MULTI-SCALE SIMULATION OF ELECTROSTATIC CHANNELING

By

Yuanchao Liu

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Materials Science and Engineering—Doctor of Philosophy

2018

ABSTRACT

MULTI-SCALE SIMULATION OF ELECTROSTATIC CHANNELING

By

Yuanchao Liu

One-pot multistep catalysis, also called tandem catalysis, denotes stepwise chemical reactions over sequential active sites in a single vessel. Such an approach enables efficient synthesis of various product molecules from simple precursors, and complete utilization of the energy stored in each chemical bond. Therefore, such cascade catalysis is of great significance to the manufacture of fine chemicals, the production of pharmaceutical intermediates, and electrocatalytic devices. A key factor limiting these processes is the mass transport of reaction intermediates, which in nature is found to be largely facilitated by substrate channeling, in which intermediate molecules are transferred directly to a subsequent active site instead of equilibrating to bulk environment. Electrostatically bound diffusion represents one such channeling mechanism, in which charged intermediates are transported along an oppositely charged pathway.

Naturally occurring electrostatic channeling has been studied for decades, but the challenge of controlling molecular-level interactions along with catalytic kinetics has hindered its application to artificial cascades. In this work, multi-scale simulations by a combination of molecular dynamics (MD) and kinetic Monte Carlo (KMC) are utilized to quantify the overall kinetics of artificial cascades, aiming to further explore the channeling mechanism and reveal potential limitations for future cascade design. Taking advantage of this hierarchical model, the molecular complexity may be fully considered and a wide range of time and length scales can be effective covered to bridge the gap between microstructures and kinetic events.

Specifically, a hopping surface diffusion mode is demonstrated by molecular dynamics simulation, wherein charged intermediate molecules are shuttled along a cationic oligopeptide bridge that is proposed to covalently conjugate hexokinase (HK) and glucose-6-phosphate dehydrogenase (G6PDH). Strong experimental evidence for the occurrence of electrostatic channeling is provided by ionic-strength dependent studies, via both simulations and experimental stop-flow lag time analysis. Specifically, simulations suggest that a balance between surface adsorption and diffusion is required for optimal channeling efficiency. To further quantify the energy associated with each elementary step in channeling process, advanced sampling methods are employed to study interactions on the cascade surface. Transition state theory is used to calculate the hopping energy barrier via a temperature-dependent study of hopping frequency. Desorption energy is calculated by umbrella sampling, further revealing ionic strength-independent Stern layer diffusion under the protection of an ionic strength-dependent diffuse layer. Intermediate hopping from the peptide bridge to its binding site on G6PDH is analyzed by transition pathway theory, which provides sufficient sampling of the interactive pathways on this 2D flexible surface. The leakage in this process is evaluated by probability analysis.

Finally, overall cascade kinetics is quantified by the KMC method, integrating the MD and experiment parameters. The KMC model enables direct comparison with stop-flow lag time analysis, by evaluating the product evolution over the entire experimental time scale, including the pre-steady state. The KMC results provide good agreement with experiment in terms of ionic strength dependence. Moreover, KMC reveals several key parameters limiting overall cascade kinetics, including the strength of surface interactions, the length of channeling pathway, and the energy barrier around the artificial interface of synthetic cascade.

This dissertation is dedicated to my family.

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my advisor, Dr. Scott Calabrese Barton, for his continuous support on my Ph.D. study and research, for his patience, motivation, and immense knowledge. He is a good mentor, actively guiding me towards independent thinking and positive communication; He is a good collaborator, proficiently working with me on a great deal of code work on our attempts to new modeling tools; He is also a good friend, always sharing various social/academic activities and much fun stuff with us.

Secondly, I gratefully acknowledge the funding support from the Army Research Office MURI (#W911NF1410263) via the University of Utah. My sincere thanks also goes to our project partners and their co-authorship on my peer-reviewed publications. Specifically, great thanks to Dr. Shelley D. Minteer, Dr. Matthew S. Sigman and Dr. David P. Hickey from University of Utah, for their support of experimental works. Thanks also go to Dr. Plamen Atanassov and Dr. Ivana Matanovic from University of New Mexico, for their help on our simulation work.

Besides my advisor, I would like to thank the rest of my committee members: Dr. Yue Qi, Dr. Wei Lai, Dr. Alex Dickson and Dr. Richard Lunt, for their insightful comments and encouragement, but also for the hard question which incented me to widen my research from various perspectives. Particular thanks to Dr. Alex Dickson for his great help on the advanced sampling techniques in my dissertation.

Last but not the least, I would like to thank my parents for raising me up, giving me an optimistic and positive attitude towards life, and shaping me as a strong and independent personality. I would also like to thank my girlfriend for her companionship and care during my doctoral career..

Computational work in support of this research was performed at Michigan State University’s High Performance Computing Facility. Thanks to Dr. Xiaoge Wang and Dr. Chun-min Cheng for their help on software compiling and job management; I thank my fellow labmates, Dr. Cenk Gumeci, Jacob Anibal, Alex Miraba, Kanchan Chavan and Erica Earl, for the stimulating discussions and for all the fun we have had together in the last four years. Lastly, I would to thank my friends at MSU, including Yuxi Ma and Mingmin Wang.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Introduction

1.1 Overview

One-pot multistep catalysis, also called tandem catalysis, involves a series of catalysts in a single vessel.[1] As compared to conventional stepwise reaction, tandem catalysis eliminates the isolation and purification processes that usually requires high process costs, result in yield losses, and generate waste. Additionally, these cooperative reactions can effectively protect unstable intermediates and reduce unnecessary side reactions. As a result, multistep catalysis enables high chemical conversion efficiency, making complex product molecules from simple and accessible precursors. Therefore, tandem catalysis is of great significance to the manufacture of fine chemicals and the production of pharmaceutical intermediates.[2] Taking advantage of complete energy extraction from various chemical bonds in complex molecule structures, multistep reactions also have great potential to be applied to electrocatalytic devices, such as biofuel cells.[3] Multi-enzyme biofuel cells demonstrate a higher energy (1.4~3 fold) delivery due to the deep oxidation by tandem catalysis.[4,5]

Figure 1-1 shows an example cascade of various catalysts anchored on a DNA scaffold, including metallic, organic, and bio/enzymatic catalysts. Besides finding proper catalysts and the conditions that fully functionalize each reaction site[1,6], the mass transport of reaction intermediates is a key factor impacting the overall efficiency of this tandem process.[7] Due to Brownian effects, the produced intermediate molecules tend to move in a random and non-directional mode. In spite of a clustering of sequential catalysts, intermediate molecules still have a high probability to equilibrate to bulk media, where a side reaction and unfavorable binding can occur. Therefore, the flux control of reaction intermediates is a key challenge to the development of advanced catalytic cascades.

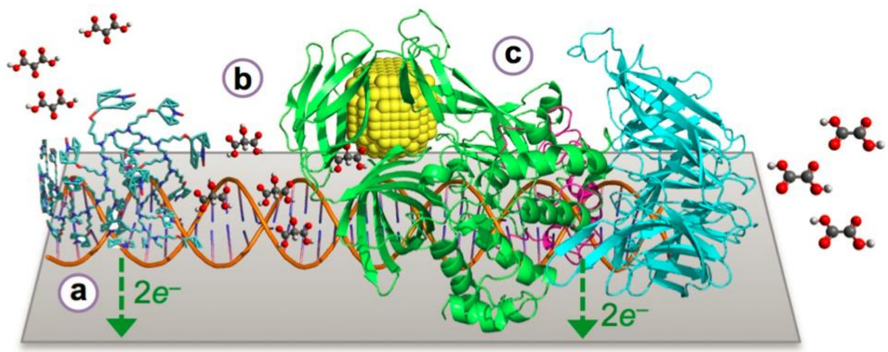


Figure 1-1 Conceptual cluster of various catalysts anchored on a DNA scaffold. (a) Tartronic acid oxidation to mesoxalic acid at TEMPO catalyst; (b) facilitated transport of mesoxalic acid along DNA; (c) oxidation of mesoxalic acid to oxalic acid via catalysis by PtRu nanoparticle – AldDH adduct.

1.2 Substrate Channeling in Cascade Catalysis

Over billions of years of evolution, nature has developed very efficient catalytic pathways to perform chemical reactions in the cell stepwise, utilizing transient super-molecular complexes termed metabolons.[8,9] By these one-pot multistep catalyses, a wide range of complex biomolecules are synthesized from simple precursors, and energy is generated from carbohydrate oxidation through a metabolic pathway. Although the cell has a complex chemical environment, these well-defined cascades are able to maintain high catalytic efficiency and prevent the undesired bonding of reaction intermediates to unproductive active sites. The key was found to be substrate channeling, which is defined as the direct transfer of reaction intermediates from one catalytic active site to the next, without equilibrating into bulk media.[7,10,11] Therefore, substrate channeling may have promising applications in chemical manufacturing and electrocatalytic devices. For example, with cascade catalysis, energy devices based on organic fuels may accommodate more fuel categories and more fully utilize the energy stored in high molecular weight fuels.[3] These features enable biofuel cells with high energy density and biosensors with enhanced sensitivity.[4,5]

Under most conditions, intermediate diffusion is fast compared to reaction rate, making proximity alone insufficient to realize effective channeling. To illustrate this, the product concentration around a single active site is calculated by equation 1-1,[12] where *c* is the concentration at radial distance, *r*, from a point active site, *t* is time, *D* is diffusion coefficient and τ is the average time between reaction events (k^{-1}, where *k* is turnover frequency).

$$c(r,t) = \sum_{i=0}^{i=t/\tau-1} \frac{1}{\left(4\pi D(t-i\tau)\right)^{3/2}} exp\left(-\frac{r^2}{4D(t-i\tau)}\right) \qquad \text{1-1}$$

Figure 1-2 shows resulted concentration profile at steady state ($t = 10^4\ \tau$) with different *k/D* values. In a typical system with the diffusion of small molecule in aqueous solution, the *k/D* value is 0.01 μm^2. As shown in Figure 1-2, the spatial distribution of product concentration is quite uniform in most conditions, indicating that proximity alone has very little impact on the local concentration of reaction intermediate in cascade reactions. Brownian dynamics simulations also indicate that the concentration profile for most small intermediate molecule is raised only to a distance around 1 nm from a neutral catalytic active site.[13] Recent experimental work on covalently conjugated enzymes also indicated that proximity alone did not have impact on the activity improvement, as shown in Figure 1-2b.[14] These results suggest that channeling by proximity alone only works when sequential active sites are extremely close to each other. However, intermediate diffusion can be enhanced and well-controlled through the incorporation of a functional surface or tunnel. For example, the channeling distance was proposed to be increased 10-fold in the presence of electrostatic guidance.[15]

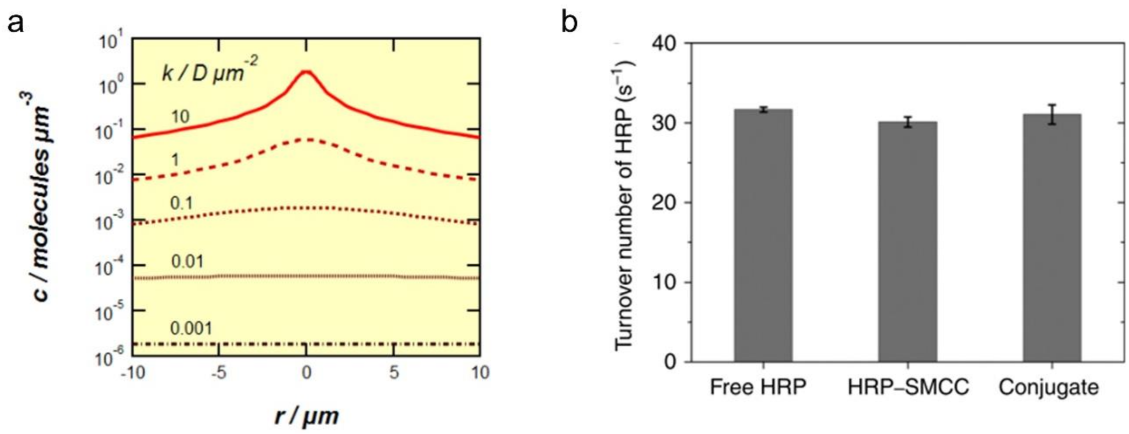


Figure 1-2 (a) Three dimensional diffusion field near a point source at steady state ($t = 10^4\ \tau$), where k is reaction rate and D is diffusivity. (b) Comparison of experimental turnover frequency of free enzyme couple and covalently conjugated enzyme couple.[12,14]

Figure 1-3 shows three natural channeling mechanisms, each featuring a functional surface or tunnel with reasonable spatial organization. A well-studied example of an intra-molecular tunnel is tryptophan synthase, Figure 1-3a, which produces tryptophan through a two-step reaction using indole as the intermediate.[16] The crystal structure reveals a hydrophobic tunnel (2.5-3 nm) connecting the two active sites, which physically confines the indole intermediate and transports it to the second active site.[17] As compared to intra-molecular tunneling, the pyruvate dehydrogenase complex (Figure 1-3b) creates a pathway by covalent bonding, where the acetyl group (intermediate) is transferred by a lipoamide swing arm. Electrostatic guidance refers to interactions between a charged intermediate and an oppositely charged pathway on cascade surface, and provides possibly the most straightforward channeling method for application in synthetic cascades. As shown in Figure 1-3, the crystal structure of MD-CS (malate dehydrogenase–citrate synthase) shows a positive surface (purple area) bridging the two active sites, providing a diffusion path for negatively charged oxaloacetate (intermediate).

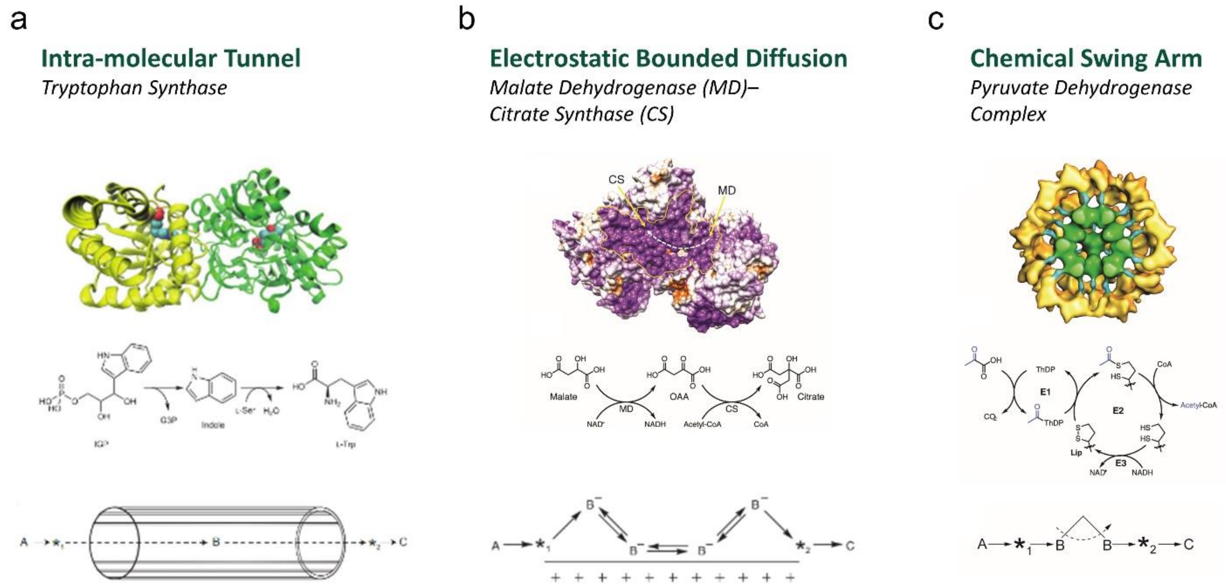


Figure 1-3 Mechanisms of substrate channeling in nature.[7,16,18,19] (a) The crystal structure of traptophan synthase (PDB:1A5S) with α-subunit shown in yellow and β-subunit shown in green. (b) The crystal structure of malate dehydrogenase–citrate synthase with possible channeling pathway marked in dashed yellow line. (c) The crystal structure of eukaryotic pyruvate dehydrogenase complex with E1 enzymes in yellow, E2 core structures in green and linker in between marked in blue.

In past decades, attempts were made to apply substrate channeling in artificial systems.[20–22] For example, multiple enzyme species were anchored on DNA scaffold to increase the effect of cluster and spatial orientation.[20] A good example of artificial chemical swing arm is a glucose-6-phoshate dehydrogenase (G6pDH) and malate dehydrogenase (MDH) cascade spatially organized on a DNA scaffold where a NAD+ functionalized single-stranded DNA swing arm is used to channel NADH to the next active site.[21] Examples of synthetic electrostatic channeling and intra-molecular tunnel are quite limited. Local effectiveness of reaction intermediates was studied by introducing favorable/unfavorable electrostatic interactions between intermediate and cascade scaffold.[23,24] A decreased catalytic efficiency was found with either too low or too high binding constants.

So far, introducing substrate channeling artificially is complicated by both the precise molecular-level control and the lack of experiment methods to detect the occurrence of such transient phenomenon (~ ns).[7] Therefore, new methodology (modeling/experiment) is greatly needed to aid the design of channeling pathways and the quantification of overall catalytic kinetics. Natural channeling mechanisms utilize intramolecular tunnels[16,17,25], chemical swing arms[19], spatial organization[12,26,27] and electrostatic guidance[15,18,28–30] to facilitate bound diffusion of reaction intermediates. These mechanisms highlight a significant distinction between substrate channeling, which features bound or restricted diffusion, and active site proximity.

1.3 Experimental Measurement

Modern biochemical analyses of substrate channeling are indirect measurements, tracking the bulk intermediate evolution instead of the actual channeling molecules. Examples are transient time (τ) analyses, isotope dilution and enrichment studies, cascade resistance to a competing side reaction, and cascade resistance to a reaction inhibitor.[7] Transient time analysis, as an example, is defined as the time required to reach steady-state flux of reaction intermediate. If a reactant “A” is converted to intermediate “B” and finally converted to “C”, the product time course can be observed graphically when only “A” is input to the system, as shown in Figure 1-4a. Transition time, also called lag time, is obtained by extrapolation of the steady state line back to the time axis. There are several explanations of the physical meaning of lag time; one of them is the average time for intermediate to reach the sequential active site. From the statistical point of view, on the other hand, lag time is also the time required for bulk intermediate concentration to reach a constant value, such that the reaction rate of the second site equals that of the first site, and a steady state is achieved.

For perfect channeling, τ is zero. With insufficient leaky channeling, τ tends to shift to the value for free standing enzyme couple, (section 1.5.1) depending on the extent of leakage. However, it is important to know that all these methods are indirect measurements. As a consequence, current experimental measurements work well with perfect channeling, but are less clear for leaky channeling and free diffusion. Also, owing to the phenomena observed, it is difficult to evaluate the contribution of channeling compound to bulk diffusion. Therefore, substrate channeling is a complex phenomenon that is difficult to model or analyze by single technique. A combination of hierarchical modeling and experimental measurement would be preferred to understand channeling mechanisms, and further design and quantify cascade reactions. From the perspective of simulation, the model should be efficient but have enough complexity to capture molecular-level effects and evaluate the contribution of individual event. This will be discussed in detail in Chapter 3.

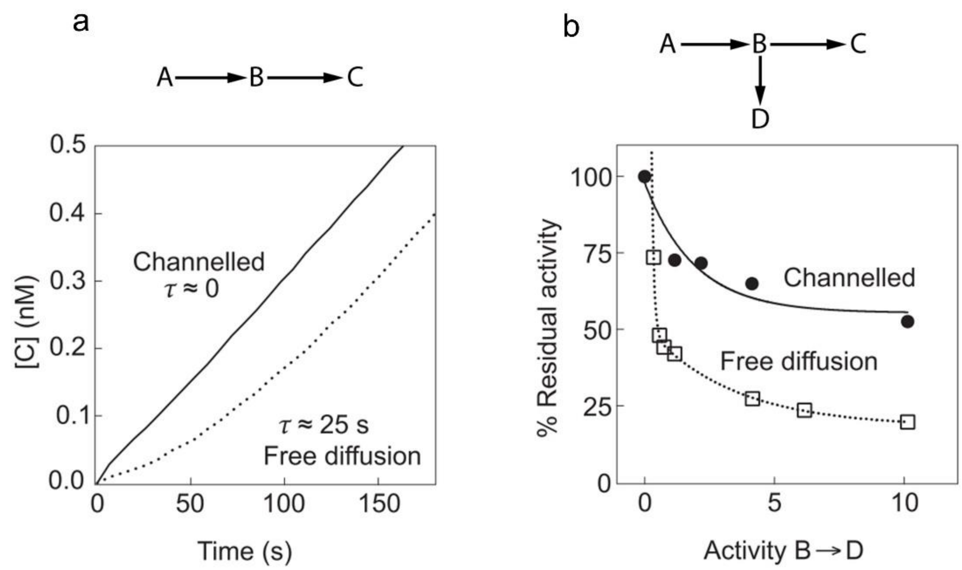


Figure 1-4 Reaction scheme and example data for transient time (τ) analysis.[7]

Another significant technique is the use of competing reactions, wherein bulk intermediates are consumed by a side reaction in bulk. The observed yield loss is thus related to the extent of substrate channeling, as shown in Figure 1-4b.[31] When substrate channeling is high, the effect of competing reaction is minimal. Without channeling, the yield can be significantly reduced. Details of these analysis techniques will be described in the following simulation section.

1.4 Intermediate Transport in Electrostatic Channeling

Electrostatic channeling is bounded diffusion of charged intermediates along an oppositely charged surface, a non-specific pathway applicable to any charged intermediate. It is a more general mechanism, as compared to structure-specific intramolecular tunnels and chemical swing arms. In past decades, electrostatic channeling studies mainly focused on the bi-functional enzyme TS-DHFR (thymidylate synthase-dihydrofolate reductase) and the TCA cycle supercomplex MDH-CS (malate dehydrogenase-citrate synthase), as shown in Figure 1-3b.

Specifically, experimental work showed strong evidence of improved kinetics by lag time analysis.[31,32] The occurrence of substrate channeling was also supported by apparent resistance to competing bulk reactions. By structural characterization (Figure 1-3), both the TS-DHFR and MDH-CS complexes were found to have positively charged surfaces located between the two active sites.[18,28] Moreover, the channeling efficiency could be significantly disabled by either a neutralization of the channeling pathway[22,29] or an increased ionic strength (IS)[29,33]. This evidence gives strong qualitative support for experimental results.

Recently, we have reported the first cascade with artificially introduced electrostatic channeling (details in Chapter 2 and Chapter 3), composed of hexokinase (HK) and glucose-6-phosphate dehydrogenase (G6PDH) covalently conjugated by a cationic oligopeptide bridge.[33] In this work, a "hopping" surface diffusion mode was built by molecular dynamics (MD) simulation, and studies of ionic strength dependence (MD and Exp.) on lag time provided strong evidence of the occurrence of electrostatic channeling.

1.5 Kinetic Quantification

Kinetic quantification is a key to bridge the gap between microstructure information (Figure 1-3) and kinetic macroscale phenomena (Figure 1-4). It not only helps to understand the channeling mechanism, but also accesses time and length scales (~ns and ~nm) inaccessible to current experiment techniques. More importantly, it is able to identify potential rate limitations in elementary steps, thus providing guidelines for future cascade design.

1.5.1 Analytical methods

Generally, analytical quantification is based on an assumption of steady-state conditions, where the intermediate flux is time independent, requiring an equivalent production and consumption of reaction intermediate. If $V_{\mathrm{max},1}$ denotes the reaction rate of fully saturated enzyme 1 (E1) and V_2 denotes the actual rate of enzyme 2 (E2) at steady state, their relationship can be expressed by equation 1-2. Usually, $V_{\mathrm{max},2} \geq V_{\mathrm{max},1}$; otherwise the system will never reach steady state.

$$V_{\mathrm{max},1} = V_2 \tag{1-2}$$

Michaelis-Menten kinetics is applied to a reversible binding/unbinding of substrate, S, to enzyme, E, followed by a non-reversible reaction to product, P. The whole process is shown by equation 1-3, where k_{des}, k_{ads} and k_{cat} are the rate constants for substrate desorption, adsorption and reaction, respectively.

$$\mathrm{E} + \mathrm{S} \xleftrightarrow{k_{\mathrm{des}},\, k_{\mathrm{ads}}} \mathrm{E} \cdot \mathrm{S} \xrightarrow{k_{\mathrm{cat}}} \mathrm{E} + \mathrm{P} \tag{1-3}$$

In a free standing system, as shown by equation 1-4 and 1-5, product evolution rate on second enzyme, V_2, can be correlated to the fully saturated E2 rate through intermediate concentration, [I], and Michaelis constant, K_{M}.

$$V_2 = \frac{dP_2}{dt} = \frac{V_{\mathrm{max},2} \times [\mathrm{I}]}{K_{\mathrm{M},2} + [\mathrm{I}]} \tag{1-4}$$

$$K_{\mathrm{M}} = \frac{k_{\mathrm{cat}} + k_{\mathrm{des}}}{k_{\mathrm{ads}}} \quad 1\text{-}5$$

By combining equation 1-2 and 1-4, steady state intermediate concentration, $[I]_{\mathrm{ss}}$, can be correlated to three experimental accessible parameters as shown by equation:

$$V_{\mathrm{max,1}} = \frac{V_{\mathrm{max,2}} \times [I]_{\mathrm{ss}}}{K_{\mathrm{M,2}} + [I]_{\mathrm{ss}}} \quad 1\text{-}6$$

By definition, lag time for uncoupled enzymes (Figure 1-4a), τ_{free}, is the time required for bulk intermediate concentration to reach a value making the reaction on E2 equal to that of E1, as shown by equation:

$$[I]_{\mathrm{ss}} = V_{\mathrm{max,1}} \times \tau_{\mathrm{free}} \quad 1\text{-}7$$

To introduce the turnover frequency (TOF), $k_{\mathrm{cat,1}}$, as the rate constant for reaction, equation 1-7 can also be expressed by

$$[I]_{\mathrm{ss}} = [E] \times k_{\mathrm{cat,1}} \times \tau_{\mathrm{free}} \quad 1\text{-}8$$

Through equation 1-6, 1-7 and 1-8, the lag time can be derived as follows:

$$\tau_{\mathrm{free}} = \frac{K_{\mathrm{M,2}}}{V_{\mathrm{max,2}} - V_{\mathrm{max,1}}} \quad 1\text{-}9$$

$$\tau_{\mathrm{free}} = \frac{K_{\mathrm{M,2}}}{[\mathrm{E}] \cdot \left(k_{\mathrm{cat,2}} - k_{\mathrm{cat,1}}\right)} \quad 1\text{-}10$$

Therefore, the lag time for free standing system mainly depends on the concentration of enzyme couple, the TOF difference between two enzyme species and the Michaelis constant of enzyme-2. This can be used as an indication of the maximum lag time for the study of a channeling system.

When introducing substrate channeling, the lag time, τ_{ch}, can be expressed as follows:[34]

$$\tau_{\mathrm{ch}} = \frac{p_{\mathrm{c}}}{k_{\mathrm{cat},2} + k_{\mathrm{des},2}} + \frac{K_{\mathrm{M},2}(1 - p_{\mathrm{c}}p_{\mathrm{r}})}{V_{\mathrm{max},2} - V_{\mathrm{max},1}(1 - p_{\mathrm{c}}p_{\mathrm{r}})} \qquad \text{1-11}$$

$$p_{\mathrm{r}} = \frac{k_{\mathrm{cat},2}}{k_{\mathrm{cat},2} + k_{\mathrm{des},2}} \qquad \text{1-12}$$

where p_{c} is the probability for intermediate molecules to reach second enzyme through channeling, and p_r is the reaction probability for readily adsorbed intermediates as compared to unbinding from E2. Since the first item in equation 1-11 is usually smaller than 1 and $V_{\mathrm{max},1}$ is usually much smaller than $V_{\mathrm{max},2}$, equation 1-11 can be simplified as follows:

$$\tau_{\mathrm{ch}} = \frac{K_{\mathrm{M},2}(1 - p_{\mathrm{c}}p_{\mathrm{r}})}{V_{\mathrm{max},2}} \qquad \text{1-13}$$

As compared to free standing lag time, therefore, the lag time reduction of channeling system is only determined by the mass transport-related channeling efficiency, p_{c}, and ligand binding/unbinding kinetics-related leakage, p_{r}.

Although analytic methods can give a basic idea on the components of channeling parameters, it is not able to extimate the channeling efficiency, p_c, the key parameter to overall kinetics. We are therefore motivated to use computational techniques to obtain such estimation.

1.5.2 Continuum modeling

Electrostatic channeling has been studied via continuum modeling, where the distribution of substrates/intermediates are represented by a continuous media.[15,35] With pre-determined spatial factors (*e.g.*, boundary conditions) and system parameters (*e.g.*, diffusivity and reaction rate), the intermediates migrate under the impact of concentration gradient (*e.g.*, Fick's law) and external electric force field, as shown in Figure 1-5. Based on this, a steady state flux is used to evaluate the channeling effect by long-range electrostatic interaction. McCammon's group has reported the electrostatic confinement model with reflecting boundary conditions. The results of diffusion limited system show that the attractive electrostatic force between intermediate and both enzyme is able to give the maximal channeling efficiency, which is similar to a compartment model.[15] Recently, our group reported a more specific two-step cascade model with more detail and complexity, as shown in Figure 1-5. By setting zero boundary conditions, the model is mimicking a competing reaction system (Figure 1-4b), where the yield percentage is employed to evaluated the channeling efficiency. In addition, by fully incorporating the range of diffusion and reaction rate constants, the contribution of electrostatic channeling was also identified at high Damkohler number, which also means a diffusion limited region.[35]

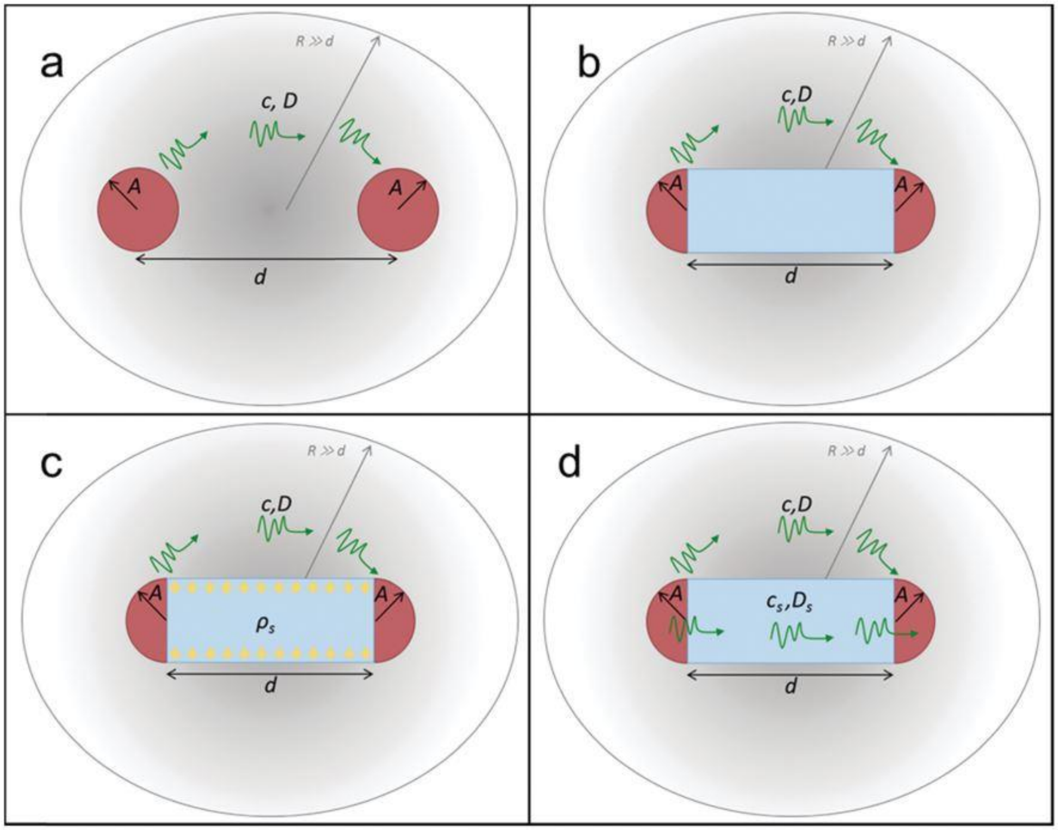


Figure 1-5 Schematic of two sequential enzyme-catalyzed reactions (diffusion limited) in continuum model.[35] (a) Base model without bridge between two spherical active sites. (b) Model with volumetric bridge. (c) Electrostatic model with positive charged bridge. (d) Adsorption model with surface diffusion on bridge.

However, continuum modeling relies heavily on parameters estimated from experiment results and other simulations. Moreover, the impact of molecular-level interactions cannot be assessed at these length scales.

1.5.3 Molecular simulation

Modern molecular simulation techniques offer powerful tools to test scientific hypotheses and predict macroscopic properties. In addition, simulation can build understanding of mechanisms behind experimental phenomena and study the time and length scale inaccessible to experiment. In recent years, Quantum Mechanism (QM), Molecular Dynamics (MD) and Kinetic

Monte Carlo (KMC) simulations are increasing significant in the area of heterogeneous catalysis[25,36–41], as shown in Figure 1-6.[42] Based on the Schrödinger equation, QM deals with electron structure, which is able to calculate the activation energy and frequency factor for elementary chemical reactions. However, because of its intrinsic simulation load, QM cannot represent systems with large length and time scale. Using classical mechanics, MD represents explicit interactions between atoms with mutual potential energy (bonded and non-bonded). Due to larger time and length scale, MD is capable of representing the ensemble properties, the interaction between large biomolecules and the diffusion coefficient of small molecules. Therefore, MD simulation is a suitable tool to study the interaction between intermediate molecules and bio-molecular catalysts (*e.g.*, enzyme) or scaffold (*e.g.*, DNA and poly peptide). As a result, binding and transport behavior of reaction intermediates can be obtained for each catalytic system. As compared to conventional MD, reactive force field (ReacFF) [43] is originally developed for hydrocarbon oxidation in MD and still under development for other systems.

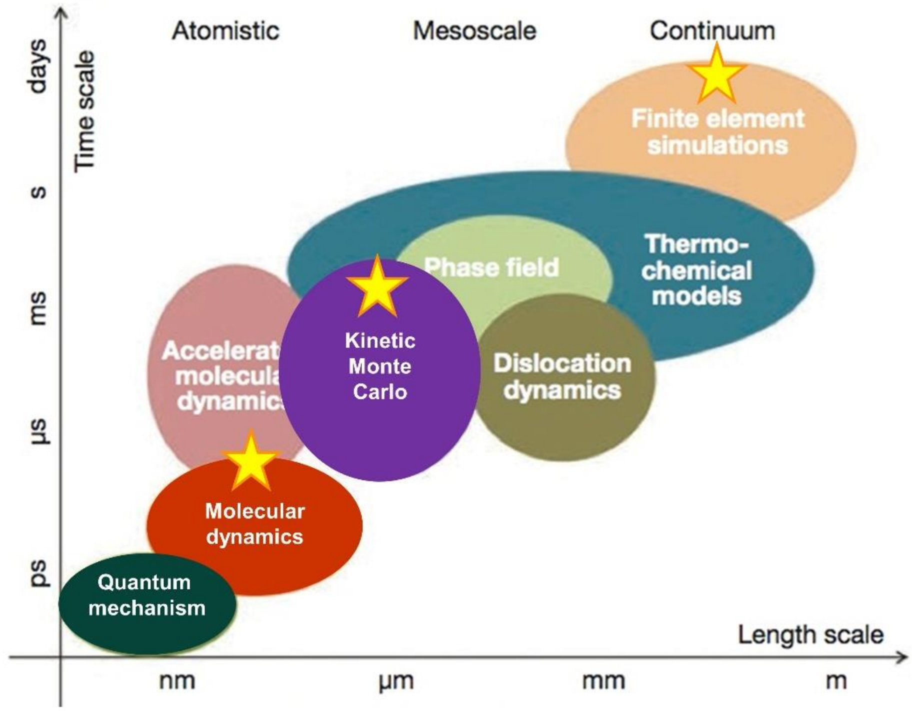


Figure 1-6 Time and length scale of different simulations techniques.[42]

Over past two decades, molecular simulation has aided the study of electrostatic channeling mechanisms in enzymatic super-complexes. This has been demonstrated in the case of naturally occurring enzyme complexes (such as the Krebs cycle and the electron transport chain). First molecular simulation of electrostatic channeling were reported by McCammon's group, using Brownian simulations to estimate transport efficiency in terms of probability that the explicit intermediate reaches the vicinity (~0.7 nm) of the second active site.[13,29,30,44] These modeling results were further integrated by an analytical approach representing the second active site with Michaelis kinetics[34]. By combining with experimental results, it is demonstrated that restricted diffusion of intermediates caused by electrostatic interactions improved transfer efficiencies up to ~80% (compared to < 10% for that caused by non-electrostatic interaction). Low channeling leakage and high binding affinity at the second active site were highlighted as key factors impacting overall kinetics.

These simulations represent a refinement of the continuum model, but were still coarse-grained, in that the intermediate was treated as charged sphere migrating in an electric field generated by cascade surface. In addition to long-range electrostatic interaction, charged and polarized intermediates can interact strongly with channeling surface through hydrogen bond interaction. However, consideration of such local interactions in electrostatic models has been largely absent. Therefore, MD simulation is a good candidate to study such interactions due to its detailed representation of various energy terms in channeling system.

In this context, molecular simulation has emerged as a technique that enables the study of electrostatic interactions that may be inaccessible to experimental detection limits or impractical for computationally expensive techniques, such as density functional theory. Additionally, these simulations provide a platform for exploring possible substrate/intermediate combinations by describing their interaction with a computationally-optimized electrostatic surface.

1.5.4 Kinetic Monte Carlo model

Despite several advantage, molecular dynamics has intrinsic limitations which can be complemented by kinetic Monte Carlo. As shown by Table 1, regular MD is not able to represent reactions, and so it is not by itself able to characterize the overall reaction kinetics.

Table 1 Kinetic Monte Carlo (KMC) vs. Molecular Dynamics (MD)

Feature	MD	KMC
Reaction	No	Yes
Explicit representation	atom	event
Time representation	finite, constant and small time steps	discrete time step depending on individual random events
Governing factor	Potential energy *e.g.*, L-J, Coulomb	Transition State Theory *e.g.*, $\Delta G_{act}, f$
Rare events	advanced sampling techniques are required	yes

Moreover, because MD calculates the trajectory of individual molecules, governed by local potential energy, it cannot well represent rare events, as shown in Figure 1-7.[36] That is, the MD system tends to oscillate around an energy minimum and meanwhile get the probability to hop to the neighboring energy minimum exponentially depending on the energy barrier in between. As a result, large amount simulation resource is wasted when MD system is stuck at such state, making MD simulation unable to well represent rare events. Enhanced sampling techniques[45] are usually required to accelerate MD systems to overcome the energy barriers separating the meta-states, in order to acquire longer time scale and accurate sampling probability. Current enhanced sampling approaches includes Replica-exchange molecular dynamics[46,47], Metadynamics[48,49], Umbrella Sampling[50], and Transition Path Analysis by Markov State Model[51–57].

Nonetheless, MD simulations are still expensive and not capable to cover the time and length scale of real word experimental system. In comparison, KMC simulation uses given rate constants to randomly determine the occurrence of specific events and corresponding time evolutions, instead of performing the actual oscillating process in MD simulations. The probability of executed events depends on the weight of rate constants that are determined by the energy barrier of such events. More detail of KMC method can be found in Chapter 5.

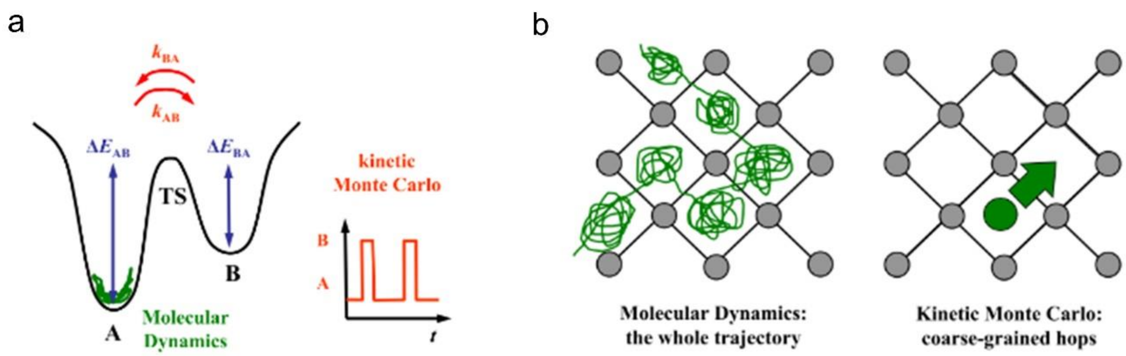


Figure 1-7 Rare event in Molecular Dynamics and Kinetic Monte Carlo.[36]

As a simplified model represented by explicit events and rate constants, KMC is able to quantify overall kinetics, achieving the time scale of experimental systems. Although KMC is a coarse-grained method, it allows tunable complexity to mimic experimental conditions, such as interactive events and changing external environment. For example, when oxygen repulsion is considered in CO oxidation on a RuO_2 surface, the KMC result is much superior to the classical microkinetics theory which assumes a homogeneous surface distribution and thus underestimates the complexity of local conditions.[41] In spite of KMC's advantages, it is very important to know that KMC *per se* cannot predict energy barriers or frequency factors needed to calculate the rate constants via transition state theory (TST). These key parameters have to be measured by experiment or calculated by QM and/or MD. Therefore, hierarchical simulation is a proper approach to model a channeling system. By the combination of MD and KMC, a bridge can be built between the molecular level design and the macroscopic experimental phenomena.[58]

From the aspect of applications, KMC models are widely used for catalytic oxidation and diffusion of small molecules on lattice surfaces.[41,59–63] For example, DFT parameterized *ab initio* KMC was employed to study the impact of catalyst composition, surface structure, lateral interactions, and environment on the overall catalytic performance of CO and NO oxidation.[62] Moreover, experimentally based KMC was used to study the CO oxidation on a RuO_2 surface,[60] and a comparison between DFT and experimental parameters was made to identify the significant impact of lateral repulsion of adsorbent.[41] Besides quantum mechanics, a KMC model combined with MD simulation was used to study the epitaxial growth of fcc and hcp islands on a fcc (111) surface, exhibiting great simulation acceleration. As a result, boundary pinning effects by the island of adsorbed atom was revealed, which is usually difficult to be elucidated by conventional KMC methods.[64] Another good example of model hybridization is a combination of first-principle MD and KMC simulations, where the impact of rotational coupling of the side groups was shown to influence the proton conduction in proton exchange membrane fuel cells (PEMFC).[65]

1.6 Overview of Dissertation

The goal of this work is using multi-scale simulations to further understand the mechanism of electrostatic channeling, quantitatively calculate the cascade kinetics and develop design rules for future artificial cascades. The model construction particularly focuses on the surface diffusion of reaction intermediates, and the quantification approaches are aiming to fully consider the molecular complexity of channeling pathways. Enhanced sampling techniques are utilized to cover the ergodicity of the channeling system, and give a rational representation of all possible events with different frequency and probability. Specifically, molecular dynamics (MD) simulations enable the calculation of energy-determined surface equilibrium constants and surface diffusivity, and a kinetic Monte Carlo (KMC) model integrated all rate constants from MD (*e.g.*, surface diffusion and desorption rate) and experiments (*e.g.*, turnover frequency), to estimate the product evolution on experimental time scales.

The cascade and reaction scheme employed in this work is the conversion of glucose to phospho-6-gluconolactone by hexokinase (HK) and glucose-6-phosphate dehydrogenase (G6PDH), covalently conjugated by a cationic oligopeptide bridge. The system is simulated and validated by comparison to stopped-flow lag time analysis. This section is in collaboration with Dr. Shelley D. Minteer's lab at University of Utah.

In Chapter 2, MD simulation is used to study the surface interaction between cationic oligopeptide (Lysine, Arginine and Histine) and various negatively charged intermediate species (oxalate, glyoxylic acid and glucose 6-phosephate). A "hopping" surface diffusion model is built, and the MD and experiment studies of ionic strength dependence on lag time (HK-G6PDH cascade) provides strong evidence of the occurrence of electrostatic channeling. In addition, the balance between surface adsorption and diffusion is studied to indicate an optimization of channeling efficiency. (1st paper published on *ACS Catalysis*[33])

In Chapter 3, regular MD simulation and umbrella sampling are used to quantify the energy barrier for surface hopping and desorption. A Stern layer diffusion under the protection of diffuse layer is revealed by ionic strength dependence study. As a result, the thermodynamic parameters of an individual hop is quantified in great detail. This helps to further understand the channeling mechanism and also give support to the rate constants for kinetic models.

In Chapter 4, the leakage between peptide bridge and enzyme binding pocket is considered. On this complicated 2D surface with multiple possible pathways, probability analysis and a Markov State model are used to map the energy profile in this area. Then, this energy landscape is parameterized into KMC rate constants to involve this additional leakage before final reaction.

In Chapter 5, the overall kinetics of HK-G6PDH cascade was quantified by a combination of MD and KMC simulation. Specifically, KMC simulation enable the integration of all rate constants to model the evolution of bulk intermediate, particularly from the pre-steady state. This allows for a direct comparison to experiment results. From the resulting lag time, the leaking probability of each individual hop and the length of channeling bridge are found to be the key factors for overall kinetics. The barrier between bridge and second enzyme not only caused leakage by itself but also exacerbated the leakage on channeling bridge. The KMC lag time agrees much better to experiment as compared to the results only with leakage on peptide bridge. (2nd paper submitted to *ACS Catalysis*, in revision)

In summary, the multi-scale simulation in this work enable a detailed quantification approach, fully considering the molecular-level interactions in electrostatic channeling process via strong surface diffusion mechanism. This is of great significance to understand the micro-scale mechanism and identify potential limitations to be improved for cascade design. In future, highly interactive work between such modeling technique and experiment is greatly needed for the molecular-level design of synthetic cascade with channeling efficiency comparable to or even better than natural cascade.

Chapter 2 Surface Diffusion by Molecular Dynamics Simulation*

2.1 Introduction

The phenomenon of electrostatic channeling is demonstrated naturally by coupled enzymes, such as dihydrofolate reductase-thymidylate synthase (TS–DHFR) and malate dehydrogenase- citrate synthase (MD-CS) cascades. As for structure characterization, XRD diffraction results reveal a charged surface between active sites for both cascades. Theoretically, charged intermediate molecules interact with cascade surfaces through long-range electrostatic forces and local, short-range interactions via Van der Waals forces and hydrogen bonds. Therefore, besides the migration under the electric field generated by charged cascade surface, intermediate molecules also potentially perform "surface diffusion" at closed proximity to a channeling pathway. However, the channeling process occurs at nanometer and nanosecond scale, which is inaccessible to current experimental techniques all of which rely on capturing bulk intermediate evolution at real world time and length scales.[7] Given the fact that *in-situ* observation is currently impossible, molecular simulation combined with experimental kinetics is an optimal approach to understanding the channeling mechanism on cascade surface, particularly through a surface diffusion mode.

Using the mechanism of electrostatic guidance as an inspiration, we envisioned the use of a charged oligopeptide bridge as a molecular construct to facilitate restricted diffusion of a

---

* The content of this chapter has been published on *ACS Catalysis*, in collaboration with University of Utah. Experimental work in this chapter is in collaboration with David P. Hickey, Dr. Shelley D. Minteer and Dr. Matthew S. Sigman at University of Utah.

Y. Liu, D. P. Hickey, J.-Y. Guo, E. Earl, S. Abdellaoui, R. D. Milton, M. S. Sigman, S. D. Minteer and S. Calabrese Barton, "Substrate Channeling in an Artificial Metabolon: A Molecular Dynamics Blueprint for an Experimental Peptide Bridge", *ACS Catalysis*, 7, 2486–2493 (2017). doi:10.1021/acscatal.6b03440.

charged intermediate between two sequential catalytic sites. In this chapter, we describe the use of MD simulations to design and optimize a series of theoretical cationic α-helix peptides, and quantify their ability to transport charged intermediates across the theoretical surface. These optimized theoretical conditions were applied to design and construct a simple template for preparing artificial enzyme complexes capable of inducing restricted diffusion of a charged reaction intermediate between two sequential enzymes in a metabolic pathway. Stop-flow lag time analysis was used to experimentally examine the channeling efficiency and validate the MD results. These studies provide the basis for utilizing MD simulations to strategically design synthetic substrate channeling cascades and develop a detailed understanding of the range and limitations of artificial electrostatic substrate channeling.

2.2 Simulation Methods

GROMACS 5.1.1[66–72] was used as the MD simulation package. CHARMM36[73,74] was employed as the force field governing whole simulation and CGenFF (CHARMM General Force Field) [75] was used to generate topology for small intermediate molecules. Periodic boundary conditions (PBC) were applied to all MD simulations. The peptide/intermediate couple was first solvated with tip3p water molecules in a dodecahedral box (~ 100 nm^3) and then neutralized by Na^+ or Cl^- ions. The system energy was then minimized using a steepest descent algorithm, followed by a 0.1 ns NVT and 1.0 ns NPT equilibration process with position restraints for each molecule. Finally, the MD simulation was performed under NPT ensembles for 50 ns, repeating 10 times for each system. For all MD simulations, the system temperature was coupled at 300 K by velocity-rescale thermostat [76] and the pressure was stabilized at 1 bar by Parrinello-Rahman barostat.[77] Particle-Mesh Ewald (PME) algorithm with cubic interpolation (0.16 ns grid space) was used to calculated long-range electrostatic interaction.[78,79] Cutoff values for both short-range electrostatic and van der Waals interactions were set to 1 nm. A Verlet cutoff scheme[80] is used to calculated non-bonded interactions on a GPU accelerator.[81]

Initial α-helix peptide structures was generated by Avogadro 1.1.1[82] with torsion angles Φ = -60 ° and Ψ = -45 °. Neutral terminals (-COOH and –NH_2 instead of –COO^- and –NH_3^+) were applied to the helical backbone for MD simulations. From resulting backbone, the alpha-carbon of the N-terminus and alpha-carbon of the C-terminus (two points in total) were pinned to immobilize the peptide. The fraction of basic residues was tuned by adding alanine (Ala) between basic residues (Arg, Lys). Accordingly, the peptide chain used in simulations were named based on the number of basic residues (Lys, His, Arg) and the number of interstitial Ala. For example, the nomenclature K4-A3 indicates a peptide with four Lys residues in total with three Ala between each Lys, resulting in the structure, "…KAAAKAAAK…". Finally, four extra Ala were appended to the terminal Lys at each end, in order to provide pinning sites and to maintain the α-helical structure.

Electrostatic potential maps were prepared using APBS and PDB2PQR,[83–85] and VMD 1.9.2 (Visual Molecular Dynamics)[86] was used for visualization. Intermediate molecule structures (.mol2 file) were downloaded from ZINC database.[87,88]

The ionic strength of MD system was tuned with explicit ions. For these simulations, box size was increased to 325 nm^3, to access lower ionic strength than the initial case. Counter ions were added after neutralization until the ionic strength reached a target value. For 0 mM ionic strength, all counter ions were removed, resulting in a non-neutral system.

Surface interaction was evaluated by the fraction of adsorption time, f_{ads}, as shown by following equation, where t_{ads} is the time for intermediate molecule to stay on cascade surface, when the short-range coulombic energy is less that 50 kJ mol^{-1}. Similarly, t_{des} corresponds to the desorption time.

$$f_{\mathrm{ads}} = \frac{t_{\mathrm{ads}}}{t_{\mathrm{ads}} + t_{\mathrm{des}}} \tag{2-1}$$

Adsorption energy, E_{ads}, is calculated based on the radial distribution function (RDF)[38,89] of the intermediate atoms around peptide surface, which is defined as the probability to find the intermediate molecule at a specific distance away from the reference surface. Using the intermediate's trajectories in MD simulation, the RDF was calculated by taking the distance *r* between substrate atoms and the closest peptide atom. Using the GROMACS function *gmx_rdf*, the probability density p_i was obtained versus *r*. Then p_i was normalized by bin (Δr = 5 pm) and summed using:

$$P_i = \left(\frac{p_i}{\Delta r \cdot \sum p_i}\right) \tag{2-2}$$

where $\sum P_i \cdot \Delta r = 1$. The energy level of a specific radial position was calculated by equation 2-3, where G_0 was taken at 1.0 nm away from charged peptide system and 1.5 away from non-charged peptide.

$$\Delta G_i = G_i - G_0 = -RT \cdot \ln\left(\frac{P_i}{P_0}\right) \tag{2-3}$$

Finally, the adsorption energy was obtained by taking the integral summation of relative energy differences, weighted by normalized probability density:

$$\Delta G_{ads} = \int_0^{r_{\max}} P_i \cdot \Delta G_i \cdot dr \quad \text{2-4}$$

To quantify the diffusion rate, mean square displacement (MSD) was used to evaluate the mobility of intermediate in bulk or on α-helix peptide surface. The relationship between MSD and diffusion constant, D, is shown in equation 2-5, where *n* is 2 for one dimension and 6 for three dimensions. In addition, τ is the time interval (lag time) corresponding to the displacement.

$$msd(\tau) = n \times D \times \tau \quad \text{2-5}$$

Two diffusion modes were assumed to calculate the MSD. For bulk diffusion, 3-D Brownian motion was assumed and the MSD was obtained from MD trajectories by taking the mean square displacement according to equation 2-5, where τ is lag time and r_j is the position of the intermediate's mass center at simulation time *j*.

$$msd(\tau) = \frac{\Sigma_0^{jmax+1-\tau}\left(r_{j+\tau} + r_j\right)^2}{j_{max} + 1 - \tau} \quad \text{2-6}$$

It should be noted that the terminology "lag time" for MSD calculation is mathematically independent from the "lag time" used in the Experiment Section. Unless specified, the average bulk diffusivity, D_{avg}, was calculated by fitting the MSD vs τ curve between 200-1000 ps. Since the MSD was taken over the entire simulation region, it contains the average of adsorption and desorption components. Therefore, a subscript "avg" was appended to distinguish it from pure bulk diffusivity, but when the interaction is very weak (*e.g.*, on His peptide), D_{avg} can represent the bulk diffusivity. As a result, the D_{avg} can reflect the average mobility of intermediate in peptide system.

When intermediates were closely associated with peptide surface, surface diffusivity, D_{surf}, can be calculated using a 1-D hopping diffusion mode, where the intermediate molecule can only move to one of its two nearest-neighbor sites (1-D) with equal probability. The MSD for hopping mode can be calculated according to equation 2-7, where L_{hop} is the hopping distance to the nearest charged residue and Γ is the hoping rate (s^{-1}).

$$msd(\tau) = L_{\mathrm{hop}}^2 \times \Gamma \times \tau \qquad 2\text{-}7$$

By combining equation 2-5 and 2-7, one dimensional surface diffusivity, D_{surf}, can be correlated to L_{hop} and Γ, so that

$$D_{surf} = \frac{1}{2} \cdot L_{jump}^2 \cdot \Gamma \qquad 2\text{-}8$$

As for hopping rate, the coulombic energy diagram was first slightly smoothed with convolution of a scaled window (win = 21). From coulombic energy distribution, two to three peaks (*e.g.*, -150, -280 and -400 kJ mol^{-1} for the oxalate-Lys couple) could be identified, corresponding to the plateau in coulombic energy diagram. Detailed figures can be found in APPENDIX A. Number of hops, N_{hop}, was taken as half of the number of energy shifts between energy plateaus, excluding the shift involving 0 kJ mol^{-1} (which indicates desorption). After that, average hoping rate (Γ_{avg}) was obtained by dividing N_{hop} by simulation time (50 ns). To exclude the impact of the varying desorption period, Γ (the hoping rate upon adsorption) was calculated by further dividing Γ_{avg} by f_{ads}.

In this way, D_{surf} can be compared to f_{ads} and E_{ads} to describe the relationship between surface mobility and desorption probability. In order to obtain the hopping distance, L_{hop}, the rise along helical axis was taken from a generally accepted value (0.15 nm per residue), which is then multiplied by the numbers of residues to yield L_{hop}.

2.3 Results and Discussion

2.3.1 Surface hopping mechanism

The adsorption of small charged molecules on biomolecular surfaces is due to local potential energy minima caused by non-specific electrostatic interactions about the bio-interface. Mechanistically, local potential energy minimization competes with kinetic exchange between adsorbed molecule and water molecules, resulting in a dynamic adsorption and desorption process that allows for bound diffusion across the charged biological surface.

Figure 2-1 shows the chemical structures of charged intermediate species and the oppositely charged amino acid side chains. As shown in Figure 2-1a, oxalate and glucose 6-phosphate (G6P) molecules both carry -2 charge, but oxalate has a symmetric charge distribution due to the two identical carboxylate groups, and G6P is polarized, with negative charge concentrated on its phosphate group. Glyoxylic acid has similar geometry to oxalate, with only -1 charge. As for the amino acid side chains in Figure 2-1b, Arginine and Lysine both have +1 net charge as indicated by their high pKa values. But Arginine has a much stronger polarization degree on its guanidine group, as shown by the electrostatic potential map in Figure 2-1c.[33]

Therefore, Arginine and Lysine are expected to have the same long range electrostatic impact but a very different local interactions.[40] In contrast, Histidine has no net charge but a high polarization degree. Figure 2-1f-h show several example of Lysine-Alanine peptides in an α-helix conformation. The charge density is tuned by the neutral Alanine residues (Figure 2-1b) placed between adjacent LYS side chains.

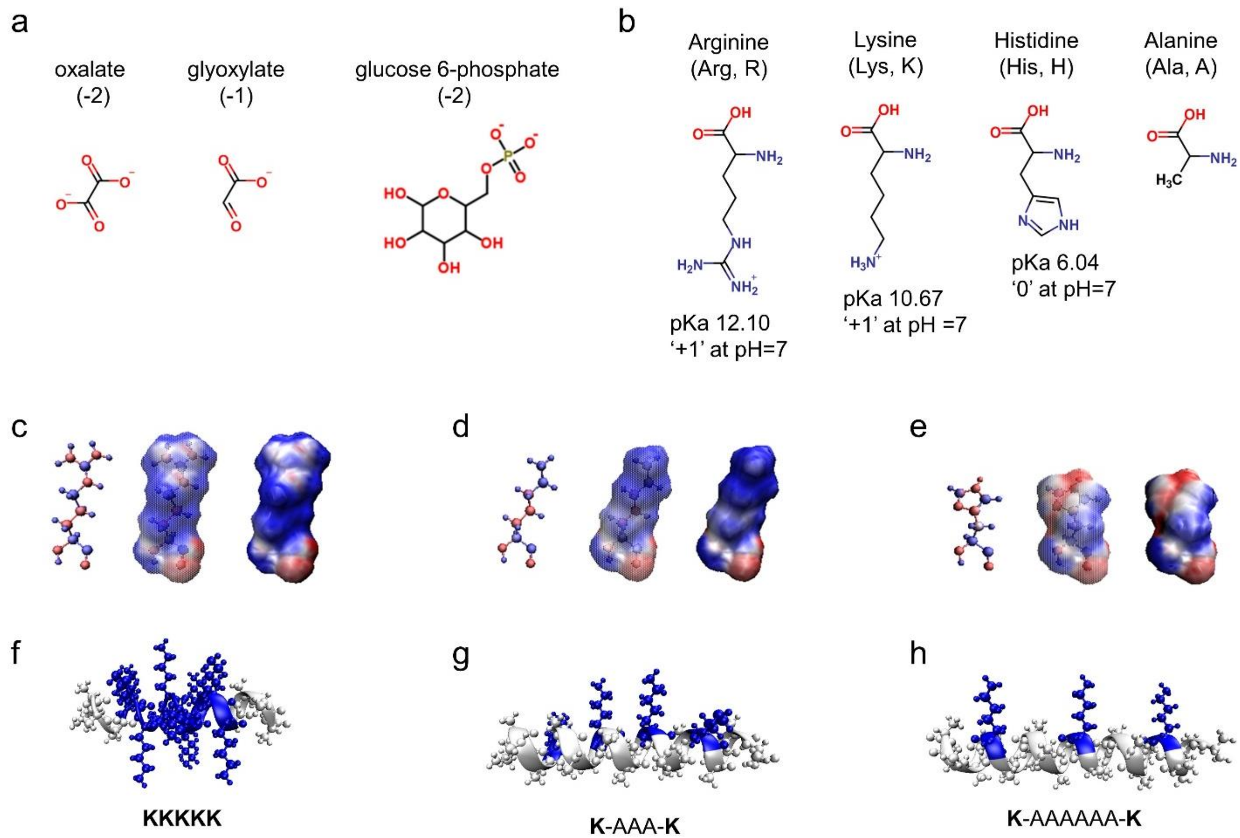


Figure 2-1 (a) Intermediate species. (b) Amino acid side chain species. (c-e) electrostatic potential map of Arginine, Lysine and Histidine. (f-h) a-helix Lysine peptide with various charge density.

With the mechanism of electrostatic channeling as a pretext, we used MD simulations to study the bound diffusion of a common small biological intermediate, oxalate, across a charged peptide surface. The adsorption energy, E_{ads}, and average coulombic energy, E_{coul}, between the intermediate and peptide were chosen as metrics indicating to an effective substrate channeling.

Specifically, E_{coul} can be calculated for every frame of the simulation, and its time average indicates the degree of electrostatic interaction between charges. E_{ads} is calculated based on the proximal density of oxalate around peptide surface (radial distribution function, RDF), and describes interactions caused by hydrogen bonding and polarization of the interacting groups.

RDF data provides information about the proximity of an electrostatic interaction based on both discrete hydrogen bonding interactions and cumulative adsorption energy. Here, we use the RDF of oxalate about a polypeptide chain to determine the probability that oxalate will be stabilized at a given distance from the electrostatic peptide surface. As shown in Figure 2-2, corresponding RDF diagrams for oxalate with the charged amino acid residues of interest indicate a high density of interactions occurring at ~0.15 nm for cationic Arg- and Lys-containing peptides. A minimum distance of 0.145 nm is also observed, which corresponds closely to the minimum van der Waals' distance (~0.15) associated with hydrogen bonding between anionic oxalate and either the ε-ammonium of Lys or the guanidinium of Arg ($NH^+ \cdots^{-} O$).[90] These results indicate a fast adsorption process and a possible contacting surface diffusion process. However, peptide surfaces containing His residues primarily interact with oxalate at a distance between 0.5 and 1.5 nm, which extends beyond the short-range cut-off distance for dissociation (1.2 nm). Therefore, HIS residues do not create coulombic forces and do not promote electrostatic channeling.

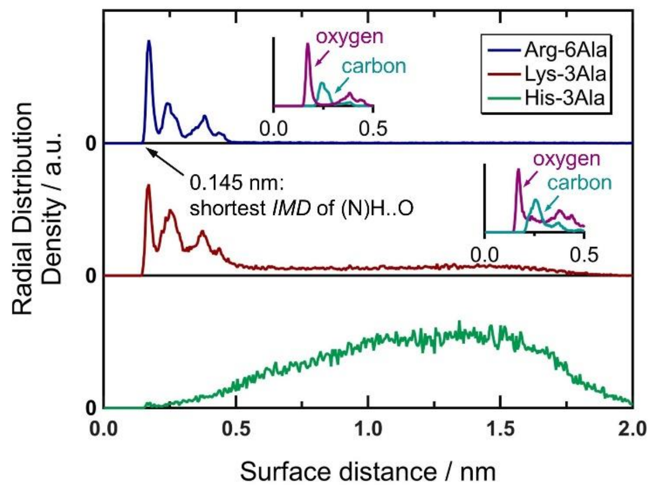


Figure 2-2 Radial distribution function (RDF) diagrams of oxalate around the peptide surface and the corresponding diagrams for oxygen atoms and carbon atoms individually (inset).

Whereas the RDF demonstrates oxalate's proximity to peptide surface, surface mobility is illustrated by short-range coulombic diagram, as shown in Figure 2-3a. MD simulations describing coulombic energy during diffusion of oxalate along a polyarginine peptide chain reveal several discrete energy states that dictate its surface mobility. The corresponding coulombic energy diagram (Figure 2-3a) suggests that oxalate primarily exists in either a singly- or dually-bound ionic conformation with adjacent Arg residues (Figure 2-3b) with coulombic energy of -190 kJ mol^{-1} or -390 kJ mol^{-1}, respectively, or in an unbound state of coulombic energy 0 kJ mol^{-1}. As oxalate travels across the peptide surface, the coulombic energy diagram exhibits a period of rapid flux in energy between each state (highlighted as the shaded region of coulombic diagram, Figure 2-3a). During this process, the randomly oriented kinetic exchange between oxalate and either water molecules or adjacent charged sites results in oxalate "hopping" to a neighboring site (Figure 2-3b). This energy-hopping region of coulombic energy diagram serves as a qualitative indication that diffusion is largely confined to the oligopeptide surface. Additionally, the combined RDF minimum distance (0.145 nm)[90] and coulombic energy results suggest a diffusion mechanism that utilizes distinct hydrogen bonding interactions between oxalate and basic amino acid residues, which adds significant detail to previously reported mechanisms of locally restricted diffusion within an electrostatic field.[29]

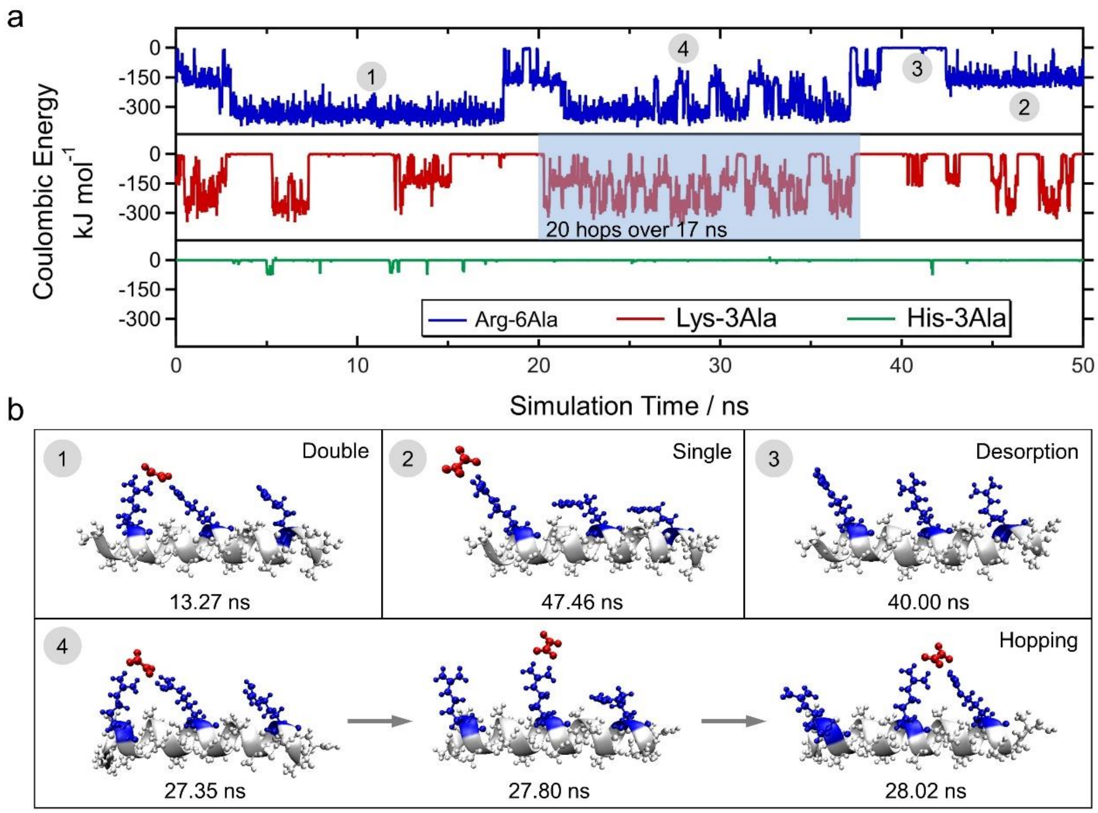


Figure 2-3 (a) Short-range Coulombic energy diagrams between oxalate (-2) and peptide composed of combinations of alanine and arginine, Lysine, and Histidine. The shaded area in Lys peptide figure (red) represents a frequent surface diffusion process. (b) Representative MD simulations corresponding to the plateau region in the Coulombic energy diagram between ARG side chains (blue) and oxalate (red).

One of the primary benefits of MD simulations is the opportunity to rapidly explore several theoretical electrostatic peptide bridges that would be impractical to screen experimentally. Using coulombic energy diagrams as described above, we sought to identify optimal cationic peptide residues to facilitate electrostatic channeling of an anionic intermediate. By tuning the type of charged amino acid (Arg, Lys, His), we studied the electrostatic interactions of various peptide structures with a dianionic oxalate molecule. Despite being highly polarized, neutral His ($pK_a = 6.04$) peptides do not allow for ionic stabilization and therefore do not enable any adsorption of oxalate. Both Arg and Lys have the same +1 net charge, but Lys has a much weaker polarization degree and H-bonding capability as compared to Arg.[40] Additionally, the volume of Lys's ε-amino group is smaller than Arg's guanidinium group and thus is not able to sterically block bulk water molecules as effectively. This results in stronger interactions between oxalate and water molecules, and increased kinetic exchange with the bulk media. Consequently, oxalate molecules displayed a shorter adsorption time fraction (f_{ads}) on Lys-containing peptides than their Arg counterparts (Figure 2-3a).

Despite exhibiting smaller f_{ads}, Lys peptides display a unique transport phenomenon in the exceptionally high number of energy hops exhibited in coulombic diagram. Highlighted by the shaded area of the energy diagram in Figure 2-3a, the coulombic energy hops 20 times over the 17 ns timeframe and moves frequently on the peptide surface. In this case, the Lys residues are separated by three alanine (Ala) residues, with 0.6 nm between adjacent Lys residues along the α-helix axis (assuming a normal helical increment of 0.15 nm per residue). This results in a surface diffusivity of 2.1×10^{-6} cm^2 s^{-1}, which is comparable to the typical bulk diffusivity of small molecules ($10 - 20 \times 10^{-6}$ cm^2 s^{-1}). In contrast, Arg peptides interact too strongly with oxalate, anchoring the intermediate instead of allowing transport. These data suggest that Lys appears to be an effective choice of charged amino acid residue for design of an electrostatic channel. Detailed discussion on surface diffusion constant can be seen in Section 2.3.3.

2.3.2 Impact of intermediate species

With the insights afforded above by coulombic energy diagrams, we sought to identify characteristics that allow various charged intermediates to be efficiently shuttled between conjugated active sites. To this end, MD simulations were performed on three common biological intermediates, glucose-6-phosphate (G6P), oxalate, and glyoxylate (Figure 2-4a-c). As shown in Figure 2-1, both oxalate and G6P possess a -2 charge, however, G6P has a much more localized charge with respect to its molecular volume.

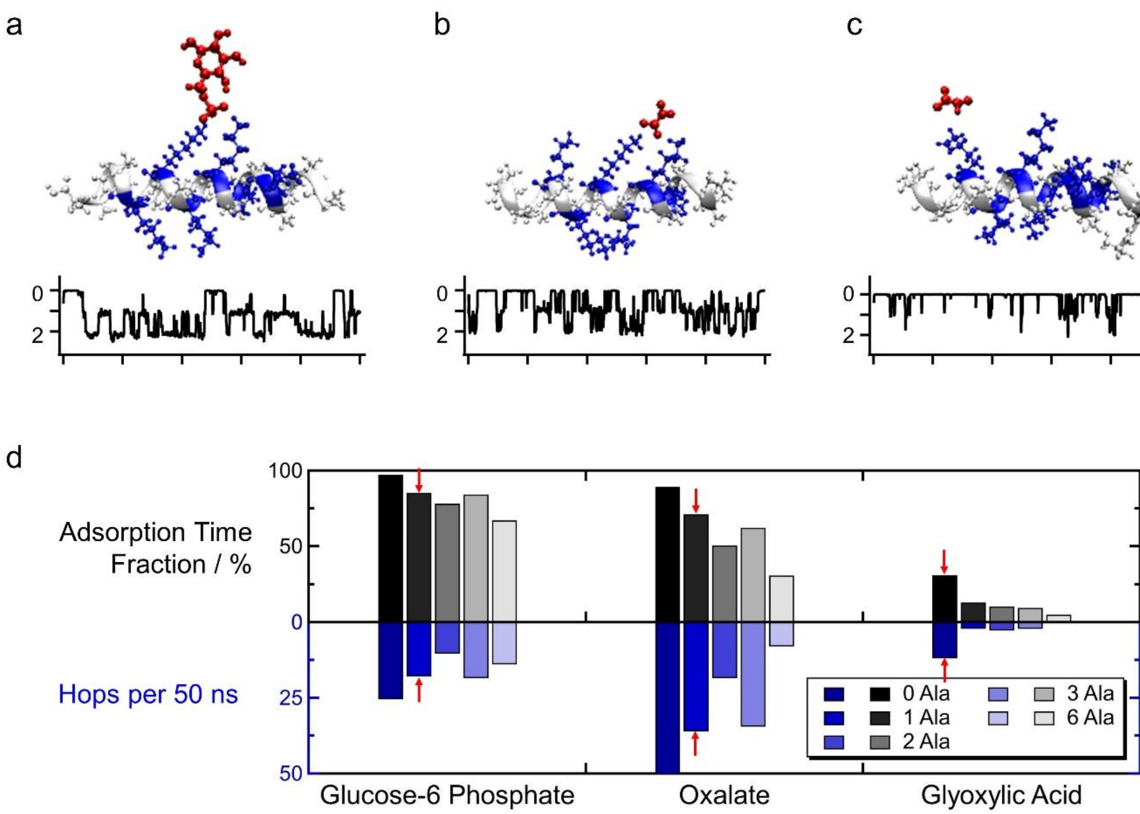


Figure 2-4 System configuration and coulombic energy diagram for glucose 6-phosphate (a), oxalate (b) and glyoxylate (c). The inserts are simplified coulombic energy diagram, where "0" and "2" on vertical axis stands for the energy level of dual association and desorption. (d) Adsorption time fraction and hopping times of intermediate molecules on peptides with various Lys fraction over the entire 50 ns simulation. The bar marked by red arrows corresponds to the coulombic diagram in above figures. Numerical data can be found in APPENDIX B

Coulombic energy diagrams of oxalate and G6P (Figure 2-4a and b) indicate that they both readily adsorb (f_{ads}= 71 $\pm$7 % and 85 $\pm$8 %, respectively) and diffuse across the peptide bridge through a double association mechanism. However, surface interactions changed dramatically in the case of the singly charged glyoxylic acid (Figure 2-4c). Despite having a similar structure to that of oxalate, hardly any dual association (1.2 $\pm$0.6 % time fraction) is observed in corresponding coulombic diagram and the intermediate remains either singly associated with Lys or desorbed (f_{ads} = 12 $\pm$4 %). Significant desorption of glyoxylic acid arises because its single anionic site does not allow for simultaneous coordination between adjacent Lys residues. Therefore, glyoxylate depends on a dissociative jumping mechanism between residues that increases the probability of desorption from peptide surface. This suggests that singly-charged reaction intermediates should be avoided as substrate channeling targets due to a high propensity for dissociation with the charged peptide chain. Residues with smaller positive charge (*e.g.*, + 0.5) might be an interesting subject to study the dual association and transportation of such intermediates. However, the natural HIS residue is only 10% charged given its pKa value at 6.04, and it is not supposed to be a good candidate. Therefore, synthetic amino acid residues might be able to give more inspiration to this issue. Also, further experiment is needed to validate the channeling behavior of singly charged intermediate molecule, and this has the potential to add more details and evidence to the surface diffusion mechanism and design rules.

2.3.3 Balance between surface adsorption and mobility

In an ideal electrostatic channeling system, the intermediate molecule should have a strong adsorption and thus a long adsorption time. At the same time, the intermediate molecule should be very mobile on cascade surface to reach the sequential site. However, as discussed above, strong interaction may decrease intermediate mobility on peptide surface. To further study the relationship between surface adsorption and mobility, adsorption energy and diffusivity were quantified for oxalate on Lys peptides. Simulated electrostatic interactions between oxalate and peptide bridge were described in terms of the absorption energy (E_{ads}, described above), the fraction of time that intermediate is adsorbed, f_{ads}, and the diffusivity along the surface of peptide chain, D_{surf}.

As shown in Figure 2-5, MD simulations of both oxalate and G6P were analyzed in terms of $\mathrm{E_{ads}}$, $\mathrm{D_{surf}}$, and $\mathrm{f_{ads}}$, where $\mathrm{E_{ads}}$ was controlled by varying the ratio of charged Lys residues to neutral Ala residues (*e.g.*, Figure 2-1f-h). These simulations demonstrate that $\mathrm{E_{ads}}$ for oxalate is enhanced with increasing Lys fraction, from ~0 kJ mol^{-1} to 8.91 kJ mol^{-1} (Figure 2-5a). As a result, $\mathrm{f_{ads}}$ increased gradually from 31%, reaching a maximum of 90% for oxalate, while $\mathrm{f_{ads}}$ for G6P ranged from 67% to 97%. This suggests that a higher fraction of Lys is required to prevent oxalate from desorbing from peptide surface, while G6P is not likely to desorb under any of the conditions studied. Additionally, we found that both oxalate and G6P maintained a high Γ (~1 times per ns) for all Lys-based peptides with the exception of $\mathrm{Lys_1}$-$\mathrm{Ala_6}$, wherein the separation of neighboring Lys residues is too great to allow for a dual association diffusion mechanism. This resulted in a consistently high $\mathrm{D_{surf}}$ for most conditions. However, according to equation 2-8, $\mathrm{L_{jump}}$ affects $\mathrm{D_{surf}}$ exponentially, where $\mathrm{D_{surf}}$ decreases substantially for the peptide chains where Lys fraction is too high. This is due to persistent double and triple association of the intermediate to proximal Lys residues that dramatically slow diffusion across the peptide chain.

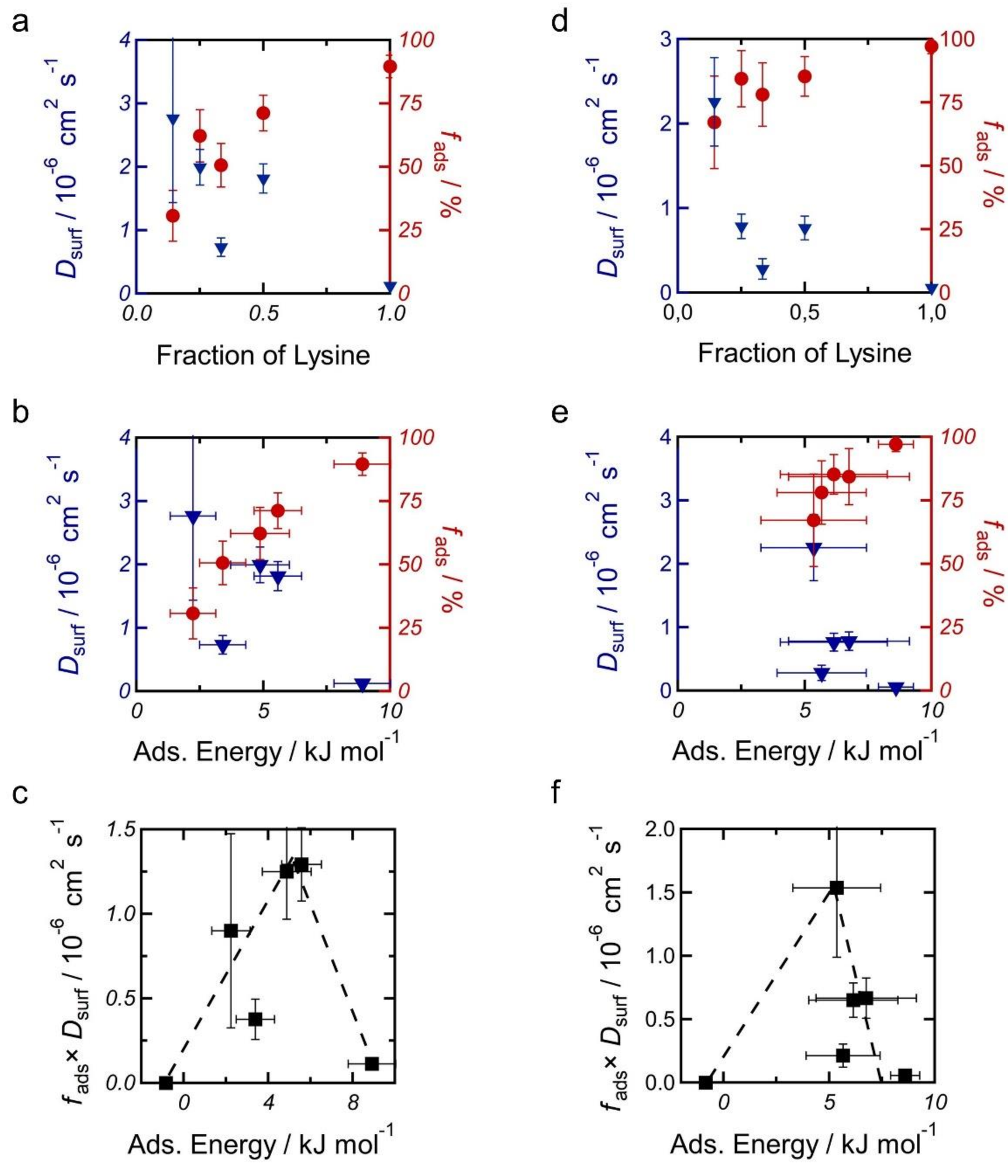


Figure 2-5 Effect of the Lys fraction (a, d) and thermodynamic adsorption energy, E_{ads}, (b, e) on surface diffusivity, D_{surf}, and adsorption time fraction, f_{ads}. Effect of E_{ads} on transport efficiency, characterized by $f_{\mathrm{ads}} \times D_{\mathrm{surf}}$ (c, f). Panels a−c correspond to oxalate, and panels d−f are for G6P. Error bars for all parameters represent the standard deviation of 10 parallel simulations for each individual system.

In order to provide an optimization metric to balance adsorption energy with surface mobility, we simply took the product of $f_{\mathrm{ads}} \times D_{\mathrm{surf}}$ as a measure of transport efficiency. The resulting plots of $f_{\mathrm{ads}} \times D_{\mathrm{surf}}$ vs E_{ads} (Figure 2-5 c, f) indicates that transport efficiency reaches maximum for peptides where the Lys fraction is slightly less than saturated (i.e., Lys_6-Ala_1). The simulated transport efficiency for a Lys-saturated peptide (Lys_{10}-Ala_0) was decreased due to the insufficient surface mobility, while the low transport efficiency of Lys_1-Ala_6 was due to poor adsorption time. Collectively, MD simulations indicate that an electrostatic channeling surface with high (but less than saturated) charge density and a polyanionic intermediate would strike a balance between surface mobility and adsorption energy to allow for efficient electrostatic substrate channeling.

2.3.4 Qualitative comparison with experiment by stop-flow lag time analysis

In this section, MD simulation results are compared with experiments, in terms of the occurrence of channeling and the dependence of ionic strength. In cooperation with our partners from University of Utah, we selected the reaction of glucose with HK and G6PDH as a model system to introduce an artificial channeling bridge, as shown in Figure 2-6a.

This enzymatic cascade utilizes HK with adenine triphosphate (ATP) to phosphorylate glucose to G6P, which is subsequently oxidized by G6PDH with nicotinamide adenine dinucleotide phosphate ($NADP^+$) as a terminal oxidant (Figure 2-6b). The reaction rate on each active site (~0.01-0.1 s^{-1}) is much slower than the diffusion rate of small ligand molecules (~10^{-5} cm^2 s^{-1}), including cofactor, reactant, intermediate and product. Therefore, it can be assumed that there is no delay on bulk concentration change when turnover occurs at any active sites. Therefore, by measuring the absorbance corresponding to NADPH formation, we were able to indirectly monitor the activity of HK or directly measure the activity of G6PDH. Additionally, this reaction sequence provides G6P as a charged intermediate, which allows for the comparison of our experimental findings with those suggested by MD modeling. Based on the simulated oligopeptide chains, we synthesized an electrostatic bridge consisting of a poly(Lys) oligopeptide, as shown in Figure 2-6a. Corresponding synthesis detail can be found in our published work.[4,33]

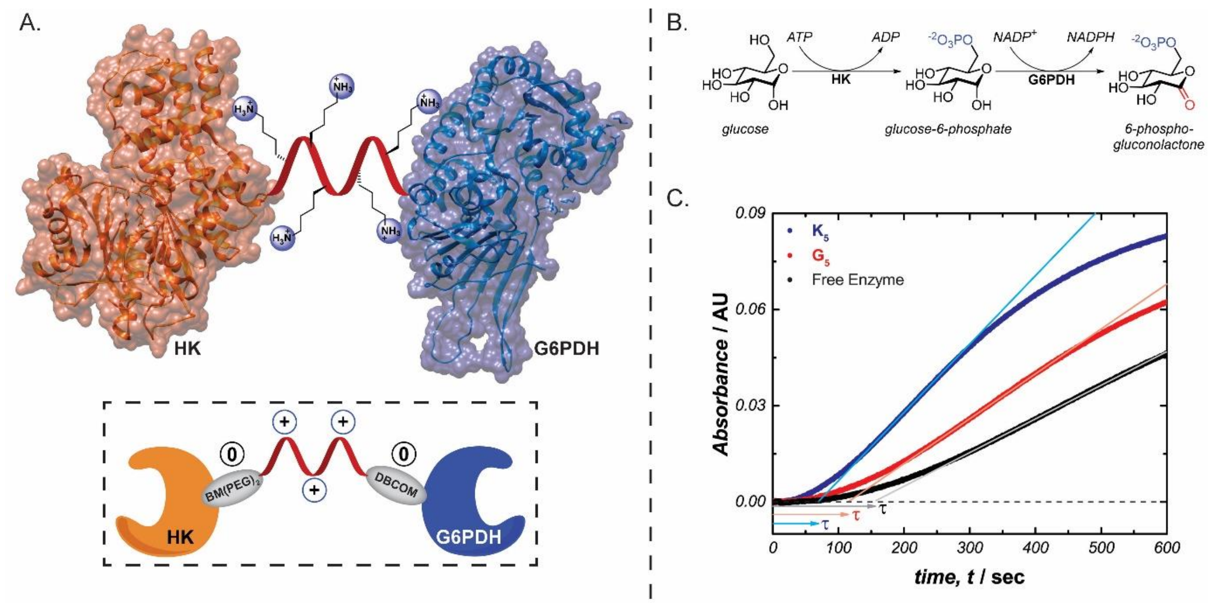


Figure 2-6 (a) Illustration of the proposed channeling complex using a poly(lysine) bridge as an electrostatic surface between hexokinase (HK; PDB 3VF6) and glucose-6-phosphate dehydrogenase (G6PDH; PDB 4LGV). (b) Experimental reaction scheme used to study electrostatic channeling of the charged intermediate (glucose-6-phosphate) across a cationic peptide bridge. (c) Sample absorbance plot highlighting the determination of experimental lag time (τ) for complexes containing a 4 nm cationic bridge (K5), neutral bridge (G5), or free enzymes.[33]

As discussed in Chapter 1.3, one of the most common methods for studying substrate channeling involves measuring the transition time, also called lag time, required to reach steady-state flux of the reaction intermediate (G6P).[22] The lag time, τ, was determined experimentally by stopped-flow injection analysis, in which the absorbance of a solution containing both enzymes was measured following an injection of glucose. From this plot, the maximum change in absorbance is extrapolated to the time axis (τ) at which absorbance = 0, which is exemplified in Figure 2-6c showing the experimental results of three types of enzymatic couple, including free standing enzymes (no channeling), neural bridge (proximity only) and charged bridge (proposed channeling). The result on charged peptide bridge shows much smaller lag time than its uncoupled counterparts, indicating that the electrostatic surface of the peptide facilitates transport of reaction intermediate (G6P). In addition, the lag time of enzymes coupled by neutral bridge stayed at the same level of free standing enzymes, which rolled out the contribution of proximity alone.

Theoretically, the ions in solution can shield the electrostatic interaction between charges, which should undermine the channeling efficiency in this case. As a result, the degree of transport efficiency could be adjusted by ions strength, which broadens the range of comparison between simulation prediction and experiment measurement. In order to further demonstrate the impact of electrostatic interaction, τ is measured with variable ionic strength, as shown in Figure 2-7. Specifically, when IS level is increased from 0 mM to 100 mM, the lag time for K5 system almost increases to the same level with non-channeling system, approving the contribution of electrostatic interaction to the intermediate transport at low ionic strength.

However, the K15 case has a more concentrated charge but a lower channeling efficiency as compared to its K5 counterpart. The quantitative reason of this will be demonstrated in Chapter 5.

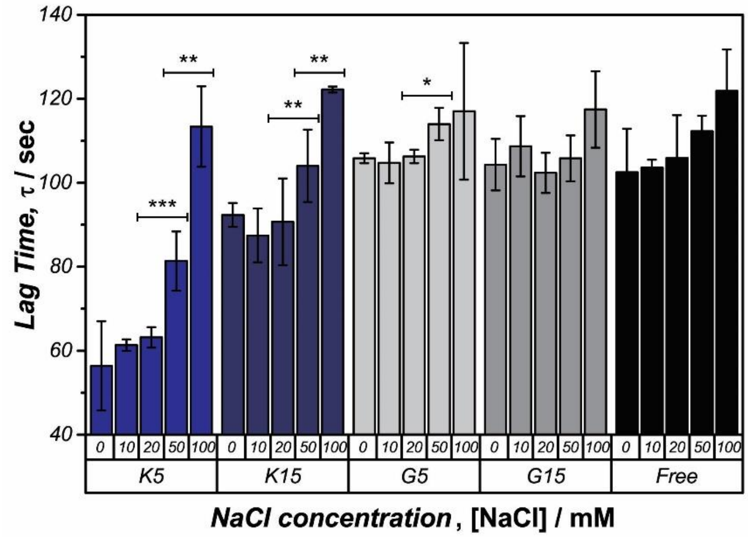


Figure 2-7 Lag time of the free enzymes and 5xLysine (K5), 15xLysine (K15), 5xGlycine (G5), and 15xGlycine (G15) complexes using 1.4 mM citrate with variable ionic strength, where the ionic strength is controlled by the concentration of NaCl. Experiments were performed by adding 275 mM glucose to 10 μgmL−1 of enzyme (complex) at pH 7.0 and 37 ℃. Error bars represent one standard deviation; (*) 0.05 > p > 0.01; (**) 0.01 > p > 0.001; (***) 0.001 > p. [ref]

The dependence of electrostatic interaction on ionic strength is also studied by MD, where ionic strength were tuned with explicit ions. The box size was controlled at 325 nm^3, and extra counter ions were added after neutralization, until the ionic strength reached target values. For 0 mM ionic strength, all counter ions were removed, resulting in a non-neutral system. MD simulation on ionic strength shows that adsorption time, f_{ads}, generally experiences a decreasing trend with increasing ionic concentration, where the fully saturated Lys peptide exhibits a slight decrease on f_{ads} (Δ = 4.3%) and the peptide with moderate Lys density has a severe decrease (Δ = 40%) on f_{ads}, as shown in Figure 2-8a. This suggests that electrostatic interaction and thus channeling efficiency tend to decrease with increased ion concentration, particularly for the less saturated Lys bridge. These simulation results qualitatively agree well with experiment results in Figure 2-7.

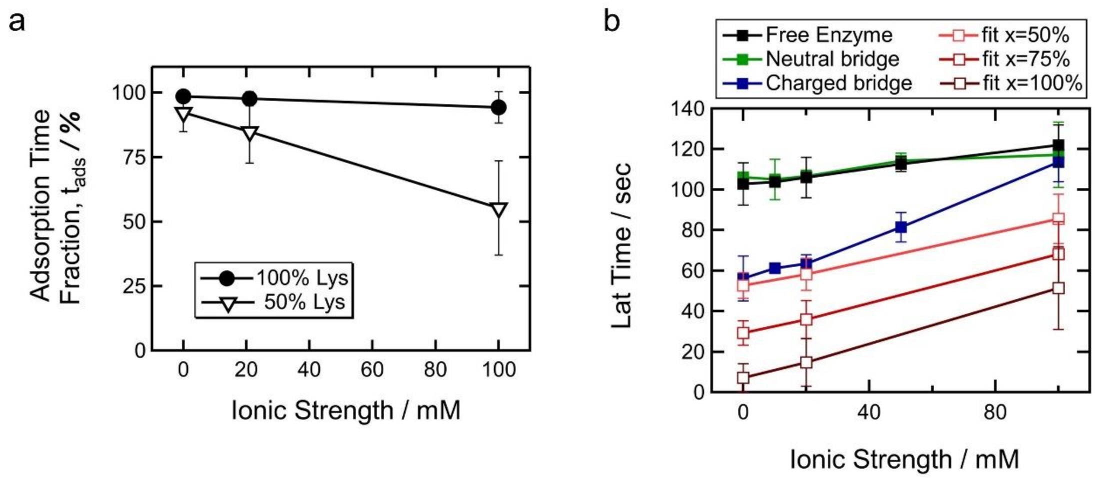


Figure 2-8 (a) MD Simulation on ionic strength, based on the system with G6P and Lys peptide. The theoretical "100 % Lys" was calculated from Lys-0Ala peptide (K10), and "50 % Lys" was calculated from Lys-1Ala peptide (K6-A1). (b) Plot of "Free", "G5" and "K5" in Figure 2-7, and analytical fit on "K5" lag time according to equation 2-9.

In order to further compare MD and experiment results, the simulated lag time of K5, τ_x, is analytically derived as shown in equation 2-9, where τ_0 is the lag time of free standing experimental lag time, c is probability of successfully channeled intermediate, x is the percent of coupled enzyme pairs as compared to their non-coupled counterparts, and V_1, V_2 is the maximum reaction rate of two enzyme. Given the fact that $V_2 \gg V_1$, equation 2-9 is simplified to its final expression:

$$\frac{\tau_x}{\tau_0} = \frac{1 - xc}{1 + xc \cdot \frac{V_1}{V_2 - V_1}} \approx 1 - xc \qquad \text{2-9}$$

$$c = p_c \cdot \frac{k_{cat}}{k_{cat} + k_2} \qquad \text{2-10}$$

By assuming c is equal to f_{ads} in Figure 2-8a, the fitted MD lag time and experiment lag time are shown in Figure 2-8b. Although the lag times agree well at low ionic strength when assuming a 50% contamination of uncoupled enzyme pair, the IS dependence trends show a significant discrepancy between the simulation and experiment results. This is due to the absent of quantitative parameters that can be directly correlated the rate constants of reaction on each active site.

Nonetheless, by combining equation 2-9 and Figure 2-7, the $x \cdot c$ value is around 50% percent at zero IS level. That means the channeling probability is between 50% and 100% at low IS level, which suggested a non-perfect but still very effective electrostatic channeling.

2.4 Summary

MD simulations were used to explore electrostatic interactions between cationic α-helix peptide surface and negatively charged reaction intermediates. Oxalate molecules was found to undergo a surface diffusion mechanism across cationic poly(Arg) and poly(Lys) peptide, whereby oxalate hops between discrete hydrogen bonding interactions between proximal charged residues. By varying the composition of simulated peptide bridge and the characteristics of the diffusing intermediate, we were able to define several rules for designing electrostatic substrate channels. Specifically, Lys residues were found to provide a balance of intermediate adsorption and surface diffusivity that allow for efficient electrostatic channeling while preventing dissociation of the intermediate into the bulk. Additionally, simulations suggest that a dianionic intermediate is required for the double associative diffusion mechanism that prevents desorption from the peptide surface.

Using these simulation-derived design principles as a foundational blueprint, we synthesized an enzyme complex by covalently conjugated HK and G6PDH by a poly(Lys) bridge. The synthetically cross-linked enzyme complex was shown to facilitate electrostatic substrate channeling by decreasing the lag time required to reach steady state with respect to the intermediate from 102 ± 10 sec for a mixture of coupled enzymes to 56 ± 11 sec for the Lys-bridged enzymes. The study of synthetic channeling complexes allowed us to identify low ionic strength as ideal experimental conditions to observe electrostatic substrate channeling.

Current MD and experiment results are able to give a strong support of the occurrence of artificially introduced electrostatic channeling, and also qualitatively agree well to each other. However, to further quantitatively compare these results and identify potential limitations in this channeling process, detailed MD study and advanced sampling technique are needed.

2.5 Acknowledgements

We greatly acknowledge the experiment support from Dr. David Hickey in Dr. Shelley Minteer's group at University of Utah. Oligonucleotides were synthesized by the DNA/Peptide Facility, part of the Health Sciences Center Cores at the University of Utah.

Chapter 3 Quantification of Thermodynamic Parameters*

3.1 Introduction

Surface adsorption of small ligand molecules is driven by a local energy minimum, and surface diffusion can be seen as a stochastic process on a given energy landscape. The charged surface of natural cascade[18,28] promotes surface diffusion along with intermediate migration under a proximal electric field created by the cascade surface. Continuum modeling is able to give a good description of the intermediate migration with long range electrostatic interaction.[15,35] However, there is no representation of the surface interaction. Brownian dynamics gives a more detailed description of the electric field and also explicit intermediate molecules,[29] but potential local h-bond interactions are underestimated. For kinetic quantification, rate constants are greatly needed and are the key to evaluate overall kinetics.[34] Therefore, detailed quantification of thermodynamics parameters, such as hopping energy barrier and desorption energy, is of great significance to further understand the channeling mechanism and quantify overall kinetics.

In this chapter, by using MD simulations, the energy barrier of an individual hop is calculated by transition state theory and desorption energy is calculated by umbrella sampling. By combining these two energy terms, as shown in Figure 3-1, the probability for intermediate to traverse the bridge is quantified in detail.

* The content of this chapter has been published on *ACS Catalysis* as a full paper.

Y. Liu, I. Matanovic, D. P. Hickey, S. D. Minteer, P. Atanassov and S. Calabrese Barton, "Cascade Kinetics of an Artificial Metabolon by Molecular Dynamics and Kinetic Monte Carlo", 8, 7719-7726 (2018). doi: 10.1021/acscatal.8b01041

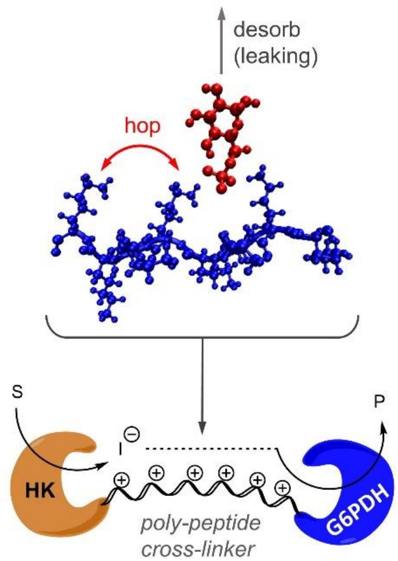


Figure 3-1 Poly-lysine peptide as an electrostatic bridge to transfer reaction intermediate (G6P, glucose 6-phosphate) from HK (hexokinase) to G6PDH (glucose 6-phosphate dehydrogenase).22

3.2 Model Description

A description of the basic molecular dynamics model on the peptide bridge is given in Chapter 2.

3.2.1 Transition state theory study

Glucose 6 phosphate's (G6P's) hopping rate (k_{hop}) on the peptide surface was calculated by the short-range coulombic energy change between the levels of dual- (-400 kJ/mol) and single- (-200 kJ/mol) association configuration, as shown in Figure 2-3. According to transition state theory (TST) the rate constant, k, is related to the energy barrier ΔG by an Arrhenius expression:

$$k = A \cdot \exp\left(-\frac{\Delta G}{RT}\right) \tag{3-1}$$

where A is the frequency factor, R is the gas constant and T is the absolute temperature. Therefore, A and ΔG could be calculated by fitting $\ln k$ as a function of $1/T$, which will be further discussed in following section. In order to do this, the MD system was set at 10 temperatures (10, 15, 20, 25, 30, 37, 40, 45, 50, 55 ^{o}C). At each point, 10 parallel 100-ns simulations were conducted to calculate the average hopping rate, $k_{hop,}$ and its standard deviation. It should be noted that ΔG and A values were calculated by fitting all of the 10×10 data points, rather than the average at each temperature. This enables a sufficient evaluation of the uncertainty of hopping rate. For ionic strength (IS) dependence study, IS values were set at 0, 20, 40, 70, and 120 mM explicitly represented by Na^+ and Cl^- ions. The IS value here represents additional ion concentration beyond that required for neutralization of MD system.

3.2.2 Umbrella sampling

From the MD trajectory of TST study, a dual association configuration was extracted for umbrella sampling.[50] After that, the peptide was first restrained at a reference position and the readily adsorbed intermediate molecules (G6P) were pulled away perpendicularly from the peptide surface for 200 ns. In this process, a spring constant of 1000 kJ mol^{-1} and a pulling rate of 0.01 nm ps^{-1} were used, as shown in Figure 3-2a. The two ε-ammonium nitrogen atoms were selected as the reference group and the whole G6P molecule was selected as the pull group. As a result, a maximum distance of 2 nm was obtained for the center-of-mass (COM) of G6P as a reference to its original position. From pulling trajectories, 15 frames with a COM increment of 0.1 nm were attempted to be selected as the initial configuration for each window in umbrella sampling. Such spacing distance allow sufficient overlap between the probability distribution within neighboring windows. This will be further discussed in section 3.2.2

In order to optimize the frame selection for umbrella sampling, two reaction coordinates, x_1 and x_2, were defined as equations:

$$x_1 = |\vec{x}_{\mathrm{G6P}} - \vec{x}_{\mathrm{ref}}| \quad 3\text{-}2$$

$$x_2 = |(\vec{x}_{\mathrm{G6P}} - \vec{x}_{\mathrm{ref}}) \cdot \vec{v}| \quad 3\text{-}3$$

where $\vec{x}_{\mathrm{g6p}}$ and $\vec{x}_{\mathrm{ref}}$ are the coordinates of the COM of G6P molecule and associated LYS's ε-ammonium nitrogen atoms, respectively. $\vec{v}$ is the given vector perpendicular to peptide surface in pulling process. In other words, x_2 is the projection of x_1 on $\vec{v}$. Figure 3-2c shows an example on the time course of x_1 and x_2. A certain difference can be seen between these values, which is due to the intermediate's lag response to the harmonic spring force during pulling process. Therefore, the pulling process was repeated 50 times to pick the 15 frames with $|x_1 - x_2|$ less than 0.01 nm. Usually, 20 parallel pulling enable at least one appropriate frame for each window.

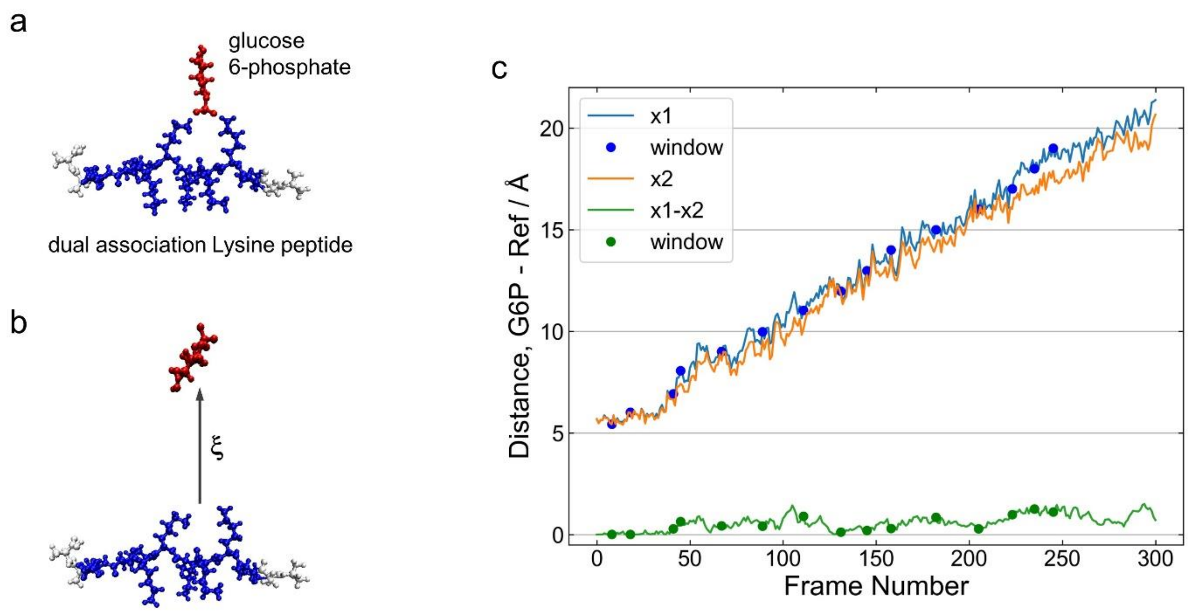


Figure 3-2 (a) System configuration of a dual association mode. (b) An assumed desorption state by pulling G6P 1.5 nm away from the surface.

For Umbrella Sampling, above mentioned reaction coordinates x_1 and x_2 were used for GROMACS "COM-distance" and "COM-direction" modules, as shown by Figure 3-3 and equation:

$$V_{\text{bias}} = \frac{1}{2} \cdot k \cdot (x - x_0)^2 \tag{3-4}$$

where V_{bias} is the biased potential applied to the intermediate molecule as its reaction coordinate, x, deviates from the energy minimum position, x_0. k is the spring constant at 500 kJ mol^{-1}. Usually, 10 ns MD simulations was conducted for each window to collect G6P's trajectory in umbrella sampling.

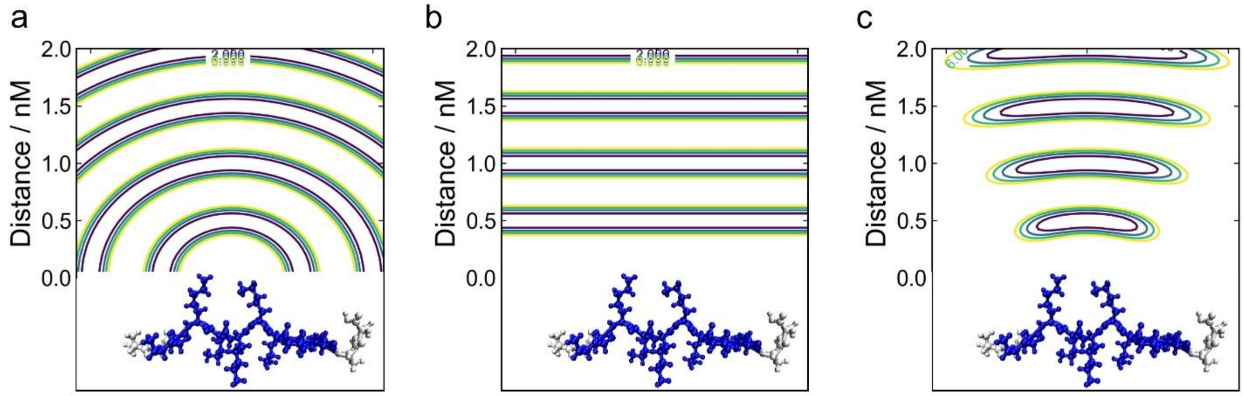


Figure 3-3 Gromacs biased potential for umbrella sampling. (a) *gmx_distance* restraints the pull group at a specific distance away from the reference group. (b) *gmx_direction* restraints the pull group at a plane perpendicular to given vector and also a specific distance away from reference group. (c) "*gmx_distance"* + "*gmx_direction*" allows the pull group to be restraint at the dome area above reference group. The x-axis for all figures are at the same scale to y-axis.

In order to calculate the potential of mean force (PMF)[91] as an indication of sorption energy, the sampling results were combined by weighted histogram analysis method (WHAM)[92,93] using Grossfield-WHAM code[94]. As shown in Figure 3-4, a 2-D PMF with non-independent reaction coordinates, x_1 and x_2, was obtained. That is, x_1 was above mentioned GROMACS "COM-distance" and x_2 was "COM-direction". From the 2D PMF plot, a smooth gradient can be seen from the proximity region to desorption region. The ragged bottom of the band is due to the insufficient sampling on the dome edge in Figure 3-3c, where x_1 and x_2 were too far away from their energy minimum. Therefore, this region was dropped when calculating the 1-D PMF along x_1. The average on x_2 was calculated by only taking the area above the yellow line in Figure 3-4, where $|x_1 - x_2| \leq 0.1$ nm.

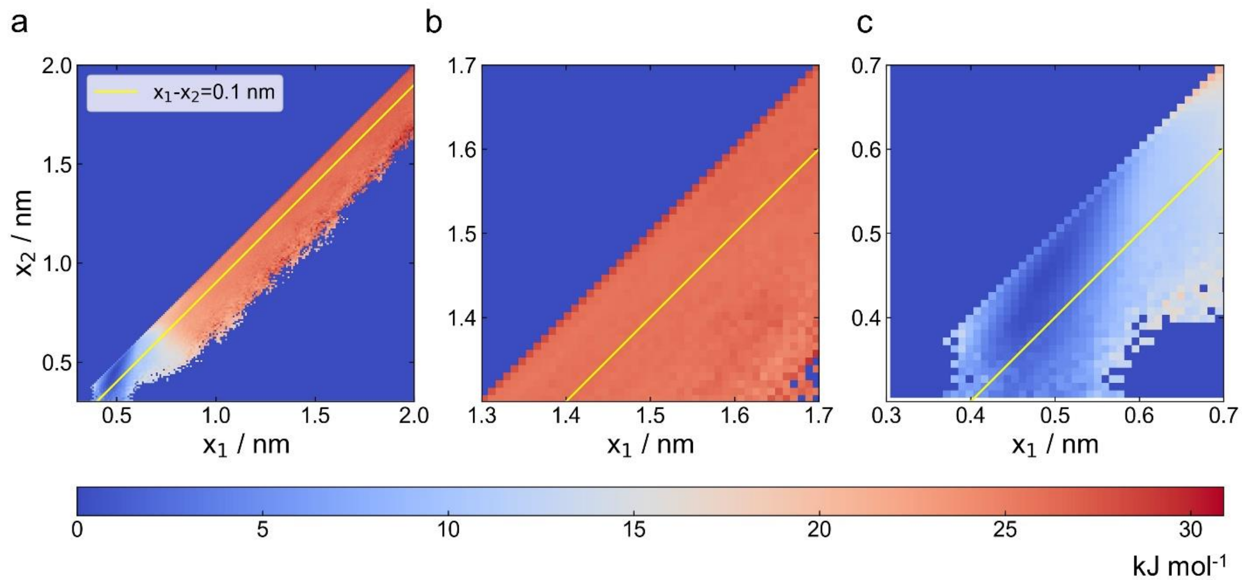


Figure 3-4 Two-dimensional PMF calculated by Grossfield WHAM code.

3.3 Results and Discussion

3.3.1 Energy barrier of surface hopping

In Chapter 2, charged oligopeptides were demonstrated to be an effective channeling bridge between sequential enzymes. MD studies on G6P molecule interacting with a charged polylysine bridge found that the G6P associates with two lysine side chains by dynamic exchange with surrounding water molecules, and surface diffusion was achieved through hopping between neighboring association sites (Figure 3-1). Additionally, surface hopping was indicated by short-range coulombic energy changes between discrete levels (Figure 2-3a and Figure 3-5a). By counting these energy level changes, the hopping frequency, k_{hop}, can be calculated. MD studies over a temperature range (Figure 3-5b) yields the hopping energy barrier, G_{hop}, via the Arrhenius equation:

$$k_{\mathrm{hop}} = A \cdot \exp\left(-\frac{G_{\mathrm{hop}}}{RT}\right) \tag{3-5}$$

where A is the frequency factor, R is the gas constant and T is temperature. The hopping mechanism involves significant intermolecular contact through hydrogen-bond interactions. In order to hop to the next association site, the G6P molecule dissociates from one LYS residue while still bonded to the other LYS as a swing arm. During this process, an energy penalty has to be overcome until G6P touches another LYS and thus reaches a new energy minimum. This is a Stern layer diffusion[95] that is less impacted by the ionic shielding from the bulk environment. As shown in Figure 3-5c, despite ionic strength variation from 0 mM to 120 mM, the energy barrier remains fairly consistent at 12 $\pm$0.5 kJ/mol. Therefore, a value at 12 kJ/mol was used for all IS conditions in later KMC parameterization.

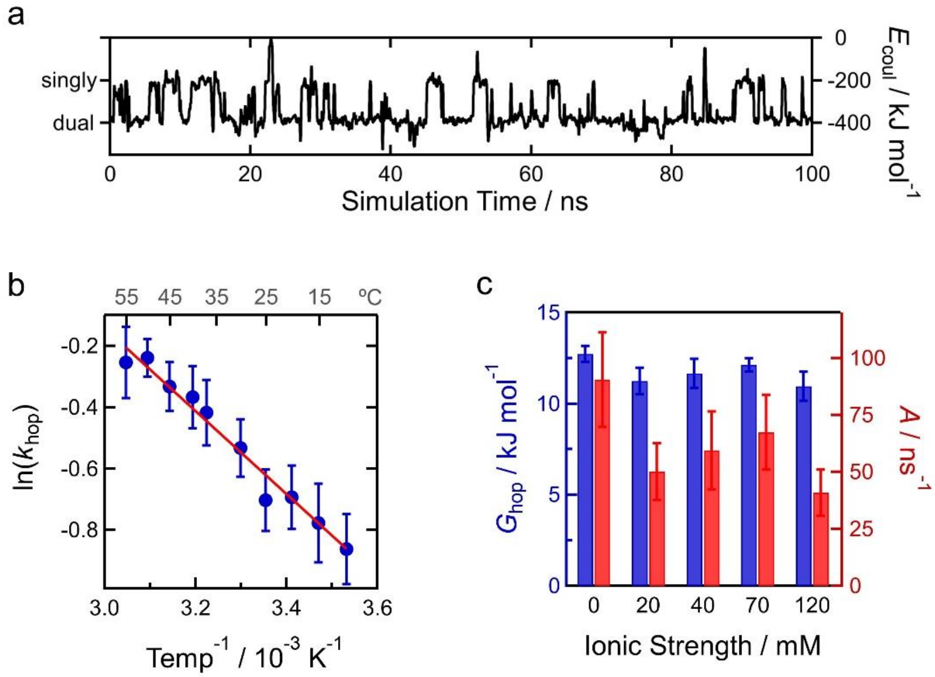


Figure 3-5 (a) Short-range coulombic diagram showing G6P hopping as indicated by energy fluctuation between single- and dual-association configurations. (c) Temperature dependence of surface hopping rate. (d) Ionic strength dependence on hopping energy barrier and frequency factor.

3.3.2 Desorption energy

Desorption energy was calculated by Umbrella Sampling.[50] Figure 3-6 summarizes the IS dependence study on the 1-D energy profile near the peptide surface. The probability distributions in Figure 3-6a indicate a sufficient overlap between neighboring windows, allowing an effective combination of each relative energy profiles. Figure 3-6b shows the energy profile of G6P's desorption from dual association to the bulk environment. Theoretically, a desorbing intermediate molecule has to traverse the double layer near the charged oligopeptide, comprising the above mentioned Stern layer and a diffuse later created by long-range electrostatic interactions. At low ionic strength, the energy drop can be well separated into Stern layer and diffuse layer components, as shown by the blue and orange arrows in Figure 3-6b. With increasing ionic strength, the long-range electrostatic interaction is gradually screened out and the corresponding energy region becomes a plateau just beyond the Stern layer. In contrast, the Stern layer potential well is less impacted by changes in ionic strength. This agrees well with the ionic strength independence study on hopping energy barrier Figure 3-5.

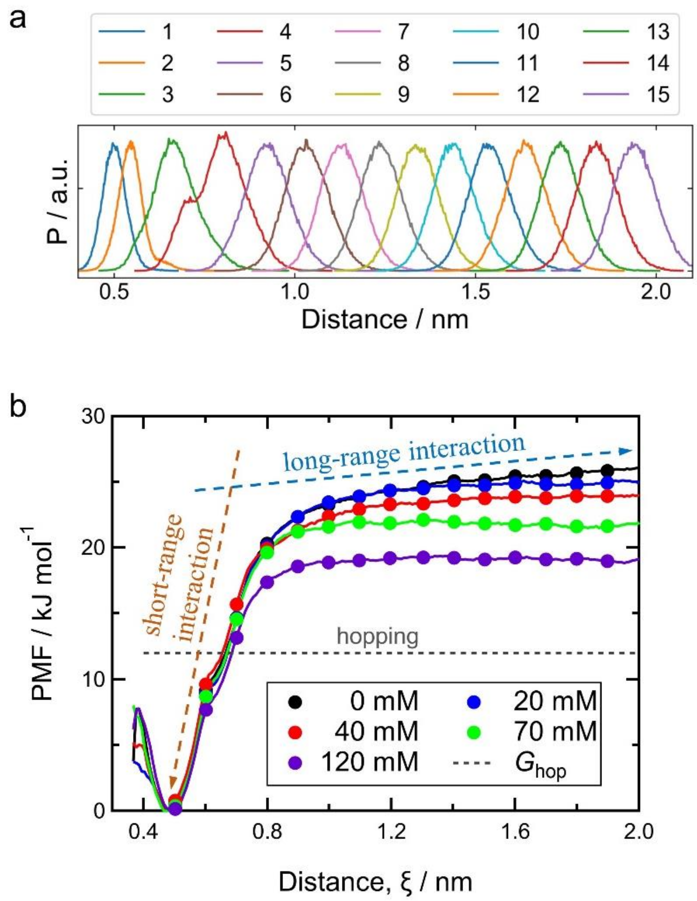


Figure 3-6 (a) Biased probability distribution of G6P molecule in each window. (b) 1-D potential of mean force by Umbrella Sampling. The points shows the biased energy minimum for each sampling window.

From Figure 3-6b, desorption energy, G_{des}, was calculated by taking the difference between the energy minimum and the energy average between 1.6 and 2.0 nm. Consequently, hopping of an adsorbed intermediate molecule requires a climb up the IS-independent energy hill (dashed line in Figure 3-6b) while simultaneously having a probability to reach a desorption state that is dependent on bulk ionic strength. Assuming a Boltzmann distribution, the leaking probability of each individual hop depends on the difference between G_{hop} and G_{des}:

$$\frac{p_{\mathrm{des}}}{p_{\mathrm{hop}}} = \exp\left(-\frac{G_{\mathrm{des}} - G_{\mathrm{hop}}}{RT}\right) \qquad \text{3-6}$$

At high IS, the leaking probability is 1/16, which means one out of 16 hops results in desorption. However, the low-IS leaking probability is calculated as 1/189, an order of magnitude lower than the 120 mM case. This means the hopping is much less prone to desorption at low IS and the intermediate is more likely to traverse the bridge. Table 2 summarizes the energy constants calculated by MD simulations, which will be used to parameterize the KMC model.

Table 2 Energy barriers and corresponding rate constants.

IS mM	$\boldsymbol{\lambda_{debye}}$ nm	$\boldsymbol{G_{hop}^{BrBr}}$ kJ/mol	$\boldsymbol{G_{des}^{BrBr}}$ kJ/mol	$\boldsymbol{\Delta G^{BrBr}}$ kJ/mol	$\boldsymbol{\frac{k_{hop}^{BrBr}}{k_{des}^{BrBr}}}$
0	9.8	12	25.5	13.5	189
20	2.2	12	24.9	12.9	146
40	1.6	12	23.85	11.9	99
70	1.2	12	21.7	9.7	43
120	0.9	12	19.14	7.1	16

IS: ionic strength

λ_{debye}: Debye length

G_{hop}^{BrBr}, k_{hop}^{BrBr}: energy barrier and rate constant for hopping from one dual association site to neighboring site on LYS bridge

G_{des}^{BrBr}, k_{des}^{BrBr}: energy barrier and rate constant for desorption from one dual association site

ΔG^{BrBr}=G_{des}^{BrBr}-G_{hop}^{BrBr}: energy difference from hopping transition state level to desorption level

3.4 Summary

The G6P intermediate molecule was found to undergo a surface diffusion mode under the protection of the electric double layer created by the charged peptide surface. Specifically, surface hopping occurs within a Stern layer due to local hydrogen bond interaction that is less impact by the ionic environment of bulk media. The actual value for hopping energy barrier approximate the energy difference between a dual and singly association mode. Above Stern layer, the long range electrostatic interaction creates a diffuse layer to further protect the surface diffusion. However, this layer could be largely shielded at high ionic strength. Collectively, the overall leaking probability for each individual hop depends on the energy difference between the hopping transition state and desorption state.

This detailed quantification on thermodynamic parameters not only further explore the surface diffusion mechanism, but also give a reasonable support for the rate constants that can be used to parameterize KMC model.

3.5 Acknowledgements

The analysis of umbrella sampling in this chapter is under the help of my committee member, Dr. Alex Dickson. My adviser, Dr. Scott Calabrese Barton, greatly contribute to the build-up of the python code for KMC simulation.

Chapter 4 Transition Pathway from Bridge to Binding Pocket*

4.1 Introduction

Natural cascades typically include 2D channeling surfaces between sequential active sites.[18,28] Similarly, the area between the artificial bridge and its enzyme binding site is a flexible 2D surface with multiple pathways and corresponding leakage. Figure 4-1 shows a proposed channeling pathway from peptide bridge to G6P's binding site on G6PDH. As discussed in Chapter 2 and Chapter 3, the surface diffusion is a stochastic process on an energy landscape with discrete potential wells. As compared to the 1D hopping on the peptide bridge, with random motion either forward or backward, 2D surface diffusion includes multiple pathways over a network of hopping sites. As a result, the overall leakage cannot be simply estimated by thermodynamic parameters.

Transition pathway analysis using a Markov state model (MSM) is a good approach to analyze this process.[51,53,54,57,96] MSM is widely used in the area of protein folding and ligand unbinding,[53,56,57,97,98] where the transition pathway is complicated by spatial configuration with multiple degrees of freedom (*e.g.*, inter-atomic distances and dihedral angles) and highly interconnected states of the system. Using detailed configuration discretization and combining states with related physical meanings (*e.g.*, desorption or binding states), MSM converts multi-dimensional molecular dynamics transitions to elementary pathways, allows the selection of dominant and long-timescale pathways, and finally human readable pathways.[57,99] As a result,

---

* The content of this chapter has been published on *ACS Catalysis* as a full paper. The work on molecular docking in this chapter is in collaboration with Dr. Ivana Matanovic and Dr. Plamen Atanassov at University of New Mexico

Y. Liu, I. Matanovic, D. P. Hickey, S. D. Minteer, P. Atanassov and S. Calabrese Barton, "Cascade Kinetics of an Artificial Metabolon by Molecular Dynamics and Kinetic Monte Carlo", 8, 7719-7726 (2018). doi: 10.1021/acscatal.8b01041

the energy landscape[53,98] and committor probability [57] to minimum-energy basins can be calculated.

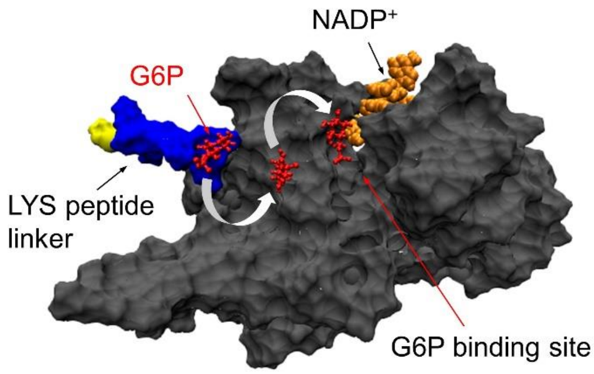


Figure 4-1 Proposed channeling pathway from the last dual association site on poly-Lysine peptide bridge to the G6P's binding pocket on G6PDH.

In this chapter, we further map the channeling pathway between the bridge and intermediate active site of G6PDH using a probability analysis based on transition state theory. This result is then validated by Transition Path Theory (TPT) with a Markov state model. This analysis helps to elucidate the intermediate's ergodicity, transition pathways, and energy barriers in channeling from peptide bridge to intermediate's binding pocket on G6PDH.

4.2 Model Description

Basic MD parameters follow the settings as discussed in Chapter 2.

4.2.1 Selection of complex structure

The crystal structure of G6PDH from *Saccharomyces cerevisiae* (experimental enzyme) is not yet available.[33] In order to obtain a relevant enzyme crystal structure, a comparison was made between this experimental G6PDH and the four species with available crystal structure (Table 3). Due to charge dissimilarity, Mycobacterium Avium G6PDH (Ma_G6PDH) and

Leuconostoc Mesenteroides G6PDH (Lm_G6PDH) were ruled out. The secondary cofactor binding site of human G6PDH (Homo sapiens G6PDH) structure is very different from yeast G6PDH (Saccharomyces Cerevisiae G6PDH) used in experiment. Moreover, Trypanosoma Cruzi G6PDH (Tc_G6PDH) has multiple CYS linker sites and available complex structure with both substrate and cofactor readily adsorbed. Therefore, Tc_G6PDH was selected to study the hopping from bridge to downstream enzyme active site. CYS-528 was selected as the linker site, as it was fully solvent exposed and close to the G6P binding site.

4.2.2 Hybrid topology at enzyme/linker interface

In order to conduct MD simulation of the bridge-G6PDH complex, force field parameters are required for all atoms and bonds. Generally, as discussed in section 2.2, CHARMM36 all-atom force field[73,74] provides the topology for all standard residues, such as protein and DNA components. In addition to this, CGenFF (CHARMM General Force Field)[75] can be used to calculate parameters for arbitrary, non-standard small molecules, such as ligands.

Here, these two databases are combined to generate topology for the interface between the bridge and its linker[33] and standard peptide/enzyme.[†] As shown in Figure 4-2, the linker molecule was capped with CYS residues on each side using *ChemAxon Marvin Sketch*. Then, the topology of whole complex was generated by CGenFF. The interface topology (bonds, angles, dihedrals) was used for the entire complex in MD simulation. The independent standard and non-standard sections were generated by CHARMM and CGenFF, respectively. The capped linker

[†] Q&A of CGenFF web page: https://mackerell.umaryland.edu/~kenno/cgenff/faq.php#hybrid

structure in Figure 4-2 was applied to the HK-bridge segment and similar approach was also used for bridge-G6PDH interface.

Table 3 Comparison of experimental G6PDH[33] and available crystal structures.

Enzyme	**Sc_G6PDH**	**Hs_G6PDH**	**Tc_G6PDH**	**Ma_G6PDH**	**Lm_G6PD**
Organism	Saccharomyces Cerevisiae	Homo Sapiens	Trypanosoma Cruzi	Mycobacterium Avium	Leuconostoc Mesenteroides
Species	Eukaryote yeast	Eukaryote human	Eukaryote parasite	Prokaryote bacteria	Prokaryote bacteria
Similarity[a]	--	48%	49%	34%	35%
Similarity +[b]	--	64%	64%	53%	54%
Total charge	-4.6	-1.1	1.1	-18.4	-20.5
PDB index	none	2BH9	5AQ1	4LGV	1E7Y
Ligands[c]	--	NADP/G6P	NADP&G6P	no	NADP&G6P
CYS number	single	multi	multi	single	no
CYS proximity[d]	good	medium	very good	medium	--

[a] Similarity: residue similarity to Sc_G6PDH; [b] Similarity +: positive similarity to Sc_G6PDH; [c] Ligands: substrate and cofactor availability in crystal structure; [d] CYS proximity: the proximity of CYS residue to G6P binding site.

To stabilize the complex, the terminal HK CYS of the bridge was first pulled away from the position-restrained G6PDH, extending the LYS bridge. Subsequently, the HK CYS side was position restrained while the rest of complex was allowed to stabilize over a 100 ns MD simulation. The resulting complex was used to initialize probability analysis.

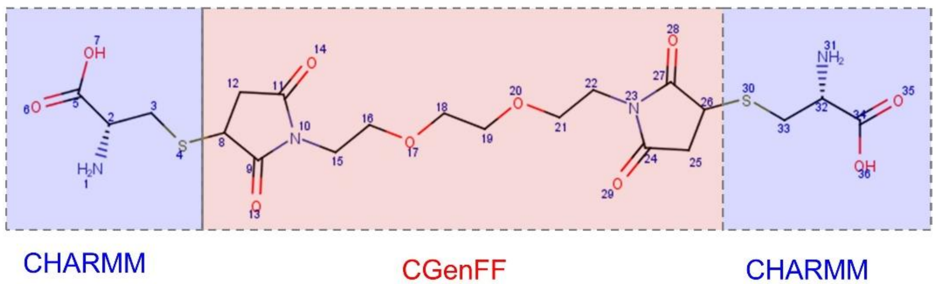


Figure 4-2 Chemical structure of linker molecule capped with CYS residue on each side, CYS-(BM-(PEG)$_2$)-CYS. The whole complex structure was processed by CGenFF to obtain the topology parameters for the interface between standard residue segment shaded by blue color and the non-standard residue segment shaded in red color. In actual MD simulation, the topology for pure standard residue segment was taken from CHARMM, and CGenFF parameters were applied to the pure non-standard segment and the interface topology.

4.2.3 *Intermediate initialization on complex structure*

To choose a G6P release point, molecular docking was used to identify favorable binding areas between the LYS bridge and G6P binding pocket that could be ruled out as leading to long-term adsorption, as shown in Figure 4-3. The docking simulations were performed using AutoDock Vina[100,101] by Dr. Ivana Matanovic working at Dr. Plamen Atanassov's group at University of New Mexico. Docking of G6P to G6PDH in the presence of the LYS bridge was performed multiple times using a range of areas (Figure 4-3a) between LYS bridge and the G6PDH active site. As a result, the rest of the area between the bridge and the active site was considered as a potential transition state area, where the position of G6P can be adjusted to find the points as shown in Figure 4-5.

Subsequently, 500 parallel simulations of 2 ns duration were conducted with velocity regeneration at each ionic strength (0, 20, 40, 70, 120 mM). At 10 ps per frame, these simulations generated 10^5 frames per ionic strength, to be used in analyzing the channeling behavior. The identification of starting points will be described in Section 4.3.1.

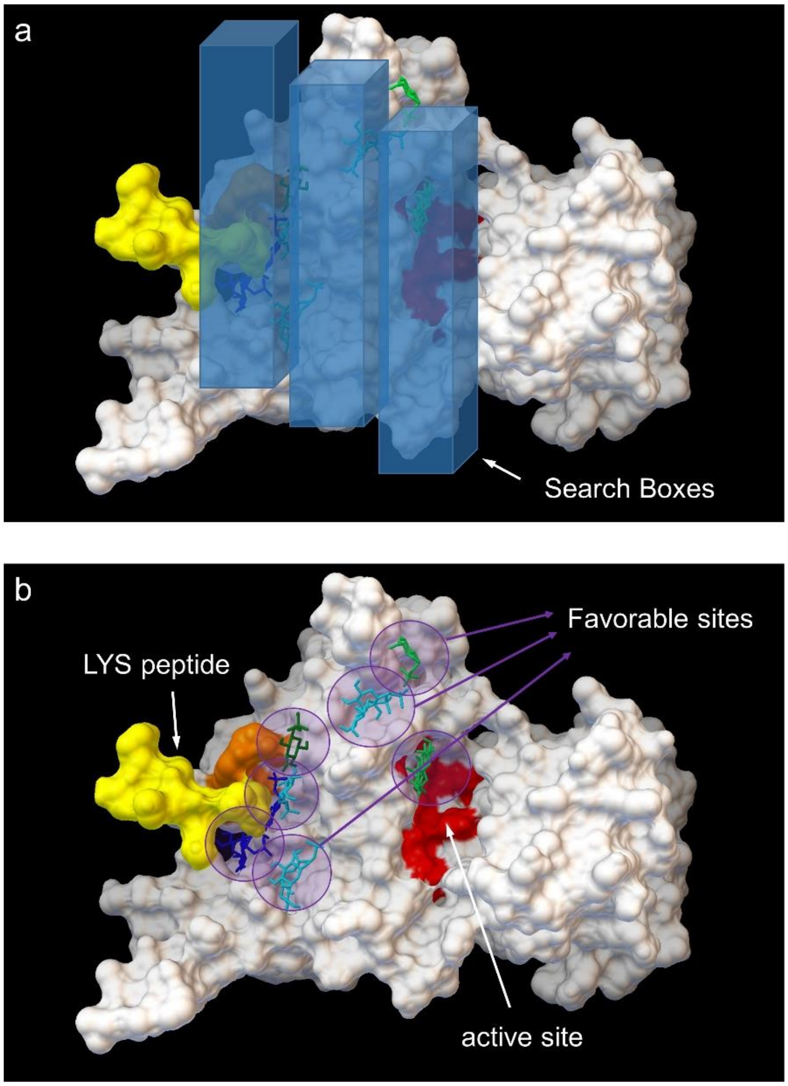


Figure 4-3 The results of molecular docking simulations for the binding of G6P to G6PDH (circled positions) in the area between LYS bridge and G6P binding pocket on G6PDH. (a) Search boxes. (b) Several favorable binding sites were seemed as potential kinetic trapping spots in short MD simulation and were avoid in initializing the parallel simulations.

4.2.4 Markov State Model on transition pathway on flexible 2D surface

Python package *MSMBuilder*[102] is used to conduct the MSM analysis in this work.

In order to analyze the trajectory of intermediate molecule on a flexible 2D surface, the atom coordinates in each 10^5-frame MD trajectory were featurized into a vector with representative coordinates, $X(t)$:

$$X(t) = [x_1 \quad x_2 \quad x_3] \qquad 4\text{-}1$$

where t is time step, x_1 is the G6P's distance to bridge surface, x_2 is the G6P's distance to complex surface and x_3 is the G6P's distance to the COM of its binding pocket on G6PDH. Python package *MDAnalysis*[103,104] is used for the trajectory featurization. The actual distance to a reference surface was calculated by taking the distance between the COM of intermediate molecule and the closest atoms of reference group. Depending on user's requirement and specific issues to be studied, the dimension of $X(t)$ can be increased by adding features of interest such as dihedral angles, distances to charged moieties, etc., to reveal more detail in channeling process.

After trajectory featurization, the original vectors, $X(t)$, was grouped into 500 clusters/states based on their conformational similarity. As a result, the trajectory of vectors is converted into a trajectory of 500 states, $S(t)$:

$$X(t) \xrightarrow{clustering} S(t) \qquad 4\text{-}2$$

Then, a Markov State Model, $C(\tau)$, is built by counting the transition between each state in $S(t)$:

$$S(t) \xrightarrow{MSM} C(\tau) \tag{4-3}$$

where τ is the time interval used to count the transition. Specifically, $C_{i,j}(\tau)$ represent the number of counts for the system to go from state i to state j during time interval, τ. The the count matrix is converted to a transition matrix according to the following equation:

$$M_{i,j}(\tau) = \frac{C_{i,j}(\tau)}{\sum_{j=1}^{j=j\_max} C_{i,j}(\tau)} \tag{4-4}$$

This makes each row of the transition matrix, $M_{i,j}(\tau)$, sum to one. Figure 4-4 shows an example of 500×500 transition matrix, where 4.25 % of the elements are non-zero. This means that the transition matrix is usually sparse, because mutual connections only exist among directly connected states.

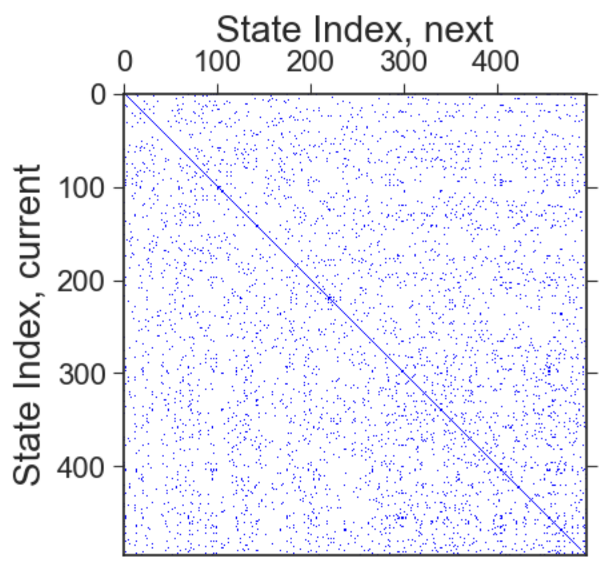


Figure 4-4 An example of transition matrix, where white color stands for zero element and blue color stands for non-zero items. In this case, there are 10,625 non-zero elements out of the total 250,000 items, resulting in a ~ 4.25% occupancy of the matrix.

Finally, three basins are defined to calculate the committor probability with Python package *CSNAnalysis*.[‡] The energy basins were defined as:

1. The *bridge basin*, including clusters where the G6P phosphate group lies within 0.3 nm of the peptide bridge surface.
2. The *pocket basin*, of clusters where the G6P phosphate group lies within 0.3 nm of the center of the G6PDH binding site.
3. The *desorption basin*, in which the G6P phosphate group is more than 2 nm away from the complex surface.

4.3 Results and Discussion

4.3.1 Channeling process

Unlike the LYS bridge with uniform potential wells (and thus precisely defined energy terms), the configuration of the peptide bridge with respect to G6PDH is flexible due to its single bond connection (Figure 4-1). Additionally, no single reference group can be defined along the multiple trajectories between the bridge and binding pocket; therefore, the leaking probability in this region cannot be easily assessed by umbrella sampling. Here, a direct probability measurement was conducted by MD to explore the leakage and estimate a corresponding energy barrier.

By combining equations 3-5 and 3-6, the rate constant ratio, $k_{\mathrm{des}}/k_{\mathrm{hop}}$, and probability ratio, $p_{\mathrm{des}}/p_{\mathrm{hop}}$, are equal and can be related to the energy difference, $G_{\mathrm{des}} - G_{\mathrm{hop}}$, according to:

[‡] *CSNAnalysis*: https://github.com/ADicksonLab/CSNAnalysis

$$\frac{k_{\text{des}}}{k_{\text{hop}}} = \frac{A \cdot exp(-G_{\text{des}}/RT)}{A \cdot exp\left(-G_{\text{hop}}/RT\right)} = exp\left(-\frac{G_{\text{des}} - G_{\text{hop}}}{RT}\right) = \frac{p_{\text{des}}}{p_{\text{hop}}} \qquad 4\text{-}5$$

In this chapter, the subscript "hop" in k_{hop} and G_{hop} refer to the hop across the energy barrier between last dual association site on peptide bridge and the intermediate's binding pocket on G6PDH.

The most direct approach to measuring $p_{\text{des}}/p_{\text{hop}}$ would be to follow the MD trajectory of the G6P molecule, starting from either the LYS bridge or G6PDH binding pocket, and track its probability to arrive at the other site. However, because of the deep potential well at each of those sites, such traversals are extremely rare on MD time scales. Alternatively, if the intermediate is initialized at a high-energy transition state (Figure 4-5 point c), it should have equal probability to reach the peptide bridge, p_{br}, and the G6PDH pocket, p_{poc}. There also exists a finite probability, p_{des}, of desorption to the bulk, depending on the energy difference between transition and desorption state, ΔG.

More generally, the G6P molecule may be released near the transition state (Figure 4-5 points b or d), where p_{br} and p_{poc} are not equal but comparable. Given point d as an example, $p_{\text{poc}}^{d} > p_{\text{br}}^{d}$ indicates the shift from transition points to enzyme pocket. Here, p_{br}^{d} as well as the desorption probability, p_{des}^{d}, can be related to their respective energy barriers, G_{hop}^{d} and G_{des}^{d}, by an Arrhenius expression:

$$\frac{p_{\text{des}}^{\text{d}}}{p_{\text{br}}^{\text{d}}} = \frac{A \cdot exp\left(-G_{\text{des}}^{\text{d}}/RT\right)}{A \cdot exp\left(-G_{\text{hop}}^{\text{d}}/RT\right)} = exp\left(-\frac{G_{\text{des}}^{\text{d}} - G_{\text{hop}}^{\text{d}}}{RT}\right) = exp\left(-\frac{\Delta G}{RT}\right) \qquad 4\text{-}6$$

This expression suggests generally that wherever the release point, the ratio of hopping and leaking probabilities is given by $\exp(-\Delta G/RT)$, where ΔG is the energy difference between a perfect transition point (point c in Figure 4-5) and the bulk energy level. Therefore, in order to calculate the value of ΔG, the G6P molecule can be released from a nearby region (Figure 4-5 points b or d) , instead of finding a single perfect transition state (point c).

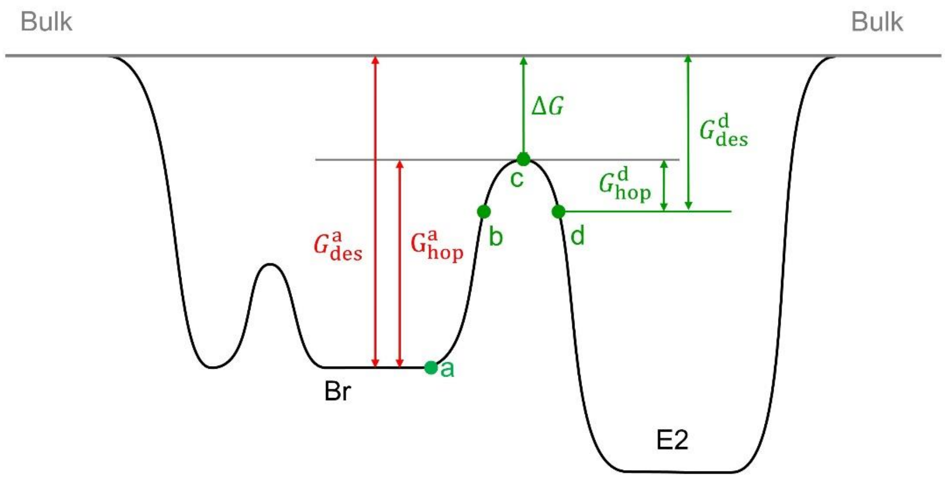


Figure 4-5 Simplified energy profile from bridge to G6P binding site on G6PDH. Point a is the last dual association site on peptide bridge; point c is the transition state at the energy barrier between bridge and G6PDH; point b and d are the two states slightly deviating from the transition state.

Probability analysis was conducted using a partial complex, including G6PDH (PDB: 5AQ1, Table 3), the LYS bridge and only one interfacial CYS residue of HK, as shown in Figure 4-6a. The HK CYS was position-restrained to mimic the existence of the HK biomolecule. Using the configuration of Figure 4-6a as the initial frame, the velocity of all atoms was regenerated and then equilibrated for 1 ns (position restraint on all G6P and complex atoms). After that, 500 parallel MD simulations (2 ns) were conducted with position restraint only on the HK CYS residue.

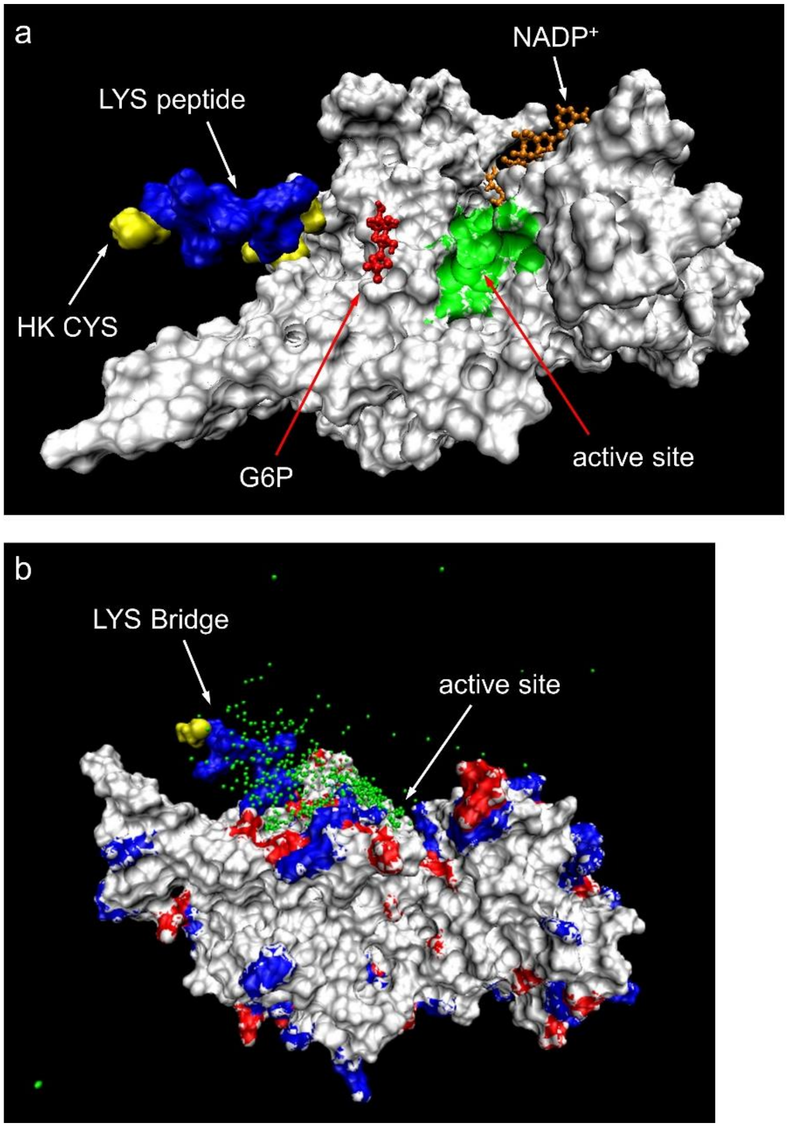


Figure 4-6 (a) System configuration when G6P was released around point c in Figure 4-5. (b) System configuration showing the distribution of G6P intermediate (green dots) in 500 parallel simulations. The complex structure is the 0 ns frame and the G6P molecules position are taken from the last frame of each parallel simulation. The coordinates are rotated to make the G6PDH secondary structure fit that of first frame.

Probability was calculated for three outcomes of G6P molecule: reaching the peptide bridge (p_{br}), reaching the G6PDH pocket (p_{poc}) or desorbing into the bulk (p_{des}). Each outcome was determined by the distance of the G6P's phosphate group to the corresponding residues. That is, when phosphate group was within 1.2 nm (short range cut-off) to the LYS bridge surface or the G6P binding pocket, it was assumed to reach the destination. If G6P molecule was 1.2 nm away from the whole complex, it was assumed to desorb. Cases where G6P was located elsewhere on the complex surface after 2 ns (~20% of all simulations) were considered incomplete channeling, and were not included in the calculation. Finally, the initial position of G6P was slightly adjusted until p_{br} is comparable to p_{poc}. Figure 4-7 shows that the probability became fairly consistent when the simulation was repeated more than 200 times. Therefore, all the probability results in the main text were calculated from 500 parallel simulations.

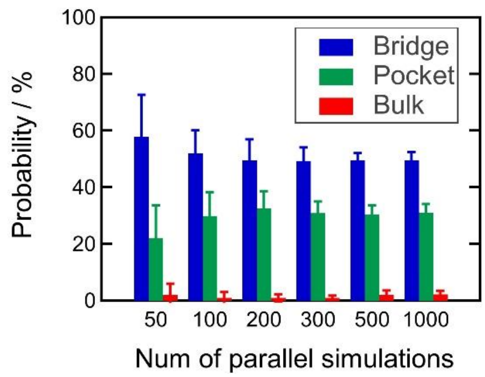


Figure 4-7 Probability results as a function of number of parallel simulations. The resulting probability depends on the intermediate releasing point. In this set of simulations, G6P molecule was released at a single point around the perfect transition state area, and then parallel simulations was conducted to study the convergence of probabilities.

Using the release point yielding comparable values of p_{br} and p_{poc}, a IS dependence study was conducted via 500 parallel simulations. For example, at IS=20 mM (Figure 4-8), p_{br}, p_{poc} and p_{des} were found to be ~53.5%, 43.4% and 3.0%, respectively, resulting in a leaking probability $\frac{p_{\mathrm{des}}}{p_{\mathrm{poc}}} = 7.14\%$. Using this value in equation 4-6, we calculated the resulting ΔG^{BrE2} value to be 6.9 kJ/mol. As compared to the bridge, where $\Delta G^{\mathrm{BrBr}} = G_{\mathrm{des}}^{\mathrm{BrBr}} - G_{\mathrm{hop}}^{\mathrm{BrBr}} \approx 13$ kJ/mol (Table 2), traversal from the bridge to G6PDH is less protected from leaking into the bulk.

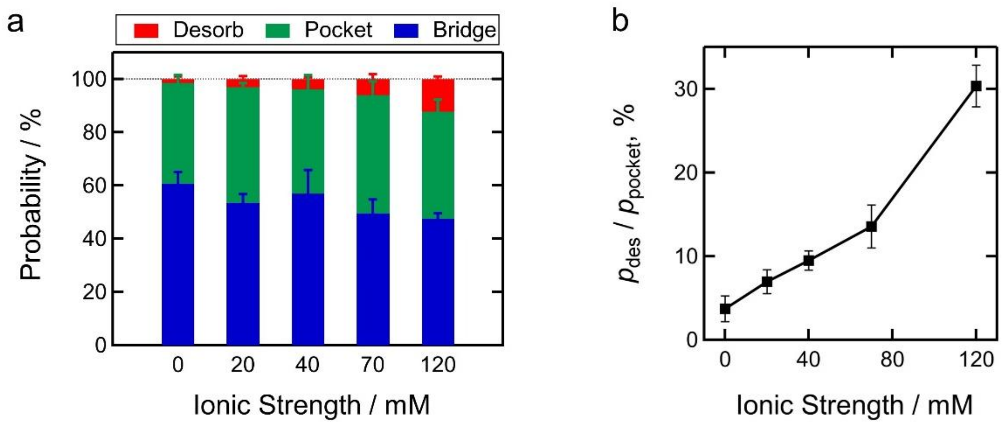


Figure 4-8 (a) Ionic strength dependence of leaking probability, showing the probability of G6P's final states (e.g, desorption, bridge, pocket). Up 500 parallel simulations were conducted to sample the rare leaking event. (b) The ration of desorption and channeling events in Fig.3b.

Specifically, the energy difference between transition state and desorption level was calculated from the resulted probability according to Figure 4-5, Figure 4-8 and equation:

$$\Delta G^{BrE2} = -RT\ln\left(\frac{p_{hop}^{BrE2}}{p_{des}^{BrE2}}\right) \qquad 4\text{-}7$$

Assuming a uniform bulk energy level, ΔG^{BrE2} can be used to correlate the transition state energy level to $G_{\mathrm{des}}^{\mathrm{BrBr}}$and $G_{\mathrm{hop}}^{\mathrm{BrBr}}$ on LYS bridge, as shown in equation below:

$$G_{des}^{BrBr} = G_{des}^{BrE2} \qquad 4\text{-}8$$

$$G_{hop}^{BrE2} = G_{des}^{BrBr} - \Delta G^{BrE2} \qquad 4\text{-}9$$

Therefore, a complete energy profile can be discerned between the LYS bridge to the transition state to G6PDH binding pocket. Figure 4-8b shows different p_{des}/p_{pocket} value as a function of ionic strength, indicating increased leakage under a more concentrated ionic environment. It should be noted that a high energy barrier tends to keep the G6P molecule on the bridge, increasing the residence time of intermediates on the bridge. Below we show that this energy barrier results in further leakage from the bridge.

As for the rate constants, since all hopping and desorption events on LYS bridge can be correlated to the last dual association site on peptide bridge, $k_{\mathrm{hop}}^{\mathrm{BrBr}}$, $k_{\mathrm{des}}^{\mathrm{BrBr}}$, $k_{\mathrm{hop}}^{\mathrm{BrE2}}$ and $k_{\mathrm{des}}^{\mathrm{BrE2}}$ are assumed to have the same frequency factor. Similarly to the KMC parameterization on peptide bridge (Table 2), rate constant ratio from bridge to E2, $k_{\mathrm{hop}}^{\mathrm{BrE2}}/k_{\mathrm{des}}^{\mathrm{BrE2}}$, was calculated from probability analysis.

4.3.2 Transition pathway analysis

Given the fact that channeling on 2D surface is complicated and there could be multiple pathways, Transition Path Theory is conducted via Markov state model to better understand the channeling process and also validate the probability analysis in above section. Figure 4-6b shows all G6P's final positions relative to bridge-G6PDH complex in 500 parallel simulations. In spite of one frame from each 200-frame trajectory, the area between peptide bridge and binding pocket was already fully covered, even including the desorption case. Therefore, the local ergodicity is good enough to build a Markov state model.

Figure 4-9 shows the network of MSM states for the 0 mM and 120 mM case in Figure 4-8a. Specifically, the node size corresponds to its stationary populations and the node color stands for its committor probability to one of the pre-defined basins. For example, the states belonging to "bridge basin" has a 100% committor probability to this basin and thus has a pure blue color. Similarly, the pure cyan nodes have an equivalent 50% probability to "pocket" and “bridge” basins. The MSM results indicate separate energy basins of peptide bridge and G6P binding site. But these two basins are connected via a highly interactive nodes that correspond to the transition area in Figure 4-5 and Figure 4-6a. It is obvious that neither of these two basins has a good connection to the bulk, because there is a large kinetic trap due to the charge density and strong local h-bond interaction between G6P’s phosphate group and LYS/ARG residues at “bridge” and “pocket”. But the transition area (cyan) shows a relative frequent interaction to desorption states. This agrees well with the leakage when G6P hops from peptide bridge to its binding site on G6PDH. At 0 mM ionic strength (Figure 4-9a), the “pocket” and “bridge” basins are strongly connected by the cyan states in between, and very little network was observed to the “bulk” basin. For the 120 mM system, however, significant shift can be observed to the bulk state, as indicated by the larger red area and more apparent purple/yellow areas. This give a good qualitative visualization on the impact of ionic strength.

In order to visualize the transition probability from bridge states to pocket or bulk states, the bridge was not defined as a basin when calculating the committor probabilities, as shown in Figure 4-10. In this way, the committor probabilities to “bulk” and “pocket” spread and compete for the original “bridge” states. Graphically, the blue color in Figure 4-9 a and b is removed in Figure 4-10, and the green and red spread to and paint the original blue states as circled by dashed oval shape. As a result, the original “bridge” states turned a very green color that is far from yellow, indicating that the intermediate on bridge was more like to go to the binding pocket (green) instead of being lost in to bulk media (red). However, when comparing the 0 mM and 120 mM ionic strength, more yellow can be observed in original bridge states in Figure 4-10a, demonstrating more leakage occurred with increasing ionic strength. This agrees well with direct probability analysis (Figure 4-8) and the three-basin committor probability result (Figure 4-9).

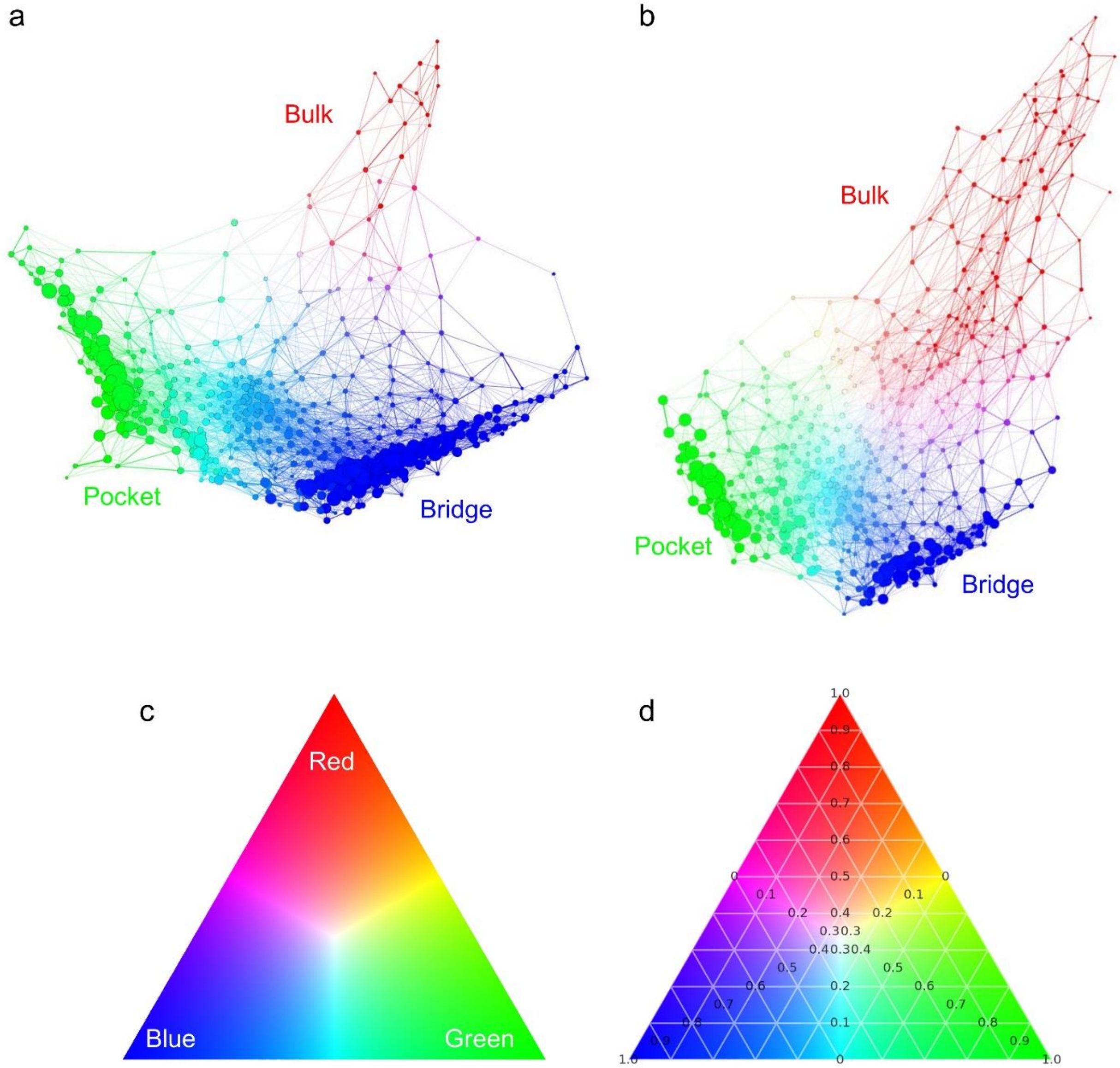


Figure 4-9 Transition pathway visualization at 0 mM (a) and 120 mM (b). The node size is based on the stationary populations of each state. The colors indicates the three-basin committor probabilities to the pre-defined basins of desorption states (red), peptide bridge states (blue) and binding pocket states (green). (c) Triangle color bar. (d) Triangle color bar with grid and probability labels.

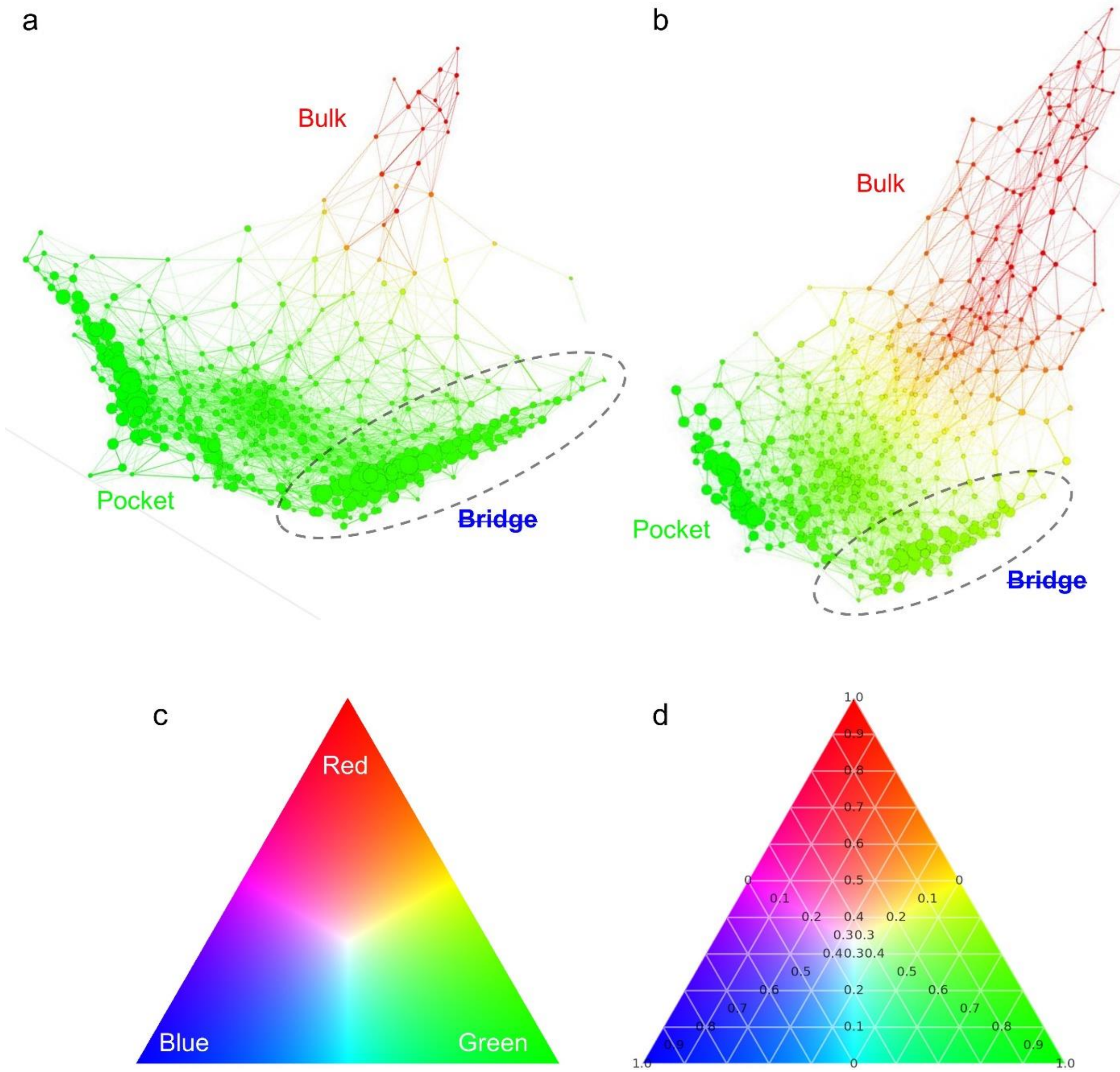


Figure 4-10 Two-basin committor probabilities for transition pathways visualization at 0 mM (a) and 120 mM (b). As compared to Figure 4-9, the “bridge” basin was not defined when calculating the committor probabilities, but the corresponding states were still recognized as an energy basin. In this way the probability to bulk/pocket states from “bridge” states can be visualized. (c) Triangle color bar. (d) Triangle color bar with grid and probability labels.

In order to quantitatively compare the direct probability analysis and MSM result, the overall leakage was calculated. By studying the states with equal committor probability to “bridge” and “basin” in Figure 4-9, their probability to “bulk” can be used to compare with the probability ratio in Figure 4-8b, making a more reasonable ΔG as illustrated in Figure 4-5.

Specifically, the states with equal committor probability to "bridge" and "pocket" ($p_{i,brg}$ and $p_{i,poc}$) were extracted, and then only the states representing surface interactions were kept, in which the distance between G6P's phosphate group and complex surface was less than 0.3 nm. Finally the overall desorption probability, $p2p$, was calculated by the summation of weighted desorption probability at each states:

$$p2p = \sum_{i} \frac{p_{i,des}}{p_{i,poc}} \cdot wt_i \qquad \text{4-10}$$

where wt_i is the weight of each selected state calculated from their stationary populations in MSM. Table 4 summarizes all energy values, key rate constant ratios and key probability ratios. With the ΔG^{BrE2} value from probability analysis and assuming a uniform bulk energy level (equation 4-8), the hopping energy barrier from bridge to G6PDH can be calculated From Eq. 4-9. Given the ionic strength at 120 mM as an example, the minimum p_{des}/p_{poc} for these MSM clusters are 30.3% (Table 4), comparable to the 31% in Figure 4-8b.

Table 4 Energy barriers, related rate constants and cascade kinetics.

		TST			TPT		
IS mM	$\boldsymbol{\lambda_{debye}}$ nm	$\boldsymbol{\frac{p_{hop}^{BrE2}}{p_{des}^{BrE2}}}$	$\boldsymbol{\Delta G^{BrE2}}$ kJ/mol	$\boldsymbol{G_{hop}^{BrE2}}$ kJ/mol	$\boldsymbol{\frac{p_{hop}^{BrE2}}{p_{des}^{BrE2}}}$	$\boldsymbol{\Delta G^{BrE2}}$ kJ/mol	$\boldsymbol{G_{hop}^{BrE2}}$ kJ/mol
0	9.8	27.0	8.5	17.0	16.6	7.24	18
20	2.2	14.4	6.9	18.0	8.7	5.56	19
40	1.6	10.5	6.1	17.7	11.2	6.22	18
70	1.2	7.4	5.2	16.5	6.1	4.66	17
120	0.9	3.3	3.1	16.0	3.3	3.26	16

IS: ionic strength

λ_{debye}: Debye length

p_{des}^{BrE2}: desorption probability for hopping from last dual association site to G6P binding pocket on G6PDH

ΔG^{BrE2}: energy difference between transition state level and desorption level, when hopping from last dual association site to G6P binding pocket on G6PDH

p_{hop}^{BrE2}, ΔG^{BrE2}: probability and energy barrier for hopping from last dual association site to G6P binding pocket on G6PDH

Figure 4-11 makes a graphical comparison between TST and TPT methods in terms of desorption probability and the energy difference between transition area energy and bulk energy level. From both plots, the TST and TPT results generally agree well, giving a good quantitative impact of ionic environment on the channeling from bridge to G6PDH binding site. Therefore, the probability analysis in above section is reasonable to parameterize KMC model to quantify the cascade kinetics.

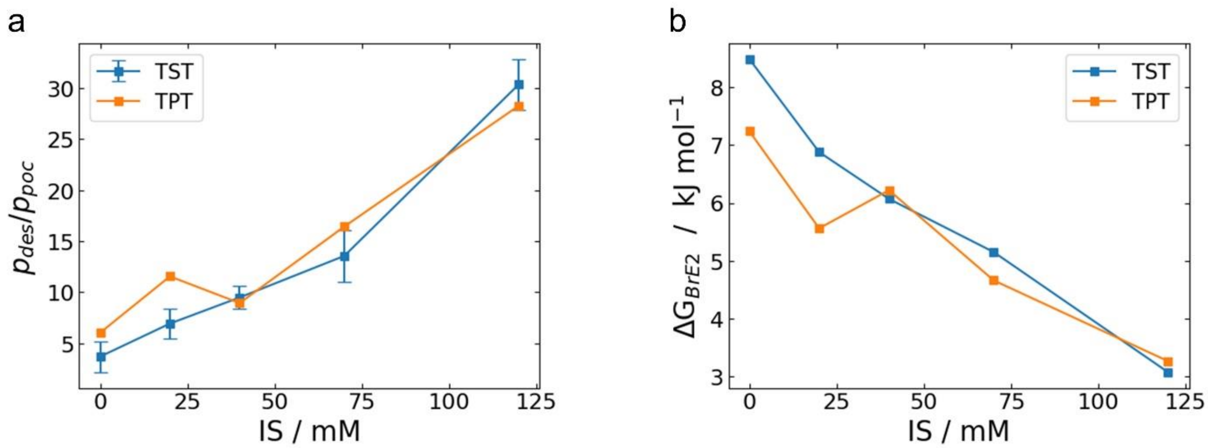


Figure 4-11 Comparison between the leakage calculated by transition state theory (TST) and transition path theory. (a) Leaking probability represented by desorption probability and the probability to reach the G6PDH binding pocket. (b) Energy difference between transition state to bulk environment, corresponding to ΔG in Figure 4-5.

4.4 Summary

Intermediate's hopping from poly-lysine peptide bridge to its binding site on G6PDH is studied by a direct probability measurement via MD simulation. By releasing the G6P molecules at a transition state area with relative high energy level, three event probabilities (p_{br}, p_{poc}, p_{des}) are used to calculate the energy difference between this state and bulk energy level. Then, this energy difference is correlated to the desorption energy on peptide bridge, in order to complete the energy profile to G6P's active site on G6PDH. Transition pathway analysis by MSM further elucidates the details of this process. Discrete potential basins are clarified by the visualization of MSM results, and the transitions states in between (cyan color) shows more propensity to desorb to the bulk as compared to arriving at the bridge and pocket basins. The leaking probability by committor probability agrees well with the direct probability analysis in Chapter 3. These probability results are used to estimate the kinetic parameters for the KMC model in Chapter 5.

4.5 Acknowledgements

The molecular docking in this chapter was conducted by Dr. Ivana Matanovic from Dr. Plamen Atanassov's research group at University of New Mexico. The work on Markov State Model was suggested by, Dr. Alex Dickson. Also, my advisor, Dr. Scott Calabrese Barton make great contribution to the python code on topology hybridization.

Chapter 5 Cascade Kinetics by Kinetic Monte Carlo Method*

5.1 Introduction

In previous chapters, thermodynamic parameters were quantified in detail by molecular dynamics simulation combined with advanced sampling methods. In order to further reveal the potential limitations of channeling efficiency, quantitative kinetics is greatly needed. A quantitative model should not only effectively cover the time and length scales from micro-structure to experiment, but also fully consider the molecular complexity of the cascade topology.

Reported work on this includes the analytical method with assumed channeling efficiency, continuum modeling focusing on the charge migration under electric field and simplified molecular simulation on the probability for intermediate molecule to reach the second active site. By solving a set of equations based on mass balance, analytical methods or micro-kinetics are able to correlate the observed cascade kinetics to channeling efficiency, which can not be measured directly by current experimental techniques, and are usually estimated by computational simulations.[34] Continuum modeling represents the intermediate distribution as a continuous field. With predetermined spatial factors and boundary conditions, the intermediate migration is governed by concentration gradients and the electric field created by the cascade surface.[15,35] As a result, steady state flux can be calculated to determine the yield as an indication of channeling efficiency. Recent molecular simulations were conducted under Brownian

---

* The content of this chapter has been published on *ACS Catalysis* as a full paper.

Y. Liu, I. Matanovic, D. P. Hickey, S. D. Minteer, P. Atanassov and S. Calabrese Barton, "Cascade Kinetics of an Artificial Metabolon by Molecular Dynamics and Kinetic Monte Carlo", 8, 7719-7726 (2018). doi: 10.1021/acscatal.8b01041

dynamics, wherein the transport efficiency was estimated by the probability of simplified but explicit intermediate molecules to reach the vicinity (0.7 nm) of second active site.[29,30,44] This probability term was integrated by the above-mentioned analytical approaches to quantify the cascade kinetics.

However, these models are not able to adequately represent the surface interaction between charged and polarized intermediate species and channeling surface. Additionally, the quantification algorithm relies heavily on the estimation of channeling efficiency and ignore the complexity of intermediate's interactions with cascade surface.

Because of its focus on sequences of events rather than stepping through time, Kinetics Monte Carlo (KMC) model has been widely used for simulating chemical reaction and diffusion on the lattice surface of inorganic catalysts.[41,59–63] Basic events include surface adsorption, desorption, hopping and reaction, which are common to intermediate channeling. Specifically, DFT parameterized KMC, also called first-principle KMC, is used to study the effect of catalyst composition, surface structure, lateral interactions, and operating conditions on the overall catalytic performance of chemical (*e.g.*, CO or NO) oxidation/reduction on catalyst lattice surface.[62] In addition, experimentally based KMC was used to study the CO oxidation on RuO_2 surface,[60] and a comparison between DFT and experimental parameters was made to reveal the significant impact of lateral interaction of adsorbed species.[41] Besides quantum mechanics, a combination of MD and KMC simulations was also used to study epitaxial growth of fcc and hcp islands on fcc (111) surface, showing great simulation acceleration and a boundary pinning effect by adsorbed atom islands that is difficult to reveal by conventional KMC methods.[64] Another multi-level example is a hybrid of first-principle MD and KMC simulations, showing the impact of rotational coupling of the side groups influencing the proton conduction in proton exchange membrane fuel cells (PEMFC).[65]

In this chapter, we use the KMC method to quantify the cascade kinetics of the HK-G6PDH cascade. Parameterized by MD and experimental results, the KMC model was built to estimate product evolution on the experimental time scale. This model enables a direct comparison between simulations and experiments, focusing on pre-steady state product evolution. These studies build a detailed quantitative approach, enabling us to further elucidate the range and limitations of electrostatic channeling.

5.2 Model Description

5.2.1 Kinetic Monte Carlo model

Figure 5-1 shows the two-step KMC model for cascade reactions:

$$\mathrm{glucose} + \mathrm{ATP} \xrightarrow{\mathrm{HK}} \mathrm{glucose\ 6} - \mathrm{phosphate} + \mathrm{ADP} \tag{5-1}$$

$$\mathrm{G6P} + \mathrm{NADP}^{+} \xrightarrow{\mathrm{G6PDH}} \mathrm{5\ phospho\ gluconolactone} + \mathrm{NADP} \tag{5-2}$$

As shown in the figure, the two active sites (E_1 and E_2) are connected by several discrete hopping sites. On each site, rate constants for all possible events (*e.g.*, $k_{\mathrm{hop}}^{\mathrm{left}}$, $k_{\mathrm{hop}}^{\mathrm{right}}$, k_{leak}, k_{ads} , k_{cat}) were assigned explicitly. All sites were allowed to exchange intermediate with the bulk environment. Given the fast diffusion rate of G6P (~10^{-5} cm^2 s^{-1}) as compared to the turnover frequency (TOF) of active site (~0.01-0.1 s^{-1}), the G6P was assumed to diffuse immediately into a homogeneous bulk media once it left the cascade surface. Therefore, the bulk environment was only represented by a changing intermediate concentration, which will in turn impact the adsorption rate on active sites. MD simulation of intermediate in the binding pocket is complex and usually handled by more detailed and advanced sampling techniques such as Markov State model[57]. Therefore, it is out of the scope of this work. To simplify the KMC system, hopping was assumed to be reversible between bridge sites, but was irreversible from E1 to bridge and bridge to E2. Reversible hopping on and off bridge is of potential interest for future work.

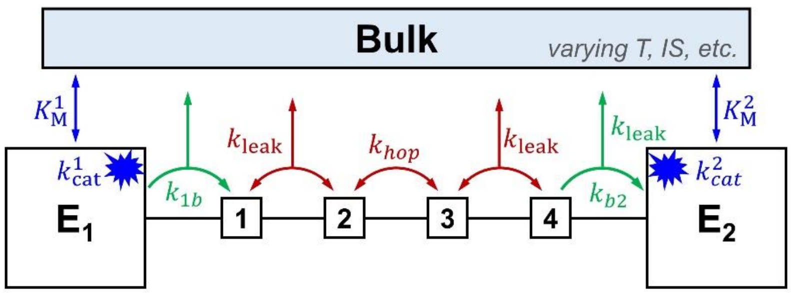


Figure 5-1 Schematic diagram of kinetic Monte Carlo model. E1 and E2 are two active sites. Sites 1-4 are discrete hopping sites representing the dual association sites on peptide bridge. Bulk are the environment with changing intermediate concentration. K_{M} is the Michaelis constant and k values are the rate constants for various events, including k_{cat} for turnover frequencies, k_{leak} for desorption from bridge, k_{hop} for hopping on bridge, k_{1b} for hopping from enzyme-1 to bridge, k_{b2} for hopping from bridge to enyme-2.

A zero k value was given if an event was disallowed. Figure 5-1 indicates all allowed events on each site. The actual rates, r, in each KMC step were calculated by taking the product of rate constant, k, and the occupancy of current and neighboring sites, Θ and Θ_{nb}, represented by either 1 or 0. As mentioned above, bulk concentration, c, was involved in this rate/occupancy product when calculating the adsorption rate. Each specific calculations can be seen as follows:

$$r_{\mathrm{ads}} = k_{\mathrm{ads}} \cdot (1 - \Theta) \cdot \mathrm{c} \qquad 5\text{-}3$$

$$r_{\mathrm{des}} = k_{\mathrm{des}} \cdot \Theta \cdot \mathrm{c} \tag{5-4}$$

$$r_{\mathrm{reaction}} = k_{\mathrm{cat}} \cdot \Theta \cdot \mathrm{c} \tag{5-5}$$

$$r_{\mathrm{hop}} = k_{\mathrm{hop}} \cdot \Theta \cdot (1 - \Theta_{nb}) \cdot \mathrm{c} \tag{5-6}$$

Having assigned all rate values (r_1, $r_2 \ldots r_n$) to all available sites, a random number (ρ_1) is generated between 0 and 1, to select a specific event according to:[36,60]

$$\sum_{i=1}^{i_0-1} r_i < \rho_1 \cdot \Gamma_{\mathrm{total}} \le \sum_{\mathrm{i}=1}^{\mathrm{i}_0} r_i \tag{5-7}$$

where Γ_{total} is the summation of all r values in current KMC step. The corresponding event i_0 was then executed and the time evolution, Δt, was calculated by:

$$\Delta t = -\ln(\rho_2)/\Gamma_{\mathrm{total}} \tag{5-8}$$

where ρ_2 is another random number. After execution, the occupancy and rate values were updated accordingly. Then, KMC simulation entered a loop until the time reached 1000 ns (steady state product evolution from E2). From the time course of product evolution, lag time, τ, was calculated by extrapolating the 500-1000s segment back to the time axis. Detailed discussion of this is provided in section 5.3.

In each KMC simulation, 100 parallel cascades were employed to enhance the event sampling and reduce the uncertainty of intermediate concentration. Normally, each KMC simulation took less than 10^7 steps, depending the parameters of KMC simulation. Finally, 5-10 parallel KMC simulations were performed to evaluate the error in lag time estimation. Table 5 shows the parameters for KMC simulation.

Table 5 KMC parameters.

Constant	Value
c_{E1E2}, cascade concentration / mol L^{-1}	$8*10^{-9}$
V, compartment volume / L	$2.07*10^{-14}$
c_{sub}, concentration of substrate for E1 / mol L^{-1}	2
$k_{\mathrm{cat}}^{\mathrm{E1}}$, TOF on E1 / molec s^{-1}	0.7
$k_{\mathrm{des}}^{\mathrm{E1}}$, desorption rate on E1, s^{-1}	0.07
$k_{\mathrm{ads}}^{\mathrm{E1}}$, adsorption rate on E1, s^{-1} M^{-1}	$7.7*10^{5}$
$K_{\mathrm{M,1}}$, Michaelis Constant of E1, mmol L^{-1}	10^{-2}
$k_{\mathrm{hop}}^{\mathrm{1b}}$, hopping rate from E1 to bridge, s^{-1}	$7*10^{15}$
$k_{\mathrm{cat}}^{\mathrm{E2}}$, TOF on E2 / molec s^{-1}	6.2
$k_{\mathrm{des}}^{\mathrm{E2}}$, desorption rate on E2, s^{-1}	0.62
$k_{\mathrm{ads}}^{\mathrm{E2}}$, adsorption rate on E2, s^{-1} M^{-1}	$1.26*10^{6}$
$K_{\mathrm{M,2}}$, Michaelis Constant of E2, mmol L^{-1}	$5.4*10^{-3}$
k_{hop}, hopping rate on bridge, s^{-1}	$k_{\mathrm{cat}}^{\mathrm{E1}} \times 100$
k_{leak}, desorption rate on bridge, s^{-1}	$k_{\mathrm{hop}} \,/\, (189, 146, 99, 43, 16)$
$k_{\mathrm{hop}}^{\mathrm{b2}}$, hopping rate from bridge to E2, s^{-1}	$k_{\mathrm{leak}} \times (27, 14.4, 10.5, 7.4, 3.3)$

Specifically, turnover frequencies (TOFs) on E1 ($k_{\mathrm{cat}}^{\mathrm{E1}}$) and E2 ($k_{\mathrm{cat}}^{\mathrm{E2}}$) was calculated by fitting experimental data for steady state product evolution of a fully saturated HK and

G6PDH.[33] The Michaelis constant for E2, $K_{\mathrm{M,2}}$, was calculated from the experiment lag time of a free-standing system, τ_{free}, according to following equation[105], where [I] is the intermediate concentration at steady state.

$$k_{\mathrm{cat}}^{\mathrm{E1}} = k_{\mathrm{cat}}^{\mathrm{E2}} \times \frac{[I]}{K_{\mathrm{M,2}} + [I]} = k_{\mathrm{cat}}^{\mathrm{E2}} \times \frac{k_{\mathrm{cat}}^{\mathrm{E1}} \cdot \tau_{\mathrm{free}}}{K_{\mathrm{M,2}} + k_{\mathrm{cat}}^{\mathrm{E1}} \cdot \tau_{\mathrm{free}}} \qquad \text{5-9}$$

Specifically, the $K_{\mathrm{M,2}}$ in Figure 5-1 is a combination of $k_{\mathrm{des}}^{\mathrm{E2}}$, $k_{\mathrm{ads}}^{\mathrm{E2}}$ and $k_{\mathrm{cat}}^{\mathrm{E2}}$, as show in following equation. In order to minimize the leakage of a readily channeled intermediate, substrate desorption rate constant on each active site, $k_{\mathrm{des}}^{\mathrm{E1}}$ and $k_{\mathrm{des}}^{\mathrm{E2}}$, was taken as 1/10 of the $k_{\mathrm{cat}}^{\mathrm{E1}}$ and $k_{\mathrm{cat}}^{\mathrm{E2}}$, respectively. The adsorption rate constants, $k_{\mathrm{ads}}^{\mathrm{E1}}$ and $k_{\mathrm{ads}}^{\mathrm{E2}}$, were calculated according to:

$$K_{\mathrm{M,2}} = \frac{k_{\mathrm{cat}}^{\mathrm{E2}} + k_{\mathrm{des}}^{\mathrm{E2}}}{k_{\mathrm{ads}}^{\mathrm{E2}}} \qquad \text{5-10}$$

As mentioned above, 100 parallel cascades were simulated within a single common bulk environment. The total volume, V, for these simulations was 21 fL, based on the experimental concentration of enzymatic cascade, [E1, E2], at 8.9 nM:

$$V = \frac{100/N_{\mathrm{A}}}{[\mathrm{E1,E2}]} \qquad \text{5-11}$$

As a result, the concentration of leaked and hopping intermediate can be calculated for each event. In this way, the degree of leakage and hopping could be compared directly with product and intermediate evolution. E1 was assumed to be saturated in glucose substrate.

Rate constants for channeling on the bridge and E2 were obtained from MD results. The hopping rate (~ ns^{-1}) on the peptide bridge is orders of magnitude higher than the TOF on each active site (~ s^{-1}), such that enzyme turnover is rate limiting. Therefore, the channeling process was actually governed by equilibrium ratios of transport rates, rather the absolute values of each transport rate constant. In the KMC simulation, k_{hop} was set at only two orders of magnitude higher than $k_{\mathrm{cat}}^{\mathrm{E1}}$, in order to improve simulation efficiency. Meanwhile, values of k_{leak} and $k_{\mathrm{hop}}^{\mathrm{b2}}$ were varied until their ratios to k_{hop} were equal to those of MD simulation results in Table 2 and Table 4. This guarantees the leaking probability for each individual KMC event is the same as the dynamic behavior observed in MD. $k_{\mathrm{hop}}^{\mathrm{1b}}$ was set to an infinite large value (7×10^{15}) to make sure the intermediate goes to bridge immediately after it is produced on E1. For the IS-dependent study, k_{leak} was varied until its ratio to k_{hop} was equal to that of MD results, as shown in Table 2, Table 4 and Table 5.

This Kinetic Monte Carlo model was built in Python[101,106], and it is available at the repository in APPENDIX D.

5.3 Results and Discussion

KMC result for singe enzyme kinetics agree perfectly with Michaelis-Menten kinetics and details are provided in APPENDIX E.

5.3.1 Cascade Kinetics by KMC model

KMC simulation was applied to the cascade model as shown in Figure 5-1. As discussed in section 1.3 and section 2.3.4, stop-flow lag time analysis is a widely-used experimental method to evaluate channeling efficiency.[7,33] Figure 5-2 shows an example of product evolution simulated by KMC that is comparable to a stop-flow experiment, and allows estimation of the lag time, τ. In this figure, four types of product evolution are presented, including non-channeling, perfect channeling and two leaking cases (K5+E2 and K15+E2). As indicated by the curve slope in Figure 5-2a, the reaction rate reaches a steady state value at the second half of the simulation. Additionally, the time course of bulk intermediate in Figure 5-2b gives a more direct visualization on such time period. That is, the intermediate concentration reaches a plateau level after 400 sec, when this concentration allows the second active site to reach a reaction rate equal to that of its upstream active site. Therefore, by extrapolating the 500-1000 ns curve Figure 5-2a to its time axis, the lag time can be calculated to compare with experiment results.

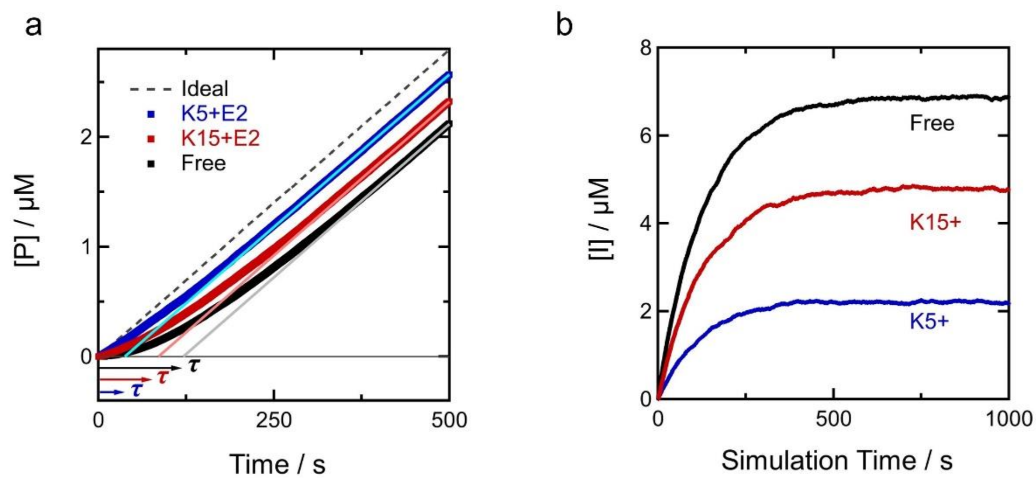


Figure 5-2 (a) A example of simulated stop-flow lag time analysis, showing the product evolution under different channeling conditions. (b) The evolution of bulk intermediate.

In previous work, an analytical expression was employed to estimate lag time based on channeling efficiency.[34] In contrast, the KMC model is capable of evaluating product evolution on experimental time scales starting from the pre-steady state, thereby allowing direct comparison with experimental results. Moreover, the contribution of any elementary step can be tracked over the entire simulation process, as exemplified by the bulk intermediate in Figure 5-2b.

5.3.2 *Quantitative comparison with experiment via ionic strength dependence*

Figure 5-4a demonstrates a comparison between experimental lag time, τ_{expt}, and KMC lag time, τ_{KMC}, as a function of ionic strength. Here, we assume 100% hopping efficiency from HK to the LYS bridge, and transport from the bridge to E2 is fast with no desorption allowed. Bridge lengths of 5 LYS residues (K5) and 15 LYS residues (K15) are represented both in experiment and simulation. These polypeptides offer 4 and 14 dual-association sites, respectively.

At low IS and short channeling distance (K5), both τ_{expt} and τ_{KMC} are much lower than that of free (uncoupled) enzyme. Based on a random walk model, the expected number of hops required to traverse the bridge is the square of the bridge site number, N. The leaking probability, P_{leak}, is the expected probability of desorption during traversal and can be expressed as:

$$P_{\mathrm{leak}} = 1 - \left(\frac{1}{1 + k_{\mathrm{des}}/\mathrm{k}_{\mathrm{hop}}}\right)^{N^2} \qquad \text{5-12}$$

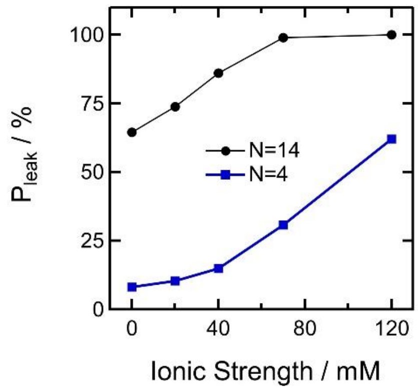


Figure 5-3 Plot of leaking probability based on equation 5-12.

As shown in Figure 5-3, P_{leak} was estimated to be 8.1% for K5 at IS=0, which means that electrostatic channeling is 92% successful under these conditions. At higher IS, the energy barrier to desorption decreases (Figure 3-6b), and P_{leak} becomes 62.1%, resulting in an increased lag time. Similarly, intermediate leakage increases for a longer channeling pathway. At the extreme case, for IS= 120 mM and N=14, desorption probability, P_{leak}, approaches 100% and both τ_{expt} and τ_{KMC} revert to values for the free-standing system.

The lag time for K15 (~55-120 sec) is consistently higher that of K5 (~ 20-60 sec). Taking advantage of KMC's capability to track all elemental steps, the time course of hopping and leakage is plotted for K5 and K15 in Figure 5-4b and c. With more hopping sites and thus a longer channeling pathway, the K15 system requires more hops that its K5 counterparts. As a result, a leakage increase was observed in Figure 5-4c. These agree well with the analytical result in equation 5-12. Therefore, the individual leaking probability and channeling distance must be minimized in an effective cascade.

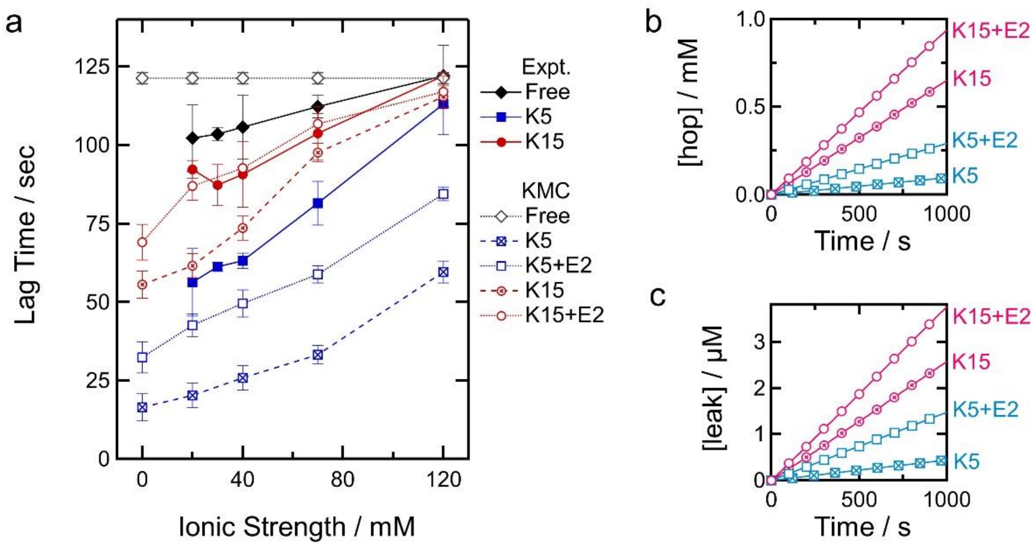


Figure 5-4 (a) Comparison of experimental lag time and KMC lag time. K5 and K15 only consider the leakage on LYS bridge. K5+E2 and K15+E2 involve the leakage from bridge to G6PDH pocket. (b) Time course of surface hopping on bridge. (c) Time course of leakage on bridge.

Based on MD simulations on the channeling from bridge to pocket, leakage between the bridge and G6PDH is included, which is depicted as K5+E2 and K15+E2 plots in Figure 5-4a. The result of leakage between the bridge and E2 is a decrease in channeling efficiency and an increase in lag time. With the existence of leakage from bridge and active site, τ_{KMC} increased to values comparable to experimental results for both bridge lengths. As discussed above, the relative high energy barrier between bridge and G6PDH pocket not only increases the downstream leakage at this site, but increases the residence time of G6P on LYS bridge, and therefore increasing the probability of desorption. As shown in Figure 5-4b, bridge hopping times for both K5+E2 and K15+E2 were higher than that of K5 and K15 over the entire simulation, which resulted in further leakage on the channeling pathway (Figure 5-4c).

The resulting IS dependence of lag time agrees well with experiment results, particularly in terms of the slope. Inclusion of hopping between the bridge and enzyme site in the KMC model generally increases the calculated lag times, bringing them closer to experimental values. Additionally, whereas experimental and KMC results for the K15+E2 case compare well, there is some discrepancy between experimental and KMC results for the K5+E2 case. This can be explained by additional leakage from HK to the LYS bridge (k_{1b} in Fig. 4a), which is not studied in this work. Such leakage is expected to affect lag times for the shorter K5 bridge more, because the longer K15 bridge itself exhibits greater leaking into the bulk.

5.3.3 Model extension to multistep cascade

In order to broaden the application of KMC, the present model can be extended to a three-step cascade:

$$\mathrm{S_1} \xrightarrow{\mathrm{E1}} \mathrm{I_1} \xrightarrow{\mathrm{E2}} \mathrm{I_2} \xrightarrow{\mathrm{E3}} \mathrm{P_3} \qquad \text{5-13}$$

where E1, E2 and E3 are three sequential active sites. Given reaction on E1 as an example, S_1 is the substrate for E1, and $\mathrm{I_1}$ is the product of E1 and also the substrate for E2. By assuming a same kinetics for the second site (E2) and third site (E3), the product evolution for the free standing system is shown in Figure 5-5a. KMC result show that the lag time of E3 is almost two times that of E2, because the two bulk intermediate species concentrations must be generated from E1. This model extension shows a potential of KMC model to study more complicated cascade systems.

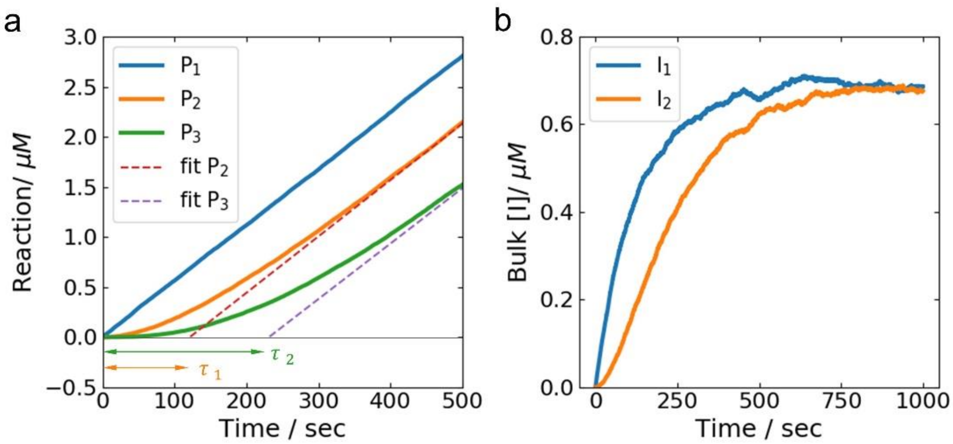


Figure 5-5 Product and intermediate evolution in three step KMC model (Free standing system). All parameters for bridge-2 and enzyme-3 are the same with that of bridge-1 and enzyme-2, respectively.

5.4 Summary

Using thermodynamic and kinetic parameters derived from molecular dynamics studies, the KMC model enables direct comparison with stop-flow lag time analysis, by evaluating the product evolution over the entire experimental time scale, particularly at pre-steady state. Moreover, it reveals several key parameters limiting overall cascade kinetics. Specifically, the lag time depends on the overall leakage of intermediate molecules, a result of joint action of the hopping and leaking probability on the bridge. A high IS environment tends to increase the desorption probability of each random hop, and a longer hopping pathway is found to dramatically decrease the likelihood that intermediates traverse the cascade surface. At given ionic strength, therefore, the length of channeling pathway and the strength of surface interaction should be balanced to achieve an efficient intermediate transport between sequential active sites.

Parameterized by these energy terms, a KMC model is utilized to evaluate the impact of the leakage between peptide bridge and active site. By including such leakage, the predicted KMC lag time matches better with experiment results. Detailed tracking of event evolution reveals that the energy barrier in this area not only resulted in leakage on the enzyme, but more likely pushes the intermediate back to the peptide bridge. This leads to longer retention time and thus more desorption on LYS bridge.

Unlike the well-defined energy terms and channeling behavior on the peptide bridge, the configuration of singly bonded peptide and G6PDH need to be further sampled by advanced techniques. In addition, the energy barrier between LYS bridge and G6PDH pocket further increases the likelihood of downstream leakage. This energy barrier should be carefully considered, because it not only caused leakage by itself but also exacerbated the leakage on channeling bridge. The leakage from HK to LYS bridge is also a possible factor to further bridge gap between simulation and experiment results.

The present modelling approach is applicable to the design of synthetic catalytic cascades, as well as natural cascades to better understand channeling mechanisms.

Chapter 6 Conclusion

This work builds up a multi-scale model to quantify the kinetics of artificial cascade, enabling a further understanding on the channeling mechanism and an indication of potential limitations for future cascade design. In this hierarchical model, molecular dynamics simulation enables a completed consideration of molecular complexity and kinetic Monte Carlo model covers a wide range of time and length scale, efficiently bridging the gap between microstructures and phenomenon kinetics.

Specifically, charged intermediate molecules were found to undergo a surface diffusion mechanism across cationic poly-Arginine and poly-Lysine peptide, which was realized by discrete hops between neighboring amino acid residues through hydrogen bonding interactions. Lys residues were found to provide a balance of intermediate adsorption and surface diffusivity that allow for efficient electrostatic channeling while preventing dissociation of the intermediate into the bulk. Additionally, simulations suggest that a dianionic intermediate is required for the double associative diffusion mechanism that prevents desorption from the peptide surface. Also, a balance between surface adsorption and mobility is required to achieve an optimal channeling efficiency. The comparison with stop-flow lag time analysis gave a strong support of the occurrence of artificially introduced electrostatic channeling.

Further molecular dynamics study demonstrates that the surface hopping is actually under the protection of the electric double layer created by the charged peptide surface. The hopping in Stern layer was less impacted by ionic screening, but the diffuse layer protection due to long-range electrostatic interaction could be shielded at high ionic strength. The leaking probability for each hop depends on the energy difference between the hopping transition state and desorption state. In addition, the energy barrier around the artificial interface between LYS bridge and G6PDH pocket further increases the likelihood of downstream leakage, making the channeling even unsafe upon concentrated ionic environment.

Using thermodynamic and kinetic parameters derived from molecular dynamics studies, the KMC model enables a direct comparison with experiment, by evaluating the product evolution over the entire experimental time scale, particularly from the pre-steady state. Moreover, it reveals the key parameters limiting overall cascade kinetics. Specifically, the number of hopping sites and strength of close-range interactions account for the leakage from the channeling bridge. The barrier between bridge and second enzyme should be carefully considered, because it not only caused leakage by itself but also exacerbated the leakage on channeling bridge.

The present modelling approach is applicable to the design of synthetic catalytic cascades, as well as natural cascades to better understand channeling mechanisms. Furthermore, for natural cascades with more stable dimer interfaces and channeling surfaces, such as TS-DHFR and MDH-CS, a complete energy profile may be mapped between active sites to reveal more precise mechanisms of the corresponding biological pathways.

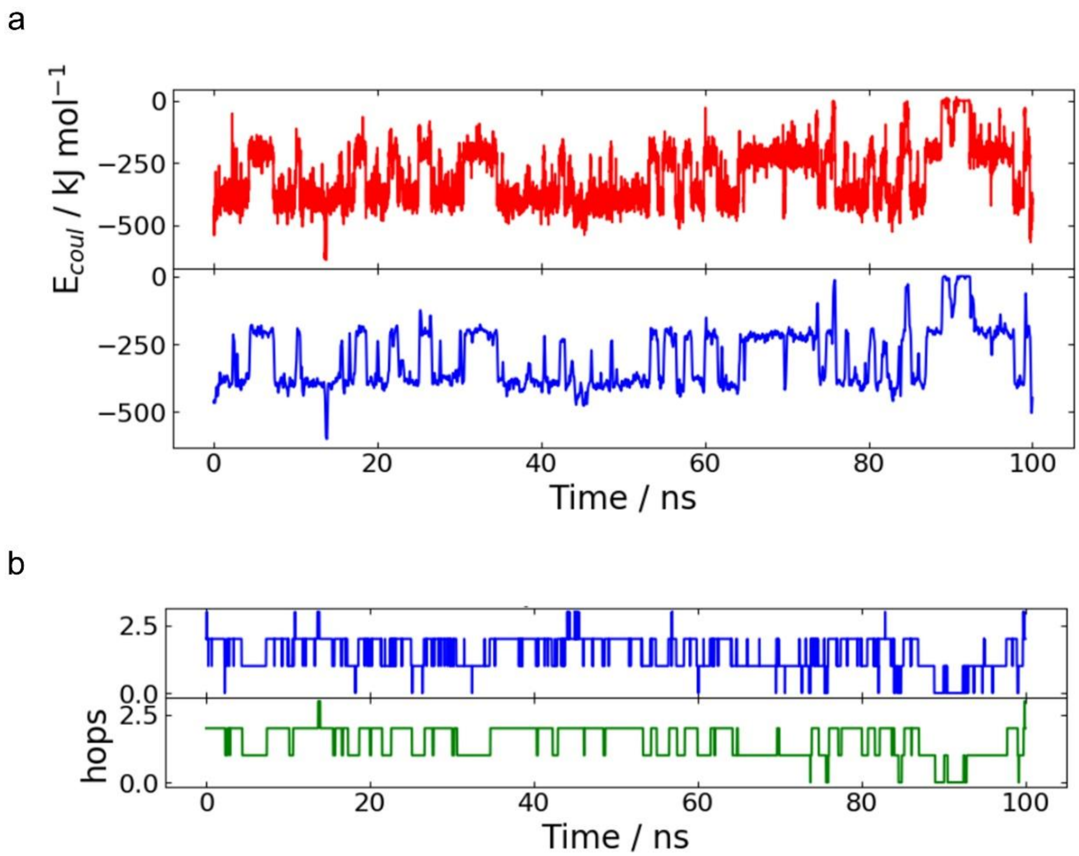
From a prospective point of view, kinetic quantification will continue to rely on seamless connection between computational techniques at different time and length scales that fully cover the molecular-level interactions and phenomenon kinetics. Continuum modeling and Kinetic Monte Carlo simulations are two powerful approaches capable of integrating such tasks. A hybridization of continuum and KMC models, either sequential or simultaneous, is promising to account for bulk concentration field and discrete surface hopping. If key meta-states can be further coarse-grained into major energy basins, Markov State models have a great potential to bridge the energy-discrete cascade surface to an energy-continuous long-range electrostatic region.

Molecular simulations are so far the most effective approach to quantify the parameters for KMC and continuum models, and also help to set up the geometry of these coarse-grained models. The key challenges of molecular simulations are the full representation of transition pathways, and dealing with kinetic traps in energy landscapes. Markov State models are still the best way to map the complex transition pathways from very elementary states to a human readable pattern. As for the kinetic traps, advanced sampling techniques, such as Umbrella Sampling[50] and Metadynamics[49,107], are required to assist the MD simulations for MSM purpose. For example, by the development and combination of various computational simulations, sampling techniques and even experimental crystal structures, it is possible to build a complete MSM that fully cover the transitions from perfect upstream binding states, to unbinding states, to channeling pathways, to downstream unbinding states and finally to perfect binding states at second active site. This promising pattern will allow us to further understand the channeling mechanisms and more precisely quantify cascade kinetics, both of which will enhance the design of synthetic cascades for various applications.

APPENDICES

APPENDICES

APPENDIX A. Method of counting hopping frequency



(a) An example of original coulombic energy diagram (red) and corresponding smoothed curve (blue). (b) Evolution of discrete energy levels for counting the number of hops. Blue curve was normalized from the original coulombic diagram and green curve was from smoothed coulombic diagram. 37 hops were counted from the green curve.

APPENDIX B. Various intermediates performance on LYS peptide surface

A summary of calculated diffusion parameters for all theoretical peptide chains and intermediates studied is provided in Table 6

Table 6 Properties of Oxalate intermediates channeled by Lys-Ala peptides.

	Peptide	N_{Ala}	R:R_{tot}	E_{ads} / kJ mol^{-1}	E_{coul} / kJ mol^{-1}	t_{ads} / %	D_{avg} / 10^{-6} cm^2 sec^{-1}	L_{jump} / nm	Γ_{avg} / ns^{-1}	Γ/ ns^{-1}	D_{surf} / 10^{-6} cm^2 sec^{-1}	t_{ads}*D_{surf}
Oxalate	K10	0	1:1	-8.91	-243	89.5	2.62	0.15	1.01	1.13	0.13	0.11
	K6-A1	1	1:2	-5.58	-142	71.2	5.16	0.60**	0.72	1.01	1.82	1.29
	K5-A2	2	1:3	-3.40	-97.3	50.7	8.59	0.45	0.37	0.73	0.74	0.38
	K4-A3	3	1:4	-4.87	-126	62.3	6.75	0.60	0.69	1.11	2.00	1.25
	K4-A6	6	1:7	-2.23	-47.0	30.7	11.96	1.05	0.16	0.50	2.77	0.90
	His	3	0	0.84	-0.51	0.92	15.97	--	--	--	0.00	0.00
G6P	K10	0	1:1	-8.59	-266	97.2	1.77	0.15	0.51	0.52	0.059	0.06
	K6-A1	1	1:2	-6.14	207	85.3	2.65	0.60**	0.36	0.42	0.764	0.65
	K5-A2	2	1:3	-5.66	-189	78.1	3.32	0.45	0.21	0.28	0.281	0.21
	K4-A3	3	1:4	-6.75	-195	84.4	3.01	0.60	0.37	0.44	0.784	0.67
	K4-A6	6	1:7	-5.36	-127	67.2	4.62	1.05	0.28	0.41	2.26	1.54
	His	3	0	0.84	-1.22	2.97	9.33	--	--	--	0.00	0.00
Glyoxylate	K10	0	1:1	-1.75	-44.5	31.0	15.30	--	--	--	--	--
	K6-A1	1	1:2	-0.22	-16.5	12.9	18.88	--	--	--	--	--
	K5-A2	2	1:3	-0.13	-14.4	10.4	19.90	--	--	--	--	--
	K4-A3	3	1:4	0.08	-12.3	9.31	19.79	--	--	--	--	--
	K4-A6	6	1:7	0.25	-5.86	4.81	22.56	--	--	--	--	--
	His	3	0	0.62	-1.06	1.58	22.23	--	--	--	--	--

Explanation of column headings: N_{ala}: number of Ala residues between two Lys. R/R_{total}: fraction of Lys, equal to $1/(1+N_{\mathrm{ala}})$. E_{ads}: adsorption energy calculated from radial distribution function (RDF). E_{coul}: average coulomb energy over entire simulation. t_{ads}: adsorption time fraction. D_{avg}: average diffusivity calculated by MSD based on MD trajectory. L_{hop}: hoping distance. Γ_{avg}: average hoping rate (*e.g.*, single-double-single association times) over a 50 ns simulation. Γ: hoping rate during adsorption, equal to $\Gamma_{\mathrm{avg}}/f_{\mathrm{ads}}$. D_{surf}: surface diffusivity calculated based on Γ and L_{hop}, assuming a 1D hopping diffusion mode. **special jumping distance due to α-helix structure.

APPENDIX C. Vector transformation and dimension reduction

Then, the dimension of the vector, $X(t)$, can be selectively transformed and reduced by independent component analysis (ICA)[108]. In order to better understand the vector transformation, we first introduce the principal component analysis (PCA).

μ_i is the time average of x_i as shown by equation:

$$\mu_i = \mathrm{E}(x_i)|_t \tag{6-1}$$

Then $\tilde{x}_i$ is used to denote the deviation of x_i from its time average, μ_i:

$$\tilde{x}_i = x_i - \mu_i \tag{6-2}$$

The covariance matrix, Σ, can be denoted as:

$$\Sigma_{ij} = \mathrm{cov}(x_i, x_j) = \mathrm{E}(\tilde{x}_i \cdot \tilde{x}_j) \tag{6-3}$$

The matrix version is expressed by the cross product of $\tilde{X}$ and its matrix transpose, $\tilde{X}^{\mathrm{T}}$:

$$\Sigma = \mathrm{E}(\tilde{X} \times \tilde{X}^{\mathrm{T}}) \tag{6-4}$$

After that, the relationship of the i^{th} eigenvalue, λ_i, and eigenvector, V_i, of covariance matrix can be expressed as follows:

$$\Sigma V_i = \lambda_i V_i \tag{6-5}$$

Then the matrix version of equation 6-5 can be expressed as follows, by assuming $d \times d$ matrix U as the matrix of all independent eigenvectors, V_i, and $d \times d$ matrix Λ as the diagonal matrix of all eigenvalues, λ_i.

$$\Sigma \cdot \mathrm{U} = \mathrm{U} \cdot \Lambda \tag{6-6}$$

$$\mathrm{U} = [V_1 \quad \dots \quad V_d] \tag{6-7}$$

$$\Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_d \end{bmatrix} \tag{6-8}$$

Finally, the transformation is conducted by projecting the original featurized vectors, $X(t)$, on the eigenvectors. The eigenvectors with higher eigenvalues have a larger variance of data projection. As for the dimension reduction, the eigenvector matrix U is ordered according to a descending order of egenvectors, λ_i. Then, a sub set of eigenvector matrix, $\mathrm{U_{sub}}$, is used instead as following equation.

$$Y(t) = X(t) \cdot \ \mathrm{U_{sub}} \tag{6-9}$$

PCA is able to find the projecting vectors with largest variance, and time-structure independent components analysis (tICA)[97,109,110] can be used to find the slowest-relaxing degree of freedom, which is more useful to analyze the system with relative slow surface hopping from bridge to enzyme pocket. Specifically, time-lagged covariance matrix $\mathrm{C}(\tau)$ is introduced as follows, where the covariance is taken with a time difference, τ, between the i, j entries of featurized vectors, $X(\mathrm{t})$

$$\mathrm{C}_{ij}(\tau) = \mathrm{E}\left(\tilde{x}_{i,t} \cdot \tilde{x}_{j,t+\tau}\right)|_{t=1 \dots N-\tau-1} \quad \text{6-10}$$

$$\mathrm{C}(\tau) = \mathrm{E}\left(\tilde{X}_t \times \tilde{X}_{t+\tau}{}^{\mathrm{T}}\right)|_t \quad \text{6-11}$$

Obviously, the $\mathrm{C}(\tau)$ is equal to Σ when lag time is zero.

$$\mathrm{C}(0) = \Sigma \quad \text{6-12}$$

Similarly to PCA method, the “eigenvector matrix” can be calculated by solving following generalized eigenfunction. Finally, the transformation and dimension reduction can be conducted by equation 6-9.

$$\mathrm{C}(\tau) \cdot \mathrm{U} = \mathrm{C}(0) \cdot \mathrm{U} \cdot \Lambda = \Sigma \cdot \mathrm{U} \cdot \Lambda \quad \text{6-13}$$

Collectively, when coding the above mentioned process or using *MSDbuilder*, the dimension of reduced freedom is required to select the $\mathrm{U}_{\mathrm{sub}}$ in equation 6-9. For tICA, additionally, a lag time, τ, usually in form of the steps of minimum time interval, is applied at the very beginning of the analysis.

APPENDIX D. Code repository

All Python codes for KMC models and data analysis can be accessed from this GitLab repository:

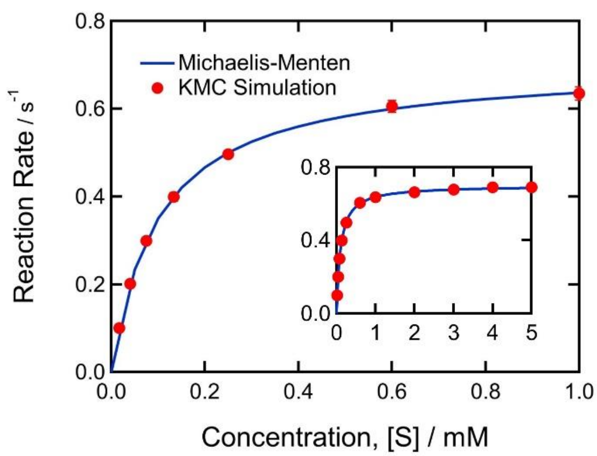
https://gitlab.msu.edu/scbgroup/yuanchao-dissertation-code.git

APPENDIX E. Mono enzyme kinetics

To integrate all possible events and corresponding rate constants, we employed Kinetic Monte Carlo for quantification of cascade kinetics. But firstly, KMC simulation was applied to reproduce Michaelis-Menten kinetics (equation 6-14),[111] which indicates a linear reaction rate increase within low substrate concentration region ($[S] \ll K_M$) and a constant rate value at high concentration region ($[S] \gg K_M$):

$$r = \frac{k_{cat} \cdot [S]}{K_M + [S]}; \quad K_M = \frac{k_1 + k_{cat}}{k_2} \tag{6-14}$$

where r is the TOF of single enzyme molecule, $[S]$ is substrate concentration, K_M is Michaelis constant, k_2 is the desorption rate from enzyme and k_1 is the adsorption rate onto enzyme. The figure below shows the dependence of reaction rate on substrate concentration, where the rate constants are taken from E1 (K_M=0.1 mM). The KMC results agrees well with conventional Michaelis equation, which proves KMC's feasibility on mono-enzyme kinetics.



KMC calculation Michaelis-Menten kinetics on uncoupled enzyme. Error bars represent standard deviation of 10 calculations.

REFERENCES

REFERENCES

1. C. A. Denard, J. F. Hartwig and H. Zhao, "Multistep One-Pot Reactions Combining Biocatalysts and Chemical Catalysts for Asymmetric Synthesis", *ACS Catal.*, **3**, 2856–2864 (2013). doi:10.1021/cs400633a.

2. Y. Wang, H. Ren and H. Zhao, "Expanding the boundary of biocatalysis: design and optimization of in vitro tandem catalytic reactions for biochemical production", *Crit. Rev. Biochem. Mol. Biol.*, **53**, 115–129 (2018). doi:10.1080/10409238.2018.1431201.

3. M. J. Moehlenbrock, T. K. Toby, L. N. Pelster and S. D. Minteer, "Metabolon Catalysts: An Efficient Model for Multi-enzyme Cascades at Electrode Surfaces", *ChemCatChem*, **3**, 561–570 (2011). doi:10.1002/cctc.201000384.

4. M. J. Moehlenbrock, M. T. Meredith and S. D. Minteer, "Bioelectrocatalytic Oxidation of Glucose in CNT Impregnated Hydrogels: Advantages of Synthetic Enzymatic Metabolon Formation", *ACS Catal.*, **2**, 17–25 (2012). doi:10.1021/cs200482v.

5. M. J. Moehlenbrock, T. K. Toby, A. Waheed and S. D. Minteer, "Metabolon Catalyzed Pyruvate/Air Biofuel Cell", *J. Am. Chem. Soc.*, **132**, 6288–6289 (2010). doi:10.1021/ja101326b.

6. J. M. Sperl and V. Sieber, "Multienzyme Cascade Reactions—Status and Recent Advances", *ACS Catal.*, **8**, 2385–2396 (2018). doi:10.1021/acscatal.7b03440.

7. I. Wheeldon, S. D. Minteer, S. Banta, S. C. Barton, P. Atanassov and M. Sigman, "Substrate channelling as an approach to cascade reactions", *Nat. Chem.*, **8**, 299–309 (2016). doi:10.1038/nchem.2459.

8. P. A. Srere, "The metabolon", *Trends Biochem. Sci.*, **10**, 109–110 (1985).

9. C. Vélot, M. B. Mixon, M. Teige and P. A. Srere, "Model of a quinary structure between krebs TCA cycle enzymes: A model for the metabolon", *Biochemistry*, **36**, 14271–14276 (1997). doi:10.1021/bi972011j.

10. M. K. Geck and J. F. Kirsch, "A Novel, Definitive Test for Substrate Channeling Illustrated with the Aspartate Aminotransferase/Malate Dehydrogenase System †", *Biochemistry*, **38**, 8032–8037 (1999). doi:10.1021/bi983029c.

11. C. Singleton, T. P. Howard and N. Smirnoff, "Synthetic metabolons for metabolic engineering", *J. Exp. Bot.*, **65**, 1947–1954 (2014). doi:10.1093/jxb/eru050.

12. J. Fu, M. Liu, Y. Liu, N. W. Woodbury and H. Yan, "Interenzyme Substrate Diffusion for an Enzyme Cascade Organized on Spatially Addressable DNA Nanostructures", *J. Am. Chem. Soc.*, **134**, 5516–5519 (2012). doi:10.1021/ja300897h.

13. P. Bauler, G. Huber, T. Leyh and J. A. McCammon, "Channeling by Proximity: The Catalytic Advantages of Active Site Colocalization Using Brownian Dynamics", *J. Phys. Chem. Lett.*, **1**, 1332–1335 (2010). doi:10.1021/jz1002007.

14. Y. Zhang, S. Tsitkov and H. Hess, "Proximity does not contribute to activity enhancement in the glucose oxidase–horseradish peroxidase cascade", *Nat. Commun.*, **7**, 13982 (2016). doi:10.1038/ncomms13982.

15. C. Eun, P. M. Kekenes-Huskey, V. T. Metzger and J. A. McCammon, "A model study of sequential enzyme reactions and electrostatic channeling", *J. Chem. Phys.*, **140**, 105101 (2014). doi:10.1063/1.4867286.

16. T. R. Schneider, E. Gerhardt, M. Lee, P.-H. Liang, K. S. Anderson and I. Schlichting, "Loop Closure and Intersubunit Communication in Tryptophan Synthase", *Biochemistry*, **37**, 5394–5406 (1998). doi:10.1021/bi9728957.

17. M. F. Dunn, V. Aguilar, P. Brzović, W. F. Drewe, K. F. Houben, C. a Leja and M. Roy, "The tryptophan synthase bienzyme complex transfers indole between the alpha- and beta-sites via a 25-30 A long tunnel.", *Biochemistry*, **29**, 8598–8607 (1990). doi:10.1021/bi00489a015.

18. F. Wu and S. Minteer, "Krebs Cycle Metabolon: Structural Evidence of Substrate Channeling Revealed by Cross-Linking and Mass Spectrometry", *Angew. Chemie Int. Ed.*, **54**, 1851–1854 (2015). doi:10.1002/anie.201409336.

19. Z. H. Zhou, D. B. McCarthy, C. M. O'Connor, L. J. Reed and J. K. Stoops, "The remarkable structural and functional organization of the eukaryotic pyruvate dehydrogenase complexes", *Proc. Natl. Acad. Sci.*, **98**, 14802–14807 (2001). doi:10.1073/pnas.011597698.

20. J. Müller and C. M. Niemeyer, "DNA-directed assembly of artificial multienzyme complexes", *Biochem. Biophys. Res. Commun.*, **377**, 62–67 (2008). doi:10.1016/j.bbrc.2008.09.078.

21. J. Fu, Y. R. Yang, A. Johnson-Buck, M. Liu, Y. Liu, N. G. Walter, N. W. Woodbury and H. Yan, "Multi-enzyme complexes on DNA scaffolds capable of substrate channelling with an artificial swinging arm", *Nat. Nanotechnol.*, **9**, 531–536 (2014). doi:10.1038/nnano.2014.100.

22. B. Bulutoglu, K. E. Garcia, F. Wu, S. D. Minteer and S. Banta, "Direct Evidence for Metabolon Formation and Substrate Channeling in Recombinant TCA Cycle Enzymes", *ACS Chem. Biol.*, **11**, 2847–2853 (2016). doi:10.1021/acschembio.6b00523.

23. J.-L. Lin and I. Wheeldon, "Kinetic Enhancements in DNA–Enzyme Nanostructures Mimic the Sabatier Principle", *ACS Catal.*, **3**, 560–564 (2013). doi:10.1021/cs300766d.

24. Y. Gao, C. C. Roberts, J. Zhu, J.-L. Lin, C. A. Chang and I. Wheeldon, "Tuning Enzyme Kinetics through Designed Intermolecular Interactions Far from the Active Site", *ACS*

Catal., **5**, 2149–2153 (2015). doi:10.1021/acscatal.5b00130.

25. R. Amaro and Z. Luthey-Schulten, "Molecular dynamics simulations of substrate channeling through an α–β barrel protein", *Chem. Phys.*, **307**, 147–155 (2004). doi:10.1016/j.chemphys.2004.05.019.

26. C. C. Roberts and C. A. Chang, "Modeling of Enhanced Catalysis in Multienzyme Nanostructures: Effect of Molecular Scaffolds, Spatial Organization, and Concentration", *J. Chem. Theory Comput.*, **11**, 286–292 (2015). doi:10.1021/ct5007482.

27. O. Idan and H. Hess, "Origins of Activity Enhancement in Enzyme Cascades on Scaffolds", *ACS Nano*, **7**, 8658–8665 (2013). doi:10.1021/nn402823k.

28. D. R. Knighton, C.-C. Kan, E. Howland, C. A. Janson, Z. Hostomska, K. M. Welsh and D. A. Matthews, "Structure of and kinetic channelling in bifunctional dihydrofolate reductase–thymidylate synthase", *Nat. Struct. Biol.*, **1**, 186–194 (1994). doi:10.1038/nsb0394-186.

29. A. H. Elcock, M. J. Potter, D. a Matthews, D. R. Knighton and J. A. McCammon, "Electrostatic Channeling in the Bifunctional Enzyme Dihydrofolate Reductase-thymidylate Synthase", *J. Mol. Biol.*, **262**, 370–374 (1996). doi:10.1006/jmbi.1996.0520.

30. N. Wang and J. A. McCammon, "Substrate channeling between the human dihydrofolate reductase and thymidylate synthase", *Protein Sci.*, **25**, 79–86 (2016). doi:10.1002/pro.2720.

31. C. Lindbladh, M. Rault, C. Hagglund, W. C. Small, K. Mosbach, L. Buelow, C. Evans and P. A. Srere, "Preparation and kinetic characterization of a fusion protein of yeast mitochondrial citrate synthase and malate dehydrogenase", *Biochemistry*, **33**, 11692–11698 (1994). doi:10.1021/bi00205a004.

32. M. Trujillo, R. G. K. Donald, D. S. Roos, P. J. Greene and D. V. Santi, "Heterologous Expression and Characterization of the Bifunctional Dihydrofolate Reductase−Thymidylate Synthase Enzyme of Toxoplasma gondii", *Biochemistry*, **35**, 6366–6374 (1996). doi:10.1021/bi952923q.

33. Y. Liu, D. P. Hickey, J.-Y. Guo, E. Earl, S. Abdellaoui, R. D. Milton, M. S. Sigman, S. D. Minteer and S. Calabrese Barton, "Substrate Channeling in an Artificial Metabolon: A Molecular Dynamics Blueprint for an Experimental Peptide Bridge", *ACS Catal.*, **7**, 2486–2493 (2017). doi:10.1021/acscatal.6b03440.

34. A. H. Elcock, G. A. Huber and J. A. McCammon, "Electrostatic Channeling of Substrates between Enzyme Active Sites: Comparison of Simulation and Experiment", *Biochemistry*, **36**, 16049–16058 (1997). doi:10.1021/bi971709u.

35. E. Earl and S. Calabrese Barton, "Simulation of intermediate transport in nanoscale scaffolds for multistep catalytic reactions", *Phys. Chem. Chem. Phys.*, **19**, 15463–15470 (2017). doi:10.1039/C7CP00239D.

36. K. Reuter, First-Principles Kinetic Monte Carlo Simulations for Heterogeneous Catalysis: Concepts, Status, and Frontiers, in *Model. Simul. Heterog. Catal. React.* Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany (2011), pp. 71–111.

37. M. Karplus and J. A. McCammon, "Molecular dynamics simulations of biomolecules", *Nat. Struct. Biol.*, **9**, 646–652 (2002). doi:10.1038/nsb0902-646.

38. V. P. Raut, M. A. Agashe, S. J. Stuart and R. A. Latour, "Molecular Dynamics Simulations of Peptide−Surface Interactions", *Langmuir*, **21**, 1629–1639 (2005). doi:10.1021/la047807f.

39. H. D. Herce and A. E. Garcia, "Molecular dynamics simulations suggest a mechanism for translocation of the HIV-1 TAT peptide across lipid membranes", *Proc. Natl. Acad. Sci.*, **104**, 20805–20810 (2007). doi:10.1073/pnas.0706574105.

40. L. Li, I. Vorobyov and T. W. Allen, "The different interactions of lysine and arginine side chains with lipid membranes", *J. Phys. Chem. B*, **117**, 11906–11920 (2013). doi:10.1021/jp405418y.

41. F. Hess and H. Over, "Kinetic Monte Carlo simulations of heterogeneously catalyzed oxidation reactions", *Catal. Sci. Technol.*, **4**, 583 (2014). doi:10.1039/c3cy00833a.

42. M. Stan, J. C. Ramirez, P. Cristea, S. Y. Hu, C. Deo, B. P. Uberuaga, S. Srivilliputhur, S. P. Rudin and J. M. Wills, "Models and simulations of nuclear fuel materials properties", *J. Alloys Compd.*, **444**, 415–423 (2007).

43. A. C. T. van Duin, S. Dasgupta, F. Lorant and W. A. Goddard, "ReaxFF: A Reactive Force Field for Hydrocarbons", *J. Phys. Chem. A*, **105**, 9396–9409 (2001). doi:10.1021/jp004368u.

44. Y. M. Huang, G. A. Huber, N. Wang, S. D. Minteer and J. A. McCammon, "Brownian dynamic study of an enzyme metabolon in the TCA cycle: Substrate kinetics and channeling", *Protein Sci.*, **27**, 463–471 (2018). doi:10.1002/pro.3338.

45. R. C. Bernardi, M. C. R. Melo and K. Schulten, "Enhanced sampling techniques in molecular dynamics simulations of biological systems", *Biochim. Biophys. Acta - Gen. Subj.*, **1850**, 872–877 (2015). doi:10.1016/J.BBAGEN.2014.10.019.

46. Y. Sugita and Y. Okamoto, "Replica-exchange molecular dynamics method for protein folding", *Chem. Phys. Lett.*, **314**, 141–151 (1999). doi:10.1016/S0009-2614(99)01123-9.

47. D. J. Earl and M. W. Deem, "Parallel tempering: Theory, applications, and new perspectives", *Phys. Chem. Chem. Phys.*, **7**, 3910 (2005). doi:10.1039/b509983h.

48. A. Laio and M. Parrinello, "Escaping free-energy minima.", *Proc. Natl. Acad. Sci. U. S. A.*, **99**, 12562–6 (2002). doi:10.1073/pnas.202427399.

49. A. Barducci, G. Bussi and M. Parrinello, "Well-Tempered Metadynamics: A Smoothly

Converging and Tunable Free-Energy Method", *Phys. Rev. Lett.*, **100**, 020603 (2008). doi:10.1103/PhysRevLett.100.020603.

50. J. Kästner, "Umbrella sampling", *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, **1**, 932–942 (2011). doi:10.1002/wcms.66.

51. V. S. Pande, K. Beauchamp and G. R. Bowman, "Everything you wanted to know about Markov State Models but were afraid to ask", *Methods*, **52**, 99–105 (2010). doi:10.1016/j.ymeth.2010.06.002.

52. G. R. Bowman, An Overview and Practical Guide to Building Markov State Models, in Springer, Dordrecht (2014), pp. 7–22.

53. A. C. Pan and B. Roux, "Building Markov state models along pathways to determine free energies and rates of transitions", *J. Chem. Phys.*, **129**, 064107 (2008). doi:10.1063/1.2959573.

54. J.-H. Prinz, H. Wu, M. Sarich, B. Keller, M. Senne, M. Held, J. D. Chodera, C. Schütte and F. Noé, "Markov models of molecular kinetics: Generation and validation", *J. Chem. Phys.*, **134**, 174105 (2011). doi:10.1063/1.3565032.

55. M. Biswas, B. Lickert and G. Stock, "Metadynamics Enhanced Markov Modeling of Protein Dynamics", *J. Phys. Chem. B*, **122**, 28 (2018). doi:10.1021/acs.jpcb.7b11800.

56. A. Dickson and C. L. Brooks, "Native States of Fast-Folding Proteins Are Kinetic Traps", *J. Am. Chem. Soc.*, **135**, 4729–4734 (2013). doi:10.1021/ja311077u.

57. S. D. Lotz and A. Dickson, "Unbiased Molecular Dynamics of 11 min Timescale Drug Unbinding Reveals Transition State Stabilizing Interactions", *J. Am. Chem. Soc.*, **140**, 618–628 (2018). doi:10.1021/jacs.7b08572.

58. D. D. Vvedensky, "Multiscale modelling of nanostructures", *J. Phys. Condens. Matter*, **16**, R1537–R1576 (2004). doi:10.1088/0953-8984/16/50/R01.

59. S. Völkening and J. Wintterlin, "CO oxidation on Pt(111)—Scanning tunneling microscopy experiments and Monte Carlo simulations", *J. Chem. Phys.*, **114**, 6382 (2001). doi:10.1063/1.1343836.

60. A. Farkas, F. Hess and H. Over, "Experiment-Based Kinetic Monte Carlo Simulations: CO Oxidation over RuO 2 (110)", *J. Phys. Chem. C*, **116**, 581–591 (2012). doi:10.1021/jp204703p.

61. A. Chatterjee and D. G. Vlachos, "An overview of spatial microscopic and accelerated kinetic Monte Carlo methods", *J. Comput. Mater. Des.*, **14**, 253–308 (2007). doi:10.1007/s10820-006-9042-9.

62. M. Stamatakis and D. G. Vlachos, "Unraveling the Complexity of Catalytic Reactions via Kinetic Monte Carlo Simulation: Current Status and Frontiers", *ACS Catal.*, **2**, 2648–2663

(2012). doi:10.1021/cs3005709.

63. D.-J. Liu, A. Garcia, J. Wang, D. M. Ackerman, C.-J. Wang and J. W. Evans, "Kinetic Monte Carlo Simulation of Statistical Mechanical Models and Coarse-Grained Mesoscale Descriptions of Catalytic Reaction–Diffusion Processes: 1D Nanoporous and 2D Surface Systems", *Chem. Rev.*, **115**, 5979–6050 (2015). doi:10.1021/cr500453t.

64. P. Zoontjens, T. P. Schulze and S. C. Hendy, "A Hybrid Kinetic Monte Carlo-Molecular Dynamics Method for Modeling Epitaxial Growth", 1–6 (2007).

65. G. Kabbe, C. Wehmeyer and D. Sebastiani, "A Coupled Molecular Dynamics/Kinetic Monte Carlo Approach for Protonation Dynamics in Extended Systems", *J. Chem. Theory Comput.*, **10**, 4221–4228 (2014). doi:10.1021/ct500482k.

66. H. J. C. Berendsen, D. van der Spoel and R. van Drunen, "GROMACS: A message-passing parallel molecular dynamics implementation", *Comput. Phys. Commun.*, **91**, 43–56 (1995). doi:10.1016/0010-4655(95)00042-E.

67. E. Lindahl, B. Hess and D. van der Spoel, "GROMACS 3.0: a package for molecular simulation and trajectory analysis", *J. Mol. Model.*, **7**, 306–317 (2001). doi:10.1007/s008940100045.

68. B. Hess, C. Kutzner, D. van der Spoel and E. Lindahl, "GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation", *J. Chem. Theory Comput.*, **4**, 435–447 (2008). doi:10.1021/ct700301q.

69. S. Pronk, S. Pall, R. Schulz, P. Larsson, P. Bjelkmar, R. Apostolov, M. R. Shirts, J. C. Smith, P. M. Kasson, D. van der Spoel, B. Hess and E. Lindahl, "GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit", *Bioinformatics*, **29**, 845–854 (2013). doi:10.1093/bioinformatics/btt055.

70. M. J. Abraham, T. Murtola, R. Schulz, S. Páll, J. C. Smith, B. Hess and E. Lindahl, "GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers", *SoftwareX*, **1**–**2**, 19–25 (2015). doi:10.1016/j.softx.2015.06.001.

71. D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark and H. J. C. Berendsen, "GROMACS: Fast, flexible, and free", *J. Comput. Chem.*, **26**, 1701–1718 (2005). doi:10.1002/jcc.20291.

72. S. Páll, M. J. Abraham, C. Kutzner, B. Hess and E. Lindahl, Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS, in *Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* (2015), pp. 3–27.

73. B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan and M. Karplus, "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations", *J. Comput. Chem.*, **4**, 187–217 (1983).

doi:10.1002/jcc.540040211.

74. R. B. Best, X. Zhu, J. Shim, P. E. M. Lopes, J. Mittal, M. Feig and A. D. MacKerell, "Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone ϕ, ψ and Side-Chain χ 1 and χ 2 Dihedral Angles", *J. Chem. Theory Comput.*, **8**, 3257–3273 (2012). doi:10.1021/ct300400x.

75. K. Vanommeslaeghe, E. Hatcher, C. Acharya, S. Kundu, S. Zhong, J. Shim, E. Darian, O. Guvench, P. Lopes, I. Vorobyov and A. D. Mackerell, "CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields", *J. Comput. Chem.*, **31**, 471–690 (2009). doi:10.1002/jcc.21367.

76. G. Bussi, D. Donadio and M. Parrinello, "Canonical sampling through velocity rescaling", *J. Chem. Phys.*, **126**, 014101 (2007). doi:10.1063/1.2408420.

77. M. Parrinello and A. Rahman, "Polymorphic transitions in single crystals: A new molecular dynamics method", *J. Appl. Phys.*, **52**, 7182–7190 (1981). doi:10.1063/1.328693.

78. C. L. Wennberg, T. Murtola, B. Hess and E. Lindahl, "Lennard-Jones Lattice Summation in Bilayer Simulations Has Critical Effects on Surface Tension and Lipid Properties", *J. Chem. Theory Comput.*, **9**, 3527–3537 (2013). doi:10.1021/ct400140n.

79. C. L. Wennberg, T. Murtola, S. Páll, M. J. Abraham, B. Hess and E. Lindahl, "Direct-Space Corrections Enable Fast and Accurate Lorentz–Berthelot Combination Rule Lennard-Jones Lattice Summation", *J. Chem. Theory Comput.*, **11**, 5737–5746 (2015). doi:10.1021/acs.jctc.5b00726.

80. S. Páll and B. Hess, "A flexible algorithm for calculating pair interactions on SIMD architectures", *Comput. Phys. Commun.*, **184**, 2641–2650 (2013). doi:10.1016/j.cpc.2013.06.003.

81. C. Kutzner, S. Páll, M. Fechner, A. Esztermann, B. L. de Groot and H. Grubmüller, "Best bang for your buck: GPU nodes for GROMACS biomolecular simulations", *J. Comput. Chem.*, **36**, 1990–2008 (2015). doi:10.1002/jcc.24030.

82. M. D. Hanwell, D. E. Curtis, D. C. Lonie, T. Vandermeersch, E. Zurek and G. R. Hutchison, "Avogadro: an advanced semantic chemical editor, visualization, and analysis platform", *J. Cheminform.*, **4**, 17 (2012). doi:10.1186/1758-2946-4-17.

83. T. J. Dolinsky, J. E. Nielsen, J. A. McCammon and N. A. Baker, "PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations", *Nucleic Acids Res.*, **32**, W665–W667 (2004). doi:10.1093/nar/gkh381.

84. T. J. Dolinsky, P. Czodrowski, H. Li, J. E. Nielsen, J. H. Jensen, G. Klebe and N. A. Baker, "PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations", *Nucleic Acids Res.*, **35**, W522–W525 (2007). doi:10.1093/nar/gkm276.

85. N. A. Baker, D. Sept, S. Joseph, M. J. Holst and J. A. McCammon, "Electrostatics of nanosystems: Application to microtubules and the ribosome", *Proc. Natl. Acad. Sci.*, **98**, 10037–10041 (2001). doi:10.1073/pnas.181342398.

86. W. Humphrey, A. Dalke and K. Schulten, "VMD: Visual molecular dynamics", *J. Mol. Graph.*, **14**, 33–38 (1996). doi:10.1016/0263-7855(96)00018-5.

87. J. J. Irwin, T. Sterling, M. M. Mysinger, E. S. Bolstad and R. G. Coleman, "ZINC: A Free Tool to Discover Chemistry for Biology", *J. Chem. Inf. Model.*, **52**, 1757–1768 (2012). doi:10.1021/ci3001277.

88. J. J. Irwin and B. K. Shoichet, "ZINC--a free database of commercially available compounds for virtual screening.", *J. Chem. Inf. Model.*, **45**, 177–82 (2005). doi:10.1021/ci049714+.

89. R. J. Ferreira, M.-J. U. Ferreira and D. J. V. A. dos Santos, "Do adsorbed drugs onto P-glycoprotein influence its efflux capability?", *Phys. Chem. Chem. Phys.*, **17**, 22023–22034 (2015). doi:10.1039/C5CP03216D.

90. R. Rowland and R. Taylor, "Intermolecular nonbonded contact distances in organic crystal structures: Comparison with distances expected from van der Waals radii", *J. Phys. Chem.*, (1996).

91. B. Roux, "The calculation of the potential of mean force using computer simulations", *Comput. Phys. Commun.*, **91**, 275–282 (1995). doi:10.1016/0010-4655(95)00053-I.

92. M. Souaille and B. Roux, "Extension to the weighted histogram analysis method: combining umbrella sampling with free energy calculations", *Comput. Phys. Commun.*, **135**, 40–57 (2001). doi:10.1016/S0010-4655(00)00215-0.

93. S. Kumar, J. M. Rosenberg, D. Bouzida, R. H. Swendsen and P. A. Kollman, "THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method", *J. Comput. Chem.*, **13**, 1011–1021 (1992). doi:10.1002/jcc.540130812.

94. A. Grossfield, WHAM: the weighted histogram analysis method. Version 2.0.9, n.d. . Available at: http://membrane.urmc.rochester.edu/content/wham. [Accessed: 17-Aug-2017].

95. O. Stern, "The theory of the electrolytic double-layer", *Z. Elektrochem*, **30**, 1014–1020 (1924).

96. G. Pérez-Hernández, F. Paul, T. Giorgino, G. De Fabritiis and F. Noé, "Identification of slow molecular order parameters for Markov model construction", *J. Chem. Phys.*, **139**, 015102 (2013). doi:10.1063/1.4811489.

97. C. R. Schwantes and V. S. Pande, "Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9", *J. Chem. Theory Comput.*, **9**, 2000–2009 (2013). doi:10.1021/ct300878a.

98. L. J. Lapidus, S. Acharya, C. R. Schwantes, L. Wu, D. Shukla, M. King, S. J. DeCamp and V. S. Pande, "Complex Pathways in Folding of Protein G Explored by Simulation and Experiment", *Biophys. J.*, **107**, 947–955 (2014). doi:10.1016/J.BPJ.2014.06.037.

99. A. Dickson and C. L. Brooks, "Quantifying Hub-like Behavior in Protein Folding Networks", *J. Chem. Theory Comput.*, **8**, 3044–3052 (2012). doi:10.1021/ct300537s.

100. O. Trott and A. J. Olson, "AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading", *J. Comput. Chem.*, **31**, 671–690 (2009). doi:10.1002/jcc.21334.

101. M. F. Sanner, "Python: a programming language for software integration and development.", *J. Mol. Graph. Model.*, **17**, 57–61 (1999).

102. K. A. Beauchamp, G. R. Bowman, T. J. Lane, L. Maibaum, I. S. Haque and V. S. Pande, "MSMBuilder2: Modeling Conformational Dynamics on the Picosecond to Millisecond Scale", *J. Chem. Theory Comput.*, **7**, 3412–3419 (2011). doi:10.1021/ct200463m.

103. N. Michaud-Agrawal, E. J. Denning, T. B. Woolf and O. Beckstein, "MDAnalysis: A toolkit for the analysis of molecular dynamics simulations", *J. Comput. Chem.*, **32**, 2319–2327 (2011). doi:10.1002/jcc.21787.

104. R. J. Gowers, M. Linke, J. Barnoud, T. J. E. Reddy, M. N. Melo, S. L. Seyler, J. Domański, D. L. Dotson, S. Buchoux, I. M. Kenney and O. Beckstein, "MDAnalysis: A Python Package for the Rapid Analysis of Molecular Dynamics Simulations", 98–105 (2016).

105. Y. Liu, I. Matanovic, D. P. Hickey, S. D. Minteer, P. Atanassov and S. C. Barton, "Cascade Kinetics of an Artificial Metabolon by Molecular Dynamics and Kinetic Monte Carlo", *ACS Catal.*, **8**, 7719–7726 (2018). doi:10.1021/acscatal.8b01041.

106. F. Perez and B. E. Granger, "IPython: A System for Interactive Scientific Computing", *Comput. Sci. Eng.*, **9**, 21–29 (2007). doi:10.1109/MCSE.2007.53.

107. O. Valsson, P. Tiwary and M. Parrinello, "Enhancing Important Fluctuations: Rare Events and Metadynamics from a Conceptual Viewpoint", *Annu. Rev. Phys. Chem*, **67**, 159–84 (2016). doi:10.1146/annurev-physchem-040215-112229.

108. What is Independent Component Analysis?, in *Indep. Compon. Anal.* John Wiley & Sons, Inc., New York, USA (n.d.), pp. 145–164.

109. L. Molgedey and H. G. Schuster, "Separation of a mixture of independent signals using time delayed correlations", *Phys. Rev. Lett.*, **72**, 3634–3637 (1994). doi:10.1103/PhysRevLett.72.3634.

110. C. R. Schwantes, D. Shukla and V. S. Pande, "Markov State Models and tICA Reveal a Nonnative Folding Nucleus in Simulations of NuG2", *Biophys. J.*, **110**, 1716–1719 (2016). doi:10.1016/J.BPJ.2016.03.026.

111. H. S. Fogler, *Essentials of chemical reaction engineering*, Prentice Hall (2011).