.

[37] J. B. Klauda, R. M. Venable, J. A. Freites, J. W. O'Connor, D. J. Tobias, C. Mondragon-Ramirez, I. Vorobyov, A. D. MacKerell Jr., and R. W. Pastor, "Update of the CHARMM all-atom additive force field for lipids: Validation on six lipid types," *J. Phys. Chem. B*, vol. 114, no. 23, pp. 7830–7843, 2010.

[38] R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig, and A. D. MacKerell, "Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone , ψ and Side-Chain $\chi 1$ and $\chi 2$ Dihedral Angles," *Journal of Chemical Theory and Computation*, vol. 8, pp. 3257–3273, sep 2012.

[39] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten, "Scalable molecular dynamics with namd," vol. 26, pp. 1781–1802, 2005.

[40] J. K. Reid, "On the method of conjugate gradients for the solution of large sparse systems of linear equations," in *Large Sparse Sets of Linear Equations* (J. K. Reid, ed.), pp. 231–254, London: Academic Press, 1971.

[41] G. J. Martyna, D. J. Tobias, and M. L. Klein, "Constant pressure molecular dynamics algorithms," *J. Chem. Phys.*, vol. 101, no. 5, pp. 4177–4189, 1994.

[42] S. E. Feller, Y. Zhang, R. W. Pastor, and B. R. Brooks, "Constant pressure molecular dynamics simulation: The Langevin piston method," *J. Chem. Phys.*, vol. 103, no. 11, pp. 4613–4621, 1995.

[43] T. Darden, D. York, and L. G. Pedersen, "Particle mesh Ewald: An $N\cdot\log(N)$ method for Ewald sums in large systems," *J. Chem. Phys.*, vol. 98, no. 12, pp. 10089–10092, 1993.

[44] S. Dang, L. Sun, Y. Huang, F. Lu, Y. Liu, H. Gong, J. Wang, and N. Yan, "Structure of a fucose transporter in an outward-open conformation.," *Nature*, vol. 467, pp. 734–738, Oct. 2010.

Conclusion

We have described a combination of multiple MD based techniques that allows for a rigorous characterization of energetics and kinetics of large-scale conformational changes in proteins. The methodology is based on biased, nonequilibrium collective variable based simulations including nonequilibrium pulling, string method with swarms of trajectories, bias-exchange umbrella sampling, and rate estimation techniques. By developing this effective methodology, we have shown that, although this methodology was performed on the membrane transporter protein GlpT, this methodology can be effective tool at describing conformational transitions in various other systems as well. The overall approach remains the same, there is a requirement for a strong understanding of the conformational change that the protein undergoes. Next, the employment of nonequilibrium work relations by defining collective variables that accurately describe the conformational change and optimizing the specific pulling parameters that induce the conformational change with least amount of nonequilibrium work (the derived pathway is the much more energetically favored pathway for the protein conformational change). Path optimization and sampling have shown to be effective tools at further characterizing the conformational change of proteins. This methodology or specific aspects of the methodology were employed for the comparative analysis that was conducted to induce and characterize the conformational transitions of the two classes of homologous proteins presented here. The first class of proteins were the SARS-CoV-1 and SARS-CoV-2 spike proteins and the second class being the three MFS transporters GluT1, GkPOT, and GlpT.

By employing the methodology presented here, we have demonstrated that the SARS-CoV-1 and SARS-CoV-2 spike proteins exhibit differential dynamic behavior by characterizing the large-scale conformational changes associated with the active-to-inactive transition pathways using biased molecular dynamics simulations. More specifically, the transition pathways were first determined through the use of SMD, further refined by employing two sets of SMwST, and effectively sampled the transition pathways using BEUS to determine the associated free energy barriers along the pathways. The amount of detail provided here will allow for the testing of various experiments and even the trapping of specific intermediate states along the transition pathways.

The study performed here primarily sheds light on the conformational dynamics of the SARS-CoV-1 and SARS-CoV-2 spike proteins. While differences in the dynamic behavior of these spike proteins almost certainly contribute to differences in how transmissible or infectious each could be, factors such as spike protein glycosylation and the behavior of other viral proteins must be considered in order for a more complete hypothesis to be determined, additional experimental and computational studies are required in order to investigate the differential infectivity and transmissibility of SARS-CoV-1 and SARS-CoV-2. Our simulations provide valuable insight into the dynamic behavior of the CoV-1 and CoV-2 spike proteins in regards to the active to inactive conformational transition pathway. However, having performed simulations using the non-glycosylated spike proteins of CoV-1 and CoV-2, it has been hypothesized that the glycan chains could play an important role in the conformational change in the RBD [1, 2].

The three classes of MFS transporters examined here are shown to be quite different in terms of the timescale of the IF-OF transition. We do, however, observe various similarities between the three systems including the involvement of specific TM helices that undergo rotational changes along the transition pathways. These include, the H1/H7 helices (periplasmic/extracellular gating) and the H5/H11 helices (cytoplasmic/intracellular gating). Additionally, these three systems contain an interbundle salt bridge that forms between the H1 and H7. The formation of these interbundle salt bridges have been known to stabilizes MFS transporters in the IF state. We observe a difference in the location in the interdomain saltbridge for GluT1. The residue pairings are consistent with the same helices, H1/H7, but the part of the helices where this saltbridge forms is on the extracellular side of the protein, unlike GkPOT and GlptT where the saltbridge forms in the central transmembrane domain. Furthermore, there are a number notable differences between these three systems when comparing the interhelical angles between H1/H7 and H5/H11, with GkPOT having a less pronounced rotational change between H1 and H7 than GluT1 and GlpT. All three proteins undergo a similar degree of rotational change between H5/H11. The presented study provides a detailed picture of the similarities and differences associated with the IF$\leftrightarrow$OF conformational transition of three transporters from three distinct classes of MFS superfamily.

References

[1] L. Casalino, Z. Gaieb, J. A. Goldsmith, C. K. Hjorth, A. C. Dommer, A. M. Harbison, C. A. Fogarty, E. P. Barros, B. C. Taylor, J. S. McLellan, E. Fadda, and R. E. Amaro, "Beyond Shielding: The Roles of Glycans in the SARS-CoV-2 Spike Protein," *ACS Central Science*, vol. 6, no. 10, pp. 1722–1734, 2020.

[2] T. Sztain, S.-H. Ahn, A. T. Bogetti, L. Casalino, J. A. Goldsmith, R. S. McCool, F. L. Kearns, J. Andrew McCammon, J. S. McLellan, and R. E. Amaro, "A glycan gate controls opening of the SARS-CoV-2 spike protein. https://www.biorxiv.org/content/10.1101/2021.02.15.431212v1," *bioRxiv*, 2021.

Appendix

Supplementary Material for Atomic-level characterization of the conformational transition pathways in SARS-CoV-1 and SARS-CoV-2 spike proteins

Dylan Ogden[1] and Mahmoud Moradi[1] *

[1]Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, AR 72701

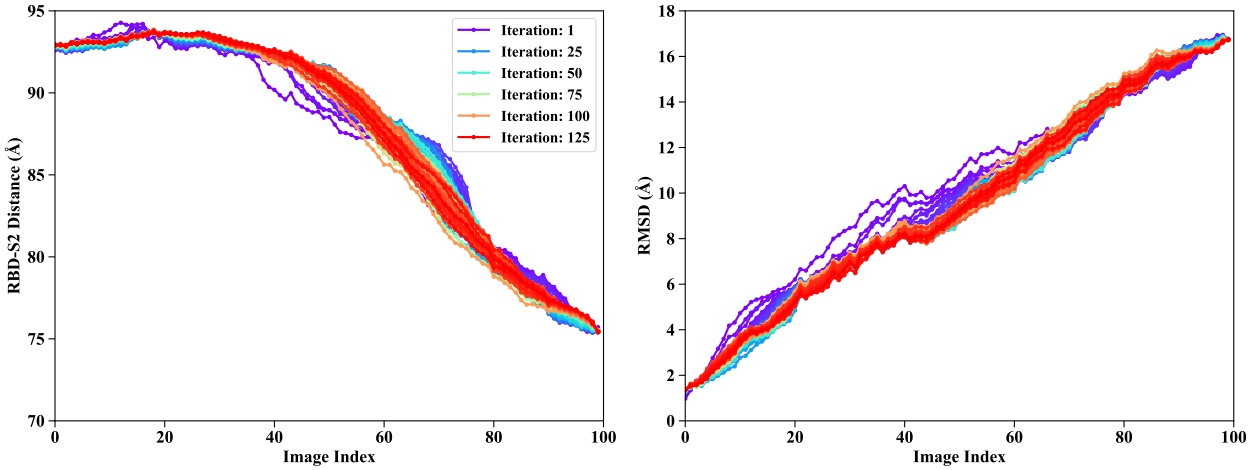

Figure S1: **SMwST set 1: Global conformational changes of the CoV-1 spike protein** RBD-S2 center of mass distance and C-α RMSD of the active protomer. The reference structure for RMSD is an equilibrated strutcure prior to SMD. Each line represents a completed iteration of the SMwST simulations and each point is the average value of the 20 copies for each image. 125 iterations were conducted.

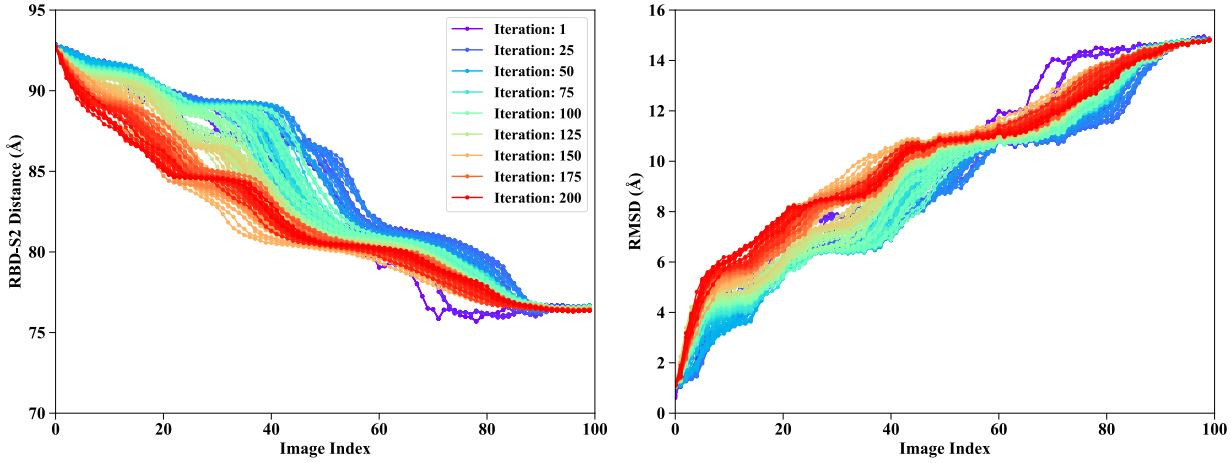

Figure S2: **SMwST set 1: Global conformational changes of the CoV-2 spike protein** RBD-S2 center of mass distance and C-α RMSD of the active protomer. The reference structure for RMSD is an equilibrated structure prior to SMD. Each line represents a completed iteration of the SMwST simulations and each point is the average value of the 20 copies for each image. 200 iterations were conducted.

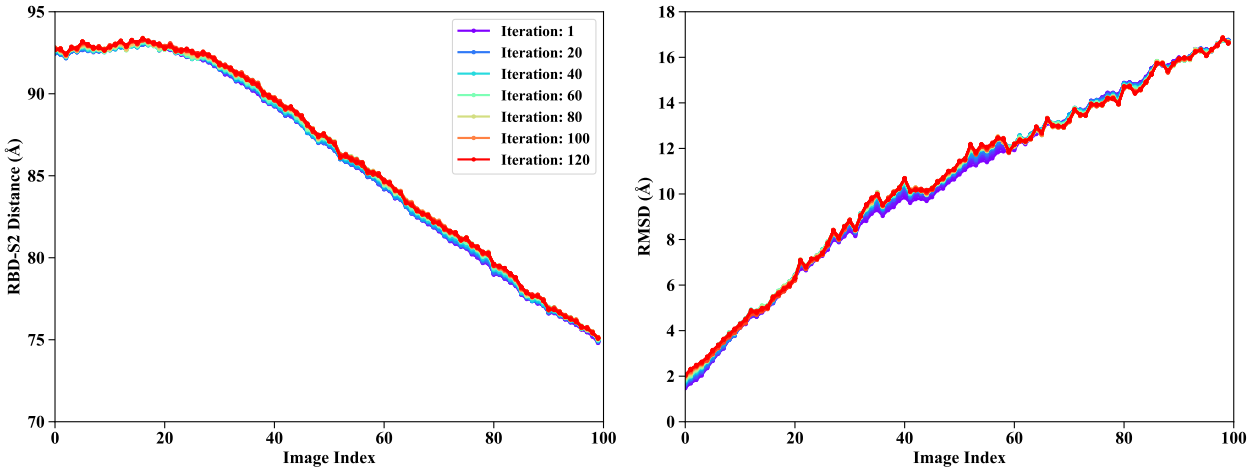

Figure S3: **SMwST set 2: Global conformational changes of the CoV-1 spike protein** RBD-S2 center of mass distance and C-α RMSD of the active protomer. The reference structure for RMSD is an equilibrated structure prior to SMD. Each line represents a completed iteration of the SMwST simulations and each point is the average value of the 20 copies for each image. 120 iterations were conducted.

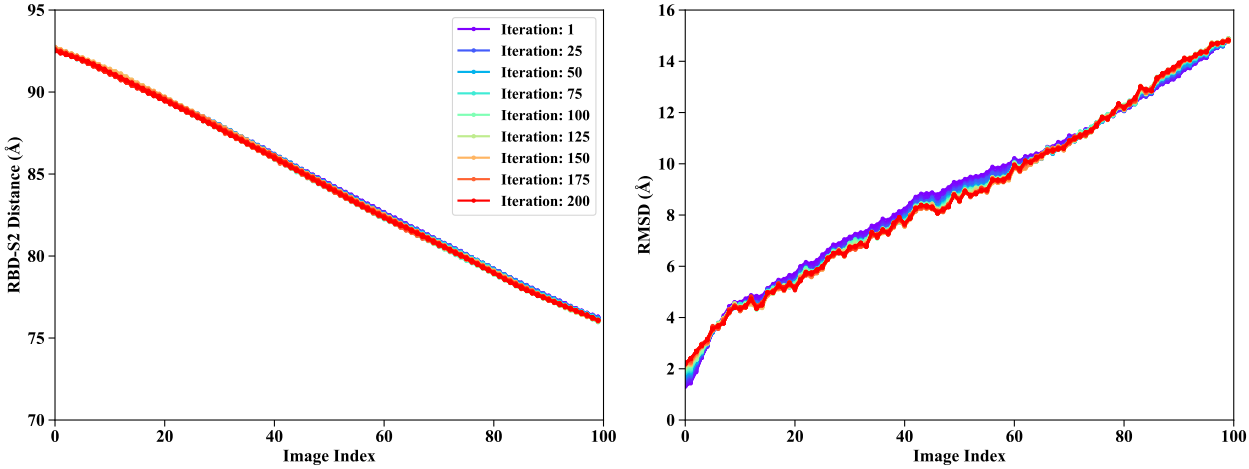

Figure S4: **SMwST set 2: Global conformational changes of the CoV-2 spike protein** RBD-S2 center of mass distance and C-α RMSD of the active protomer. The reference structure for RMSD is an equilibrated structure prior to SMD. Each line represents a completed iteration of the SMwST simulations and each point is the average value of the 20 copies for each image. 200 iterations were conducted.

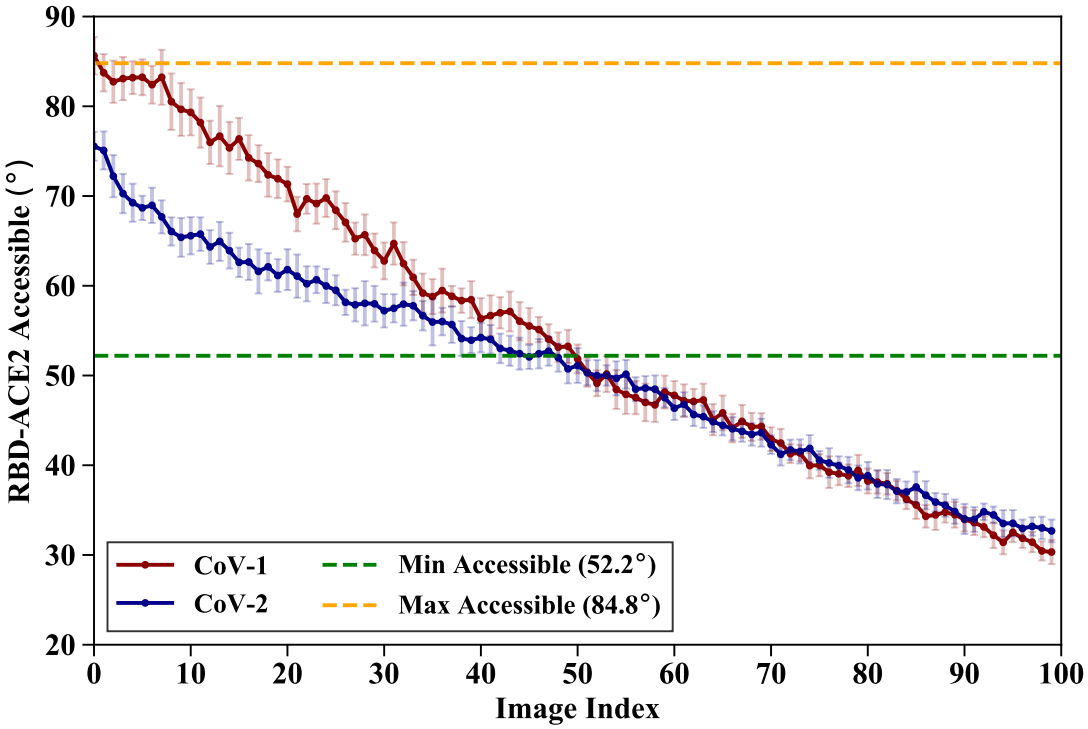

Figure S5: **ACE2 Accessibility** ACE2 Accessibility three-point angle composed of residues D392, T608, V973 (C-α atoms) for CoV-1 and D405, V622, V991 for CoV-2. Each point is the average value of the 20 copies for each image and the error bars are defined by the standard deviation of the 20 copies for each image. The upper and lower limits are defined by Peng et al. indicating the degree of RBD activation upon which ACE2 is accessible for binding with the receptor binding motif (RBM) of the active protomer.

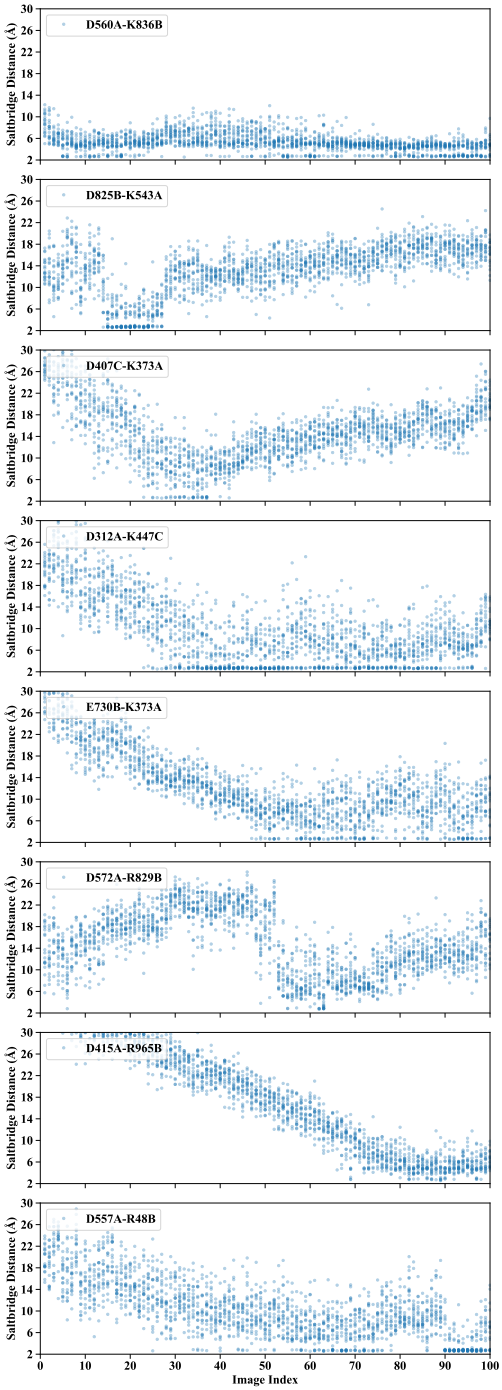

Figure S6: **Minimum salt-bridge donor-acceptor distance of key residues for CoV-1** Minimum salt-bridge distance between the donor and acceptor atoms of key residues along the transition pathway. Each point is the average value of the 20 copies for each image.

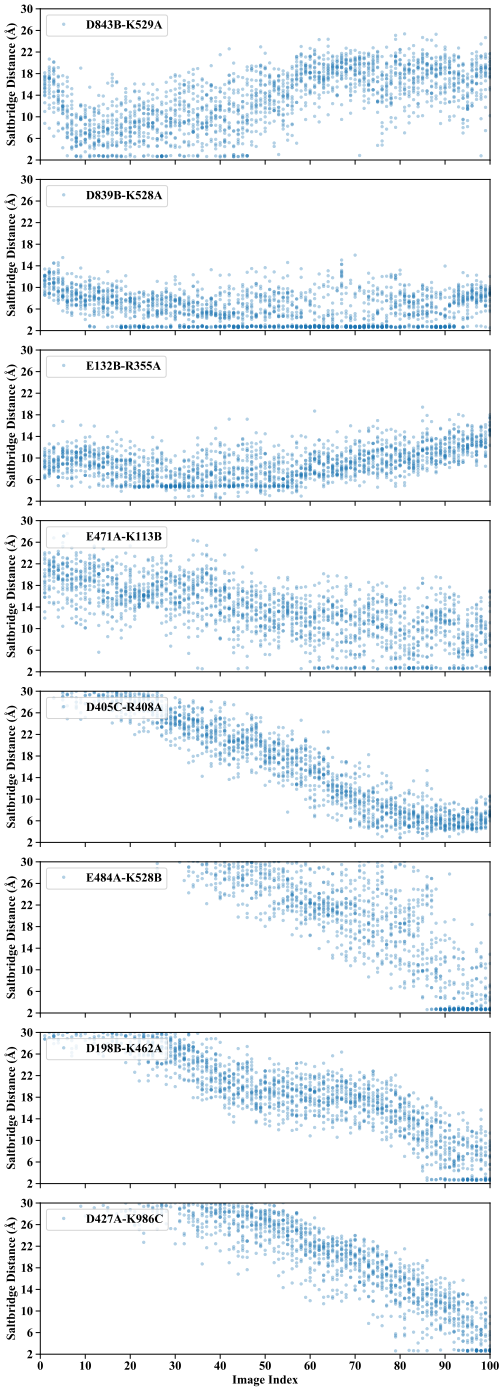

Figure S7: **Minimum salt-bridge donor-acceptor distance of key residues for CoV-2** Minimum salt-bridge distance between the donor and acceptor atoms of key residues along the transition pathway. Each point is the average value of the 20 copies for each image.

Supplementary Material for Conformational Transition Pathways in Major Facilitator Superfamily Transporters

Dylan Ogden[1], Kalyan Immadisetty[1], and Mahmoud Moradi[1] *
[1]Department of Chemistry and Biochemistry, University of Arkansas, Fayetteville, AR 72701

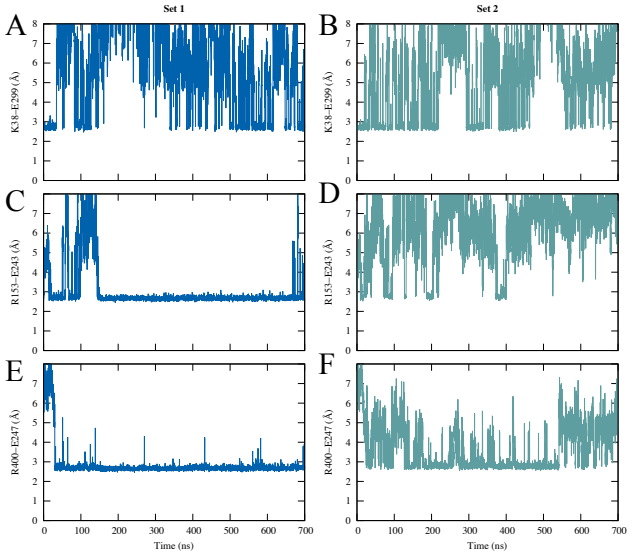

Figure S1: GluT1 minimum donor-accept distance measurements between key inter-bundle salt-bridge forming amino acids in both simulation sets. (A-B) K38 and E299. (B-C)

Nonequilibrium Pulling (NEP) Simulations

Optimizing the NEP protocol on UP-E310:Apo GKPOT

We considered N-terminal (helices 1-6) and C-terminal (helices 7-12) regions of POTs for the NEP simulations and ignored HA (residues 225-250) and HB(residues 258-280).

For the NEP simulations involving the unprotonated-E310:apo form of GKPOT we have used the 140ns equilibrated structure as the starting structure for NEP simulations

Table S1: Various NEP protocols tested on the UP-E310:*apo* form of GKPOT

Index	**Force Constant(K)**	**Simu. Time(ns)**	**Work (kcal/mol)**
1,7	10,000	5	200
	10,000	20	142
	10,000	100	
	5000	20	120
	4000	20	115
	3000	20	100
	2000	20	72
	1000	20	53
	500	20	28
1,8	10,000	20	175
1,9	10,000	20	14
1,11	10,000	20	190
1,12	10000	20	170
2,8	10,000	5	230
2,8	10,000	20	205
3,9	10,000	5	13
3,9	10,000	20	6
5,11	10,000	20	180
5,11	10,000	5	225

Table S2: Various NEP protocols tested on the UP-E310:*apo* form of GKPOT

Index	Force Constant(K)	Simu. Time(ns)	Work (kcal/mol)
6,12	10,000	5	240
6,12	10,000	20	160
7,2	10,000	20	190
7,3	10,000	20	140
7,4	10,000	20	140
7,5	10,000	20	132
7,6	10,000	20	155
	10,000	100	
4,5,11	10,000	20	215
1,7,2,8	10,000	20	300
	10,000	100	
	5000	20	230
	5000	100	
	4000	20	250
	3000	20	225
	2000	20	180
	1000	20	120
1,7,3,9	10,000	20	180
1,7,5,11	10,000	20	225
3,6,9,12	10,000	20	180

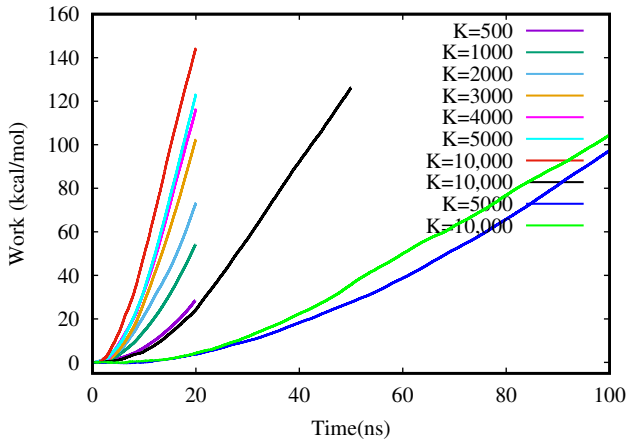

Figure S2: **Optimizing the 1,7 colvars on UP:*Apo* GkPOT.**

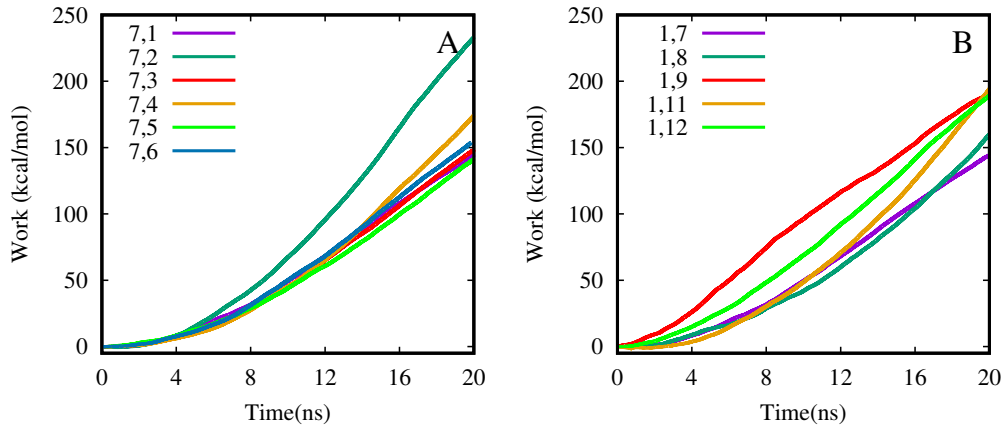

Figure S3: **Comparing the helix-1 and helix-7 colvars combinations on UP:*Apo* GkPOT.**

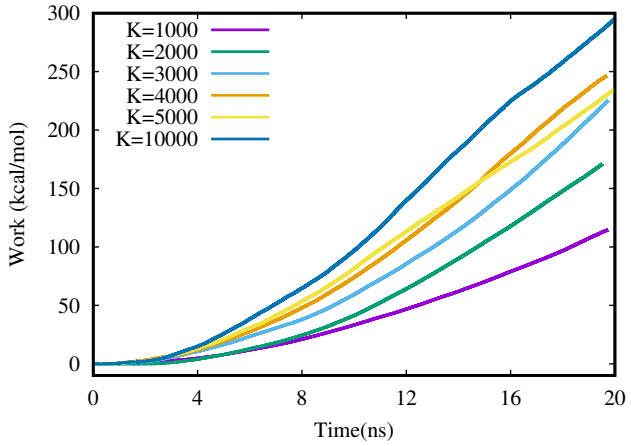

Figure S4: **Application of 1,7,2,8 colvars on UP:*Apo* with different force constants.**

List of other published papers

1. Govind Kumar, V., **Ogden, D. S.**, Isu, U. H., Polasa. A, Losey, J., & Moradi, M. (2022). Prefusion Spike Protein Conformational Changes Are Slower in SARS-CoV-2 than in SARS-CoV-1 Journal of Biological Chemistry., https://doi.org/10.1016/j.jbc.2022.101814

2. Chen, H., **Ogden, D.**, Pant, S., Cai, W., Tajkhorshid, E., Moradi, M., ... Chipot, C. (2022). A Companion Guide to the String Method with Swarms of Trajectories: Characterization, Performance, and Pitfalls. Journal of Chemical Theory and Computation, https://doi.org/10.1021-/acs.jctc.1c01049