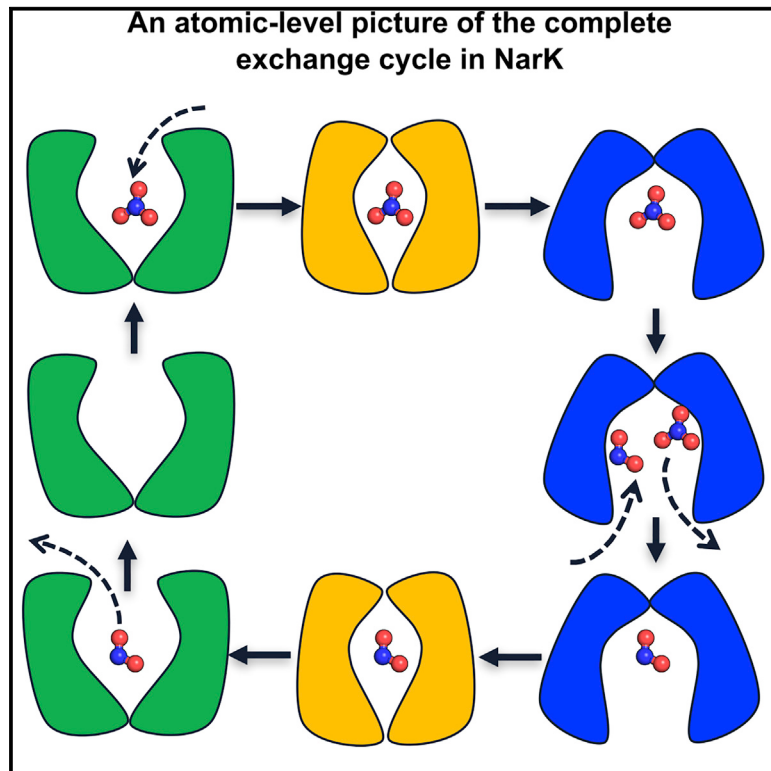


Structure

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

Graphical abstract



Authors

Jiangyan Feng, Balaji Selvam,
Diwakar Shukla

Correspondence

diwakar@illinois.edu

In brief

Feng et al. explored the structural and mechanistic basis of MFS antiporters through the study of a bacterial $\text{NO}_3^-/\text{NO}_2^-$ antiporter, NarK. Extensive MD simulations revealed a complete atomistic picture of NarK function as an antiporter. This work presents a general working model for all MFS antiporters.

Highlights

- MD simulations reveal the complete $\text{NO}_3^-/\text{NO}_2^-$ exchange cycle for NarK
- Paired basic residues restrict conformational changes of unbound protein
- NO_3^- and NO_2^- switch differently in IF and OF states
- A microscopic model for substrate recognition, binding, exchange, and translocation



Theory

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

Jiangyan Feng,¹ Balaji Selvam,¹ and Diwakar Shukla^{1,2,3,4,5,6,7,*}

¹Department of Chemical and Biomolecular Engineering, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

²Department of Plant Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

³Center for Biophysics and Quantitative Biology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

⁴NIH Center for Macromolecular Modeling and Bioinformatics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

⁵Beckman Institute for Advanced Science and Technology, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

⁶Center for Digital Agriculture, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

⁷Lead contact

*Correspondence: diwakar@illinois.edu

<https://doi.org/10.1016/j.str.2021.03.014>

SUMMARY

Major facilitator superfamily (MFS) proteins operate via three different mechanisms: uniport, symport, and antiport. Despite extensive investigations, the molecular understanding of antiporters is less advanced than that of other transporters due to the complex coupling between two substrates and the lack of distinct structures. We employ extensive all-atom molecular dynamics simulations to dissect the complete substrate exchange cycle of the bacterial $\text{NO}_3^-/\text{NO}_2^-$ 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 occurs only in the presence of substrate, which weakens the electrostatic repulsion and stabilizes the transporter. 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 the MFS.

INTRODUCTION

Major facilitator superfamily (MFS) transporters move a wide spectrum of biologically relevant substrates (nutrients, drugs, waste, etc.) across cell membranes (Marger and Saier, 1993; Pao et al., 1998; Reddy et al., 2012; Quistgaard et al., 2016). This superfamily is ubiquitous in all kingdoms of life and represents the largest and most diverse group of secondary active transporters (Law et al., 2008b). Malfunction of MFS proteins has been associated with a multitude of diseases, such as cancer, type 2 diabetes, and Alzheimer's disease (Augustin, 2010; Cura and Carruthers, 2011; Smith et al., 2013). In plants, MFS transporters mediate macronutrient (C, N, and P) uptake and extrusion of deleterious compounds (Niño-González et al., 2019; Selvam et al., 2019; Cheng et al., 2019). 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 antiporters (Forrest et al., 2011). Uniporters transport a single species of substrate across the membrane. 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 mechanisms is well advanced for symporters and uniporters, exemplified by the *Escherichia coli* lactose/ H^+ symporter (LacY) (Kachback, 2005, 2015; Madej, 2014). By contrast, molecular insights are less advanced for antiporters due to the lack of crystal structures representing different stages of the transport cycle (Law et al., 2008b). Among all the MFS proteins with known structures, most of them are symporters, and the outward-facing (OF) conformation of MFS antiporters has yet to be captured (Saier et al., 2015). Recently, computational studies of a few antiporters have advanced our understanding of the antiport mechanism (Law et al., 2008b; Moradi et al., 2015; Alhadeff and Warshel, 2015; Takemoto et al., 2018; Okazaki et al., 2019). However, these studies were able to simulate the transport of only one substrate in the transport cycle, even with enhanced sampling techniques and state-of-the-art supercomputing.

Many important structural and mechanistic aspects of the 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 an antiporter distinguish and switch two cargos? (4) How is substrate

binding coupled 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 (Zheng et al., 2013; Fukuda et al., 2015). NarK is a $\text{NO}_3^-/\text{NO}_2^-$ antiporter from *E. coli* that couples an outward flow of internal NO_2^- to the uptake of NO_3^- into cell (Fukuda et al., 2015). It belongs to the $\text{NO}_3^-/\text{NO}_2^-$ porter (NNP) family of the MFS. NNP proteins mediate the high-affinity translocation of NO_3^- and NO_2^- in both prokaryotes and eukaryotes (i.e., archaea, bacteria, fungi, yeast, algae, and plants) (Forde, 2000). NO_3^- is the most important source of mineral nitrogen for plants, and NO_3^- availability greatly limits crop yields (Plett et al., 2018). 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 (Niño-González et al., 2019). 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 (Mitchell, 1957; Jardetzky, 1966; Drew and Boudker, 2016). Switching among OF, occluded (OC), and inward-facing (IF) states alternatively exposes 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. Extensive sets of molecular dynamics (MD) simulations (Karplus and McCammon, 2002; Feng et al., 2019; Moffett and Shukla, 2018) have been successfully combined with Markov state models (MSMs) (Shukla et al., 2015; Husic and Pande, 2018) to reveal the dynamics and conformational transitions of various membrane proteins, including transporters (Selvam et al., 2018, 2019; Cheng et al., 2019; Mittal and Shukla, 2017, 2018a, 2018b; Chan et al., 2020). MSMs stitch massive parallel short MD trajectories together to build a kinetic network model that describes long-timescale protein dynamics (Shukla et al., 2015; Husic and Pande, 2018).

This work explores the structural and mechanistic principles that characterize MFS antiporters through the study of a bacterial $\text{NO}_3^-/\text{NO}_2^-$ antiporter, NarK. By combining extensive ($\sim 300 \mu\text{s}$) unbiased all-atom MD simulations and MSMs, we explored the complete substrate exchange cycle for NarK and characterized the underlying free-energy landscapes. Simulation results suggest that NarK adopts the conserved MFS fold and follows the common rocker-switch mechanism, governed by helix bending around highly conserved glycine residues. What distinguishes the NarK antiporter from symporters is the inaccessible energy barrier between IF and OF states without a bound substrate. Two highly conserved and positively charged arginine residues (R89 and R305) in the central binding site restrict the closure of unbound protein, thereby ensuring the exchange of NO_3^- and NO_2^- . Substrate binding (NO_3^- or NO_2^-) is required to weaken the electrostatic repulsion and drive the

conformational switch. We further identified key residues involved in substrate recognition, binding, exchange, and translocation, and provided a detailed model of the complete exchange process. This work provides important information both for understanding NO_3^- uptake by NNP transporters and for elucidating the antiport mechanism within the MFS.

RESULTS

NarK follows the canonical rocker-switch mechanism

All-atom MD simulations of NarK were initiated from the NO_3^- -bound OC crystal structure (PDB: 4U4W; Fukuda et al., 2015). In all these simulations, the transporter was inserted into an explicit lipid bilayer and solvated in a box, with dimensions of $76 \times 76 \times 102 \text{ \AA}$, containing TIP3P (transferable intermolecular potential with three points) water molecules. The simulations were carried out under isothermal-isobaric conditions (300 K and 1 atm). With a total of $\sim 300 \mu\text{s}$ unbiased MD simulations, the entire transport cycle of NarK and all the functionally relevant states were characterized. The MSM is a popular technique for extracting kinetic information of protein dynamics from MD simulation data (Husic and Pande, 2018; Shukla et al., 2015). All the simulation data were used to construct an MSM, and the MSM hyperparameters were selected systematically using a genetic algorithm technique (Chen et al., 2018). The MSM estimation reweights the MD trajectories such that the equilibrium kinetics and distribution among sampled configurations can be recovered. Simulation and MSM construction details are summarized in the STAR methods. Because only a finite quantity of simulation data can be obtained to construct the MSM model, the properties computed from the MSM will be statistically uncertain. Using a Gibbs sampling procedure, Bayesian MSMs construct a sample of reversible transition matrices and quantify the statistical uncertainties for all observables derived from the MSM (Noé, 2008; Trendelkamp-Schroer et al., 2015). To estimate the uncertainties, a Bayesian MSM (Noé, 2008; Trendelkamp-Schroer et al., 2015) was estimated with 100 samples and 95% confidence interval using PyEMMA2.5.7 (Scherer et al., 2015) (see the STAR methods for details). In the crystal structure (PDB: 4U4W; Fukuda et al., 2015), the 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 closing of the pore channel strictly determine the specific state, such as IF, OC, and OF (Selvam et al., 2018).

The free-energy landscape reveals the thermodynamic and kinetic information about the antiporter transport mechanism (Figure 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 predict that these crystal structures (Fukuda et al., 2015) 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 $\sim 13.2 \text{ \AA}$ apart in the crystallized IF state (PDB: 4U4T, 4U4V; Fukuda et al., 2015). However, the free-energy landscape suggests that the intracellular distance between TM4 and TM10

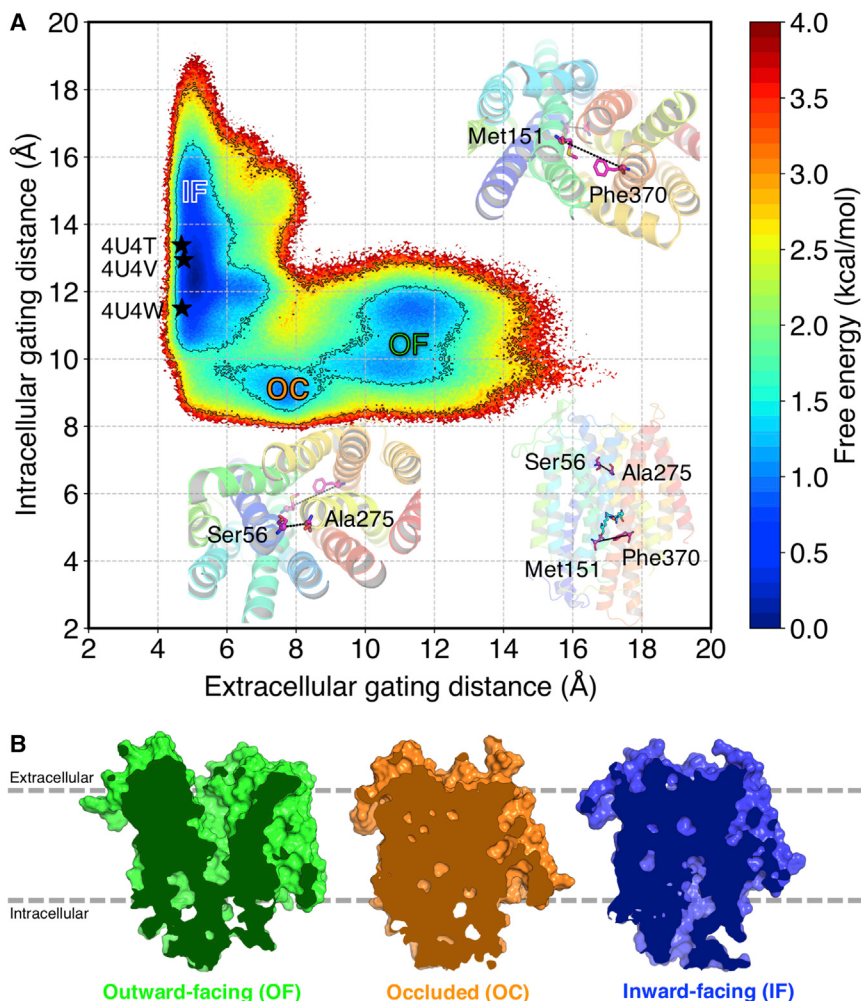


Figure 1. Alternating access cycle of NarK

(A) MSM-weighted free-energy landscape plot of NarK. The $\sim 300 \mu\text{s}$ MD simulation data were projected onto the two-dimensional space of the Ser56C α (TM1)-Ala275C α (TM7) and Met151C α (TM4)-Phe370C α (TM10) distances. Three crystal structures are labeled as black stars.

(B) Cross sections of surface representations of the representative structures, sampled randomly from the corresponding OF, OC, and IF states. The difference between the intracellular and the extracellular gating distances (Δd) was used as the selection criterion for the OF ($\Delta d \sim -1.3 \pm 0.01 \text{ \AA}$), OC ($\Delta d \sim 3.3 \pm 0.01 \text{ \AA}$), and IF ($\Delta d \sim 10.8 \pm 0.01 \text{ \AA}$) states. The extracellular gating distance is 12.3, 5.1, and 4.4 $\text{\AA}$ for the OF, OC, and IF structures shown. The intracellular distance is 11.0, 8.6, and 15.2 $\text{\AA}$ for the OF, OC, and IF structures shown. The structures are viewed from the membrane plane.

See also Figures S1, S4, and S6.

can increase up to $\sim 15.0 \text{ \AA}$ for the energetically accessible IF state (Figures 1A, S1A, and S1B). As TM4 and TM10 move toward each other, NarK adopts an OC state with intracellular distance reduced to $\sim 9.0 \text{ \AA}$, while the extracellular distance increases to $\sim 7.5 \text{ \AA}$. According to these metrics, at least, the previously reported OC state (PDB: 4U4W; Fukuda et al., 2015) may actually represent a partially occluded IF state. The intracellular side can be further closed as the intracellular distance decreases from ~ 11.5 to $\sim 9.0 \text{ \AA}$ (Figures 1A, S1C, and S1D). Finally, the extracellular distance increases to $\sim 12.0 \text{ \AA}$ and the antiporter adopts the OF state. The free-energy barrier for the transition from IF to OC state is $\sim 1.5 \pm 0.1 \text{ kcal/mol}$, and for the subsequent transition to OF state it is $\sim 1.0 \pm 0.1 \text{ kcal/mol}$. The total free-energy barrier for one complete cycle of NarK from IF to OF state is $\sim 2.5 \pm 0.1 \text{ kcal/mol}$. The relatively low energy barrier suggests that NarK can easily interconvert between IF, OC, and OF states during the translocation process. This observation is consistent with the high-affinity uptake of $\text{NO}_3^-/\text{NO}_2^-$ by NarK (Forde, 2000; Yan et al., 2013).

To gain structural insights into the free-energy landscape, three representative structures were randomly sampled from the corresponding OF, OC, and IF states (Figure 1B). The molecular surfaces for these structures suggest that the substrate

translocation pathway is the pore saddled by the N- and C-terminal domains. The N and C domains change their relative positions to alternately expose the substrate binding site to opposite sides of the membrane. These structural features reinforce the notion that NarK follows the rocker-switch mechanism (Drew and Boudker, 2016). The simulated IF and OC structures show good agreement with experimental NarK structures (PDB: 4U4V, 4U4T, 4U4W; Fukuda et al., 2015) (Figure S1). 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 ($\sim 2.0 \text{ \AA}$) and narrower for OC state ($\sim 2.5 \text{ \AA}$). The predicted OF structure was compared with a fucose/ H^+ symporter (FucP, PDB: 3O7Q; Dang et al., 2010) due to the absence of OF structures in MFS antiporters (Figures S1E and S1F). The OF conformation superimposes well on the OF structure of FucP and exhibits a similar opening at the extracellular side (Figures S1E and S1F). These findings suggest that the rocker-switch mechanism is shared among all MFS proteins, irrespective of their particular function 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 the root-mean-square fluctuation (RMSF) of each C α atom in NarK during IF to OC and OC to OF transition (Figures 2A–2F and S2). The intracellular tips of TM5, TM10, and TM11 show higher flexibility ($\sim 2.5 \text{ \AA}$) during the IF to OC transition (Figures 2E and 2F). The extracellular tips of TM1, TM2, and TM7 undergo significant

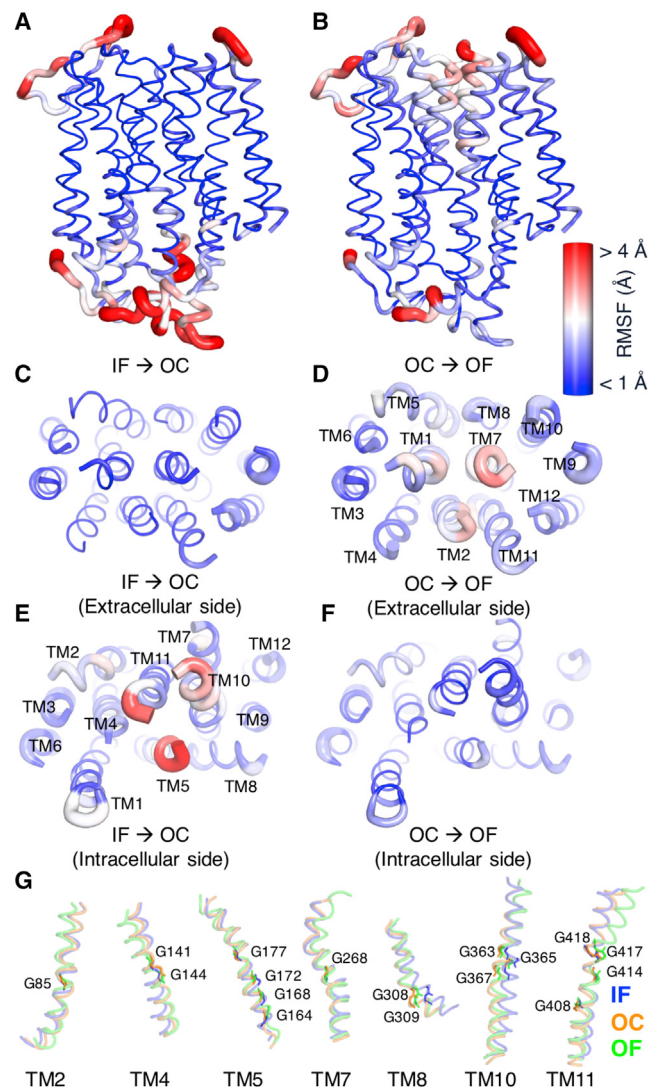


Figure 2. Global conformational changes

(A–F) Root-mean-square fluctuation (RMSF) data were mapped onto the predicted IF structure of NarK for the IF to OC (A, C, and E) and OC to OF transitions (B, D, and 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 and D) and intracellular (E and F) sides with transmembrane helices (TMs) labeled are shown.

(G) Structural comparisons among IF, OC, and OF states. The superimposition was performed with PyMOL (Schrödinger, 2015) and for the helices containing conserved glycine residues: TM2, TM4, TM5, TM7, TM8, TM10, and TM11. The helices are shown in tube representations and colored based on the conformational states. Relevant glycine residues are labeled and shown as colored sticks.

See also Figure S2.

rearrangements (~ 2.5 Å) during the transition from OC to OF state (Figures 2C and 2D). The other helices (TM3, TM4, TM6, TM8, TM9, and TM12) represent a stable structural element that displays limited conformational changes (~ 1.0 Å) during the transport cycle. Previous structural analysis hypothesized that the N bundle remains rigid while the TM7, TM10, and TM11 helices of the C bundle move toward the N bundle to close

the intracellular vestibule (Fukuda et al., 2015). In simulations, both the N (TM5) and the C domains (TM10 and TM11) move toward each other, with a higher magnitude (~ 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 the intracellular gate (Fukuda et al., 2015). By contrast, the extracellular tip of TM7 exhibits large fluctuations (~ 2.5 Å) during the OC to OF transition, whereas the motions in the intracellular tip are negligible (~ 1.0 Å) during the entire transport cycle (Figures 2C, 2D, and 2G). This strongly indicates that TM7 is critical for the opening of the extracellular instead of the intracellular gate. The simulations also shed light on the conformational transitions from OC to OF state, which were previously unknown due to the lack of an OF structure. The extracellular halves of TM1 and TM2 of the 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 (Figures 2A–2F and S2). 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. The bending of TM10 and TM11 was also observed in the previous structural studies of NarK (Zheng et al., 2013; Fukuda et al., 2015), its closest homolog NarU (Yan et al., 2013), and many other MFS proteins (Abramson, 2003; Lemieux et al., 2004; Yin, 2006; Dang et al., 2010; Solcan et al., 2012), implying significance of the flexibility of TM10 and TM11 within the MFS. 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) (Unkles et al., 2012; Zheng et al., 2013; Yan et al., 2013; Fukuda et al., 2015). It is interesting to note that large structural fluctuations in helices are located at or close to the conserved glycine residues (Figure 2G). These highly conserved glycine residues serve as pivot points for helix bending. The significance of glycine is highlighted by previous mutations of glycine to alanine that abolish the NO_3^- uptake activity (Fukuda et al., 2015). 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 the N- and C-terminal domains in three different conformational states (IF, OC, and OF). Ten sets of 1,000 MD configurations were randomly selected for each state according to the difference between the intracellular and the extracellular gating distances (Δd): OF, $\Delta d \sim -1.3 \pm 0.01$ Å; OC, $\Delta d \sim 3.3 \pm 0.01$ Å; and IF, $\Delta d \sim 10.8 \pm 0.01$ Å. By comparing the average contact frequency in different states, we identified six layers of hydrophobic and polar interactions along the transport pathway (Figure 3). Many of these interactions were not reported in the previous studies that relied on the structural overlays of IF and the partially occluded IF state (Zheng et al., 2013; Fukuda et al., 2015).

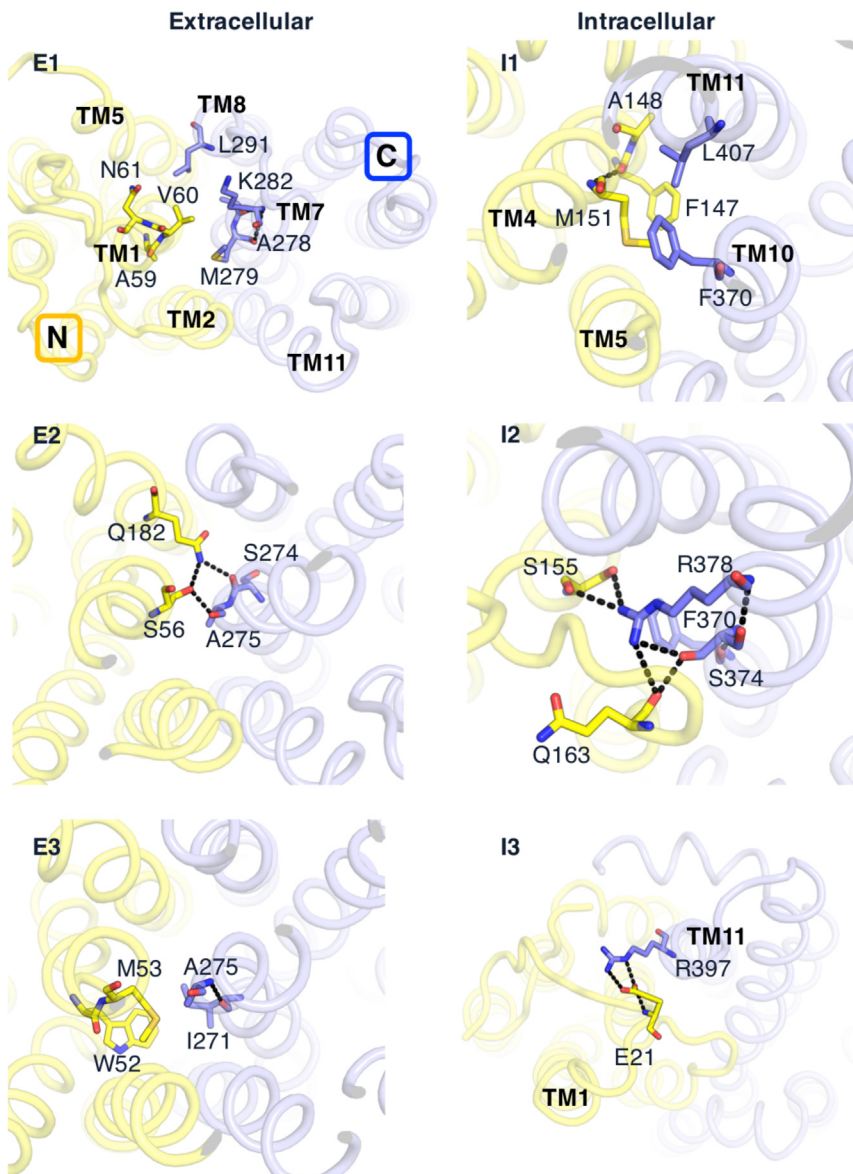


Figure 3. Extensive hydrophobic and polar interactions along the transport pathway

The three left images represent the close-up views of the extracellular gating interactions seen from the extracellular side. The three right images show the close-up views of the intracellular gating interactions seen from the intracellular side. N- and C-domain helices are colored yellow and light blue, respectively. The residues involved in each interaction layer are depicted by stick models and colored following the same scheme. See also Figure S3.

[TM1], I271 [TM7], and A275 [TM7]) mediate extensive van der Waals contacts and stabilize the extracellular gate (Figure 3, E3). W52 is invariant among all the prokaryotic NNP family transporters. The W50A mutation in NarU (corresponding to W52 in NarK) completely abolishes 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 the OC states.

Consistent with the experimental structure (PDB: 4U4W; Fukuda et al., 2015), the intracellular gate consists of three layers of interactions (Figure 3, I1–I3). 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 (Figure 3, I1). F147 and F370 are invariant among all NarK homologs in bacteria and eukaryotes. Previous functional analysis of NarK and its closest ho-

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 (Figure 3, E1–E3). 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 (Figure 3, E1). All of these residues are conserved in NarK's closest homolog (NarU), except A59, suggesting a similar hydrophobic layer in NarU (Yan et al., 2013). In layer E2, S56 (TM1), Q182 (TM5), S274 (TM7), and A275 (TM7) form an extensive network of hydrogen bonds, facilitating the close packing of gating helices (Figure 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 (Yan et al., 2013). The hydrophobic residues in layer E3 (W52 [TM1], M53

homolog (NarU) further supports the functional relevance (Yan et al., 2013; Fukuda et al., 2015). The F147A mutant shows decreased NO_3^- -transport activity in NarK (Fukuda et al., 2015), and the F367A mutation in NarU (Yan et al., 2013) (corresponding to F370 in NarK) abrogates binding to both NO_3^- and NO_2^- . In the previously reported layer I2 (Fukuda et al., 2015), 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 (Figure 3, I2). The S155–A400 interaction is observed in around 88% and 93% of IF and OF states. The R378–F158–K160 interaction occurs in approximately 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 state. Simulations suggest that S155 (TM4) and Q163 (TM5) form an extensive hydrogen bond network with S374 and R378 from TM10

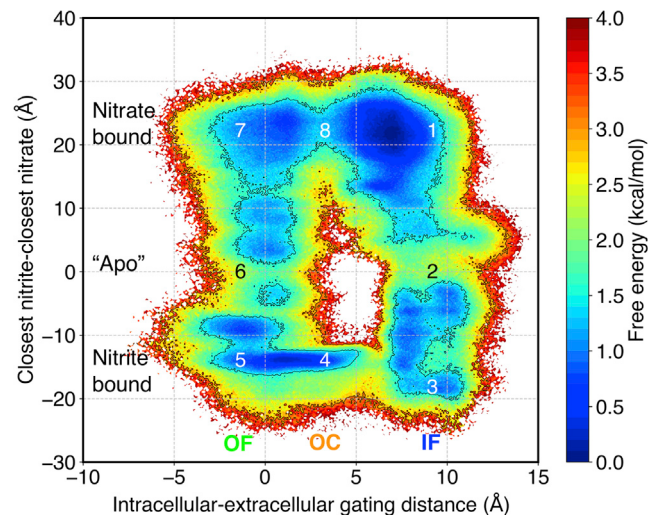


Figure 4. MSM-weighted free-energy landscape of the entire $\text{NO}_3^-/\text{NO}_2^-$ exchange cycle

The difference between the intracellular and the extracellular gating distances (x) 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 (y) shows different substrate binding conditions. The conformational states were depicted as (1) NO_3^- -bound IF (x = 9.2 Å, y = 22.0 Å), (2) “apo” IF (x = 9.2 Å, y = -0.2 Å), (3) NO_2^- -bound IF (x = 9.2 Å, y = -18.4 Å), (4) NO_2^- -bound OC (x = 3.3 Å, y = -14.2 Å), (5) NO_2^- -bound OF (x = -1.3 Å, y = -14.2 Å), (6) “apo” OF (x = -1.3 Å, y = -0.2 Å), (7) NO_3^- -bound OF (x = -1.3 Å, y = 22.0 Å), and (8) NO_3^- -bound OC (x = 3.3 Å, y = 22.0 Å).

See also Figures S4 and S6.

(Figure 3, I2). In the partially occluded IF state of NarK (PDB: 4U4W; Fukuda et al., 2015), Q163 on unbent TM5 points away from the central pathway, and thus the extensive polar interactions of layer I2 are missing (Figure S3). 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 (Figure 3, I2). Near the intracellular surface, E21 of the N-terminal loop forms a salt bridge with R397 on TM11 and blocks the access to the substrate binding site from the cytoplasm (Figure 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 one another? A major difference between a symporter and an antiporter is whether an empty transporter changes conformation from IF to OF: symporters do, but antiporters do not (Zhang et al., 2015). How is such a mechanistic scheme implemented in an antiporter? We projected all the simulation data onto a two-dimensional free-energy landscape to understand the whole NO_3^- and NO_2^- exchange cycle (Figures 4 and S4). To monitor the conformational transitions from IF to OF state, the difference between the intracellular and the extracellular gating distances was selected as the x coordinate. Previous mutational analysis demonstrates that R89 and R305 directly bind substrates (Fukuda et al., 2015). The difference between

the closest distances from NO_2^- and NO_3^- to the center of R89 and R305 was selected as the y coordinate to track the substrate binding mode.

The free-energy landscape exhibits eight distinct conformational states constituting the whole NO_3^- and NO_2^- exchange cycle (Figure 4). The white region in the center implies an inaccessible energy barrier between IF and OF states without a bound substrate. The high energy barrier excludes the alternation between the OF and the 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. Whereas the substrate-bound states correspond to the energy minima, substrate-unbound states are located in the relatively higher energy regions ($\sim 2.0 \pm 0.1$ kcal/mol). Substrate binding lowers the energy by $\sim 1.5 \pm 0.1$ kcal/mol in both IF and OF states and enhances the probability of conformational switching between them.

Moreover, the free-energy landscape enables the quantitative comparison between the NO_3^- and the NO_2^- 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 $\sim 1.0 \pm 0.1$ kcal/mol with bound NO_3^- . When NO_2^- is bound, the OC and IF states are separated by a barrier of $\sim 2.5 \pm 0.1$ kcal/mol. According to the Arrhenius equation (Connors, 1990), the $\sim 1.5 \pm 0.1$ kcal/mol energy difference corresponds to an ~ 12 -fold binding affinity difference between NO_3^- and NO_2^- . This is consistent with the 10-fold higher binding affinity for NO_3^- ($\sim 33 \mu\text{M}$) over NO_2^- ($\sim 373 \mu\text{M}$) in NarK’s closest homolog, NarU (Yan et al., 2013). The computational analyses suggest that NO_3^- favors the transition from OC to IF state 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 (Law et al., 2008b). To gain structural insights into this capability, we compared eight representative structures randomly sampled from the center of corresponding metastable states in Figure 4. Each free-energy basin corresponds to a distinct substrate binding mode. In agreement with previous crystal structures (Fukuda et al., 2015), NO_3^- and NO_2^- bind to NarK at exactly the same position (Figure 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 (Figure 5B). The functional significance of these residues has also been confirmed by mutational analysis directly performed on NarK (Fukuda et al., 2015). Two conserved and positively charged residues, R89 from TM2 and R305 from TM8, directly bind to substrates (Fukuda et al., 2015; Unkles et al., 2004). Mutation of either arginine into lysine decreases the transport activity of NarK. Without bound substrate, the electrostatic repulsion between R89 and R305 prevents the 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 *E. coli* sn-glycerol-3-phosphate transporter (GlpT) (Huang, 2003), *E. coli* hexose-6-phosphate:phosphate antiporter (UhpT) (Ambudkar et al., 1990; Auer et al., 2001),

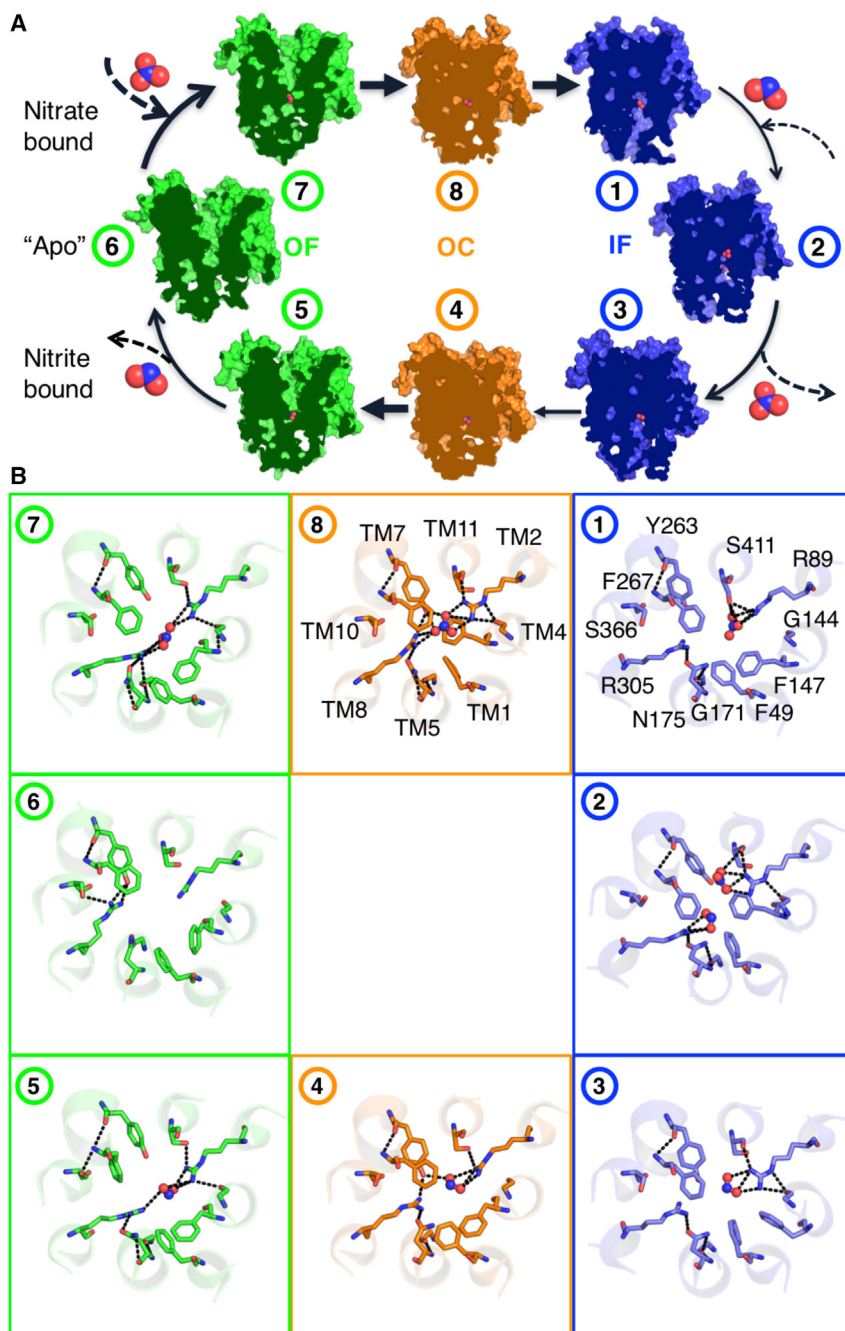


Figure 5. NO_3^- and NO_2^- switching mechanism

(A) Cross section of surface representations of the representative structures from the intermediate states along the exchange cycle. The conformations shown were randomly sampled from the center of the eight metastable states in Figure 4. The bound substrates (NO_3^- and NO_2^-) are shown in ball-and-stick form.

(B) Close-up view of the binding sites visualizing the residues coordinating the binding of NO_3^- and NO_2^- . The dotted black lines represent hydrogen bonds among residues. Substrates and binding-pocket residues are represented in ball-and-stick and stick form, respectively.

See also Figure S5.

of NarK tightly. This may accelerate the OC to IF transition and lead to the $\sim 1.5 \pm 0.1$ kcal/mol difference in Figure 4.

The next question is how the antiporter switches the substrates. A common hypothesis for exchangers is the “ping-pong” mechanism (West, 1997). 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 the IF state (state 2 in Figure 5B), the binding of both substrates is preferred ($\sim 65\%$) over the unbound conformation ($\sim 35\%$) (Figure S5A). In comparison, $\sim 90\%$ of structures adopt an unbound conformation in the OF state (state 6 in Figure 5B) (Figure S5B). This is likely due to the relative spacing between R89 and R305. In the IF state, the shortest distance between the R89 and the R305 side chains is ~ 9.0 Å. R89 and R305 could not simultaneously coordinate the small anion (~ 1.96 Å for NO_3^-) (states 1 and 3 in Figure 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 (states 5 and 7 in Figure 5B). From these data, we

and the oxalate:formate transporter from *Oxalobacter formigenes* (OxIT) (Hirai et al., 2002; Hirai and Subramaniam, 2004). 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 (states 4 and 8 in Figure 5B). The additional oxygen atom of NO_3^- allows the formation of additional hydrogen bonds with R89 and R305 and connects the two halves

propose a state-dependent exchange model. In the 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 the 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

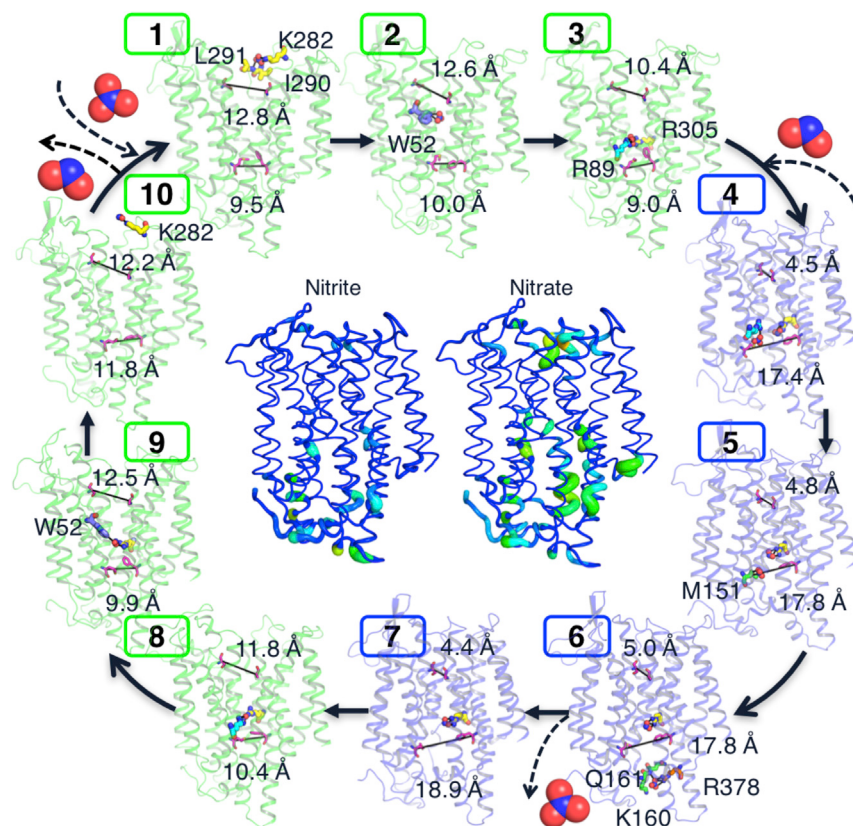


Figure 6. Substrate recognition, binding, exchange, and translocation

The center shows the MSM-weighted 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 the 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 the inside gates and triggers the conformational transitions from IF to OF. (9–10) NO_2^- follows the same translocation pathway with NO_3^- and is released to the outside. For clarity, only R89 and R305 are shown in the binding site to illustrate the substrate binding. The dotted black lines represent the distances between gating residues (magenta sticks). Substrates and relevant residues are represented as ball-and-stick models.

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 (Figure 5B). In the IF and OF states, there is no interaction between Y263 and substrate in the binding pocket. In the OC states (states 4 and 8 in Figure 5B), however, the phenol hydroxyl group of Y263 hydrogen bonds with both the 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 (Fukuda et al., 2015). Y263F and R305K mutants in NarK completely abolished transport activity even under high isopropyl- β -D-thiogalactoside (IPTG) concentration, which is used to induce protein expression (Fukuda et al., 2015), whereas the R89K mutant rescued the NO_3^- -uptake activity under the highest IPTG concentration (Fukuda et al., 2015). These results support the idea 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 (Figure 5B): (1) In IF or OF state, 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 tighter substrate binding to the transporter. (3) The binding energy released by this stronger

interaction is used to overcome the energy barrier for IF to OC or OF to OC transitions. The compact hydrogen bond network at the binding pocket turns on the helix bending motions. Because these helices constitute the N and C bundles, the bending motions directly result in global conformational changes. Taken 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 (center of Figure 6). The exchange between NO_3^- from the outside and NO_2^- from the inside can be described by the following steps (Figure 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 inward to make contact with the backbones of L291 and I290 on TM8 (Figure 6, [1]). (2) After transient coordination with K282 and nearby L291 and I290, NO_3^- rapidly diffuses deep into the pore channel through W52 (Figure 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 an 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 enable the outward halves of the core domains to approach each other and facilitates the conformational change

from the OF to the OC state. At this juncture, the extracellular gating distance decreases to ~ 10.4 Å (Figure 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 (Figure 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 (Figure 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 (Figure 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 (Figure 6, [7]). (8) At 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) (Figure 6, [8]). (9) NO_2^- follows the same pathway as NO_3^- to be released to the outside solution. NO_2^- is dissociated from the binding site through W52 (Figure 6, [9]). (10) NO_2^- is released to the outside solution mediated by the positively charged K282 on the outward surface (Figure 6, [10]).

DISCUSSION

In an attempt to decipher the molecular origin of MFS antiporters, we studied a bacterial $\text{NO}_3^-/\text{NO}_2^-$ 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 the conformational transitions. Paired basic residues were also found to be critical for substrate binding in several other MFS antiporters (i.e., GlpT, UhpT, and OxiT; Huang, 2003; Ambudkar et al., 1990; Auer et al., 2001; Hirai et al., 2002; Hirai and Subramaniam, 2004). This suggests that paired basic residues in the binding pocket may be the common structural determinant for the exchange function in antiporters that transport anionic substrates.

- (3) How does the antiporter distinguish and switch two cargos? Breakage and formation of hydrogen bonds rearrange 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 the 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 the 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 ($\sim 2.5 \pm 0.1$ kcal/mol) required to complete the cycle from IF to OF state. Another structural feature is the gating residues lining up the translocation pathway. Previous studies of the GlpT antiporter suggested that salt bridges are important for gating on both sides of the binding pocket (Law et al., 2008a). 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 Figure 3) occlude the substrate and require only modest local structural changes to break or form.

To summarize, by performing extensive all-atom MD simulations, the present study provides the complete picture of the NarK antiporter function (Figure 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.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

● KEY RESOURCES TABLE

- **RESOURCE AVAILABILITY**
 - Lead contact
 - Materials availability
 - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
- **METHOD DETAILS**
 - Molecular dynamics (MD) simulation
 - Adaptive sampling
 - Markov state model (MSM) construction
- **QUANTIFICATION AND STATISTICAL ANALYSIS**

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.str.2021.03.014>.

ACKNOWLEDGMENTS

The 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 an NSF Early Career Award (MCB 1845606). J.F. was supported by a Chia-chen Chu fellowship and a Harry G. Drickamer graduate research fellowship from the University of Illinois, Urbana-Champaign, IL, USA.

AUTHOR CONTRIBUTIONS

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

DECLARATION OF INTERESTS

The authors declare no competing financial interests.

Received: July 15, 2020

Revised: January 7, 2021

Accepted: March 19, 2021

Published: April 8, 2021

REFERENCES

- Abramson, J. (2003). Structure and mechanism of the lactose permease of *escherichia coli*. *Science* **301**, 610–615.
- Alhadeff, R., and Warshel, A. (2015). Simulating the function of sodium/proton antiporters. *Proc. Natl. Acad. Sci. U S A* **112**, 12378–12383.
- Ambudkar, S.V., Anantharam, V., and Maloney, P.C. (1990). Uhpt, the sugar phosphate antiporter of *escherichia coli*, functions as a monomer. *J. Biol. Chem.* **265**, 12287–12292.
- Auer, M., Kim, M.J., Lemieux, M.J., Villa, A., Song, J., Li, X.-D., and Wang, D.-N. (2001). High-yield expression and functional analysis of *escherichia coli* glycerol-3-phosphate transporter. *Biochemistry* **40**, 6628–6635.
- Augustin, R. (2010). The protein family of glucose transport facilitators: its not only about glucose after all. *IUBMB Life* **62**, 315–333.
- Beauchamp, K.A., Bowman, G.R., Lane, T.J., Maibaum, L., Haque, I.S., and Pande, V.S. (2011). MSMBuilder2: modeling conformational dynamics on the picosecond to millisecond scale. *J. Chem. Theory Comput.* **7**, 3412–3419.
- Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A., and Haak, J.R. (1984). Molecular dynamics with coupling to an external bath. *J. Chem. Phys.* **81**, 3684–3690.
- Berman, H.M. (2000). The protein data bank. *Nucleic Acids Res.* **28**, 235–242.
- Bowman, G.R., Ensign, D.L., and Pande, V.S. (2010). Enhanced modeling via network theory: adaptive sampling of markov state models. *J. Chem. Theory Comput.* **6**, 787–794.
- Braun, E., Gilmer, J., Mayes, H.B., Mobley, D.L., Monroe, J.I., Prasad, S., and Zuckerman, D.M. (2019). Best practices for foundations in molecular simulations [article v1. 0]. *Living J. Comput. Mol. Sci.* **1**, <https://doi.org/10.33011/live-coms.1.1.5957>.
- Bussi, G., Donadio, D., and Parrinello, M. (2007). Canonical sampling through velocity rescaling. *J. Chem. Phys.* **126**, 014101.
- Case, D., Babin, V., Berryman, J., Betz, R., Cai, Q., Cerutti, D., Cheatham, T., Darden, T., Duke, R., Gohlke, H., et al. (2014). *Amber 14* (University of California San Francisco).
- Chan, M.C., Selvam, B., Young, H.J., Procko, E., and Shukla, D. (2020). The substrate import mechanism of the human serotonin transporter. *ChemRxiv* **569**, <https://doi.org/10.26434/chemrxiv.9922301.v2>.
- Chen, Q., Feng, J., Mittal, S., and Shukla, D. (2018). Automatic feature selection in markov state models using genetic algorithm. *J. Comput. Sci. Educ.* **9**, 14–22.
- Cheng, K.J., Selvam, B., Chen, L.-Q., and Shukla, D. (2019). Distinct substrate transport mechanism identified in homologous sugar transporters. *J. Phys. Chem. B* **123**, 8411–8418.
- Chodera, J.D., and Noé, F. (2014). Markov state models of biomolecular conformational dynamics. *Curr. Opin. Struct. Biol.* **25**, 135–144.
- Connors, K.A. (1990). *Chemical Kinetics: The Study of Reaction Rates in Solution* (Wiley-VCH Verlag GmbH).
- Cura, A.J., and Carruthers, A. (2011). Role of monosaccharide transport proteins in carbohydrate assimilation, distribution, metabolism, and homeostasis. *Compr. Physiol.* **2**, 863–914.
- Dang, S., Sun, L., Huang, Y., Lu, F., Liu, Y., Gong, H., Wang, J., and Yan, N. (2010). Structure of a fucose transporter in an outward-open conformation. *Nature* **467**, 734–738.
- Drew, D., and Boudker, O. (2016). Shared molecular mechanisms of membrane transporters. *Annu. Rev. Biochem.* **85**, 543–572.
- Essmann, U., Perera, L., Berkowitz, M.L., Darden, T., Lee, H., and Pedersen, L.G. (1995). A smooth particle mesh ewald method. *J. Chem. Phys.* **103**, 8577–8593.
- Feng, J., Chen, J., Selvam, B., and Shukla, D. (2019). Computational microscopy: Revealing molecular mechanisms in plants using molecular dynamics simulations. *Plant Cell* **31**, tpc.119.tt1219.
- Forde, B.G. (2000). Nitrate transporters in plants: structure, function and regulation. *Biochim. Biophys. Acta* **1465**, 219–235.
- Forrest, L.R., Krämer, R., and Ziegler, C. (2011). The structural basis of secondary active transport mechanisms. *Biochim. Biophys. Acta* **1807**, 167–188.
- Fukuda, M., Takeda, H., Kato, H.E., Doki, S., Ito, K., Maturana, A.D., Ishitani, R., and Nureki, O. (2015). Structural basis for dynamic mechanism of nitrate/nitrite antiport by NarK. *Nat. Commun.* **6**, 1–12.
- Hirai, T., Heymann, J.A., Shi, D., Sarker, R., Maloney, P.C., and Subramaniam, S. (2002). Three-dimensional structure of a bacterial oxalate transporter. *Nat. Struct. Biol.* **9**, 597–600.
- Hirai, T., and Subramaniam, S. (2004). Structure and transport mechanism of the bacterial oxalate transporter OxIT. *Biophys. J.* **87**, 3600–3607.
- Huang, Y. (2003). Structure and mechanism of the glycerol-3-phosphate transporter from *escherichia coli*. *Science* **301**, 616–620.
- Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: Visual molecular dynamics. *J. Mol. Graph.* **14**, 33–38.
- Hunter, J.D. (2007). Matplotlib: a 2d graphics environment. *Comput. Sci. Eng.* **9**, 90–95.
- Husic, B.E., and Pande, V.S. (2018). Markov state models: from an art to a science. *J. Am. Chem. Soc.* **140**, 2386–2396.
- Jardetzky, O. (1966). Simple allosteric model for membrane pumps. *Nature* **211**, 969–970.
- Jo, S., Kim, T., Iyer, V.G., and Im, W. (2008). CHARMM-GUI: a web-based graphical user interface for CHARMM. *J. Comput. Chem.* **29**, 1859–1865.
- Kaback, H.R. (2005). Structure and mechanism of the lactose permease. *C. R. Biol.* **328**, 557–567.

- Kaback, H.R. (2015). A chemiosmotic mechanism of symport. *Proc. Natl. Acad. Sci. U S A* 112, 1259–1264.
- Karplus, M., and McCammon, J.A. (2002). Molecular dynamics simulations of biomolecules. *Nat. Struct. Biol.* 9, 646–652.
- Kräutler, V., Van Gunsteren, W.F., and Hünenberger, P.H. (2001). A fast shake algorithm to solve distance constraint equations for small molecules in molecular dynamics simulations. *J. Comput. Chem.* 22, 501–508.
- Law, C.J., Almqvist, J., Bernstein, A., Goetz, R.M., Huang, Y., Soudant, C., Laaksonen, A., Hovmöller, S., and Wang, D.-N. (2008a). Salt-bridge dynamics control substrate-induced conformational change in the membrane transporter GlpT. *J. Mol. Biol.* 378, 828–839.
- Law, C.J., Maloney, P.C., and Wang, D.-N. (2008b). Ins and outs of major facilitator superfamily antiporters. *Annu. Rev. Microbiol.* 62, 289–305.
- Lemieux, M., Huang, Y., and Wang, D.-N. (2004). Glycerol-3-phosphate transporter of *Escherichia coli*: structure, function and regulation. *Res. Microbiol.* 155, 623–629.
- Loncharich, R.J., Brooks, B.R., and Pastor, R.W. (1992). Langevin dynamics of peptides: the frictional dependence of isomerization rates of *n*-acetylalanine-*n*-methylamide. *Biopolymers* 32, 523–535.
- Madej, M.G. (2014). Function, structure, and evolution of the major facilitator superfamily: the LacY manifesto. *Adv. Behav. Biol.* 2014, 1–20.
- Maier, J.A., Martinez, C., Kasavajhala, K., Wickstrom, L., Hauser, K.E., and Simmerling, C. (2015). ff14sb: improving the accuracy of protein side chain and backbone parameters from ff99sb. *J. Chem. Theory Comput.* 11, 3696–3713.
- Marger, M.D., and Saier, M.H. (1993). A major superfamily of transmembrane facilitators that catalyze uniport, symport and antiport. *Trends Biochem. Sci.* 18, 13–20.
- Martínez, L., Andrade, R., Birgin, E.G., and Martínez, J.M. (2009). PACKMOL: a package for building initial configurations for molecular dynamics simulations. *J. Comput. Chem.* 30, 2157–2164.
- McGibbon, R.T., Beauchamp, K.A., Harrigan, M.P., Klein, C., Swails, J.M., Hernández, C.X., Schwantes, C.R., Wang, L.-P., Lane, T.J., and Pande, V.S. (2015). MDTraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophys. J.* 109, 1528–1532.
- McGibbon, R.T., Hernández, C.X., Harrigan, M.P., Kearnes, S., Sultan, M.M., Jastrzebski, S., Husic, B.E., and Pande, V.S. (2016). Osprey: Hyperparameter optimization for machine learning. *J. Open Source Softw.* 1, 34.
- Mitchell, P. (1957). A general theory of membrane transport from studies of bacteria. *Nature* 180, 134–136.
- Mittal, S., and Shukla, D. (2017). Predicting optimal DEER label positions to study protein conformational heterogeneity. *J. Phys. Chem. B* 121, 9761–9770.
- Mittal, S., and Shukla, D. (2018a). Maximizing kinetic information gain of markov state models for optimal design of spectroscopy experiments. *J. Phys. Chem. B* 122, 10793–10805.
- Mittal, S., and Shukla, D. (2018b). Recruiting machine learning methods for molecular simulations of proteins. *Mol. Simulat.* 44, 891–904.
- Moffett, A.S., and Shukla, D. (2018). Using molecular simulation to explore the nanoscale dynamics of the plant kinome. *Biochem. J.* 475, 905–921.
- Moradi, M., Enkavi, G., and Tajkhorshid, E. (2015). Atomic-level characterization of transport cycle thermodynamics in the glycerol-3-phosphate:phosphate antiporter. *Nat. Commun.* 6, 1–11.
- Niño-González, M., Novo-Uzal, E., Richardson, D.N., Barros, P.M., and Duque, P. (2019). More transporters, more substrates: the arabidopsis major facilitator superfamily revisited. *Mol. Plant* 12, 1182–1202.
- Noé, F. (2008). Probability distributions of molecular observables computed from markov models. *J. Chem. Phys.* 128, 244103.
- Noé, F., and Nüske, F. (2013). A variational approach to modeling slow processes in stochastic dynamical systems. *Multiscale Model. Sim.* 11, 635–655.
- Okazaki, K., Wöhlert, D., Warnau, J., Jung, H., Yildiz, Ö., Kühlbrandt, W., and Hummer, G. (2019). Mechanism of the electroneutral sodium/proton antiporter PaNhaP from transition-path shooting. *Nat. Commun.* 10, 1–10.
- Pao, S.S., Paulsen, I.T., and Saier, M.H. (1998). Major facilitator superfamily. *Microbiol. Mol. Biol. Rev.* 62, 1–34.
- Plett, D.C., Holtham, L.R., Okamoto, M., and Garnett, T.P. (2018). Nitrate uptake and its regulation in relation to improving nitrogen use efficiency in cereals. *Semin. Cell Dev. Biol.* 74, 97–104.
- Quistgaard, E.M., Löw, C., Guettou, F., and Nordlund, P. (2016). Understanding transport by the major facilitator superfamily (MFS): structures pave the way. *Nat. Rev. Mol. Cell Biol.* 17, 123–132.
- Reddy, V.S., Shlykov, M.A., Castillo, R., Sun, E.I., and Saier, M.H. (2012). The major facilitator superfamily (MFS) revisited. *FEBS J.* 279, 2022–2035.
- Roe, D.R., and Cheatham, T.E. (2013). PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* 9, 3084–3095.
- Saier, M.H., Reddy, V.S., Tsu, B.V., Ahmed, M.S., Li, C., and Moreno-Hagelsieb, G. (2015). The transporter classification database (TCDB): Recent advances. *Nucleic Acids Res.* 44, D372–D379.
- Scherer, M.K., Trendelkamp-Schroer, B., Paul, F., Pérez-Hernández, G., Hoffmann, M., Plattner, N., Wehmeyer, C., Prinz, J.-H., and Noé, F. (2015). Pyemma 2: a software package for estimation, validation, and analysis of markov models. *J. Chem. Theory Comput.* 11, 5525–5542.
- Schneider, T., and Stoll, E. (1978). Molecular-dynamics study of a three-dimensional one-component model for distortive phase transitions. *Phys. Rev. B* 17, 1302.
- Schrödinger, LLC (2015). The PyMOL molecular graphics system (Version 1.8) (Schrödinger, LLC). <https://pymol.org/2/>.
- Selvam, B., Mittal, S., and Shukla, D. (2018). Free energy landscape of the complete transport cycle in a key bacterial transporter. *ACS Cent. Sci.* 4, 1146–1154.
- Selvam, B., Yu, Y.-C., Chen, L.-Q., and Shukla, D. (2019). Molecular basis of the glucose transport mechanism in plants. *ACS Cent. Sci.* 5, 1085–1096.
- Shamsi, Z., Cheng, K.J., and Shukla, D. (2018). Reinforcement learning based adaptive sampling: REAPing rewards by exploring protein conformational landscapes. *J. Phys. Chem. B* 122, 8386–8395.
- Shamsi, Z., Moffett, A.S., and Shukla, D. (2017). Enhanced unbiased sampling of protein dynamics using evolutionary coupling information. *Sci. Rep.* 7, 12700.
- Shirts, M.R. (2013). Simple quantitative tests to validate sampling from thermodynamic ensembles. *J. Chem. Theory Comput.* 9, 909–926.
- Shukla, D., Hernández, C.X., Weber, J.K., and Pande, V.S. (2015). Markov state models provide insights into dynamic modulation of protein function. *Acc. Chem. Res.* 48, 414–422.
- Smith, D.E., Cléménçon, B., and Hediger, M.A. (2013). Proton-coupled oligopeptide transporter family SLC15: physiological, pharmacological and pathological implications. *Mol. Aspects Med.* 34, 323–336.
- Solcan, N., Kwok, J., Fowler, P.W., Cameron, A.D., Drew, D., Iwata, S., and Newstead, S. (2012). Alternating access mechanism in the POT family of oligopeptide transporters. *EMBO J.* 31, 3411–3421.
- Takemoto, M., Lee, Y., Ishitani, R., and Nureki, O. (2018). Free energy landscape for the entire transport cycle of triose-phosphate/phosphate translocator. *Structure* 26, 1284–1296.e4.
- Trendelkamp-Schroer, B., Wu, H., Paul, F., and Noé, F. (2015). Estimation and uncertainty of reversible markov models. *J. Chem. Phys.* 143, 11B601_1.
- Unkles, S.E., Karabika, E., Symington, V.F., Cecile, J.L., Rouch, D.A., Akhtar, N., Cromer, B.A., and Kinghorn, J.R. (2012). 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.* 447, 35–42.
- Unkles, S.E., Rouch, D.A., Wang, Y., Siddiqi, M.Y., Glass, A.D.M., and Kinghorn, J.R. (2004). Two perfectly conserved arginine residues are required for substrate binding in a high-affinity nitrate transporter. *Proc. Natl. Acad. Sci. U S A* 101, 17549–17554.

- Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A., and Case, D.A. (2004). Development and testing of a general amber force field. *J. Comput. Chem.* 25, 1157–1174.
- West, I.C. (1997). Ligand conduction and the gated-pore mechanism of transmembrane transport. *Biochim. Biophys. Acta* 1331, 213–234.
- Yan, H., Huang, W., Yan, C., Gong, X., Jiang, S., Zhao, Y., Wang, J., and Shi, Y. (2013). Structure and mechanism of a nitrate transporter. *Cell Rep.* 3, 716–723.
- Yin, Y. (2006). Structure of the multidrug transporter EmrD from *Escherichia coli*. *Science* 312, 741–744.
- Zhang, X.C., Zhao, Y., Heng, J., and Jiang, D. (2015). Energy coupling mechanisms of mfs transporters. *Protein Sci.* 24, 1560–1579.
- Zheng, H., Wisedchaisri, G., and Gonen, T. (2013). Crystal structure of a nitrate/nitrite exchanger. *Nature* 497, 647–651.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited data		
NarK crystal structure (NO ₃ ⁻ bound occluded)	Fukuda et al., 2015	PDB ID: 4U4W
NarK crystal structure (NO ₃ ⁻ bound IF)	Fukuda et al., 2015	PDB ID: 4U4T
NarK crystal structure (Apo IF)	Fukuda et al., 2015	PDB ID: 4U4V
FucP crystal structure (OF)	Dang et al., 2010	PDB ID: 3O7Q
Generated coordinate files (Related to Figures 1 and 5)	This paper	https://github.com/ShuklaGroup/NarK_Structure_2021_Files
Software and algorithms		
CHARMM-GUI	Jo et al., 2008	http://www.charmm-gui.org/
AMBER14	Case et al., 2014	http://ambermd.org
Packmol	Martinez et al., 2009	http://m3g.iqm.unicamp.br/packmol/home.shtml
MSMBuilder 3.6	Beauchamp et al., 2011	http://msmbuilder.org/3.6.0/
MDTraj1.7	McGibbon et al., 2015	https://mdtraj.org/1.9.4/index.html
PyEMMA2.5.7	Scherer et al., 2015	http://www.emma-project.org/latest/index.html
Osprey1.1.0	McGibbon et al., 2016	http://msmbuilder.org/osprey/1.1.0/
VMD1.9.2	Humphrey et al., 1996	https://www.ks.uiuc.edu/Research/vmd/
PyMol 3	Schrödinger, 2015	https://pymol.org/2/
matplotlib	Hunter, 2007	https://matplotlib.org/
In-house python codes	This paper	https://github.com/ShuklaGroup/NarK_Structure_2021_Files

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Diwakar Shukla (diwakar@illinois.edu).

Materials availability

This study did not generate new unique reagents.

Data and code availability

Generated coordinate files, source codes for adaptive MD simulations, MSMs hyper-parameters selection, and MSMs construction used in the paper are available at Github: https://github.com/ShuklaGroup/NarK_Structure_2021_Files. Detailed explanation of the procedures is reported in the Method Details section. All softwares and libraries used are reported in the Method Details section, together with the [key resources table](#).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

This study is computational and did not use any experimental models. The crystal structure (PDB ID: 4U4W ([Fukuda et al., 2015](#))) used as the starting coordinates for simulations was downloaded from Protein Data Bank (PDB) ([Berman, 2000](#)). The other crystal

structures (PDB ID: 4U4T, 4U4V (Fukuda et al., 2015) and 3O7Q (Dang et al., 2010)) used for structural comparison were also downloaded from PDB.

METHOD DETAILS

Molecular dynamics (MD) simulation

The NO_3^- -bound OC NarK crystal structure (PDB ID: 4U4W (Fukuda et al., 2015)) was downloaded from PDB (Berman, 2000) 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 (Jo et al., 2008). The protein was embedded in a phosphatidylcholine (POPC) bilayer, solvated in a box with dimension of $76\text{\AA} \times 76\text{\AA} \times 102\text{\AA}$ 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 (Martínez et al., 2009). The final MD system contains 51,315 atoms. The MD system was first minimized using conjugate gradient method for 10,000 steps. Next, the system was slowly heated from 0 K to 10 K, and then 10 K to 300 K over a period of 1 ns each under canonical (NVT) condition. The system was further equilibrated in isothermal-isobaric (NPT) condition for 50 ns (Braun et al., 2019). All simulations were implemented using Amber14 package (Case et al., 2014) employing Amber ff14SB (Maier et al., 2015) and GAFF (Wang et al., 2004) force field, and carried out in the NPT conditions (300K and 1 atm) maintained using a Berendsen thermostat and a Berendsen barostat (Berendsen et al., 1984). Periodic boundary conditions (PBC) were applied to all simulations, Particle Mesh Ewald was used to treat long-range non-bonded interactions (Essmann et al., 1995), and the SHAKE algorithm was used to constrain hydrogen-containing bonds (Kräutler et al., 2001). 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 Markov state model (MSM) based adaptive sampling protocol. Adaptive sampling is a widely used sampling method and has demonstrated the efficiency of capturing various biologically relevant processes and tremendous value in time and resource savings (Bowman et al., 2010; Shamsi et al., 2017, 2018). The adaptive sampling scheme used in this work is based on the counts of different states in order to discover new states quickly. The procedure of adaptive sampling is:

- (1) Run a series of short MD simulations from initial seeds in parallel. For the first round, 250 parallel simulations were launched from the crystal structure (PDB ID: 4U4W (Fukuda et al., 2015)). After the first round, around 200 parallel simulations were launched per round. The simulation length is set as 30 ns to both satisfy the Markovian assumption and maximize the sampling efficiency.
- (2) Cluster all the simulation data collected so far using the K-means algorithm with number of clusters specified as 5000. Two biologically relevant metrics are selected as clustering metrics: extracellular gating distance (Ser56C α (TM1) - Ala275C α (TM7)) and intracellular gating distance (Met151C α (TM4) - Phe370C α (TM10)).
- (3) Select around 200 structures from the least populated clusters as seeds for the next run of simulation.
- (4) Repeat steps 1-3 until the sampling reaches the convergence criterion: the free energy landscapes stop varying as rounds continue.
- (5) Build a Markov state model (see the Markov state model (MSM) Construction section in STAR methods for details) from the final data set to capture the proper thermodynamics and kinetics. MSMs will help correct any sampling bias introduced by selecting the starting conformations from each round of simulations according to the least populated clusters instead of a Boltzmann distribution.

A total of $\sim 300\text{ }\mu\text{s}$ simulation data were obtained and used for further data analysis. The in-house python codes used to perform adaptive MD sampling are available at https://github.com/ShuklaGroup/NarK_Structure_2021_Files.

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 (Chodera and Noé, 2014; Husic and Pande, 2018). The procedure of constructing a MSM is: (1) featurizing the trajectory data using a set of C α distances between residues; (2) decomposing the featurized data using the time-lagged independent component analysis (tICA) technique, which finds the slowest collective motions in the system through linear combinations of the input features (C α 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 (C α contacts, number of tICA components, and number of clusters) systematically and automatically, we employed a genetic algorithm based technique developed from our lab (Chen et al., 2018) (Tables S1 and S2). The source codes and the resulting data associated with this algorithm are available at https://github.com/ShuklaGroup/NarK_Structure_2021_Files. 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 GMRQ, the better the MSM is at capturing the slowest motions in the system (McGibbon et al., 2015; Noé and Nüske, 2013). The workflow for the genetic algorithm based search of optimal hyper-parameters consist of:

- (1) Prepare the input file which specifies four genetic algorithm parameters: *N_ITERATIONS*, *populationSize*, *percentMutation*, and *percentCrossover*. The input file used in this work *nark.inp* is available at https://github.com/ShuklaGroup/NarK_Structure_2021_Files.
- (2) Generate a pool of all possible residue pairs to explore. A total of $L(L-1)/2$ residue pairs were obtained for NarK where $L = 447$ is the protein length. The generated file *compatiblePairs.txt* is available at https://github.com/ShuklaGroup/NarK_Structure_2021_Files.
- (3) Run the genetic algorithm which automatically constructs MSMs for different combinations of hyper-parameters, evaluates the quality of MSMs with GMRQ, and generates the new generation according to the GMRQ. The algorithm stops when it reaches maximum number of iterations ($N_ITERATIONS = 40$) specified in the input file. The output files for all the 40 iterations (*iter_*_output_sets.txt*) are available at https://github.com/ShuklaGroup/NarK_Structure_2021_Files.

The convergence of GMRQ scores is shown in Figure S6A. Eventually, the highest scored combination of hyper-parameters, that is 61 C α contact distances, 10 tICA components, and 400 clusters, was chosen for the final MSM construction (details are shown in Table S3). A Markovian lag time 15 ns was chosen from the implied timescale plot to construct the MSM (Figure S6B). In-house python codes were used to construct the final MSM model (available at https://github.com/ShuklaGroup/NarK_Structure_2021_Files).

The Osprey variational cross-validation package was used in the genetic algorithm workflow to cross validate the MSMs by varying the hyper-parameters (McGibbon et al., 2016). MSMBuilder 3.6 (Beauchamp et al., 2011) was used to construct MSMs, CPPTRAJ (Roe and Cheatham, 2013) module in AMBER14 (Case et al., 2014) and MDTraj1.7 (McGibbon et al., 2015) were used to analyze simulation trajectories, and VMD1.9.2 (Humphrey et al., 1996) and PyMol 3 (Schrödinger, 2015) were used to visualize MD snapshots.

QUANTIFICATION AND STATISTICAL ANALYSIS

Data analysis were performed using MDTraj1.7 (McGibbon et al., 2015), MSMBuilder 3.6 (Beauchamp et al., 2011), PyEMMA2.5.7 (Scherer et al., 2015), along with in-house python codes (available at https://github.com/ShuklaGroup/NarK_Structure_2021_Files). Plot graphics were generated with matplotlib (Hunter, 2007).

To quantify the uncertainties of the MSM model, we estimated a Bayesian MSM using PyEMMA2.5.7 (Scherer et al., 2015). Bayesian MSMs construct a sample of reversible transition matrices using a Gibbs sampling scheme and are therefore commonly used to quantify the statistical uncertainties for all observables derived from MSM (Noé, 2008; Trendelkamp-Schroer et al., 2015). In this work, we estimated the Bayesian MSM with 100 samples and 95% confidence interval. The sample mean of the stationary distribution was then used to reweigh the free energy landscapes (Figure S4). In order to estimate the uncertainties of free energies, we computed the free energy differences between the Bayesian MSM reweighted landscapes and the maximum likelihood MSM reweighted landscapes (Figure S6). The maximum free energy deviation is less than 0.1 kcal/mol. We therefore use 0.1 kcal/mol with 95% confidence as the upper bound for the approximation errors made by modeling protein dynamics with a MSM.

A caveat for the free energy analysis is that Berendsen thermostat was used to control temperature, which may affect the distributions of the desired ensemble (Bussi et al., 2007; Shirts, 2013). To quantify the errors introduced by the thermostat, we re-ran all the simulations using Langevin thermostat with collision frequency of 2 ps^{-1} (Schneider and Stoll, 1978; Loncharich et al., 1992). The protocol of the new adaptive MD sampling is: (1) initialize $\sim 10,000$ trajectories from 400 MSM states (~ 25 trajectories per MSM state), (2) cluster all the simulation data using the K-means algorithm with number of clusters specified as 5000. Two biologically relevant metrics are selected as clustering metrics: extracellular gating distance (Ser56C α (TM1) - Ala275C α (TM7)) and intracellular gating distance (Met151C α (TM4) - Phe370C α (TM10)), (3) select around 500 structures from the least populated clusters as seeds for the second run of simulation. An aggregate simulation time of $\sim 230 \mu\text{s}$ was obtained. We reconstructed a MSM for the new simulation data and generated new free energy landscapes for comparison (Figure S4). The comparison of all four free energy landscapes shows that (1) the overall shapes of free energy landscapes do not vary, (2) the free energy differences between different states remain the same, (3) although there are minor free energy differences for some regions, the differences are much less than 0.1 kcal/mol which is the error bar of free energies determined through uncertainty test. Therefore, we suggest that the free energy analysis is still valid, because the errors caused by the thermostat choice are much less than the error bars.