How antiporters exchange substrates across the cell membrane? An atomic-level description of the complete exchange cycle in NarK.

Jiangyan Feng,[†] Balaji Selvam,[†] and Diwakar Shukla[*,†,‡,¶,§,‖,⊥]

†*Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States*

‡*Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States*

¶*Center for Biophysics and Quantitative Biology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States*

§*NIH Center for Macromolecular Modeling and Bioinformatics, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States*

‖*Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States*

⊥*Center for Digital Agriculture, University of Illinois at Urbana-Champaign, Urbana, Illinois 61801, United States*

E-mail: diwakar@illinois.edu

Abstract

Major facilitator superfamily (MFS) proteins operate via three different mechanisms: uniport, symport, and antiport. Despite extensive investigations, molecular understanding of antiporters is less advanced than other transporters due to the complex coupling between two substrates and the lack of distinct structures. We employ extensive ($\sim$300 μs) all-atom molecular dynamics simulations to dissect the complete substrate exchange cycle of the bacterial NO_3^-/NO_2^- antiporter, NarK. We show that paired basic residues in the binding site prevent the closure of unbound protein and ensure the exchange of two substrates. Conformational transition only occurs in the presence of substrate, which weakens the electrostatic repulsion and stabilizes the transporter by $\sim$1.5 kcal/mol. Furthermore, we propose a state-dependent substrate exchange model, in which the relative spacing between the paired basic residues determines whether NO_3^- and NO_2^- bind simultaneously or sequentially. Overall, this work presents a general working model for the antiport mechanism within MFS family.

Introduction

Major facilitator superfamily (MFS) transporters move a wide spectrum of biologically relevant substrates (nutrients, drugs, waste, etc.) across cell membranes.[1–4] This superfamily is ubiquitous in all kingdoms of life and represents the largest and most diverse group of secondary active transporters.[5] Malfunction of MFS proteins has been associated with a multitude of diseases, such as cancer, type 2 diabetes, and Alzheimer's disease.[6–8] In plants, MFS transporters mediate macronutrients (C, N, and P) uptake and extrusion of deleterious compounds.[9] Owing to their physiological and pathophysiological significance, MFS proteins have been popular targets for structural and mechanistic investigations. MFS transporters can be classified into three types depending on their substrate-transport mechanism: uniporters, symporters, and anitporters.[10] Uniporters transport one substrate down the concentration gradient. Symporters translocate two or more substrates in the same direction. Antiporters transport a substrate and a co-substrate in opposite directions. Our understanding of transport mechanism is well advanced for symporters and uniporters, exemplified by the *Escherichia coli* lactose/H^+ symporter (LacY).[11–13] By contrast, molecular insights are less advanced for antiporters due to the lack of crystal structures representing different stages of the transport cycle.[5] Among all the MFS proteins with known structures, most of them are symporters and the outward-facing conformation of MFS antiporters has yet to be captured.[14] Recently, computational studies of few antiporters have advanced our understanding of antiport mechanism.[5,15–18] However, these studies were only able to simulate the transport of one substrate in the transport cycle, even with enhanced sampling techniques and state-of-the-art supercomputing.

Many important structural and mechanistic aspects of antiport mechanism remain elusive: (1) Can antiporters be distinguished from other transporters (uniporters and symporters) based on their protein architecture and conformational transition mechanism? (2) How is the empty transporter prevented from changing conformations in antiporters? (3) How does antiporter distinguish and switch two cargos? (4) How is substrate binding cou-

pled with global conformational changes? (5) Which residues regulate substrate recognition, binding, and release?

NarK represents a convenient model system for studying the functions of MFS antiporters because it has been characterized in a few conformations.[19,20] NarK is a NO_3^-/NO_2^- antiporter from *Escherichia coli*, coupling an outward flow of internal NO_2^- to the uptake of NO_3^- into cell.[20] It belongs to the NO_3^-/NO_2^- porter (NNP) family of the MFS. NNP proteins mediate the high-affinity translocation of NO_3^- and NO_2^- in both prokarytoes and eukaryotes (*i.e.* archaea, bacteria, fungi, yeast, algae, and plants).[21] NO_3^- is the most important source of mineral nitrogen for plant and NO_3^- availability greatly limits crop yields.[22] Plant NNP members (e.g. *Arabidopsis thaliana* NRT2.1 and NRT2.2) account for 80% of high-affinity NO_3^- uptake, rendering them essential under low nitrogen conditions.[9] A detailed understanding of the structure, transport mechanism, and conformational changes of NNP transporters would therefore substantially guide the engineering of these transporters to enhance agricultural productivity. Recent crystal structures revealed that NarK adopts a canonical MFS fold consisting of 12 transmembrane (TM) helices. These helices are organized into two structurally similar domains, the N-domain (TM1-TM6) and the C-domain (TM7-TM12). Membrane transporters generally work by an alternating-access mechanism.[23–25] Switching among outward-facing (OF), occluded (OC), and inward-facing (IF) states alternatively expose the substrate binding site to either side of the membrane. Although static snapshots of X-ray crystallography are critical, they are insufficient to explain the mechanistic details of such dynamic transitions and their coupling to chemical events supplying the energy.

This work explores the structural and mechanistic principles that characterize MFS antiporters through the study of a bacterial NO_3^-/NO_2^- antiporter, NarK. To elucidate the transport mechanism in atomic detail, we performed extensive ($\sim$300 μs) atomic-level molecular dynamics (MD) simulations[26,27] in conjunction with Markov state models (MSMs).[28–31] We focus on answering the previously stated structural and mechanistic questions of MFS antiporters. This work provides important information both for understanding NO_3^- uptake

by NNP transporters and for elucidating the antiport mechanism within MFS family.

Results and Discussion

NarK follows the canonical rocker-switch mechanism. All-atom MD simulations of NarK were initiated from the NO_3^--bound occluded crystal structure (PDB ID: 4U4W[20]). With a total of ∼300 μs unbiased MD simulations, the entire transport cycle of NarK and all the functionally relevant states were characterized. Markov state model (MSM) is a popular technique for extracting kinetic information of protein dynamics from MD simulation data.[28,29,32] All the simulation data were used to construct an MSM and the MSM hyper-parameters were selected systematically using a genetic algorithm technique (see Methods and Materials for details).[33] In the crystal structure (PDB ID: 4U4W[20]), Ser56 (TM1) hydrogen bonds with Ala275 (TM7) at the extracellular side and closes the pore tunnel. The hydrophobic interactions between Met151 (TM4) and Phe370 (TM10) act as an intracellular gate. We projected simulation data onto these two metrics as the opening and closure of the pore channel strictly determines the specific state such as IF, OC, and OF.[30,31,34]

Free energy landscape reveals the thermodynamic and kinetic information about the antiporter transport mechanism (Fig.1A). The canonical L-shaped free energy landscape displays three distinct free energy basins corresponding to IF, OC, and OF states, in line with the rocker-switch mechanism. The three crystal structures of NarK were labeled in the free energy landscape. Simulations agree with previous experimental studies[20] by showing that these structures are all from the most stable basin of the transporter, the IF state. The measurement of extracellular and intracellular distances shows that the helices of the N and C domains are ∼4.7 and ∼13.2 Å apart in the crystallized IF state (PDB ID: 4U4T, 4U4V[20]). However, the free energy landscape suggests that the intracellular distance between TM4 and TM10 can increase up to ∼15.0 Å for the energetically accessible IF state (Fig.1A and Supplementary Fig.1). As TM4 and TM10 move towards each other, NarK adopts an OC

state with intracellular distance reduced to ∼9.0 Å while extracellular distance increased to ∼7.5 Å. Previously reported OC state (PDB ID: 4U4W[20]) actually represents the partially occluded IF state. The intracellular side can be further closed as the intracellular distance decreasing from ∼11.5 to ∼9.0 Å (Fig.1A and Supplementary Fig.2). Finally, the extracellular distance increases to ∼12.0 Å and the antiporter adopts the OF state. The free energy barrier for the transition from IF to OC state is ∼1.5 kcal/mol and for the subsequent transition to OF state is ∼1.0 kcal/mol. The total free energy barrier for one complete cycle of NarK from IF to OF state is ∼2.5 kcal/mol. The relatively low energy barrier suggests that NarK can easily interconvert between IF, OC, and OF sates during the translocation process. This observation is consistent with the high-affinity uptake of NO_3^-/NO_2^- in NarK.[21,35]

To gain structural insights into the free energy landscape, three representative structures corresponding to the free energy minima of OF, OC, and IF states were extracted (Fig.1B). The molecular surfaces for these structures suggest that the substrate translocation pathway is the pore saddled by the N- and C-terminal domains. N- and C-domains change their relative positions to alternatively expose the substrate binding site to opposite sides of the membrane. These structural features reinforce the notion that NarK follows the rocker-switch mechanism.[25] The simulated IF and OC structures show good agreement with experimental NarK structures (PDB ID: 4U4V, 4U4T, 4U4W[20]) (Supplementary Fig.1 and 2). The most pronounced differences are in the opening of intracellular gating for both IF and OC states. The opening at the intracellular side in simulated structures is wider for IF state (∼2.0 Å) and narrower in OC state (∼2.5 Å). The predicted OF structure was compared to a fucose/H^+ symporter (FucP, PDB ID: 3O7Q[36]) due to the absence of OF structures in MFS antiporters (Supplementary Fig.3). The OF conformation superimposes well on the OF structure of FucP and exhibits similar opening at the extracellular side (Supplementary Fig.3). These findings suggest that the rocker-switch mechanism is shared among all MFS proteins, irrespective of their particular functions as a uniporter, symporter, or antiporter. Therefore, the diversity of MFS transporter functions is a result of changes in a few residues

in the binding pocket and translocation pathway.

Helix bending drives opening and closing of intracellular and extracellular gates. To investigate the conformational changes underlying the rocker-switch type movement, we computed root-mean-square fluctuation (RMSF) of each Cα atom in NarK during IF to OC and OC to OF transition (Fig.2A-F, Supplementary Fig.4). The intracellular tips of TM5, TM10, and TM11 show higher flexibility ($\sim$2.5 Å) during the IF to OC transition (Fig.2E-F). The extracellular tips of TM1, TM2, and TM7 undergo significant rearrangements ($\sim$2.5 Å) during the transition from OC to OF state (Fig.2C-D). The other helices (TM3, TM4, TM6, TM8, TM9, and TM12) represent a stable structural element that displays limited conformational changes ($\sim$1.0 Å) during the transport cycle. Previous structural analysis hypothesized the N-bundle remains rigid while TM7, TM10, and TM11 helices of the C-bundle move towards the N-bundle to close the intracellular vestibule.[20] In simulations, both the N- (TM5) and C-domains (TM10 and TM11) move towards each other with a higher magnitude ($\sim$1.0 Å higher) for the C-domain to close the intracellular side. It was also suggested that TM7 is involved in the closure of intracellular gate.[20] By contrast, the extracellular tip of TM7 exhibits large fluctuations ($\sim$2.5 Å) during OC to OF transition whereas the motions in the intracellular tip are negligible ($\sim$1.0 Å) during the entire transport cycle (Fig.2 C, D, G). This strongly indicates that TM7 is critical for the opening of the extracellular instead of intracellular gate. The simulations also shed light on the conformational transitions from OC to OF states, which were previously unknown due to the lack of OF structure. The extracellular half of TM1 and TM2 from N-domain, and TM7 of the C-domain move apart to open the extracellular vestibule.

The RMSF data illustrate that only helix tips show significantly higher flexibility over the entire transport cycle from IF to OC and OC to OF (Fig.2A-F, Supplementary Fig.4). These results suggest that the conformational transition between IF and OF is governed by the internal bending and straightening motion of the helices (TM1, TM2, TM5, TM7, TM10, and TM11), rather than the rigid body tilting movement of two bundles (Supplemen-

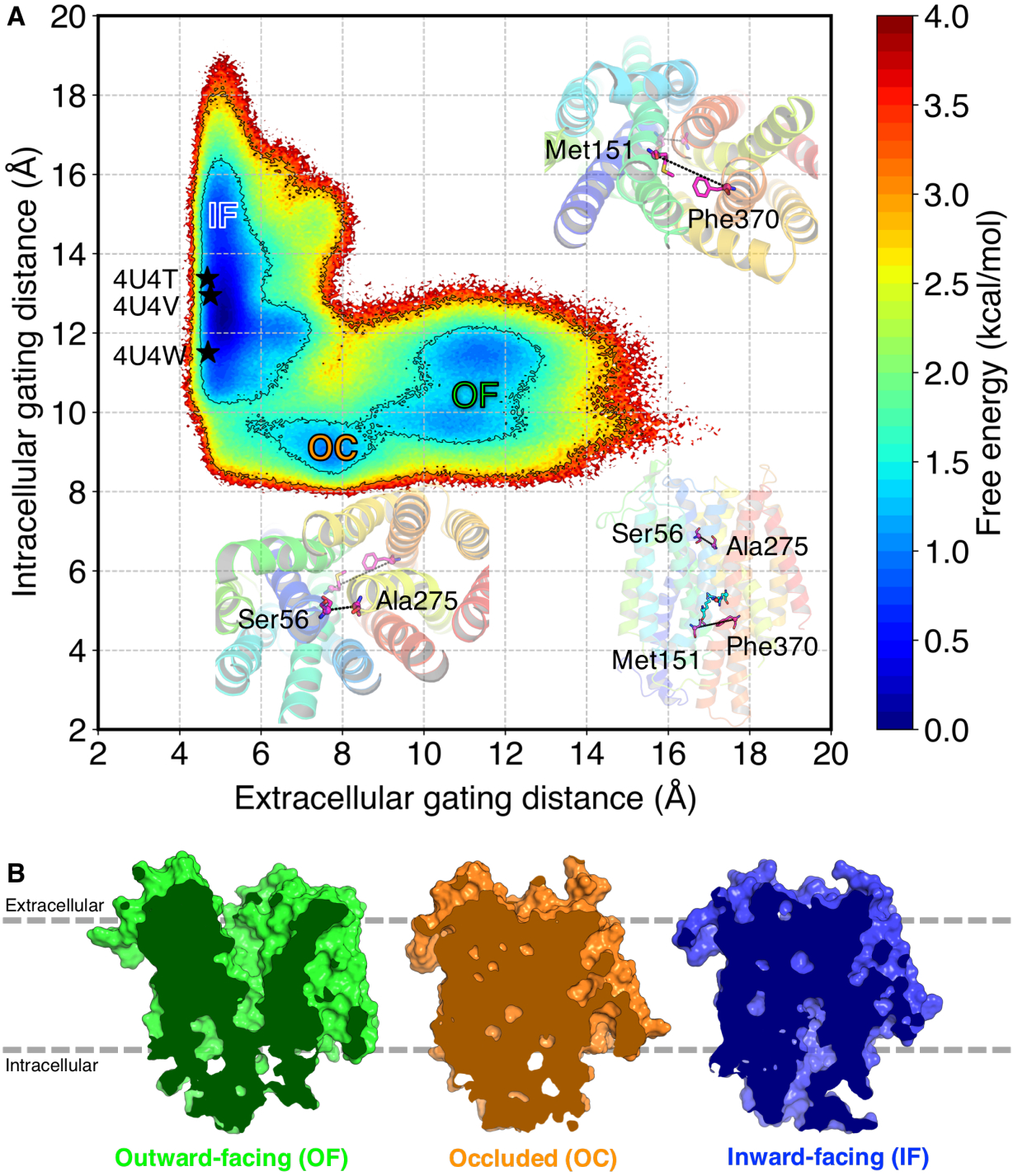


Figure 1: **Alternating access cycle of NarK.** A. MSM-weighted free energy landscape plot of NarK. The ∼300 μs MD simulation data were projected onto two-dimensional space of Ser56Cα (TM1) - Ala275Cα (TM7) and Met151Cα (TM4) - Phe370Cα (TM10) distances. Three crystal structures were labelled as black stars. B. Cross-sections of surface representations for the representative structures from the OF, OC, and IF states. The structures were viewed from the membrane plane.

tary Fig.5). The bending of TM10 and TM11 was also observed in the previous structural studies of NarK,[19,20] its closest homolog NarU,[35] and many other MFS family proteins,[36–40] implying a significance of flexibility of TM10 and TM11 within MFS family. An important feature of NNP family transporters is the presence of highly conserved glycine residues, which constitute the inner core of many TMs (TM2, TM4, TM5, TM7, TM8, TM10, and TM11).[19,20,35,41] It is interesting to note that large structural fluctuations in helices are located at or close to the conserved glycine residues (Fig.2G). These highly conserved glycine residues serve as pivot points for helix bending. The significance of glycine is underscored by previous mutations of glycine to alanine that abolish the NO_3^- uptake activity.[20] Together, our simulations revealed that the helix bending around highly conserved glycine residues governs the overall conformational change in NarK, and that these highly conserved glycine residues may support the structural flexibility in all members of the NNP family.

Hydrophobic and polar interactions lock the transport pathway. To identify the structural features that support the helix bending motions, we computed the key interactions within N- and C-terminal domains in three different conformational states (IF, OC, and OF). 10 sets of 1,000 MD configurations were randomly selected for each state. By comparing the average contact frequency in different states, we identified six layers of hydrophobic and polar interactions along the transport pathway (Fig.3). Many of these interactions were not reported in the previous studies which rely on the structural overlays of IF and partially occluded IF state.[19,20] Simulation data suggest that helix bending regulates the disruption and formation of these interactions and thus controls the opening and closing of the gates.

The exit of substrates to the extracellular side is blocked by a number of gating residues mainly in TM1 and TM7 (Fig.3 E1-3). At the extracellular side, residues A59 (TM1), V60 (TM1), N61 (TM1), A278 (TM7), M279 (TM7), K282 (TM7), and L291 (TM8) constitute the hydrophobic layer E1 (Fig.3 E1). All of these residues are conserved in NarK's closest homolog (NarU), except A59, suggesting a similar hydrophobic layer in NarU.[35] In layer E2, S56 (TM1), Q182 (TM5), S274 (TM7), and A275 (TM7) form an extensive network

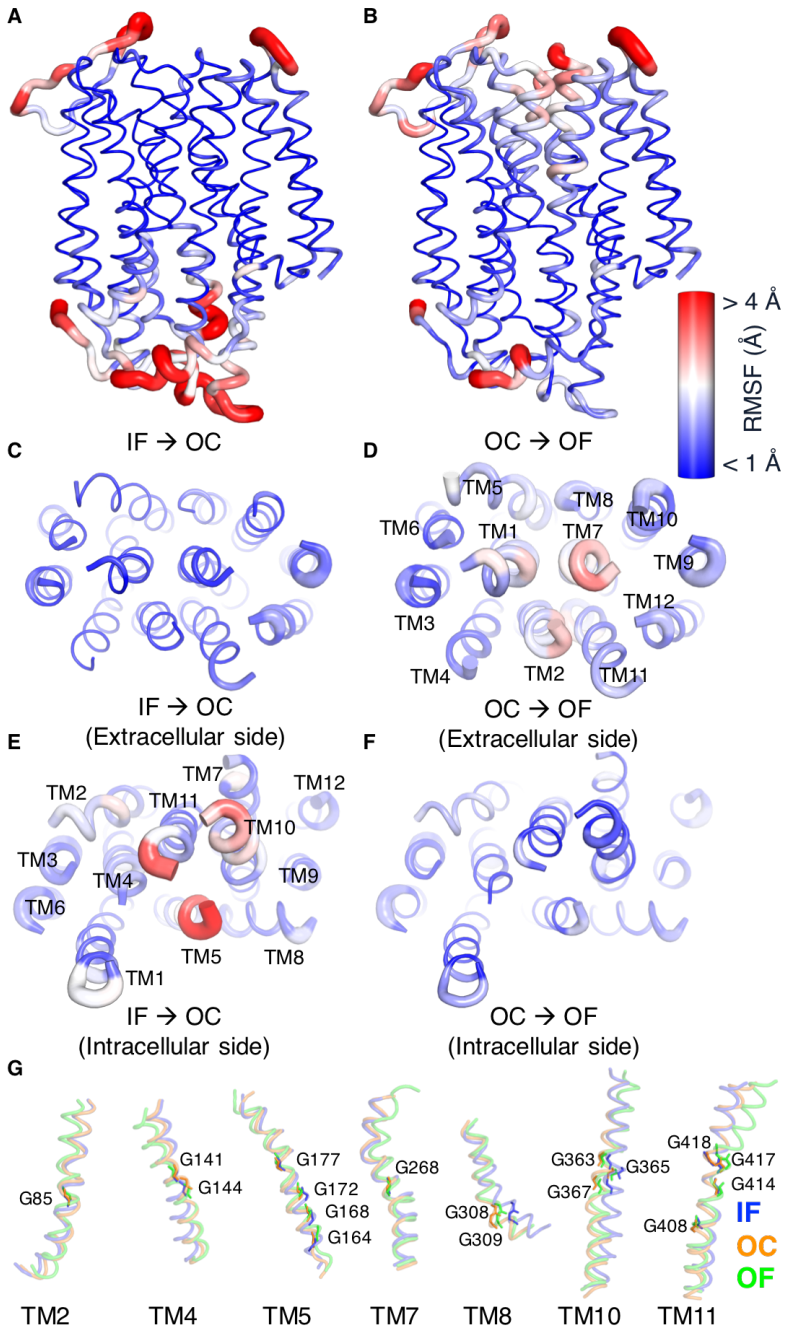


Figure 2: **Global conformational changes.** Root-mean-square fluctuation (RMSF) data were mapped onto the predicted IF structure of NarK for the IF to OC (A, C, E), and OC to OF transitions (B, D, F). Blue and red correspond to low (< 1 Å) to high (> 4 Å) RMSF values. Tube thickness corresponds to the RMSF value of each residue. Close-up views from the extracellular (C, D) and intracellular (E, F) side with transmembrane helices (TMs) labelled. G. Structural comparisons among IF, OC, and OF states. The superimposition was performed with Pymol[42] and for the helices containing conserved glycine residues: TM2, TM4, TM5, TM7, TM8, TM10, and TM11. The helices were shown in tube representations and colored based on the conformational states. Relevant glycine residues were labelled and shown as colored sticks.

of hydrogen bonds, facilitating the close packing of gating helices (Fig.3 E2). This polar interaction layer was also reported in NarU and alanine substitutions of S54 and Q180 (equivalent to S56 and Q182 in NarK) disrupted the binding reactions.[35] The hydrophobic residues in layer E3 (W52 (TM1), M53 (TM1), I271 (TM7), and A275 (TM7)) mediate extensive van der Waals contacts and stabilize the extracellular gate (Fig.3 E3). W52 is invariant among all the prokaryotic NNP family transporters. W50A mutation in NarU (corresponding to W52 in NarK) completely abolished binding to both NO_3^- and NO_2^-, implying a critical gating role in all prokaryotic NNP transporters. The extracellular gating residues are clustered in TM1 and TM7. Therefore, the bending of TM1, TM2, and TM7 can effectively control whether the extracellular gate is open or closed during the transition between the OF and OC states.

Consistent with experimental structure (PDB ID: 4U4W[20]), the intracellular gate consists of three layers of interactions (Fig.3 I1-3). In layer I1 (immediately beneath the substrate-binding pocket), residues F147, A148, and M151 of TM5 hydrophobically interact with F370 of TM10, and L407 of TM11 (Fig.3 I1). F147 and F370 are invariant among all NarK homologs in bacteria and eukaryotes. Previous functional analysis of NarK and its closest homologue (NarU) further supports the functional relevance.[20,35] F147A mutant shows decreased NO_3^--transport activity in NarK[20] and F367A mutation in NarU[35] (corresponding to F370 in NarK) abrogated binding to both NO_3^- and NO_2^-. In previously reported layer I2,[20] S155 in TM4 hydrogen bonds with A400 in TM11, and F158 and K160 in TM5 hydrogen bond with R378 in TM10. In simulations, however, the S155-A400 and R378-F158-K160 interactions are formed in both IF and OF states (Fig.3 I2). The S155-A400 interaction is observed in around 88% and 93% of IF and OF states. The R378-F158-K160 interaction occurs approximately in 75% and 100% of structures adopting the IF and OF conformations. These interactions are therefore not crucial for the conformational switch from IF to OF states. Simulations suggest that S155(TM4) and Q163 (TM5) form an extensive hydrogen bond network with S374 and R378 from TM10 (Fig.3 I2). In the partially occluded IF state

of NarK (PDB ID: 4U4W[20]), Q163 on unbent TM5 points away from the central pathway and thus the extensive polar interactions of layer I2 are missing (Supplementary Fig.6). The bending of TM5 and TM10 is crucial to form the polar interaction layer I2, thereby facilitating a much tighter packing between N and C bundles (Fig.3 I2). Near the intracellular surface, E21 of the N-terminal loop forms salt bridge with R397 on TM11 and blocks the access to the substrate binding site from the cytoplasm (Fig.3 I3). Upon the outward bending of TM11, E21 changes its salt bridge partner from R397 to K160 and allows the opening of the inside gate.

Substrate unoccupied transporter restricts conformational changes in an antiporter. Given the common conformational transition mechanism, how can uniporters, symporters, and antiporters be distinguished from each other? A major difference between a symporter and an anitporter is whether empty transporter changes conformation from IF to OF: symporters do, but antiporters do not. How is such a mechanistic scheme implemented in an antiporter? MD simulations monitor protein motions at an atomic level and thus are well-suited to answer these questions.[43,44] We projected all the simulation data onto a two-dimensional free energy landscape to understand the whole NO_3^- and NO_2^- exchange cycle (Fig.4, Supplementary Fig.7 and 8). To monitor the conformational transitions from IF to OF state, difference between the intracellular and extracellular gating distances was selected as x coordinate. Previous mutational analysis demonstrates that R89 and R305 directly bind substrates.[20] The difference between the closest distances from NO_2^- and NO_3^- to the center of R89 and R305 was selected as y coordinate to track the substate binding mode.

The free energy landscape exhibits eight distinct conformational states constituting the whole NO_3^- and NO_2^- exchange cycle (Fig.4). The white region in the center implies an inaccessible energy barrier between IF and OF states without a bound substate. The high energy barrier excludes the alternation between the OF and IF states without the aid of a substrate, thereby implementing the strict NO_3^- and NO_2^- exchanger function. Substrate binding is required in an antiporter to promote the conformational switch. While the sub-

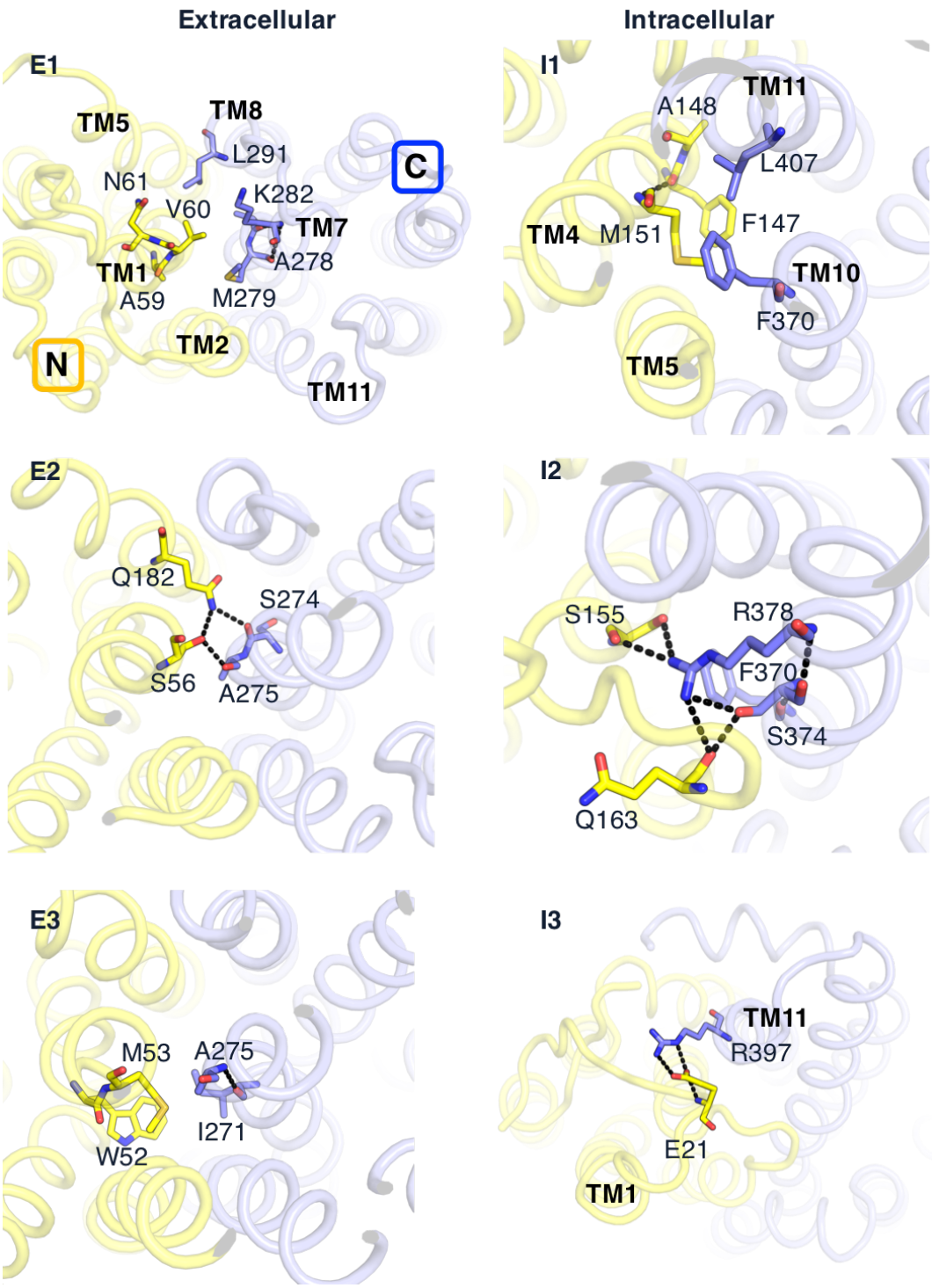


Figure 3: **Extensive hydrophobic and polar interactions along the transport pathway.** The three left panels represent the close-up views of the extracellular gating interactions seen from the extracellular side. The three right panels show the close-up views of the intracellular gating interactions seen from the intracellular side. N- and C-domain helices were colored yellow and light blue, respectively. The residues involved in each interaction layer were depicted by stick models and colored following the same scheme.

strate bound states correspond to the energy minima, substrate unbound states are located in the relatively higher energy regions (∼2 kcal/mol). Substate binding lowers the energy by ∼1.5 kcal/mol in both IF and OF states and enhances probability for the conformational switching between them.

Moreover, the free energy landscape enables the quantitative comparison between the NO_3^- and NO_3^- transport dynamics. Although NO_3^- and NO_2^- are equal in charge and differ merely by one oxygen atom, pronounced differences between them were found in the energy barrier between IF and OC states. The free energy barrier for the transition between OC and IF states is ∼1.0 kcal/mol with bound NO_3^-. When NO_2^- is bound, OC and IF states are separated by a barrier of ∼2.5 kcal/mol. According to the Arrhenius equation,[45] the ∼1.5 kcal/mol energy difference corresponds to ∼12-fold binding affinity difference between NO_3^- and NO_2^-. This is consistent with the 10-fold higher binding affinity for NO_3^- (∼33 μM) over NO_2^- (∼373 μM) in NarK's closest homolog, NarU.[35] The computational analyses suggest that NO_3^- favors the transition from OC to IF states and thus enhances the transport activity.

NO_3^- and NO_2^- switch differently in IF and OF states. One intriguing aspect of MFS antiporters as a whole is their capability to differentiate and exchange two similar substrates.[5] To gain structural insights into this capability, we extracted and compared eight representative structures corresponding to the free energy basins in Fig.4. Each free energy basin corresponds to a distinct substrate binding mode. In agreement with previous crystal structures,[20] NO_3^- and NO_2^- bind to NarK at exactly the same position (Fig.5). This binding pocket is formed by the side chains of highly conserved R89, G144, F147, F49, G171, N175, R305, S366, F267, Y263, and S411 (Fig.5B). The functional significance of these residues has also been confirmed by mutational analysis directly performed on NarK.[20] Two conserved and positively charged residues R89 from TM2 and R305 from TM8 directly bind to substrates.[20,46] Mutation of either arginine into lysine decreases transport activity of NarK. Without bound substrate, the electrostatic repulsion between R89 and R305 prevents the

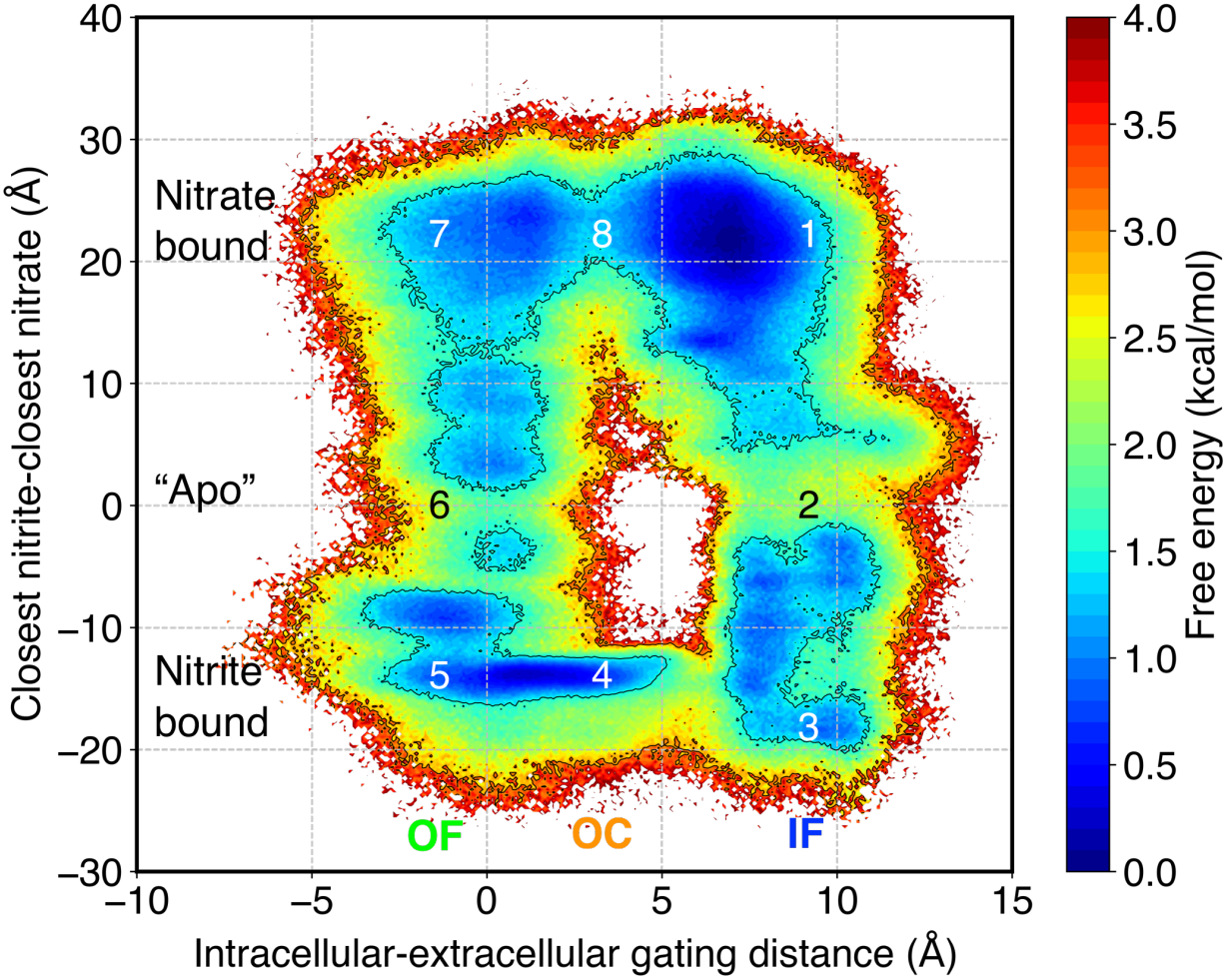


Figure 4: **MSM-weighted free energy landscape of the entire NO_3^-/NO_2^- exchange cycle.** The difference between the intracellular and extracellular gating distances differentiates IF, OC, and OF states. The difference between the closest distances from NO_2^- and NO_3^- to the center of R89 and R305 shows different substrate binding conditions. The conformational states were depicted as (1) NO_3^--bound IF, (2) *"apo"* IF, (3) NO_2^--bound IF, (4) NO_2^--bound OC, (5) NO_2^--bound OF, (6) *"apo"* OF, (7) NO_3^--bound OF, and (8) NO_3^--bound OC.

closure of the pore channel and thus implements the strict substrate exchange function in NarK. Paired basic residues are also essential for substrate binding in several other MFS antiporters such as *Escherichia coli* sn-glycerol-3-phosphate transporter (GlpT),[47] *Escherichia coli* Hexose-6-phosphate:phosphate antiporter (UhpT),[48,49] and the oxalate:formate transporter from *Oxalobacter formigenes* (OxlT).[50,51] This suggests that paired basic residues may be the structural determinant for the exchange function in antiporters transporting anionic substrates. There is a single binding pocket in NarK for two different substrates. The question is how NarK distinguishes between two similar cargo molecules such as NO_3^- and NO_2^-. The differences between NO_3^- and NO_2^- binding are strikingly clear in the OC state (state 4 and 8 in Fig.5B). The additional oxygen atom of NO_3^- allows the formation of additional hydrogen bonds with R89 and R305 and connects the two halves of NarK tightly. This may accelerate the OC to IF transition and lead to ∼1.5 kcal/mol difference in Fig.4.

The next question is how antiporter switches the substrates. A common hypothesis for exchangers is the "ping-pong" mechanism.[52] It assumes that there is only one binding site and that the protein never binds both substrates simultaneously. The antiporter must export a substrate first to import a second substrate. But NarK exhibits state-dependent substrate exchange behavior. When NarK adopts IF state (state 2 in Fig.5B), the binding of both substates is preferred (∼65 %) over the unbound conformation (∼35 %) (Supplementary Fig.9A). In comparison, ∼90 % of structures adopt unbound conformation in OF state (state 6 in Fig.5B) (Supplementary Fig.9B). This is likely due to the relative spacing between R89 and R305. In IF state, the shortest distance between R89 and R305 side chains is ∼9.0 Å. R89 and R305 could not simultaneously coordinate the small anion (∼1.96 Å for NO_3^-) (state 1, 3 in Fig.5B). During the IF to OF transition, helix bending moves R305 and R89 closer to each other by ∼2.0 Å, which allows both R89 and R305 to form hydrogen bonds with the substrate (state 5, 7 in Fig.5B). From these data, we propose a state-dependent exchange model. In IF state, NO_3^- and NO_2^- simultaneously bind to NarK. The electrostatic repulsion between the two anions facilitates the release of one substrate and exchanges the substrates.

In OF state, the first substrate leaves the binding site and then the second substrate comes in and binds the protein.

Y263 and R305 couple substrate binding and conformational changes. A fundamental question in the study of MFS proteins is how can local substrate binding initiate the global conformational changes. Comparison of the binding pockets reveals that Y263 on TM7 and R305 on TM8 are associated with substrate binding and the conformational changes during the transport cycle (Fig.5B, Supplementary Fig.10). In the IF and OF states, there is no interaction between Y263 and substrate in the binding pocket. In the OC states (state 4 and 8 in Fig.5B), however, the phenol hydroxyl group of Y263 hydrogen bonds with both R305 (TM8) guanidinium group and the substrate, thereby connecting the extensive hydrogen bond network among TM1, TM2, TM7, TM5, TM10, and TM11. The importance of Y263 and R305 is consistent with previous mutational studies.[20] Y263F and R305K mutants in NarK completely abolished the transport activity even under high isopropyl-b-D-thiogalactoside (IPTG) concentration, which is used to induce protein expression.[20] Whereas R89K mutant rescued the NO_3^--uptake activity under the highest IPTG concentration.[20] These results support that the hydrogen bond between the side chains of Y263 and R305 is critical for the transport activity.

This hydrogen bond network rearrangement involving Y263 and R305 provides a plausible mechanism of the coupling between substrate binding and conformational changes (Fig.5B, Supplementary Fig.10): (1) In IF or OF states, the anionic substrate initially binds weakly to the binding site and neutralizes the basic residues: R89 and R305. At this stage, Y263 does not participate in binding. (2) The side chain of Y263 then moves closer toward and interacts with the substrate. This elicits the tighter substrate binding to the transporter. (3) The binding energy released by this stronger interaction is used to overcome the energy barrier from IF to OC or OF to OC transitions. The compact hydrogen bond network at the binding pocket turns on the helix bending motions. Since these helices constitute the N and C bundles, the bending motions directly result in the global conformational changes. Taken

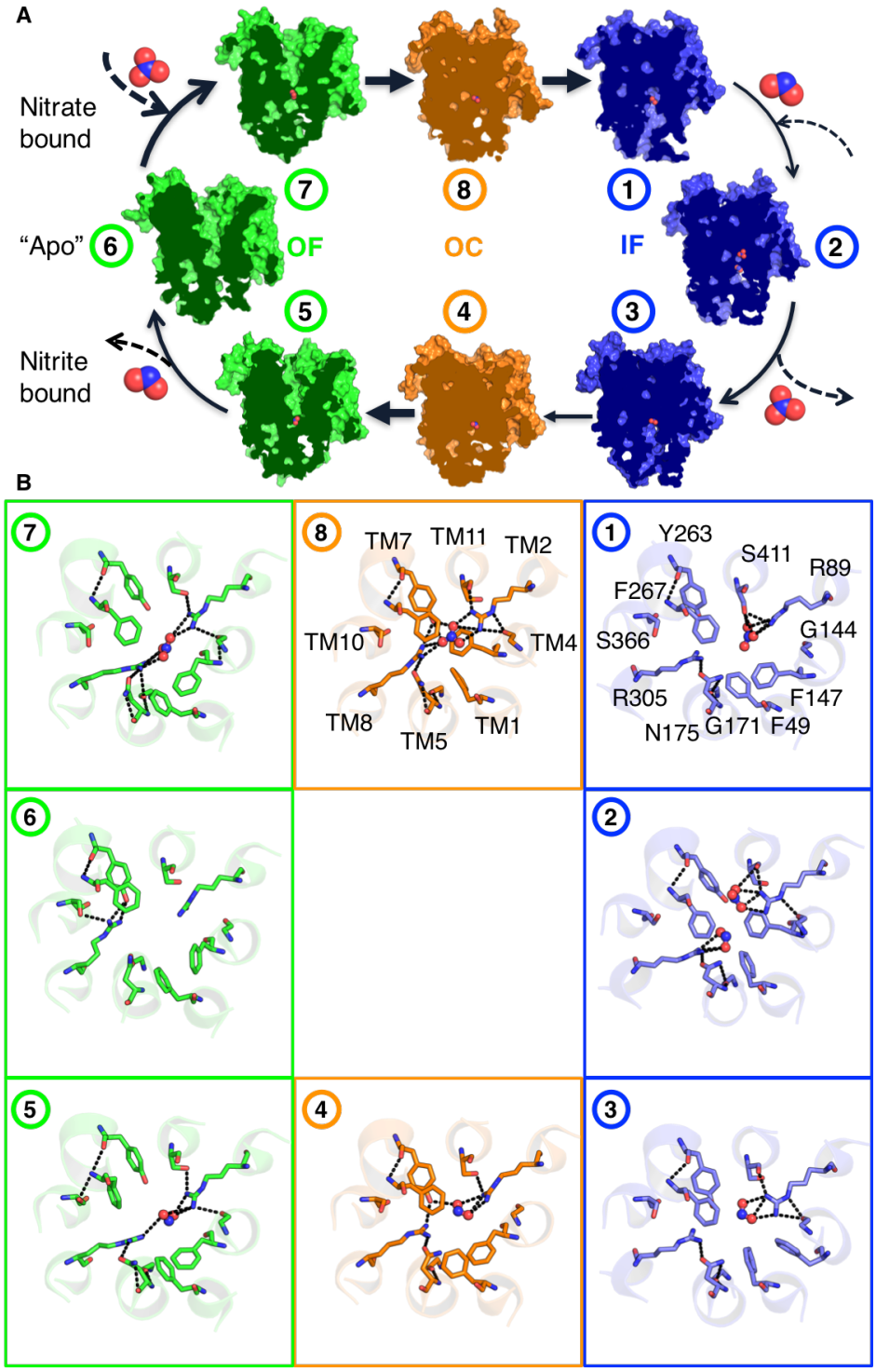


Figure 5: **NO_3^- and NO_2^- switching mechanism.** A. Cross-section of surface representations for the representative structures from the intermediate states along the exchange cycle. The number corresponds to eight conformational states in Fig.4. The bound substrates (NO_3^- and NO_2^-) were shown in ball and sticks. B. Close-up view of the binding sites visualizing the residues coordinating binding of NO_3^- and NO_2^-. The dotted black lines represent hydrogen bonds among residues. Substrates and binding pocket residues were represented in ball-and-stick and sticks, respectively.

together, Y263 and R305 play a pivotal role in the coupling mechanism of the substrate binding and global conformational changes.

A model for substrate recognition, binding, exchange, and translocation. The last question is which residues regulate substrate recognition, binding, and release? To identify key residues in substrate recognition, binding, and transport, we calculated interactions between NarK and substrates during the course of MD simulations. The contact frequency computation reveals that NO_3^- and NO_2^- take the same transport pathway (central panel in Fig.6). The exchange between NO_3^- from the outside and NO_2^- from the inside can be described by the following steps (Fig.6).

(1) At first, the extracellular gating residues are ∼12.8 Å apart and favor NO_3^- binding at the extracellular side. The positively charged residue K282 on the extracellular side of TM7 acts as a "hook", which appears to recruit the negatively charged NO_3^- and escort the NO_3^- deep inwards to make contact with the backbones of L291 and I290 on TM8 (Fig.6(1)). (2) After transient coordination with K282 and nearby L291 and I290, NO_3^- rapidly diffuses deep into the pore channel through W52 (Fig.6(2)). (3) NO_3^- binds to the two positively charged and highly conserved residues (R89 and R305) in the binding site. NO_3^- binding weakens the electrostatic repulsion between R89 and R305 and forms extensive hydrogen bond network in the binding pocket. The formation of hydrogen bonds between Y263, R305, and the substrate turns on the helix bending motions, which enables the outward halves of the core domains to approach each other and facilitates the conformational change from the OF to OC state. At this juncture, the extracellular gating distance decreases to ∼10.4 Å (Fig.6(3)). (4) The formation of three layers of interactions finally closes the extracellular gate (∼4.5 Å), which occurs with an opening of the intracellular vestibule (∼17.4 Å). Due to the outward bending of TM5, TM10, and TM11, R89 and R305 move away from each other. NO_3^- is now bound to only one of the arginine residues. This makes space for the binding of NO_2^- from the inside solution (Fig.6(4)). (5) The repulsion between NO_3^- and NO_2^- facilitates the dissociation of NO_3^- from the central binding site. NO_3^- enters the intracellular part of NarK

and forms a hydrogen bond with M151 (Fig.6(5)). (6) NO_3^- is released to the inside solution mediated by positively charged and polar residues (K160, R378, and Q161) located on the inward surface (Fig.6(6)). (7) Once NO_3^- dissociates, only NO_2^- is bound in the binding site. This allows the closure of the inside gate and leads to conformational transition from IF to OC (Fig.6(7)). (8) As the closure of the inside gate (∼10.4 Å) through the inward bending of TM5, TM10, and TM11, NO_2^- binds to both R89 and R305 which triggers the opening of the outside gate (∼11.8 Å) (Fig.6(8)). (9) NO_2^- follows the same pathway as NO_3^- to get released to the outside solution. NO_2^- is dissociated from the binding site through W52 (Fig.6(9)). (10) NO_2^- is released to the outside solution mediated by the positively charged K282 on the outward surface (Fig.6(10)).

Discussion

In an attempt to decipher the molecular origin of MFS antiporters, we studied a bacterial NO_3^-/NO_2^- antiporter, NarK. Extensive all-atom MD simulations (∼300 μs in total) allow us to characterize the unbiased dynamics along the complete substrate exchange cycle of NarK. To establish NarK as a model for understanding MFS antiporters, we focus on addressing the following structural and mechanistic questions.

(1) Can antiporters be distinguished from uniporters and symporters according to protein architecture and conformational transition mechanism? Simulation results suggest that NarK adopts the canonical MFS fold and rocker-switch mechanism. Rather than the overall protein architecture, a few residues in the binding pocket and translocation pathway result in the different MFS functions.

(2) How is the empty transporter prevented from changing conformations in antiporters? In NarK, two absolutely conserved and positively charged arginine residues (R89 and R305) in the central binding site prevent the closure of the unbound protein. Substrate (NO_3^- or NO_2^-) binding neutralizes the charge and weakens the electrostatic repulsion, thus permitting

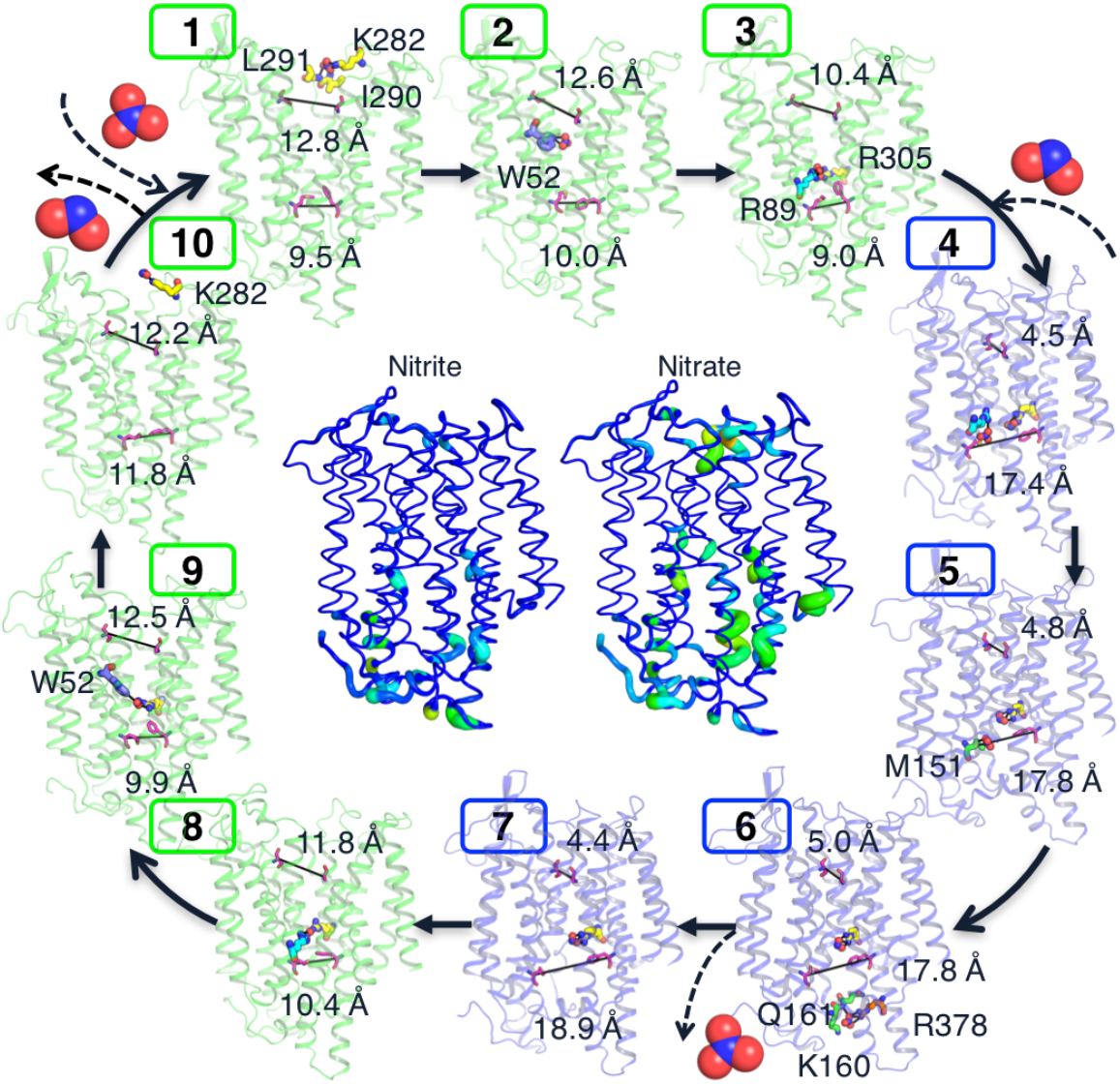


Figure 6: **Substrate recognition, binding, exchange, and translocation.** The central panel shows the MSM-weighed substrate-residue contact frequency for NO_2^- and NO_3^-. The tube thickness indicates the frequency of residue-substrate contact. (1-3) In the OF state, NO_3^- binds to NarK. (4) Structural transitions from OF to IF move R89 and R305 apart. NO_2^- from inside solution binds the other unoccupied arginine. (5-7) NO_3^- slides down to the cytosolic side. (8) The release of NO_3^- permits the closing of inside gates and triggers the conformational transitions from IF to OF. (9-10) NO_2^- follows the same translocation pathway with NO_3^- and gets released to the outside. For clarity, only R89 and R305 were shown in the binding site to illustrate the substrate binding. The dotted black lines represent the distances between gating residues (magenta sticks). Substates and relevant residues were represented as ball-and-sticks model.

the conformational transitions. Paired basic residues were also found critical for substrate binding in several other MFS antiporters (*i.e.* GlpT, UhpT, and OxlT[47–51]). This suggests that paired basic residues in the binding pocket may be the common structural determinant for the exchange function in antiporters which transport anionic substrates.

(3) How does antiporter distinguish and switch two cargos? Breakage and formation of hydrogen bonds rearranges the binding pocket to differentiate and fit two different cargo molecules. The additional hydrogen bonds formed by the extra oxygen of NO_3^- connect the two halves of NarK tighter than NO_2^- and result in higher binding affinity of NO_3^-. Regarding the switching mechanism, our results support a state-dependent substrate exchange mechanism. This is due to relative spacing between R89 and R305. Both substrates bind in IF state and the electrostatic repulsion between NO_3^- and NO_2^- speeds up the release of one substrate. By contrast, R89 and R305 in OF state are closer to each other by ∼2.0 Å and both make contacts with the substrate, leading to sequential binding.

(4) How is substrate binding coupled with global conformational changes? The coupling between substrate binding and global conformational changes is ensured through the movement of Y263 and R305 at the binding pocket. Upon substrate binding, the side chain of Y263 moves closer to interact with R305 and the substrate and forms an extensive hydrogen bond network. The shrinkage of the binding pocket triggers the helix bending motions and directly results in the global conformational changes.

(5) Which residues regulate substrate recognition, binding, and release? This work identifies all the relevant residues along the translocation pathway and explains the complete substrate recognition, binding, exchange, and release process.

Our results also identify structural features determining the high-affinity transport activity of all NNP proteins. The highly conserved glycine residues serve as the pivot points of helix bending, which is energetically efficient with a relatively low amount of energy (∼2.5 kcal/mol) required to complete the cycle from IF to OF states. Another structural feature is the gating residues lining up the translocation pathway. Previous studies of GlpT antiporter

suggested that salt bridges are important for gating on both sides of the binding pocket.[53] However, such salt bridges are lacking on both sides of the substrate-binding site of NarK. Two hydrophobic interaction layers (layer E3 and layer I1 in Fig.3) occlude the substrate and require only modest local structural changes to break or form.

Summarized, by performing the extensive all-atom MD simulations, the present study provides the complete picture of the NarK antiporter function (Fig.6). The detailed information provided in this study sheds light on the fundamental mechanism of all MFS antiporters. Furthermore, the relevant residues identified in this work can be used to engineer NNP proteins in crops to achieve higher crop productivity.

Methods and Materials

Molecular Dynamics (MD) Simulation. The NO_3^--bound OC NarK crystal structure (PDB ID: 4U4W[20]) was downloaded from Protein Data Bank (PDB)[54] and used as the starting coordinates for MD simulations. The chain termini were capped with neutral acetyl and methylamide groups. The membrane-protein MD system was built using the Membrane Builder plugin in CHARMM-GUI.[55] The protein was embedded in a phosphatidylcholine (POPC) bilayer, solvated in box with dimension of 72*72*102 with TIP3P water molecules, and neutralized by adding 22 Na^+ ions. 28 NO_3^- and 28 NO_2^- ions were randomly added to the MD system using Packmol.[56] The final MD system contains 51,315 atoms. The MD system was minimized using conjugate gradient method for 10,000 steps, slowly heated from 0 to 10 K, and then 10 to 300 K over a period of 1 ns each in canonical (NVT) and isothermal-isobaric (NPT) ensembles, and equilibrated in NPT ensemble for 50 ns. All simulations were implemented using Amber14 package[57] employing Amber ff14SB[58] and GAFF[59] force field, and carried out in the NPT conditions (300K and 1 atm) maintained using a Berendsen thermostat and a Berendsen barostat.[60] Periodic boundary conditions (PBC) were applied to all simulations, Particle Mesh Ewald was used to treat long-range non-bonded interactions,[61]

and the SHAKE algorithm was used to constrain hydrogen-containing bonds.[62] A 2 fs time step was used throughout all simulations.

Adaptive Sampling. To overcome the timescale gap between the large timescale for the NO_3^- and NO_2^- exchange processes in NarK and the short range timescale that can be explored in MD simulations, we employed a sampling strategy called adaptive sampling, which has demonstrated the efficiency of capturing various biologically relevant processes and tremendous value in time and resource savings.[63–65] The procedure of adaptive sampling is: (1) run a series of short MD simulations from initial seeds in parallel, (2) cluster simulation data on the basis of biologically relevant metrics, (3) select the structures from the least populated clusters as seeds for the next run of simulation, and (4) repeat steps 2-3 until the desired sampling is achieved. The extracellular and intracellular gating distances were used as adaptive sampling metrics for the conformational sampling of different states and the distances between the closest substrate and the center of two arginine residues (R89 and R305) in the binding pocket were used to capture the NO_3^- and NO_2^- exchange process. A total of $\sim$300 μs simulation data were obtained and used for further data analysis.

Markov state model (MSM) Construction. An issue with adaptive sampling is the introduced sampling bias as each new round of simulations starts from the least populated states, which may alter the real equilibrium population of the states. To eliminate the sampling bias, we constructed Markov state model (MSM) to statistically stitch all the short simulation data and estimate the transition probability matrix between all of the conformational states.[28,29,32] Due to the limit of the large simulation data, we observed that the difference between the raw data and the MSM weighted landscapes is small (Supplementary Fig.11-14). The procedure of constructing a MSM is: (1) featurizing the trajectory data using a set of Cα distances between residues; (2) decomposing the featurized data using the time-lagged independent component analysis (tICA) technique,[66,67] which finds the slowest collective motions in the system through linear combinations of the input features (Cα distances between residues in our case); and (3) clustering the decomposed data into

conformational microstates using Mini-batch K-means clustering algorithm. To select the optimal hyper-parameters (Cα contacts, number of tICA components, and number of clusters) systematically and automatically, we employed a genetic algorithm based technique developed from our lab (details of this algorithm are explained elsewhere[33]) (Supplementary Table 1 and 2). The whole idea is mimicking the natural selection and evolving the best combinations of these hyper-parameters based on the fitness score. The generalized matrix Raleigh quotient (GMRQ) is used as the fitness score to quantify the quality of MSM models, as GMRQ is the sum of the eigenvalues of the transition matrix estimated from MSM and the higher the GRMQ, the better the MSM is at capturing the slowest motions in the system.[68,69] The convergence of GMRQ scores is shown in Supplementary Fig.15. Eventually, the highest scored combination of hyper-parameters, that is 61 Cα contact distances, 10 tICA components, and 400 clusters, was chosen for the final MSM construction (details are shown in Supplementary Table 3, Supplementary Fig.16). A Markovain lag time 15 ns was chosen from the implied timescale plot to construct the MSM (Supplementary Fig.17). The Osprey variational cross-validation package was used in the genetic algorithm workflow to cross validate the MSMs by varying the hyper-parameters.[70] MSMBuilder 3.6[71] was used to construct MSMs, CPPTRAJ[72] module in AMBER14[57] and MDTraj1.7[68] were used to analyze simulation trajectories, and VMD1.9.2[73] and PyMol 3[42] were used to visualize MD snapshots.

Acknowledgement

Authors thank the Blue Waters sustained-petascale computing project, which is supported by the National Science Foundation (Awards OCI-0725070 and ACI-1238993) and the state of Illinois. D.S. acknowledges support from the New Innovator Award from the Foundation for Food and Agriculture Research and NSF Early Career Award (MCB 1845606). J.F. was supported by Chia-chen Chu Fellowship and Harry G. Drickamer Graduate Research

Fellowship from University of Illinois, Urbana-Champaign, IL.

Author Contributions

D.S. conceived and supervised this study. J.F. and B.S. performed simulations. J.F. analyzed the simulation data, made figures, and wrote the manuscript with inputs from D.S.

Competing Interests

The authors declare no competing financial interests.

References

(1) Marger, M. D.; Saier, M. H. A major superfamily of transmembrane facilitators that catalyse uniport, symport and antiport. *Trends Biochem. Sci.* **1993**, *18*, 13–20.

(2) Pao, S. S.; Paulsen, I. T.; Saier, M. H. Major facilitator superfamily. *Microbiol. Mol. Biol. Rev.* **1998**, *62*, 1–34.

(3) Reddy, V. S.; Shlykov, M. A.; Castillo, R.; Sun, E. I.; Saier, M. H. The major facilitator superfamily (MFS) revisited. *FEBS J.* **2012**, *279*, 2022–2035.

(4) Quistgaard, E. M.; Löw, C.; Guettou, F.; Nordlund, P. Understanding transport by the major facilitator superfamily (MFS): Structures pave the way. *Nat. Rev. Mol. Cell Biol.* **2016**, *17*, 123–132.

(5) Law, C. J.; Maloney, P. C.; Wang, D.-N. Ins and outs of major facilitator superfamily antiporters. *Annu. Rev. Microbiol.* **2008**, *62*, 289–305.

(6) Augustin, R. The protein family of glucose transport facilitators: It's not only about glucose after all. *IUBMB Life* **2010**, NA–NA.

(7) Cura, A. J.; Carruthers, A. Role of monosaccharide transport proteins in carbohydrate assimilation, distribution, metabolism, and homeostasis. *Compr. Physiol.* **2011**, *2*, 863–914.

(8) Smith, D. E.; Clémençon, B.; Hediger, M. A. Proton-coupled oligopeptide transporter family SLC15: Physiological, pharmacological and pathological implications. *Mol. Aspects Med.* **2013**, *34*, 323–336.

(9) Niño-González, M.; Novo-Uzal, E.; Richardson, D. N.; Barros, P. M.; Duque, P. More transporters, more substrates: The Arabidopsis major facilitator superfamily revisited. *Mol. Plant* **2019**, *12*, 1182–1202.

(10) Forrest, L. R.; Krämer, R.; Ziegler, C. The structural basis of secondary active transport mechanisms. *Biochim. Biophys. Acta* **2011**, *1807*, 167–188.

(11) Kaback, H. R. Structure and mechanism of the lactose permease. *C. R. Biol.* **2005**, *328*, 557–567.

(12) Madej, M. G. Function, structure, and evolution of the major facilitator superfamily: The LacY Manifesto. *Adv. Behav. Biol.* **2014**, *2014*, 1–20.

(13) Kaback, H. R. A chemiosmotic mechanism of symport. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 1259–1264.

(14) Saier, M. H.; Reddy, V. S.; Tsu, B. V.; Ahmed, M. S.; Li, C.; Moreno-Hagelsieb, G. The transporter classification database (TCDB): Recent advances. *Nucleic Acids Res.* **2015**, *44*, D372–D379.

(15) Moradi, M.; Enkavi, G.; Tajkhorshid, E. Atomic-level characterization of transport cycle thermodynamics in the glycerol-3-phosphate:phosphate antiporter. *Nat. Commun.* **2015**, *6*.

(16) Alhadeff, R.; Warshel, A. Simulating the function of sodium/proton antiporters. *Biophys. J.* **2016**, *110*, 12a.

(17) Takemoto, M.; Lee, Y.; Ishitani, R.; Nureki, O. Free energy landscape for the entire transport cycle of triose-phosphate/phosphate translocator. *Structure* **2018**, *26*, 1284–1296.e4.

(18) Okazaki, K.; Wöhlert, D.; Warnau, J.; Jung, H.; Özkan Yildiz,; Kühlbrandt, W.; Hummer, G. Mechanism of the electroneutral sodium/proton antiporter PaNhaP from transition-path shooting. *Nat. Commun.* **2019**, *10*.

(19) Zheng, H.; Wisedchaisri, G.; Gonen, T. Crystal structure of a nitrate/nitrite exchanger. *Nature* **2013**, *497*, 647–651.

(20) Fukuda, M.; Takeda, H.; Kato, H. E.; Doki, S.; Ito, K.; Maturana, A. D.; Ishitani, R.; Nureki, O. Structural basis for dynamic mechanism of nitrate/nitrite antiport by NarK. *Nat. Commun.* **2015**, *6*.

(21) Forde, B. G. Nitrate transporters in plants: structure, function and regulation. *Biochim. Biophys. Acta* **2000**, *1465*, 219–235.

(22) Plett, D. C.; Holtham, L. R.; Okamoto, M.; Garnett, T. P. Nitrate uptake and its regulation in relation to improving nitrogen use efficiency in cereals. *Semin. Cell Dev. Biol.* **2018**, *74*, 97–104.

(23) Mitchell, P. A general theory of membrane transport from studies of bacteria. *Nature* **1957**, *180*, 134–136.

(24) Jardetzky, O. Simple allosteric model for membrane pumps. *Nature* **1966**, *211*, 969–970.

(25) Drew, D.; Boudker, O. Shared molecular mechanisms of membrane transporters. *Annu. Rev. Biochem.* **2016**, *85*, 543–572.

(26) Karplus, M.; McCammon, J. A. Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* **2002**, *9*, 646–652.

(27) Feng, J.; Chen, J.; Selvam, B.; Shukla, D. Computational microscopy: Revealing molecular mechanisms in plants using molecular dynamics simulations. *Plant Cell* **2019**, *31*, tpc.119.tt1219.

(28) Chodera, J. D.; Noé, F. Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* **2014**, *25*, 135–144.

(29) Husic, B. E.; Pande, V. S. Markov state models: From an art to a science. *J. Am. Chem. Soc.* **2018**, *140*, 2386–2396.

(30) Selvam, B.; Yu, Y.-C.; Chen, L.-Q.; Shukla, D. Molecular basis of the glucose transport mechanism in plants. *ACS Cent. Sci.* **2019**, *5*, 1085–1096.

(31) Cheng, K. J.; Selvam, B.; Chen, L.-Q.; Shukla, D. Distinct substrate transport mechanism identified in homologous sugar transporters. *J. Phys. Chem. B* **2019**, *123*, 8411–8418.

(32) Shukla, D.; Hernández, C. X.; Weber, J. K.; Pande, V. S. Markov State Models Provide Insights into Dynamic Modulation of Protein Function. *Acc. Chem. Res.* **2015**, *48*, 414–422.

(33) Chen, Q.; Feng, J.; Mittal, S.; Shukla, D. Automatic feature selection in markov state models using genetic algorithm. *J. Comput. Sci. Educ.* **2018**, *9*, 14–22.

(34) Selvam, B.; Mittal, S.; Shukla, D. Free energy landscape of the complete transport cycle in a key bacterial transporter. *ACS Cent. Sci.* **2018**, *4*, 1146–1154.

(35) Yan, H.; Huang, W.; Yan, C.; Gong, X.; Jiang, S.; Zhao, Y.; Wang, J.; Shi, Y. Structure and mechanism of a nitrate transporter. *Cell Rep.* **2013**, *3*, 716–723.

(36) Dang, S.; Sun, L.; Huang, Y.; Lu, F.; Liu, Y.; Gong, H.; Wang, J.; Yan, N. Structure of a fucose transporter in an outward-open conformation. *Nature* **2010**, *467*, 734–738.

(37) Abramson, J. Structure and mechanism of the lactose permease of Escherichia coli. *Science* **2003**, *301*, 610–615.

(38) Lemieux, M.; Huang, Y.; Wang, D.-N. Glycerol-3-phosphate transporter of Escherichia coli: Structure, function and regulation. *Res. Microbiol.* **2004**, *155*, 623–629.

(39) Yin, Y. Structure of the multidrug transporter EmrD from Escherichia coli. *Science* **2006**, *312*, 741–744.

(40) Solcan, N.; Kwok, J.; Fowler, P. W.; Cameron, A. D.; Drew, D.; Iwata, S.; Newstead, S. Alternating access mechanism in the POT family of oligopeptide transporters. *EMBO J.* **2012**, *31*, 3411–3421.

(41) Unkles, S. E.; Karabika, E.; Symington, V. F.; Cecile, J. L.; Rouch, D. A.; Akhtar, N.; Cromer, B. A.; Kinghorn, J. R. Alanine scanning mutagenesis of a high-affinity nitrate transporter highlights the requirement for glycine and asparagine residues in the two nitrate signature motifs. *Biochem. J.* **2012**, *447*, 35–42.

(42) Schrödinger, LLC, The PyMOL molecular graphics system, version 1.8. **2015**,

(43) Moffett, A. S.; Shukla, D. Using molecular simulation to explore the nanoscale dynamics of the plant kinome. *Biochem. J.* **2018**, *475*, 905–921.

(44) Feng, J.; Shukla, D. Characterizing conformational dynamics of proteins using evolutionary couplings. *J. Phys. Chem. B.* **2018**, *122*, 1017–1025.

(45) Connors, K. A. *Chemical kinetics: the study of reaction rates in solution*; Wiley-VCH Verlag GmbH, 1990.

(46) Unkles, S. E.; Rouch, D. A.; Wang, Y.; Siddiqi, M. Y.; Glass, A. D. M.; Kinghorn, J. R. Two perfectly conserved arginine residues are required for substrate binding in a high-affinity nitrate transporter. *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 17549–17554.

(47) Huang, Y. Structure and mechanism of the glycerol-3-phosphate transporter from Escherichia coli. *Science* **2003**, *301*, 616–620.

(48) Ambudkar, S. V.; Anantharam, V.; Maloney, P. C. UhpT, the sugar phosphate antiporter of Escherichia coli, functions as a monomer. *J. Biol. Chem.* **1990**, *265*, 12287–12292.

(49) Auer, M.; Kim, M. J.; Lemieux, M. J.; Villa, A.; Song, J.; Li, X.-D.; Wang, D.-N. High-yield expression and functional analysis of Escherichia coli glycerol-3-phosphate transporter. *Biochemistry* **2001**, *40*, 6628–6635.

(50) Hirai, T.; Heymann, J. A.; Shi, D.; Sarker, R.; Maloney, P. C.; Subramaniam, S. Three-dimensional structure of a bacterial oxalate transporter. *Nat. Struct. Biol.* **2002**,

(51) Hirai, T.; Subramaniam, S. Structure and transport mechanism of the bacterial oxalate transporter OxlT. *Biophys. J.* **2004**, *87*, 3600–3607.

(52) West, I. C. Ligand conduction and the gated-pore mechanism of transmembrane transport. *Biochim. Biophys. Acta* **1997**, *1331*, 213–234.

(53) Law, C. J.; Almqvist, J.; Bernstein, A.; Goetz, R. M.; Huang, Y.; Soudant, C.; Laaksonen, A.; Hovmöller, S.; Wang, D.-N. Salt-bridge dynamics control substrate-induced conformational change in the membrane transporter GlpT. *J. Mol. Biol.* **2008**, *378*, 828–839.

(54) Berman, H. M. The protein data bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

(55) Jo, S.; Kim, T.; Iyer, V. G.; Im, W. CHARMM-GUI: A web-based graphical user interface for CHARMM. *J. Comput. Chem.* **2008**, *29*, 1859–1865.

(56) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. PACKMOL: A package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* **2009**, *30*, 2157–2164.

(57) Duke, R.; Gohlke, H. AMBER 14; University of California: San Francisco. **2014**,

(58) Maier, J. A.; Martinez, C.; Kasavajhala, K.; Wickstrom, L.; Hauser, K. E.; Simmerling, C. ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **2015**, *11*, 3696–3713.

(59) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.

(60) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **1984**, *81*, 3684–3690.

(61) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. A smooth particle mesh Ewald method. *J. Chem. Phys.* **1995**, *103*, 8577–8593.

(62) Kräutler, V.; Van Gunsteren, W. F.; Hünenberger, P. H. A fast SHAKE algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J. Comput. Chem.* **2001**, *22*, 501–508.

(63) Bowman, G. R.; Ensign, D. L.; Pande, V. S. Enhanced modeling via network theory: Adaptive sampling of Markov state models. *J. Chem. Theory Comput.* **2010**, *6*, 787–794.

(64) Shamsi, Z.; Cheng, K. J.; Shukla, D. Reinforcement Learning Based Adaptive Sampling: REAPing Rewards by Exploring Protein Conformational Landscapes. *J. Phys. Chem. B* **2018**, *122*, 8386–8395.

(65) Shamsi, Z.; Moffett, A. S.; Shukla, D. Enhanced unbiased sampling of protein dynamics using evolutionary coupling information. *Sci. Rep.* **2017**, *7*.

(66) Schwantes, C. R.; Pande, V. S. Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. *Journal of Chemical Theory and Computation* **2013**, *9*, 2000–2009.

(67) Schwantes, C. R.; Shukla, D.; Pande, V. S. Markov State Models and tICA Reveal a Nonnative Folding Nucleus in Simulations of NuG2. *Biophysical Journal* **2016**, *110*, 1716–1719.

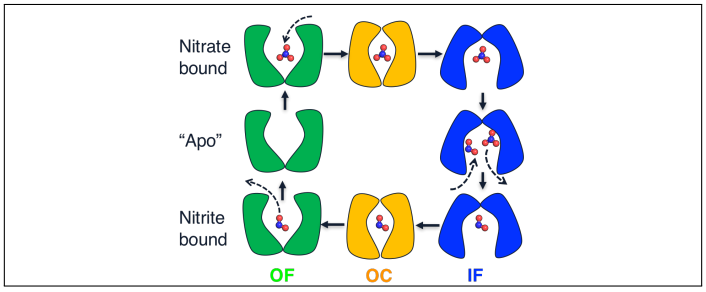
(68) McGibbon, R. T.; Pande, V. S. Variational cross-validation of slow dynamical modes in molecular kinetics. *J. Chem. Phys.* **2015**, *142*, 124105.

(69) Noé, F.; Nüske, F. A variational approach to modeling slow processes in stochastic dynamical systems. *Multiscale Model. Sim.* **2013**, *11*, 635–655.

(70) McGibbon, R. T.; Hernández, C. X.; Harrigan, M. P.; Kearnes, S.; Sultan, M. M.; Jastrzebski, S.; Husic, B. E.; Pande, V. S. Osprey: Hyperparameter optimization for machine learning. *J. Open Source Softw.* **2016**, *1*, 34.

(71) Beauchamp, K. A.; Bowman, G. R.; Lane, T. J.; Maibaum, L.; Haque, I. S.; Pande, V. S. MSMBuilder2: Modeling conformational dynamics on the picosecond to millisecond scale. *J. Chem. Theory Comput.* **2011**, *7*, 3412–3419.

(72) Roe, D. R.; Cheatham, T. E. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **2013**, *9*, 3084–3095.

(73) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graph.* **1996**, *14*, 33–38.

Graphical TOC Entry