Markov-State Transition Path Analysis of Electrostatic Channeling

Yuanchao Liu,†,* David P. Hickey,§ Shelley D. Minteer,§ Alex Dickson,*‡§ and Scott Calabrese Barton*†§

†Department of Chemical Engineering and Materials Science and ‡Department of Biochemistry and Molecular Biology, Michigan State University, East Lansing, Michigan 48824, United States
§Department of Chemistry, University of Utah, Salt Lake City, Utah 84112, United States

ABSTRACT: Electrostatic channeling is a naturally occurring approach to control the flux of charged intermediates in catalytic cascades. Computational techniques have enabled quantitative understanding of such mechanisms, augmenting experimental approaches by modeling molecular interactions in atomic detail. In this work, we report the first utilization of a Markov-state model (MSM) to describe the surface diffusion of a reaction intermediate, glucose 6-phosphate, on an artificially modified cascade where hexokinase and glucose-6-phosphate dehydrogenase are covalently conjugated by a cationic oligopeptide bridge. Conformation space networks are used to represent intermediate transport on enzyme surfaces, along with committor probabilities that assess the desorption probability of the intermediate on each segment of the channeling pathway. For the region between the peptide bridge and downstream active site, the ionic strength dependence of desorption probability by MSM agreed well with that by transition state theory. A kinetic Monte Carlo model integrates parameters from different computational methods to evaluate the contribution of desorption during each step. The approach is validated by calculation of kinetic lag time, which agrees well with experimental results. These results further demonstrate the applicability of molecular simulations and advanced sampling techniques to the design of chemical networks.

INTRODUCTION

Multistep catalytic cascades have the potential to increase the catalytic efficiency of chemical/biochemical synthesis via well-controlled reactant and intermediate flux and to enhance the output of catalytic devices such as biofuel cells and biosensors.1–5 Natural metabolism reveals a variety of cascade reactions confined in single compartments, wherein a wide range of biomolecules are synthesized from few simple precursors, and energy is efficiently extracted by deep oxidation of biofuels.4,5 A key limitation of these stepwise reactions is the mass transport of reaction intermediates, which in nature is found to be largely facilitated by substrate channeling phenomena, wherein reaction intermediate molecules are directly shuttled to downstream active sites instead of equilibrated to bulk media.6 Given the complex chemical environment of the cell, channeling increases catalytic efficiency and can prevent intermediates from binding to unproductive sites. The elucidation of these transport processes is therefore of great significance to the understanding of natural channeling mechanisms and application to synthetic cascades.

An effective reaction cascade requires an appropriate spatial organization of catalysts and a favorable pathway between sequential reactive sites. Simple proximal location of such sequential sites has so far proved insufficient to account for the increase of catalytic efficiency, and functional channeling based on molecular interactions is required.5–8 Widely recognized natural channeling mechanisms include intramolecular tunnels,9,10 electrostatic channeling,11–15 chemical swing arms,16,17 and spatial organization.18–20

Electrostatic channeling may be described as diffusion of a charged intermediate molecule within an electrostatic field generated by an oppositely charged pathway. Electrostatic channeling represents a more general mechanism for channeling of ionic substrates, as compared to structure-specific chemical swing arms and sterically bound intramolecular tunneling. So far, the study of natural electrostatic cascades has focused on the bifunctional enzyme thymidylate synthase–dihydrofolate reductase and the tricarboxylic acid (TCA) cycle supercomplex malate dehydrogenase–citrate synthase.21–24 For both natural cascades, X-ray diffraction and cross-linking mass spectrometry have revealed a positively charged pathway between the two active sites, suggesting an electrostatic channeling process.11,17 Meanwhile, cascade kinetics have been experimentally studied via lag time analysis and competing bulk reactions, both of which show a significant increase in catalytic efficiency as compared to the nonchanneled enzyme pairs.23,25 Additionally, the kinetic increase can be largely disabled by increasing the ionic strength (IS) of the surrounding environment23,25 or by neutralization of the channeling pathway.12,24 Such mechanisms were validated by our recent report on synthetic...
casades with artificially introduced electrostatic channeling, wherein hexokinase (HK) and glucose-6-phosphate dehydrogenase (G6PDH) were covalently conjugated by a cationic oligopeptide bridge.25,28

So far, experimental results provide strong evidence of the occurrence of electrostatic channeling but are also limited because of the lack of in situ insight at a molecular level. In response, various types of models were developed to further study the channeling mechanism and quantify cascade kinetics. Continuum modeling was first applied to study the channeling process, where the spatial distribution of reaction intermediates was represented by a continuous medium.13,26,27 Using well-defined geometry and transport parameters, the migration of intermediates was shown to strongly control cascade efficiency and product yield, particularly at low IS. However, molecular-level interactions cannot be well represented in the continuum model, and the cascade geometry was coarse-grained for simplicity.

The first molecular simulations of electrostatic channeling were reported by McCammon’s group, using Brownian dynamics to study the probability that simplified but explicit intermediate molecules would reach the vicinity (~0.7 nm) of a downstream active site.7,12,15,27 The resulting probabilities were applied to microkinetic models to estimate the overall cascade kinetics.23 These simulations gave a more detailed description of the channeling process than continuum modeling, but the strong hydrogen bond interactions between charged/polarized intermediate molecules and the cascade surface were not well represented. Short-range electrostatic interactions with cascade surfaces were recently shown by our work to be important in electrostatic channeling, especially at high IS.25,28

Our study on artificially modified cascades (HK–G6PDH) gave a detailed explanation of the surface diffusion via hydrogen bond interactions and its contribution to cascade kinetics.25,28 In this work, a surface “hopping” mode was established by molecular dynamics (MD) simulations, and hydrogen-bond-directed surface diffusion was observed on an energy-uniform surface, such as a fully saturated lysine (LYS) and arginine (ARG) oligopeptides. A kinetic Monte Carlo (KMC) model was used to validate the model against experimental results, pointing to the channeling pathway length and desorption during individual hops as key factors to overall channeling efficiency. In addition, an IS study using transition state theory (TST) and umbrella sampling (US) showed that the desorption probability for individual hops on a saturated LYS surface increased from 0.5% at 0 mM to 3.5% at 120 mM IS. During transport over the short (~2 nm) and nonuniform pathway between the bridge and G6PDH active site, the desorption probability was estimated to be 23% at 120 mM IS.

Markov-state models (MSMs) can be used to integrate large collections of short MD trajectories and predict long time-scale phenomena, based on the assumption that future states depend only on the current state instead of the history.7,15,29–32 Recently, MSM has been widely used to describe protein dynamics and ligand binding, where the systems include conformations with multiple degrees of freedom and highly interconnected transition pathways.33–41 As a result, a conformation space network (CSN) is usually mapped to visualize the dynamic transitions between meta-states of the system. Combined with transition path theory (TPT),35,39,42–44 the probability of the flux along each explicit transition pathway can be quantitatively calculated to enrich the MSM output, giving a complete energy landscape of the conformation space.

This ability of MSM to describe energy landscapes over multidimensional transition pathways makes it a promising approach to describe intermediate channeling in reaction cascades. Herein, we report an MSM to study the intermediate (glucose 6-phosphate, G6P) transitions on the pathway segment between the upstream HK active site and LYS bridge as well as the region from the bridge to the downstream G6PDH active site. On the basis of parallel MD simulations of intermediate trajectories around the cascade surface, an MSM transition matrix was obtained and a CSN37 was used to visualize the transition pathways during the channeling process. On the basis of TPT,35,42–44 committor probabilities were calculated to estimate the desorption probabilities along each channeling segment, which were then integrated in a KMC model and validated against experimentally measured stopped-flow lag times.

COMPUTATIONAL DETAILS

Details of MD simulations, the KMC model, and experiments can be found in the Supporting Information and our previous work.25,28

Featurization of MD Trajectories. MD trajectories for the bridge–G6PDH segment were taken from our previous work,28 wherein the G6P molecule was initialized on the surface of G6PDH, between the peptide bridge and binding pocket, such that it had comparable probability to reach either site. Starting at the release point, 500 parallel short simulations (2 ns) were conducted with randomly regenerated atom velocities and counter ions. The resulting 105 MD frames were reanalyzed using an MSM in this work.

These MD trajectories with full atom coordinates were featurized into a representative time-dependent vector. Reference groups were selected on the G6P molecule and bridge–G6PDH complex to represent the position and orientation of the G6P molecule. Specifically, the phosphorus atom and center of mass (COM) of G6P were selected, along with the COM of 15 residues on the surface of the bridge–G6PDH complex (Figure S1). Additionally, the atoms of the bridge–G6PDH complex that was nearest to G6P COM was dynamically chosen to represent the surface distance. As a result, by calculating distances between the two G6P reference groups and 16 groups on the cascade surface, a feature vector of 32 elements, \( V(t) = [v_{11}, v_{12}, v_{13}, ..., v_{32}] \), was obtained from the MD trajectories.

For the HK–bridge segment, 500 parallel, 10 ns MD trajectories were generated. In addition to the phosphorus atom and COM of G6P, 20 surface residues and one dynamically chosen nearest atom were used as reference groups. This resulted in a 42-element feature vector, \( V(t) \).

Markov-State Model. The Python package MSMBuilder was used to build MSMs.45 Using the “MiniBatchKMeans” clustering algorithm, the positions and orientations of G6P relative to the cascade surface (HK–bridge and bridge–G6PDH) in each MD frame were grouped into 1000 meta-states, \([s_{11}, s_{12}, ..., s_{1000}]\). Each meta-state contains many MD frames with similar conformations based on the feature vector. Transitions between these meta-states over a time step of duration \( \Delta t = 10 \) ps were counted, resulting in a 1000 × 1000 count matrix, \( N(\Delta t) \).

As the discrete hopping between each meta-state pair was dynamically reversible, the count matrix, \( N \), may be optionally
symmetrized using the relation $N_{\text{sym}} = (N + N^T)/2$, where $N^T$ is the transpose of $N$. Such a symmetric count matrix reflects an equilibrium condition where the counts of forward and reverse transitions between each meta-state pair are equal. Though the symmetrization step was included in this study, results are also presented for the case where the count matrix is not symmetrized (Supporting Information).

Each row of the count matrix may be normalized to obtain the discrete probability distribution of transition from a single meta-state to all other meta-states. The resulting transition probabilities were stored in a transition matrix, $M(\Delta t)$, as shown in Figure S2. If the system is known to be in a meta-state $s_0$, $M(\Delta t)$ represents the likelihood of making a transition to meta-state $s_f$ after one time step.\(^{29,40}\) Such transitions can be calculated over $n$ time steps by repeated multiplication by $M(\Delta t)$

$$s(t + n\Delta t) = s(t) \cdot M(\Delta t)^n \quad (1)$$

As $n \rightarrow \infty$, the transition matrix, $M^\infty$, converges to a stationary transition matrix, $M(\infty)$, representing a stationary distribution of system states that follows a Boltzmann distribution. Each row of $M(\infty)$ is identical, and represents the eigenvector, $\epsilon$, associated with the largest eigenvalue of $M(\Delta t)$.

A CSN of G6P channeling on each segment was rendered using Gephi software with the “Yifan Hu” layout algorithm.\(^{47}\) Each node of a CSN represents a meta-state, $s_0$, with node size proportional to its stationary population, $M(\infty) = \epsilon$. Edges connecting each node pair indicate transitions between the meta-states, with thickness representing the average transition probability $(M(\Delta t) + M(\Delta t))/2$. The CSN is thus a visual representation of the entire energy landscape sampled by the ensemble of MD simulations.

**Committer Probabilities.** Committer probabilities, based on TPT, were employed to quantify the probability of desorption, and therefore the efficiency of the channeling process.\(^{35,42-44}\) Committer probabilities are calculated by defining groups of meta-states (called basins) and determining the probability that the system, starting from a specific meta-state (node of the CSN), will reach a particular basin first—that is, before any of the other basins. These basins are considered energy minima from which the system will not emerge. For any meta-state, the committer probabilities to all basins sum to 1, as the system is ergodic and will ultimately reach one of the basins.

On the bridge–G6PDH segment, the set of meta-states wherein the phosphorus atom of G6P lay less than 0.7 nm from the bridge surface was denoted as the bridge basin. Meta-states with the phosphorus atom less than 1.4 nm from the COM of G6PDH’s binding pocket were denoted as the pocket basin, and meta-states with phosphorous atoms more than 0.7 + 1.2 = 1.9 nm above the cascade surface were defined as the desorption basin. For the HK–bridge segment, the cutoff for the binding pocket basin was 2.0 nm. A committer probability matrix, $S(\infty)$, was obtained by first zeroing each row of $M(\Delta t)$ associated with a basin, and setting the diagonal element to one, yielding a sink matrix $S(\Delta t)$. Then, $S(\infty)$ was obtained using $S(\Delta t)^n$ for $n \rightarrow \infty$.

Using row $S(\infty)$ and summing over columns associated with the three basins (bridge, pocket, desorption), a three-component committer probability vector, $p_{c3}(i)$, can be calculated for each meta-state $s_i$:

$$p_{c3}(i) = [p_{c3}^{\text{des}}(i), p_{c3}^{\text{poc}}(i), p_{c3}^{\text{brg}}(i)] \quad (2)$$

where superscripts indicate the probability of transition to a specific basin. Note that for each meta-state, $s_i$, the components of $p_{c3}(i)$ sum to 1.

Using $p_{c3}(i)$, an overall three-basin desorption probability, $p_{\text{des}}^{\text{channel}}$ was obtained. A set of states, $s_0$, were chosen which had approximately equal probability of reaching the pocket or the bridge, specifically $|p_{c3}^{\text{poc}}(k) - p_{c3}^{\text{brg}}(k)| < 2\%$. This can be considered as a set of transition states between the bridge and the pocket. Desorption probability was then calculated according to

$$p_{\text{channel}}^{\text{des}} = \sum_{k=0}^{N_b} \frac{p_{c3}^{\text{des}}(k)}{(p_{c3}^{\text{poc}}(k), p_{c3}^{\text{brg}}(k))} \epsilon_k \quad (3)$$

$$p_{\text{shrink}}^{\text{des}} = \frac{p_{\text{des}}^{\text{channel}}}{p_{\text{des}}^{\text{channel}} + p_{\text{des}}^{\text{channel}}} \quad (4)$$

where the summation is weighted by the stationary population of each meta-state, given by $\epsilon_k$.
Alternatively, a two-basin committor probability, \( p_{c_2}(i) \), was calculated by disabling one basin, typically the source of that segment of the channeling pathway. Using the HK−bridge segment as an example, only the bridge and desorption were defined as basins; two-basin committor probabilities and the resulting desorption probability were calculated for meta-states associated with the pocket by

\[
p_{c_2}(i) = [p_{c_2}^{\text{des}}(i), p_{c_2}^{\text{brg}}(i), 0]
\]

\[
p_{\text{des}}^{\text{bin}} = \sum_{\text{pocket}} p_{c_2}^{\text{des}}(i) \varepsilon_i
\]

\[\text{(5)}\]
\[\text{(6)}\]

**RESULTS AND DISCUSSION**

**Overview of Cascade Structure and Computational Methodologies.** Figure 1 shows the cascade structure (HK−bridge−G6PDH), where HK and G6PDH were covalently linked by a fully saturated peptide bridge. Our previous work showed that the energy and rate terms can be well quantified for individual hops on the energy-uniform surface of the peptide bridge.\(^{25,28}\) That is, each hop was accompanied by a certain probability of desorption, depending on the energy difference between the hopping and desorption energy barriers (Figure 1b). However, the pathway between bridge and active sites is a nonuniform 2-dimensional surface composed of flexible hydrophobic, polarized, and charged amino acid residues, necessitating more sophisticated computational approaches.

Specifically, the net charge of G6PDH is slightly positive (+1) and it is even more positive around the binding pocket. On the bridge−G6PDH segment, therefore, the G6P molecule (−2 charge) mostly experiences attractive electrostatic interactions with an oppositely charged surface. As shown in Figure 1c, the bridge−G6PDH segment is short (2 nm) and a simplified energy barrier can be assumed between the bridge and G6PDH’s binding pocket. Previously, a direct probability measurement was conducted by releasing the G6P molecule around the transition region, which resulted in a comparable probability for G6P to reach the bridge and pocket sites. By utilizing the last MD frame from each parallel simulation, a transition-state-based desorption probability was estimated and applied to a KMC model to quantify the cascade kinetics.\(^{28}\)

In contrast, the HK monomer possesses a highly negative net charge of −16. Electrostatic channeling in such an environment depends on local surface polarization, with a higher probability of desorption. Moreover, the HK−bridge segment is longer (≈4.1 nm) as shown in Figure 1a, and there exist multiple energy barriers and minima, making the HK−bridge segment more complex as compared to the bridge−G6PDH segment. As a result, the simplification to a single energy barrier, as previously applied to G6PDH, cannot be well established for G6P transport over HK.

Moreover, the previous TST desorption probability for a bridge−G6PDH segment only utilized the final frame of each parallel MD simulation to determine the channeling results, a small sample of the channeling process. To fully utilize all MD frames and give a completed description of intermediate transport, two separate MSMs are used in this work to integrate full MD trajectories and map the CSN for G6P’s channeling, first from the HK pocket to the LYS bridge (Figure...
1a), and then from the bridge to the G6PDH pocket (Figure 1c). Transition-path-based committor probabilities \footnote{35,42–44} are employed to quantify the desorption probability in each region.

**CSN for G6P on a Cascade Surface.** Figure 2 shows the CSN of the G6P molecule on the bridge–G6PDH segment in an ionic environment at 120 mM. Three major areas are circled in Figure 2a, including the “bridge” site, “pocket” site, and transition area. After MD simulations, G6P’s trajectories were collected and the density isosurfaces were mapped as shown in Figure 2b,c. Figure 2d–f shows the resulting CSNs, where each node stands for a meta-state corresponding to a group of similar G6P’s positions on a cascade surface. Node size is proportional to meta-state population and the edges indicate the transitions between nodes. (e) CSN colored by three-basin committor probabilities to each basin. (f) Two-basin CSN using “bridge” and “bulk” basins. (g,h) Two-basin CSN for IS of 40 and 120 mmol/L cases. (i) Triangle color bar where RGB components represent committor probabilities to each basin.

Figure 3. Conformational ensemble of the G6P molecule on the HK–bridge pathway at 0 mmol/L IS. (a) Top view of the HK–bridge complex structure. (b) Top view of G6P density isosurfaces corresponding to different CSN subsets. (c) Side view of G6P’s isosurfaces. (d) CSN with three predefined basins, including “bulk” in red, “pocket” in blue, and “bridge” in green. Node size is proportional to the population and the edges indicate the transitions between nodes. (e) CSN colored by three-basin committor probabilities to each basin. (f) Two-basin CSN using “bridge” and “bulk” basins. (g,h) Two-basin CSN for IS of 40 and 120 mmol/L cases. (i) Triangle color bar where RGB components represent committor probabilities to each basin.

The cyan states, for example, have roughly equal committor probabilities to green and blue basins, as shown by the triangular color bar in Figure 2i. Meanwhile, these states also have a certain committor probability to the desorption basin, \(p_{3bsn}^{des}(i)\). As stated in Methods, a three-basin desorption probability, \(p_{3bsn}^{des}\), was estimated by taking the weighted desorption probability for the meta-states with comparable committor probabilities (<2%) to bridge and pocket basins (eqs 2–4). This approach to determining desorption probability represents the basic idea in our previous work, wherein desorption probability was estimated by a direct probability analysis based on TST. This method is effective when no strong energy minima exist outside of the three basins, and the transition meta-states can be assumed to possess high energy levels.

An alternative analysis involving two-basin desorption probability was conducted by disabling the bridge as a basin. As a result, only pocket and bulk were considered as basins when calculating the committor probability. Graphically, the blue color was removed from Figure 2e and the red and green colors were renormalized to sum to one. This significantly affects the original blue region, as shown by the oval circle in Figure 2f. In this way, the color of the original bridge basin gives a qualitative visualization of the desorption probability of G6P when starting from the LYS bridge, as compared to the probability of channeling to G6PDH’s active site. For example, the yellow states in Figure 2f have equal probability of transitioning to either the desorption (red) or pocket (green) basin. As shown in Figure 2f–h, the color of the previous bridge basin states (oval) turned from green to yellow with increasing IS from 0 to 120 mM, indicating significant desorption under concentrated ionic environments. A two-
basin desorption probability, $p_{\text{des}}^{\text{3bsn}}$, was calculated according to eqs 5 and 6.

The yellow color in Figure 2f–h was useful to qualitatively visualize the desorption probability of any transition state. Therefore, we consistently color nodes such that the desorption basin is red, the source of the channeling pathway is blue, and the destination is green. Therefore, the bridge basins are colored green in the CSN of the upstream HK—bridge segment and blue in the downstream bridge—G6PDH CSN.

The HK—bridge segment is longer (~4.1 nm) as shown in Figure 3 and the length of each parallel simulation was extended to 10 ns. As compared to the CSN of bridge—G6PDH (Figure 2), one extra major cluster was found between the pocket and bridge communities, as shown in Figure 3e. This state community, which appears as cyan, was associated with the LYS-428 and ARG-423 of HK, and represents an electrostatic energy minimum on the HK surface. Such “kinetic traps” justify the MSM approach over the simplified energy barrier previously employed for the bridge—G6PDH segment.28

CSNs colored by 2-basin committor probabilities, as shown in Figure 3f, enable the visualization of IS-dependent desorption probability from the HK to the bridge. As IS increases from 0 to 120 mM, the color of the original pocket basin transitions from light green to orange, indicating an increasing affinity to bulk states (red). For example, the resulting desorption probability by eq 6 was 41% at 0 mM, which agrees well with the CSN colored by 2-basin committor probability shown in Figure 3f.

**IS Dependence and Validation by Direct Probability Measurement.** Comparing the CSNs of the HK—bridge and bridge—G6PDH segments, differences include channeling segment length (~4.1 nm vs ~2 nm) and additional energy minima on the HK surface (Figures 2e and 3e). Significantly, the bridge basin is the destination for HK—bridge transport, but the source in the bridge—G6PDH case. As previously discussed regarding Figure 1, hopping on the LYS bridge can be well represented by TST and US, whereas MSM is employed mainly to study channeling on enzyme surfaces. For this reason, application of a 2-basin desorption probability, $p_{\text{des}}^{\text{2bsn}}$, to the bridge—G6PDH segment would introduce significant desorption from the bridge, which was already parameterized by TST and US in the KMC model.

In contrast, a 3-basin desorption probability ($p_{\text{des}}^{\text{3bsn}}$) only includes desorption from a single bridge residue closest to G6PDH, which effectively avoids parameter overlap between the bridge and bridge—G6PDH segments. No significant energy minima were observed on the G6PDH surface between the bridge and pocket, indicating that intermediate transport on this area is mostly overcoming transition states at high energy levels. For these reasons, $p_{\text{des}}^{\text{3bsn}}$ from a 3-basin CSN (Figure 2e) was chosen to represent transport over the bridge—G6PDH segment.

To summarize, a 2-basin desorption probability, $p_{\text{des}}^{\text{2bsn}}$, was used for the HK—bridge segment and 3-basin desorption probability, $p_{\text{des}}^{\text{3bsn}}$, was used for the bridge—G6PDH segment. Figure 4 summarizes the IS dependence of desorption probability for G6P transport on the HK—bridge and bridge—G6PDH segments and for individual hops on the LYS bridge. Generally, all types of desorption increase with increasing IS values because of increased shielding of electrostatic interactions. Specifically, individual hops on the

**LYS bridge reflect the lowest desorption probability, because of the strong hydrogen bond interactions between G6P’s phosphate and LYS’s ε-ammonium groups. As discussed in our previous work,28 such strong interactions were able to maintain a desorption probability as low as 3.5% at high IS (120 mM). As for the bridge—G6PDH segment, the bridge is strongly charged and G6PDH has a relatively neutral net charge (+1.1). As a result, the long-range electrostatic impact of the LYS bridge was able to fully cover the transition area and even reach the binding pocket of G6PDH, providing a good electrostatic confinement of surface diffusion. As a result, desorption probability on the bridge—G6PDH segment at 0 mM can be as low as 5% as shown in Figure 4. At 120 mM, that probability increased to 25%. Through the entire IS range, the desorption probabilities as calculated by the 3-basin CSN, $p_{\text{des}}^{\text{3bsn}}$, agree with previous results from TST analysis within error.28

On the HK—bridge segment, desorption probability increased to 40−62%. One reason is the net charge of −16 on the HK subunit, which introduces long-range repulsion that frustrates electrostatic interactions with the LYS bridge. The Debye length, a measure of the range of electrostatic interactions, decreases from 9.8 nm at 1 mM IS, to 2.2 nm at 20 mM and 0.9 nm at 120 mM. Given the 4.1 nm length of the HK—bridge segment, long-range electrostatic confinement was likely negligible at higher IS, and channeling relies on local electrostatic interactions, with higher probability of desorption. Only when the intermediate approaches the bridge, (light blue cluster in Figure 3e), did desorption probability significantly decrease.

It should be noted that the channeling efficiency at low IS may be underestimated, because meta-states of more than 1.9 nm from the cascade surface were defined as the desorption basin, even though Debye lengths were greater than this distance (e.g., 9.8 nm at 1 mM). Therefore, the bridge—HK complex still exerts certain electrostatic forces on intermediate molecules in “desorption” states.

**KMC Model on Cascade Kinetics.** To further quantify the cascade kinetics, a previously developed KMC model was used to integrate probabilities generated by the MSM and other computational models.28 Figure 5a shows the schematic diagram of KMC model. Specifically, the two sequential active sites, E1 and E2, were connected by discrete hopping sites numbered from 1 to 4. Explicit rate constants were assigned to individual events, including hopping, desorption, adsorption,
and reaction. A universal bulk reservoir was employed to modulate the concentration of reaction intermediates. All event sites were allowed to exchange molecules to the bulk environment. Hopping on the bridge is reversible, but is irreversible between active sites (E₁, E₂) and the bridge.

MSM-derived desorption probability (Figure 4) was used to parameterize the KMC model. As discussed in our previous work, the intermediate surface transition rate (>1 μs⁻¹) was much faster than the turnover frequency of active sites (0.6–7 s⁻¹). Therefore, instead of their absolute values, the ratio of the hopping and desorption rate constants were considered. The MSM desorption probabilities, \( P_{\text{des}} \), were used to parameterize the ratio \( k_{\text{hop}} / k_{\text{des}} \) according to:

\[
P_{\text{des}} = \frac{k_{\text{des}}}{k_{\text{des}} + k_{\text{hop}}}
\]

Figure 5b shows IS-dependent lag time plots for the HK–bridge–G6PDH cascade with calculated results compared to the experiment. Detailed discussion of these lag time experiments as well as the TST results and calculation for the bridge alone can be found in our previous work. Briefly, the final experimental solutions (pH = 7.0) contained 1.4 mM citrate, 0.5 mM ATP, 5.6 mM MgCl₂, and 0.8 mM NAD⁺, adding up to an approximately 20 mM background IS. In the stop-flow lag time experiments, substrate-saturated kinetics at HK and observable lag times were obtained by using concentrations of 278 mM for glucose and 8 nM for the HK–bridge–G6PDH complex. Product evolution at the downstream enzyme (G6PDH) was determined spectrophotometrically by monitoring the production of NADPH at a wavelength of 340 nm.

The turnover frequency of G6PDH (≈6.2 s⁻¹) was much faster than that of upstream HK (≈0.7 s⁻¹), maximizing the capability for channeling as compared to desorption. In a reaction-limited process, such a fast downstream reaction rate results in an empty channel as described by queuing theory.

The contribution of a bridge comprising five LYS residues was studied by disallowing desorption between the bridge and active sites. The slope of the KMC lag time (bridge) agrees well with the experiment, but the absolute values were consistently lower. By including desorption in the bridge–G6PDH region, as calculated by TST (bridge–G6PDH–TST), the calculated lag time shifts upward, closer to the experimental data, whereas the slope becomes more comparable to the experiment. Replacing the TST calculation with the MSM result (bridge–G6PDH–MSM) did not significantly change the lag time plot, consistent with the desorption probability plots in Figure 4. Finally, MSM-based desorption from the HK–bridge segment is considered to yield a KMC analysis of the complete pathway (HK–bridge–G6PDH). The calculated lag time shifts further upward to the same level as the experimental lag time, with best agreement at high IS value, and underpredicted lag time at low IS.

As discussed above, the 1.9 nm definition of the desorption basin likely explains discrepancies at low IS. This boundary is controlled largely by limitations in the size of the MD box (~2–4 nm gap between cascade surface and box boundary). In the present MD simulations with full atoms, it is impractical to specify a distance between the cascade surface and box boundary greater than 10 nm. This limitation may be overcome by using a multiscale approach; for example, MD simulations could be used to build the MSM for the proximal surface diffusion and Brownian dynamics or implicit solvent MD simulations used to consider long-range electrostatics. Such a combination of techniques in a single CSN is a topic for future study.

**Design Rules for Future Cascades.** On the basis of the present analysis, cascade designs incorporating electrostatic channeling must still seek to maximize long-range electrostatic interactions as well as site–site proximity. Mutation of linker sites could be an option to realize the former goal and segment length should be well balanced to enhance adsorption but not greatly exceed the Debye length in the working environment. Where possible, any strong local energy minima on channeling surfaces should be eliminated, in order to create a continuous, favorable transition pathway and avoid significant desorption.

**CONCLUSIONS**

A Markov-state approach was applied to the simulation of surface diffusion via electrostatic channeling. The multidimensional pathways over the surface of a two-enzyme cascade were mapped from millions of physical states to a smaller distribution of meta-states in a human-readable CSN. MSM-derived intermediate transport efficiency from the LYS bridge to the binding pocket of G6PDH was found to agree well with a previous, simplified TST analysis. The longer pathway between HK and the LYS bridge introduced additional energy minima as well as overall electrostatic repulsion, resulting in significantly increased desorption.

A KMC model was used to integrate desorption probabilities as calculated by varying computational methods, and allowed comparison to experimental results. Desorption probabilities...
from different segments of the cascade (bridge, bridge—
G6PDH, and HK—bridge) were introduced independently,
elucidating the contribution of each segment to the overall
cascade efficiency. Comparison with the experiment was
poorest at low IS, where MD simulations are limited to length
scales lower than the Debye length. A combination of MD and
another, more coarse-grained molecular simulation could be a
possible way to address this issue.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge on the
ACS Publications website at DOI: 10.1021/acs.jpcc.9b02844.

Description of experimental details; parameters of MD
simulations and KMC model; crystal structure of the
cascade complex; and algorithm of MSM and committor
probability calculation (PDF)

## AUTHOR INFORMATION

### Corresponding Authors

*E-mail: alexrd@msu.edu (A.D.)
*E-mail: scb@msu.edu (S.C.B.)

### ORCID

Yuanchao Liu: 0000-0002-3669-8071
David P. Hickey: 0000-0002-5297-6895
Shelley D. Minteer: 0000-0002-5788-2249
Alex Dickson: 0000-0002-9640-1380
Scott Calabrese Barton: 0000-0002-1407-0275

### Author Contributions

The paper was written through contributions of all the authors.
All the authors have given approval to the final version of
the paper.

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We gratefully acknowledge support from the Army Research Office MURI (#W911NF1410263) via the University of Utah.
A.D. acknowledges support from the National Institutes of Health grant R01GM130794. Computational work in support
of this research was performed at Michigan State University’s High Performance Computing Facility.

## REFERENCES

(1) Denard, C. A.; Hartwig, J. F.; Zhao, H. Multistep One-Pot
Reactions Combining Biocatalysts and Chemical Catalysts for
Asymmetric Synthesis. ACS Catal. 2013, 3, 2856–2864.
(2) Sperl, J. M.; Sieber, V. Multienzyme Cascade Reactions—Status
and Recent Advances. ACS Catal. 2018, 8, 2385–2396.
(3) Wang, Y.; Ren, H.; Zhao, H. Expanding the Boundary of
Biocatalysis: Design and Optimization of in Vivo Tandem Catalytic
Reactions for Biochemical Production. Crit. Rev. Biochem. Mol. Biol.
2018, 53, 115–129.
(4) Sreer, P. A. The Metabolon. Trends Biochem. Sci. 1985, 10, 109–
110.
(5) Vélot, C.; Mixon, M. B.; Teige, M.; Sreer, P. A. Model of a
Quinary Structure between Krebs TCA Cycle Enzymes: A Model for
the Metabolon?. Biochemistry 1997, 36, 14271–14276.
(6) Wheelon, I.; Minteer, S. D.; Banta, S.; Barton, S. C.; Atanassov,
P.; Sigman, M. Substrate Channeling as an Approach to Cascade
Reactions. Nat. Chem. 2016, 8, 299–309.
(7) Bauler, P.; Huber, G.; Leyh, T.; McCammon, J. A. Channeling
by Proximity: The Catalytic Advantages of Active Site Colocalization
Using Brownian Dynamics. J. Phys. Chem. Lett. 2010, 1, 1332–1335.
(8) Zhang, Y.; Tsitikov, S.; Hess, H. Proximity Does Not Contribute
to Activity Enhancement in the Glucose Oxidase–Horseradish
Peroxidase Cascade. Nat. Commun. 2016, 7, 13982.
(9) Schneider, T. R.; Gerhardt, E.; Lee, M.; Liang, P.-H.; Anderson,
K. S.; Schlichting, I. Loop Closure and Intersubunit Communication
in Tryptophan Synthase†. Biochemistry 1998, 37, 5394–5406.
(10) Amaro, R.; Luthey-Schulten, Z. Molecular dynamics
simulations of substrate channeling through an α-β barrel protein.
Chem. Phys. 2004, 307, 147–155.
(11) Knighton, D. R.; Kan, C.-C.; Howland, E.; Janson, C. A.;
Hostomska, Z.; Welsh, K. M.; Matthews, D. A. Structure of and
kinetic channeling in bifunctional dihydrofolate reductase-thymidy-
late synthase. Nat. Struct. Biol. 1994, 1, 186–194.
(12) Elcock, A. H.; Potter, M. J.; Matthews, D. A.; Knighton, D. R.;
McCammon, J. A. Electrostatic Channeling in the Bifunctional
Enzyme Dihydrofolate Reductase-Thymidylate Synthase. J. Mol. Biol.
1996, 262, 370–374.
(13) Eun, C.; Keekenes-Huskey, P. M.; Metzger, V. T.; McCammon,
J. A. A Model Study of Sequential Enzyme Reactions and Electrostatic
Channeling. J. Chem. Phys. 2014, 140, 105101.
(14) Wu, F.; Minteer, S. Krebs Cycle Metabolon: Structural
Evidence of Substrate Channeling Revealed by Cross-Linking and
Mass Spectrometry. Angew. Chem., Int. Ed. 2015, 54, 1851–1854.
(15) Wang, N.; McCammon, J. A. Substrate Channeling between the
Human Dihydrofolate Reductase and Thymidylate Synthase. Protein
Sci. 2016, 25, 79–86.
(16) Zhou, Z. H.; McCarthy, D. B.; O’Connor, C. M.; Reed, L. J.;
Stoops, J. K. The Remarkable Structural and Functional Organization
of the Eukaryotic Pyruvate Dehydrogenase Complexes. Proc. Natl.
Acad. Sci. U.S.A. 2001, 98, 14802–14807.
(17) Fu, J.; Yang, Y. R.; Johnson-Buck, A.; Liu, M.; Liu, Y.; Walter,
N. G.; Woodbury, N. W.; Yan, H. Multi-Enzyme Complexes on DNA
Scaffolds Capable of Substrate Channelling with an Artificial Swinging
Arm. Nat. Nanotechnol. 2014, 9, 531–536.
(18) Roberts, C. C.; Chang, C.-c. A. Model of Enhanced Catalysis
in Multienzyme Nanostructures: Effect of Molecular Scaffolds, Spatial
Organization, and Concentration. J. Chem. Theory Comput. 2015, 11,
286–292.
(19) Fu, J.; Liu, M.; Liu, Y.; Woodbury, N. W.; Yan, H. Interenzyme
Substrate Diffusion for an Enzyme Cascade Organized on Spatially
Addressable DNA Nanostructures. J. Am. Chem. Soc. 2012, 134,
5516–5519.
(20) Idan, O.; Hess, H. Origins of Activity Enhancement in Enzyme
Cascades on Scaffolds. ACS Nano 2013, 7, 8658–8665.
(21) Trujillo, M.; Donald, R. G. K.; Roos, D. S.; Greene, P. J.; Santi,
D. V. Heterologous Expression and Characterization of the Bifunc-
tional Dihydrofolate Reductase–Thymidylate Synthase Enzyme
ofToxoplasma gondii†. Biochemistry 1996, 35, 6366–6374.
(22) Lindblad, C.; Rault, M.; Hagglund, C.; Small, W. C.;
Mosbach, K.; Buelow, L.; Evans, C.; Sreer, P. A. Preparation and
Kinetic Characterization of a Fusion Protein of Yeast Mitochondrial
Citrate Synthase and Malate Dehydrogenase. Biochemistry 1994, 33,
11692–11698.
(23) Elcock, A. H.; Huber, G. A.; McCammon, J. A. Electrostatic
Channeling of Substrates between Enzyme Active Sites: Comparison
of Simulation and Experiment†. Biochemistry 1997, 36, 16049–16058.
(24) Bulutoglu, B.; Garcia, K. E.; Wu, F.; Minteer, S. D.; Banta, S.
Direct Evidence for Metabolon Formation and Substrate Channeling
in Recombinant TCA Cycle Enzymes. ACS Chem. Biol. 2016, 11,
2847–2853.
(25) Liu, Y.; Hickey, D. P.; Guo, J.-Y.; Earl, E.; Abdellaoui, S.;
Milton, R. D.; Sigman, M. S.; Minteer, S. D.; Calabrese Barton, S.
Substrate Channeling in an Artificial Metabolon: A Molecular
Dynamics Blueprint for an Experimental Peptide Bridge. ACS Catal.
2017, 7, 2486–2493.
(26) Earl, E.; Calabrese Barton, S. Simulation of Intermediate Transport in Nanoscale Scaffolds for Multistep Catalytic Reactions. Phys. Chem. Chem. Phys. 2017, 19, 15463−15470.
(27) Huang, Y.-m. M.; Huber, G. A.; Wang, N.; Minteer, S. D.; McCammon, J. A. Brownian Dynamic Study of an Enzyme Metabolon in the TCA Cycle: Substrate Kinetics and Channeling. Protein Sci. 2018, 27, 463−471.
(28) Liu, Y.; Matanovic, I.; Hickey, D. P.; Minteer, S. D.; Atanassov, P.; Barton, S. C. Cascade Kinetics of an Artificial Metabolon by Molecular Dynamics and Kinetic Monte Carlo. ACS Catal. 2018, 8, 7719−7726.
(29) Pande, V. S.; Beauchamp, K.; Bowman, G. R. Everything You Wanted to Know about Markov State Models but Were Afraid to Ask. Methods 2010, 52, 99−105.
(30) Chodera, J. D.; Singhal, N.; Pande, V. S.; Dill, K. A.; Swope, W. C. Automatic Discovery of Metastable States for the Construction of Markov Models of Macromolecular Conformational Dynamics. J. Chem. Phys. 2007, 126, 155101.
(31) Noé, F.; Horenko, I.; Schütte, C.; Smith, J. C. Hierarchical Analysis of Conformational Dynamics in Biomolecules: Transition Networks of Metastable States. J. Chem. Phys. 2007, 126, 155102.
(32) 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.
(33) Prinz, J.-H.; Wu, H.; Sarich, M.; Keller, B.; Senne, M.; Held, M.; Chodera, J. D.; Schütte, C.; Noé, F. Markov Models of Molecular Kinetics: Generation and Validation. J. Chem. Phys. 2011, 134, 174105.
(34) Pérez-Hernández, G.; Paul, F.; Giorgino, T.; De Fabritiis, G.; Noé, F. Identification of Slow Molecular Order Parameters for Markov Model Construction. J. Chem. Phys. 2013, 139, 015102.
(35) Noé, F.; Schütte, C.; Vanden-Eijnden, E.; Reich, L.; Weikl, T. R. Constructing the Equilibrium Ensemble of Folding Pathways from Short Off-Equilibrium Simulations. Proc. Natl. Acad. Sci. U.S.A. 2009, 106, 19011−19016.
(36) Dickson, A.; Brooks, C. L. Quantifying Hub-like Behavior in Protein Folding Networks. J. Chem. Theory Comput. 2012, 8, 3044−3052.
(37) Lotz, S. D.; Dickson, A. Unbiased Molecular Dynamics of 11 Min Timescale Drug Unbinding Reveals Transition State Stabilizing Interactions. J. Am. Chem. Soc. 2018, 140, 618−628.
(38) Schwantes, C. R.; Pande, V. S. Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. J. Chem. Theory Comput. 2013, 9, 2000−2009.
(39) Pan, A. C.; Roux, B. Building Markov State Models along Pathways to Determine Free Energies and Rates of Transitions. J. Chem. Phys. 2008, 129, 064107.
(40) Bowman, G. R.; Beauchamp, K. A.; Boxer, G.; Pande, V. S. Progress and Challenges in the Automated Construction of Markov State Models for Full Protein Systems. J. Chem. Phys. 2009, 131, 124101.
(41) Suárez, E.; Adelman, J. L.; Zuckereman, D. M. Accurate Estimation of Protein Folding and Unfolding Times: Beyond Markov State Models. J. Chem. Theory Comput. 2016, 12, 3473−3481.
(42) Eric Vanden-Eijnden, W. E. Towards a Theory of Transition Paths. J. Stat. Phys. 2006, 123, 503−523.
(43) Metzner, P.; Schütte, C.; Vanden-Eijnden, E. Transition Path Theory for Markov Jump Processes. Multiscale Model. Simul. 2009, 7, 1192−1219.
(44) Prinz, J.-H.; Held, M.; Smith, J. C.; Noé, F. Efficient Computation, Sensitivity, and Error Analysis of Committor Probabilities for Complex Dynamical Processes. Multiscale Model. Simul. 2011, 9, 545−567.
(45) 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.
(46) Bowman, G. R. An Overview and Practical Guide to Building Markov State Models; Springer: Dordrecht, 2014; pp 7−22.
(47) Bastian, M.; Heymann, S.; Jacomy, M. Gephi: An Open Source Software for Exploring and Manipulating Networks. Third Int. AAAI Conf. Weblogs Soc. Media, 2009.
(48) Tsitkov, S.; Pesenti, T.; Palacci, H.; Blanchet, J.; Hess, H. Queueing Theory-Based Perspective of the Kinetics of “Channeled” Enzyme Cascade Reactions. ACS Catal. 2018, 8, 10721−10731.