

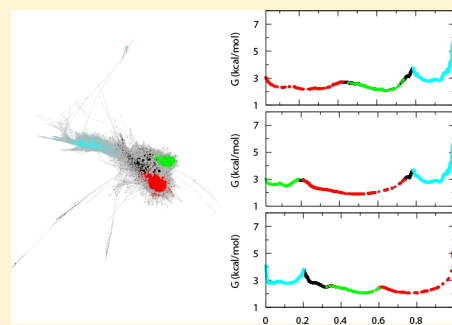
# Free Energy Surface of the Michaelis Complex of Lactate Dehydrogenase: A Network Analysis of Microsecond Simulations

Xiaoliang Pan and Steven D. Schwartz\*

Department of Chemistry and Biochemistry, University of Arizona, 1306 East University Boulevard, Tucson, Arizona 85721, United States

## Supporting Information

**ABSTRACT:** It has long been recognized that the structure of a protein creates a hierarchy of conformations interconverting on multiple time scales. The conformational heterogeneity of the Michaelis complex is of particular interest in the context of enzymatic catalysis in which the reactant is usually represented by a single conformation of the enzyme/substrate complex. Lactate dehydrogenase (LDH) catalyzes the interconversion of pyruvate and lactate with concomitant interconversion of two forms of the cofactor nicotinamide adenine dinucleotide (NADH and NAD<sup>+</sup>). Recent experimental results suggest that multiple substates exist within the Michaelis complex of LDH, and they show a strong variance in their propensity toward the on-enzyme chemical step. In this study, microsecond-scale all-atom molecular dynamics simulations were performed on LDH to explore the free energy landscape of the Michaelis complex, and network analysis was used to characterize the distribution of the conformations. Our results provide a detailed view of the kinetic network of the Michaelis complex and the structures of the substates at atomistic scales. They also shed light on the complete picture of the catalytic mechanism of LDH.

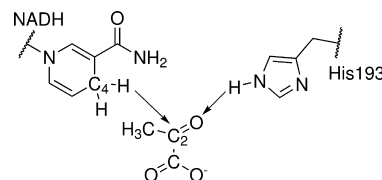


## INTRODUCTION

Emerging evidence has shown that a protein does not necessarily fold into one unique native state, but rather it may have multiple distinct “native” states that it can explore in various time scales. The process of ligand binding can be seen as a special case of protein folding, in which the loose enzyme–substrate encounter complex “folds” into catalytically competent conformations. Thus, similar considerations can be applied to ligand binding: instead of one unique Michaelis state, the encounter complex can fold into multiple interconverting Michaelis states between which the conformational equilibria are of particular interest because they perturb the distributions of the catalytically important structural parameters and thus modulate the catalytic rate.

Lactate dehydrogenase (LDH) catalyzes the interconversion of pyruvate and lactate with concomitant interconversion of two forms of the cofactor nicotinamide adenine dinucleotide (NADH and NAD<sup>+</sup>), which involves the hydride transfer between C<sub>4</sub> of the cofactor and C<sub>2</sub> of the substrate and the proton transfer between an active site histidine and the C<sub>2</sub>-bound substrate oxygen (Scheme 1). Recent experiments have shown that conformational heterogeneity exists in the Michaelis complex of LDH,<sup>1–3</sup> and the propensity toward the hydride transfer reaction may vary widely among the different substates of the Michaelis complex. However, the relationship between the substates is still elusive, and the detailed structural features of the substates are unclear. Here, we directly investigated the dynamics of the substrate-bound human heart LDH on the nanosecond-to-microsecond time scale by using all-atom

**Scheme 1.** Reaction Catalyzed by LDH from the Side of Pyruvate



molecular dynamics. The trajectories were analyzed by using network analysis to resolve the basins on the free energy landscape of the Michaelis complex and to find the time scales of the transitions between the basins. Our results not only support that there is conformational heterogeneity within the Michaelis complex of LDH but also provide a detailed view of the kinetic network of the Michaelis complex of LDH and the structures of the substates at atomistic scales.

## METHODS

**Molecular Dynamics Simulations.** The initial coordinates of the simulations were built from the X-ray crystal structure of the ternary complex of human heart LDH with the cofactor NADH and the substrate mimic oxamate (PDB ID: 1I0Z). Two asymmetric units were included in the model to form the

**Received:** February 24, 2015

**Revised:** March 31, 2015

**Published:** April 1, 2015

biological assembly, which is a homotetramer, and all the crystallographic water molecules were kept. An early simulation study of LDH has shown that it is essential to include the tetramer structure in the simulation to maintain the proper active site structure.<sup>4</sup> The substrate pyruvate was created by manually substituting the oxamate nitrogen atom with a carbon atom. The missing hydrogen atoms were added by the Protein Preparation Wizard module of Maestro,<sup>5</sup> which also predicted the protonation states of ionizable amino acid residues by exploiting the PROPKA algorithm<sup>6</sup> and the rotamers of the His, Asn, and Gln residues by optimizing the local hydrogen bond networks. His193 was predicted to be in the protonated form, consistent with the proposed protonation state of this catalytic residue in the pyruvate to lactate reaction. The resulting complex was immersed in a cubic box of pre-equilibrated TIP3P water.<sup>7</sup> The initial size of the box was chosen to be  $110 \times 110 \times 110 \text{ \AA}^3$  to ensure a minimal distance of  $10 \text{ \AA}$  between the boundary and the closest protein atom in each direction. Solvent molecules were removed if they were within  $2.8 \text{ \AA}$  of any heavy atoms of the complex. Thirty-two sodium ions were added to the system to neutralize the total net charge. These procedures led to a system of  $\sim 127\,000$  atoms for the molecular dynamics (MD) simulations.

MD simulations were conducted in periodic boundary conditions using NAMD 2.9.<sup>8</sup> The proteins and the cofactors were modeled using the CHARMM36 force field,<sup>9,10</sup> and the water molecules were described with the TIP3P model.<sup>7</sup> The parameters for pyruvate were obtained from the CHARMM General Force Field (version 2b8)<sup>11</sup> and optimized according to the CHARMM force field parametrization procedure to better reproduce the properties of the molecule of the high level ab initio calculations. All the nonbonding interactions were truncated at a cutoff of  $10 \text{ \AA}$ , and a switch function was applied starting at  $8 \text{ \AA}$  for the van der Waals interactions. The particle mesh Ewald method<sup>12</sup> was employed to treat long-range electrostatic interactions. The SHAKE algorithm<sup>13</sup> was used to constrain all bonds involving hydrogen atoms, and a time step of  $2 \text{ fs}$  was used for the MD integration. The temperature and pressure were controlled by the Langevin thermostat and the Nosé–Hoover Langevin barostat<sup>14,15</sup> as implemented in NAMD, respectively.

The system was first minimized and equilibrated under constant pressure and temperature (NPT) conditions at  $1 \text{ atm}$  and  $300 \text{ K}$  for  $500 \text{ ps}$ . During minimization and equilibration, the heavy atoms of the protein backbones were restrained to the starting structure by a harmonic potential with a  $10 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$  spring constant, and  $5 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$  was used for the heavy atoms of the protein side chains, the cofactor, and the substrate. The equilibrated box size of  $107.5 \times 107.5 \times 107.5 \text{ \AA}^3$  was used for the subsequent  $1.5 \text{ ns}$  constant volume and temperature (NVT) equilibration. During this run, the harmonic restraints were gradually released. Finally, a  $200 \text{ ns}$  NVT production run at  $300 \text{ K}$  was carried out, and the Cartesian coordinates were saved every  $20 \text{ ps}$  along the trajectory for further analysis.

**Clustering and Cut-Based FEP.** The clustering of the trajectories was carried out in the DRID metric, which showed improvements over the conventional metrics like coordinate root-mean-square deviation (cRMSD) or distance root-mean-square deviation (dRMSD) in terms of closeness to kinetic distance and calculation efficiency.<sup>16</sup> The kinetic distance between two MD snapshots is defined as the simulation time required to evolve from one to the other. DRID is based on the

pairwise distances between two groups of atoms: a set of  $n$  center atoms and a set of  $N$  reference atoms, so it does not depend on the orientation of the molecule and does not need structural alignment. The set of reference atoms works like a reference frame, and for each one of the  $n$  center atoms, the distances between the center atom and the reference atoms are collected (the distance is excluded if the center atom and the reference atom are covalently bonded). Then the distribution of the reciprocal distances is encoded to a three-dimensional vector consisting of the mean  $\mu_i \equiv [\sum_j (1/d_{ij})]/N$ , where  $d_{ij}$  is the distance between the reference atom  $j$  and the center atom  $i$  and  $N$  is the number of reference atoms; the square root of the second central moment  $\nu_i \equiv \{[\sum_j (1/d_{ij} - \mu_i)^2]/N\}^{1/2}$ ; and the cubic root of the third central moment  $\xi_i \equiv \{[\sum_j (1/d_{ij} - \mu_i)^3]/N\}^{1/3}$ . So a molecule conformation is described by a  $3n$ -dimensional vector  $(\mu_i, \nu_i, \xi_i)$ . In this study, the reference atoms were chosen to be the nonsymmetric heavy atoms in the active site residues, which include the active site loop (residues 98–110), the structure it contacts (residues 236–249), the residues interacting with the substrate (residues 138, 169, and 193), and the nicotinamide mononucleotide moiety of NADH. To account for the conformation of the protein as well as the relative position of the substrate, two sets of center atoms were used that consisted of the nonsymmetric heavy atoms in the active site loop ( $n_1 = 90$ ) and the substrate ( $n_2 = 3$ ), respectively. Each MD snapshot was encoded in a  $270$ -dimensional vector  $(\mu'_i, \nu'_i, \xi'_i)$  and a  $12$ -dimensional one  $(\mu''_i, \nu''_i, \xi''_i)$ . The geometrical distance between two conformations described by  $(\mu'_i, \nu'_i, \xi'_i)$ ,  $(\mu''_i, \nu''_i, \xi''_i)$  and  $(\tilde{\mu}'_i, \tilde{\nu}'_i, \tilde{\xi}'_i)$ ,  $(\tilde{\mu}''_i, \tilde{\nu}''_i, \tilde{\xi}''_i)$  is defined as  $\{ (1/3n_1) \sum_{i=1}^{n_1} [(\mu'_i - \tilde{\mu}'_i)^2 + (\nu'_i - \tilde{\nu}'_i)^2 + (\xi'_i - \tilde{\xi}'_i)^2] + (1/3n_2) \sum_{i=1}^{n_2} [(\mu''_i - \tilde{\mu}''_i)^2 + (\nu''_i - \tilde{\nu}''_i)^2 + (\xi''_i - \tilde{\xi}''_i)^2] \}^{1/2}$  and has units of  $1/\text{distance}$ . A modified leader algorithm was used for the clustering. In our implementation, the conformation of the first chain in the crystal structure was used as the first cluster center, and each snapshot was first compared with the center of the cluster of the previous snapshot in the trajectory it belonged to. If the distance was smaller than the cutoff, the snapshot was assigned to this cluster. If the distance was larger than the cutoff or the snapshot is the first one in the trajectory, the snapshot was compared with all the existing centers. If the distance between the snapshot and the nearest cluster was smaller than the cutoff, then the snapshot is assigned to this cluster; otherwise, the snapshot becomes a new cluster center.

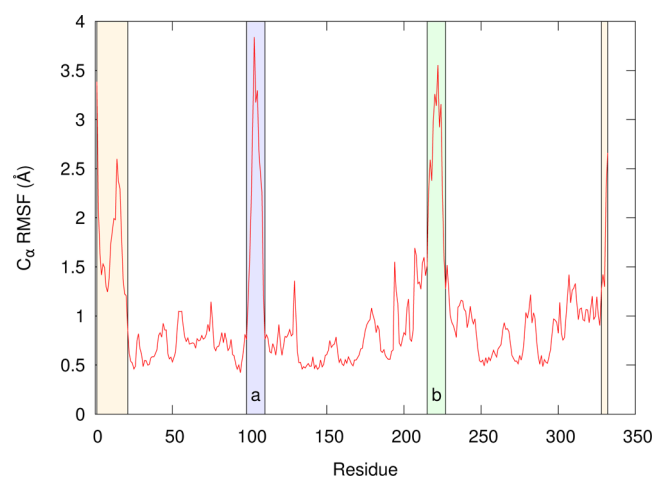
The clustering procedure provides a conformational space network<sup>17</sup> where the nodes are the clusters and the edges are the direct transitions sampled during the MD simulations with a saving frequency of  $20 \text{ ps}$ . For each node  $i$  in the transition network, the partition function is given by  $Z_i = \sum_j c_{ij}$ , i.e., the number of times the node  $i$  is visited along the trajectories, where  $c_{ij}$  is the number of direct transitions from node  $j$  to node  $i$  observed along the trajectories. Then the transition probabilities can be calculated as  $p_{ij} = c_{ij}/\sum_k c_{ik}$ . For equilibrium ensembles, detailed balance should be expected; i.e.,  $c_{ij} = c_{ji}$  should hold true for any pair of  $i$  and  $j$ . In practice, due to the finite sampling, this property is not guaranteed in the network. In this study, detailed balance was forced by setting the numbers of the transitions from node  $i$  to  $j$  and from node  $j$  to  $i$  to the arithmetic mean of them. Using the network as the input, the cut-based free energy profile (cFEP),<sup>18,19</sup> which is a barrier-preserving projection of the free energy surface, is calculated. In this study, we adopted the procedure based on the mean first passage time (MFPT). First, the most populated node is used

as a reference, and the MFPT from node  $i$  to the reference node is calculated by solving the equation  $\text{MFPT}_i = \Delta t + \sum (p_{ji} \times \text{MFPT}_j)$  with the initial boundary condition  $\text{MFPT}_{\text{reference}} = 0$ . The lag time  $\Delta t$  corresponds to the saving frequency of the trajectories. So the MFPT of a node is defined as the weighted average of the MFPT values of all the other nodes plus the lag time. Then the nodes are sorted according to their MFPT values to the reference node. For any  $\text{MFPT}_c$  (MFPT of the cutting surface) between 0 and  $\text{MFPT}_{\text{max}}$ , the set of nodes is partitioned into two groups, A and B, where A is the set of nodes with  $\text{MFPT}_i < \text{MFPT}_c$  and B is the set of nodes with  $\text{MFPT}_i > \text{MFPT}_c$  with  $Z_A = \sum_{i \in A} Z_i$  and  $Z_B = \sum_{i \in B} Z_i$ , where  $Z_A$  and  $Z_B$  are the partition function of the groups A and B. In the MFPT-based cFEP procedure, the transition barrier to leave the reference node is defined as the cutting surface with the minimal partition function that divides the network into two groups (A and B), where the nodes in group A have  $\text{MFPT}_i < \text{MFPT}_c$  and the nodes in group B have  $\text{MFPT}_i > \text{MFPT}_c$ . The partition function of the cutting surface is then  $Z_{AB} = \sum_{i \in A, j \in B} c_{ij}$ , where  $c_{ij}$  is the number of direct transitions from node  $j$  to node  $i$ . Thereby, the free energy of the barrier between groups A and B can be written as  $G = -kT \ln(Z_{AB}/Z)$ , where  $Z$  is the partition function of the entire network. Thus, a point on the cFEP is evaluated as  $(Z_A/Z, G = -kT \ln(Z_{AB}/Z))$ , where the normalized partition function  $Z_A/Z$  is used as the reaction coordinate. The result is a one-dimensional profile that preserves the barrier height between the free energy basins. The first basin consists of the set of nodes on the left side of the first barrier on the profile. The procedure can be iterated to isolate other basins. To test the robustness of the network analysis, we compared the cFEPs obtained from the conformational space networks built with different clustering cutoffs, which show that the cFEPs are converged by using cutoffs from  $3.6 \times 10^{-3}$  to  $3.9 \times 10^{-3} \text{ \AA}^{-1}$  (Figure S1, Supporting Information). The intrinsic diffusive behavior of the system, which is expected for the enzyme conformational change and the spontaneous ligand binding/unbinding, is preserved by the clustering procedure (Figure S2, Supporting Information), which is validated by calculating the cFEPs using only every other MD snapshot from the simulations<sup>20</sup> (i.e., the same simulations with a 40 ps snapshot saving frequency). The MFPT and the cFEPs were calculated by the Wordom program.<sup>21,22</sup>

## RESULTS AND DISCUSSION

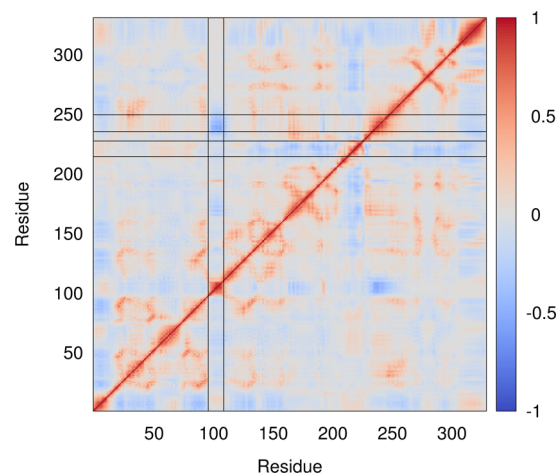
A total of 20 independent MD runs were carried out with different initial velocities. All trajectories begin with the substrate bound. Since LDH exhibits no cooperativity between the subunits of the tetramer, the coordinate trajectories of each subunit were treated as independent trajectories in the analysis. Though no artificial guiding force or biasing potential was used in the simulations, in 28 of the 80 trajectories, the substrate diffuses completely away from the surface of LDH. (The minimal distance between the carbonyl oxygen of the substrate and all the heavy atoms of the corresponding subunit is over 10 Å.) Because the focus of this study is the dynamics of the complex, for all the unbinding trajectories, only the snapshots from the beginning to the first time the hydride acceptor (the carbonyl carbon of pyruvate) is over 15 Å away from the hydride donor of NADH in the same subunit were included in the trajectory analysis. Thus, the cumulative simulation time is about 14  $\mu\text{s}$  (694 679 snapshots).

**Network Analysis.** The overall structure of LDH is quite stable throughout the simulations. The  $C_\alpha$  cRMSD from the X-ray structure is  $<2.0 \text{ \AA}$  for most of the snapshots (90.6%) and always  $<3.0 \text{ \AA}$ . The largest scale of motions are located within two loop regions (residues 98–110 and 215–227) besides the terminal regions (Figure 1), and the  $C_\alpha$  cRMSD is  $<1.5 \text{ \AA}$  for



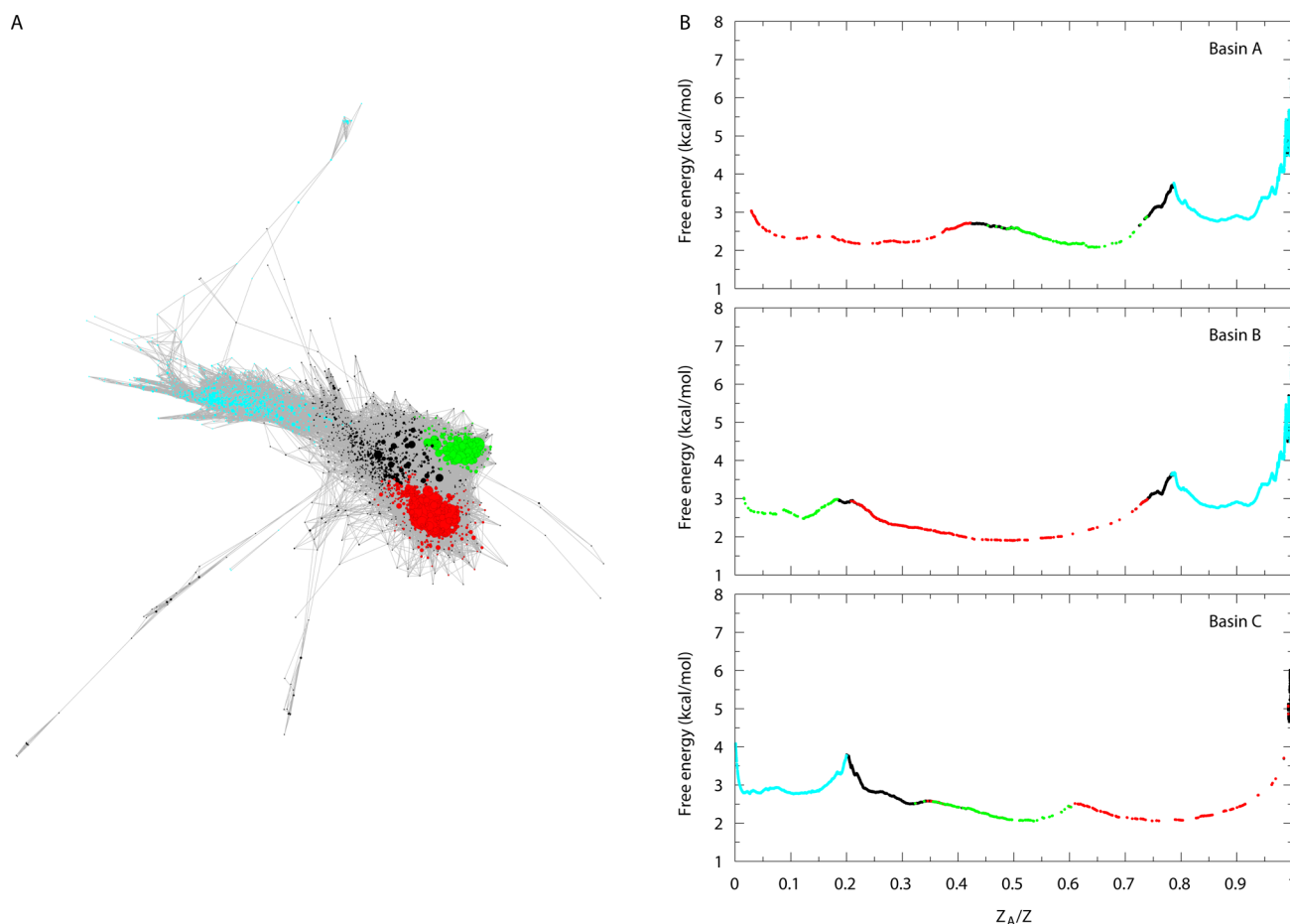
**Figure 1.** Root mean square fluctuation of the  $C_\alpha$  of LDH. The blue region (a) and green region (b) indicate the active site loop and the nonactive site loop, respectively.

97.5% of the snapshots using the same alignments as above if these regions are excluded. From the RMSF and the correlation map of the  $C_\alpha$  atom dynamics (Figure 2), the dynamics of the



**Figure 2.**  $C_\alpha$  dynamic cross-correlation map. The intersections of the active site loop (residues 98–110) with the nonactive site loop (residues 215–227) and the contacting helix (residues 236–249) are highlighted. The two loop regions show very weak dynamical correlation, while the correlation between the dynamics of the active site loop and the contacting helix is relatively strong.

enzyme can be approximately treated as two independent loop motions in a relatively rigid framework. Here we focused on the interactions between the enzyme and the substrate, so only the active site loop (residues 98–110) and the contacting helix (residues 236–249), which has correlated motions with the active site loop (Figure 2), are used to describe the conformations of the enzyme. While the commonly used distance metric cRMSD is good for the overall description of



**Figure 3.** Network analysis of the LDH–pyruvate complex. (A) Conformational space network. Each node of the network represents a cluster from the structural clustering, and only the 2069 nodes with at least 40 snapshots are shown. (B) The cFEP from the most populated node in the three basins. Basins A, B, and C and unassigned clusters are colored in red, green, blue, and black, respectively.

the protein conformations, to account for the relative positions of the substrate and the active site as well as the local conformational changes within the active site, some microscopic information, such as the interatomic distances, is needed for a detailed description of the Michaelis complex conformations. In this study, we used a newly proposed structural metric called distribution of reciprocal of interatomic distances (DRID),<sup>16</sup> which incorporates pairwise distances within the active site and substrate atoms (see the Methods) to cluster the sets of trajectories. In the DRID metric, each snapshot in the trajectories was encoded into a vector, and pairwise distances are calculated as Euclidean distances between the vectors. The clustering was done using a modified leader algorithm (see the Methods), and the cutoff was chosen to be  $3.6 \times 10^{-3} \text{ \AA}^{-1}$ , which yielded 15 580 clusters, of which 13 081 contain at least 2 snapshots, and 277 551 transitions between them were observed in the trajectories.

On the basis of the clustering, we built a conformational space network<sup>17</sup> by using clusters and transitions as nodes and edges of the network (Figure 3A) and grouped the clusters that are “kinetically close” and rapidly interconverting into basins by calculating the barrier-preserving cut-based free energy profiles (cFEPs)<sup>18,19</sup> (Figure 3B). The result of the cFEP procedure is a one-dimensional free energy profile that preserves the barrier height between the free energy basins. On the cFEP, the basin consists of the nodes on the left side of the first barrier on the profile, and the nodes on the right side can be overlapped

potentially due to similar kinetic distance from two or more basins to the reference cluster, so each basin needs a separate profile<sup>18</sup> (see the Methods for more details). Three major basins were isolated using the cFEP method. Two of them (labeled as A and B) are well-defined basins, while the third (labeled as C) is more inhomogeneous. It should be noted that not all the clusters are assigned to a basin, because most of the unassigned clusters are in the transition regions and there is no fast equilibration within them. While basins A and B can interconvert to each other directly, the interconversion between basin C and basin A or B has to pass through the transition regions. The mean first passage times (MFPTs) between the basins are summarized in Table 1, which clearly shows that there are two distinct time scales involved in the dynamics of the Michaelis complex of the enzyme.

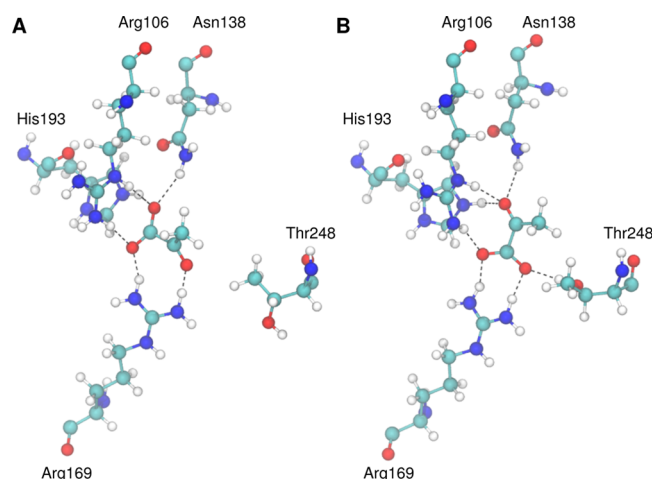
**Two Major Binding Modes.** Basins A and B actually correspond to two major binding modes of the substrate (Figure 4). While basin B corresponds to the binding mode of the crystal structure and has a statistical weight of 18.4%, basin A consists of clusters with an alternative binding mode, and the statistical weight is even larger (42.2%). Hydride donor–acceptor (D–A) distance has been shown to be a good indicator for the propensity toward the hydride transfer reaction of the conformation in our previous studies: a short D–A distance can facilitate the hydride transfer from NADH to pyruvate, which is the first chemical step of the catalytic process. Here we plotted the D–A distances distributions of



**Table 1.** Mean First Passage Time between Basins (ns)<sup>a</sup>

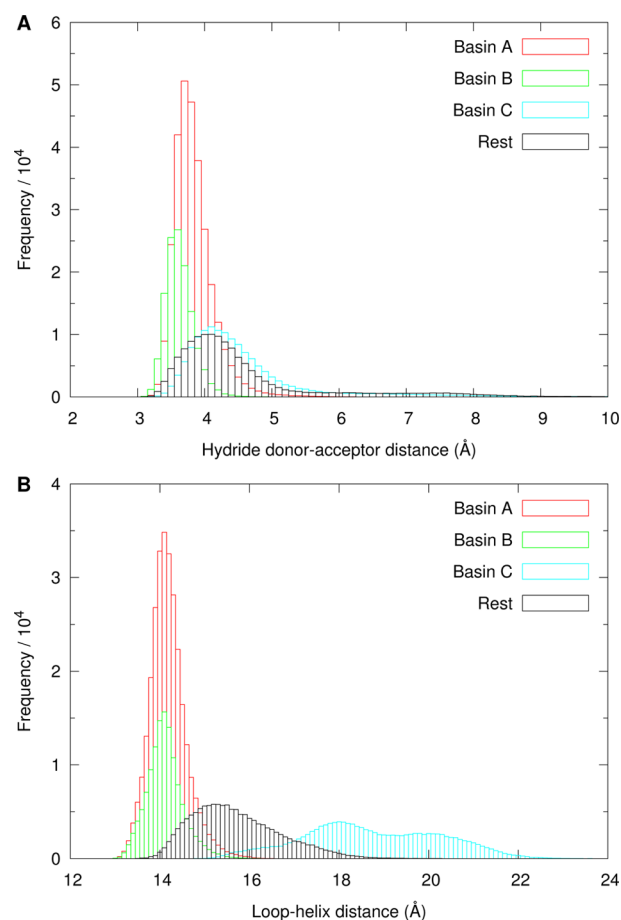
from	to basin A	to basin B	to basin C
basin A	7.0 ± 0.8	19.5 ± 0.3	104.8 ± 0.5
basin B	12.7 ± 0.2	7.5 ± 1.8	104.0 ± 0.3
basin C	24.8 ± 3.1	30.8 ± 3.1	50.0 ± 8.5

<sup>a</sup>MFPT between basins is calculated by averaging the MFPT from each cluster of the starting basin to the largest cluster of the ending basin. The values after ± are standard deviations. The diagonal elements are calculated by averaging the MFPT from each cluster (excluding the largest cluster) of the corresponding basin to the largest cluster of the same basin and represent the equilibration time within the corresponding basins.

**Figure 4.** Representative structures from (A) basin A and (B) basin B.

different basins (Figure 5A). The distributions peak at longer distances for basin A (3.7 Å) than for basin B (3.6 Å). For comparison, the D–A distances are 3.3 and 3.4 Å in the two chains of the crystal structure. Furthermore, basin A has a longer tail at larger D–A distance regions, which indicates that the substrate binds a little looser to the enzyme for conformations in basin A. So conformations in basin A are a little more flexible than those in basin B, which could explain that the conformations in basin A dominate at room temperature due to the higher entropy of basin A, while the low-temperature crystal structure only captured the energy-favored conformations in basin B. Even though the population of basin A is over twice as large as that of basin B, the latter has larger populations for snapshots with D–A distances <3.5 Å. Thus, it suggests that the conformations in basin B are more reactive toward the hydride transfer step than those in basin A, the population of which is larger than that of basin B according to D–A distance analysis, which is consistent with the experimental results.<sup>2</sup>

To understand the structural differences between basins A and B, we need to use some geometric criteria to describe the structures. While the high-dimensional metric we used in the clustering is good at differentiating the structures of the snapshots, it is better to use low-dimensional order parameters to characterize the structures for visualization purposes. Because the major and most binding-relevant motion of the protein is the opening/closing motion of the active site loop (residues 98–110), we used the loop–helix (L–H) distance, which is defined as the average of pairwise distances between the C<sub>α</sub> atoms of the active site loop (residues 98–110) and the contacting helix (residues 236–249), to characterize the overall

**Figure 5.** Distributions of key geometrical parameters for the basins: (A) hydride donor–acceptor distances and (B) loop–helix distances.

conformation of the protein. Because the distributions of the L–H distances are similar between basins A and B (Figure 5B) and basins A and B can interconvert to each other on a shorter time scale (Table 1), we conclude that the overall structures of the enzyme are similar between basins A and B, and the enzyme does not require a large structure rearrangement for the interconversion between the two major binding modes.

To account for the interactions between the substrate and the enzyme, we did a hydrogen bond analysis between the carbonyl or carboxyl oxygens of the substrate and the active site residues (Table 2). While one role of the hydrogen bonds between the substrate and the active site residues is anchoring the substrate in the right positions for the reaction to happen, the hydrogen bonds involving the carbonyl oxygen can also facilitate the hydride transfer reaction by polarizing the C<sub>2</sub>=O bond.<sup>23</sup> For basin B, the hydrogen bonds in the static picture of the crystal structure are well-conserved in the trajectories. In the three hydrogen bonds involving the carbonyl oxygen, two are relatively strong, using Asn138 and His193 as the donors and having occupations of 76.7% and 78.8%, respectively. The other hydrogen bond, which uses Arg106 as the donor, is weaker than the other two with a occupation of 46.1%. In the meantime, Arg106 can also form a strong hydrogen bond (with a occupation of 86.9%) with one of the carboxyl oxygen atoms. The hydrogen bonding between Arg169 and the carboxyl group is quite strong, as it populates in almost all the snapshots (99.3%) of the basin, which is consistent with the experimental results.<sup>1</sup> There is also a weak hydrogen bond between Thr248

Table 2. Percentage Occupation of Hydrogen Bonds

	basin A		basin B		basin C		rest	
	C=O <sup>a</sup>	COO <sup>-b</sup>	C=O	COO <sup>-</sup>	C=O	COO <sup>-</sup>	C=O	COO <sup>-</sup>
Arg106	0.4	96.6	46.1	86.9	0.1	1.3	2.4	26.1
Asn138	0.1	61.4	76.7	0.1	3.5	6.1	7.3	12.3
Arg169	90.3	54.9	0.0	99.3	7.8	80.0	9.8	64.6
His193	0.0	92.7	78.8	1.5	10.4	17.8	15.3	24.2
Thr248	8.1	0.1	0.0	26.2	3.4	53.5	2.1	35.3

<sup>a</sup>Carbonyl oxygen of the substrate pyruvate as hydrogen bond acceptor. <sup>b</sup>One of the two carboxyl oxygens of the substrate pyruvate as hydrogen bond acceptor.

and the carboxyl group, which populates in about one-quarter of the snapshots. For basin A, the substrate can be seen as rotating around its main chain by 180°. As a result, the carbonyl oxygen can only form a strong hydrogen bond with Arg169, while the carboxyl group can form hydrogen bonds with Arg106, Asn138, Arg169, and His193. In conclusion, the roles of the two active site Arg residues (Arg106 and Arg169) are anchoring the substrate as well as polarizing the C<sub>2</sub>=O bond, and they can compensate each other's role in the two binding modes. Asn138 and His193 are responsible for anchoring the substrate in basin A and polarizing the C<sub>2</sub>=O bond in basin B. The polarization of the C<sub>2</sub>=O bond has been used to characterize the propensity toward the hydride transfer reaction of the conformation and is determined by measuring the frequency of the C<sub>2</sub>=O bond using vibrational spectroscopies.<sup>23</sup> In the simulations, we used the strength of hydrogen bonding to the carbonyl oxygen of pyruvate to describe the propensity toward the hydride transfer reaction of the conformation. The hydrogen bonding to the carbonyl oxygen in basin B is stronger than that in basin A, so by using the hydrogen bond analysis, we can also conclude that the conformations in basin B are more reactive toward the hydride transfer step than those in basin A. This supports our previous calculations that showed the reactive conformation to be a minority population.<sup>24</sup>

**Encounter Complex.** While the interconversion between basins A and B is relatively fast, the conversion from basins A and B to basin C is on a much slower time scale (Table 1). The distribution of the L–H distances in basin C is more spread out than those of basin A and B (Figure 5B), which is consistent with its inhomogeneous nature. The L–H distances in basin C range from 15 to 23 Å, which indicates that conformations in basin C are in the open form. By visualizing the trajectories, we found that most of the unbinding events (in 22 of the 28 unbinding trajectories) happened when the enzyme was in the conformations of basin C, which suggested that those conformations can be seen as the encounter complex. The remaining six unbinding events happened when the enzyme was in the partially open conformations (the transition regions). This supports our previous calculations that suggested that the ligand binding could happen when the active site loop was just partially open.<sup>25</sup> The majority of the conformations in basin C have long hydride D–A distances (Figure 5A), and the hydrogen bonds to the carbonyl oxygen of the substrate are very weak (Table 2), so it is more rare for the chemical step to happen in the basin C conformations. The carboxyl group of the substrate can form a strong hydrogen bond to Arg169 (with a occupation of 80.0%), and interestingly, the hydrogen bond between the carboxyl group and Thr248 is much stronger in basin C (with a occupation of 53.5%) than those in basins A and B (with occupations of 0.1% and 26.2%, respectively)

(Table 2), which suggests that Arg169 and Thr248 play important roles for anchoring the substrate in the early stage of the binding event.

Recent experiments have found that there are three substates for the LDH·NADH·pyruvate complex, each with a different propensity toward the hydride transfer reaction, and the substates cannot interconvert directly without going through the encounter complex.<sup>3</sup> In our model shown here, the two binding modes could correspond to the two well-defined substates, basins A and B, while the transition regions could correspond to the inhomogeneous state, basin C. However, in contrast to the experimental results,<sup>3</sup> our model shows that basins A and B can interconvert either directly or by going through a transition region.

## CONCLUSIONS

In this work, we examined the free energy landscape of the Michaelis complex of LDH by using molecular dynamics and built a conformational space network of the complex. Three basins on the free energy landscape that corresponded to the three substates of the Michaelis complex—the encounter complex and the two different substrate binding modes—were located, and the structural features of the substates were analyzed on the atomistic scales. Our results support that the Michaelis complex of LDH consists of different substates that are interconverting with each other on different time scales. Interestingly, the largest substate found in this study does not correspond to the “native” structure of the LDH·NADH·pyruvate complex (the one found in the crystal structure). Thus, the conventional studies only involving the crystal structures may not get the whole picture of the catalytic reactions. Furthermore, due to the local structural rearrangements around the substrate, which is associated with the interconversion between the substates, alternative mechanisms of the chemical step that are not considered could potentially exist in the substates. So further studies can be conducted to directly explore the reactivities and reaction mechanisms of different substates of the Michaelis complex using structures from the different substates obtained in this study as starting structures.

## ASSOCIATED CONTENT

### Supporting Information

cFEPs calculated with different DRID cutoffs (Figure S1), diffusivity tests (Figure S2), and cFEPs calculated with and without applying detailed balance (Figure S3). This material is available free of charge via the Internet at <http://pubs.acs.org/>.

## ■ AUTHOR INFORMATION

## Corresponding Author

\*Phone: (520) 621-6363. E-mail: sschwartz@email.arizona.edu.

## Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

The authors acknowledge the support from National Institutes of Health Grant GM068036.

## ■ REFERENCES

- (1) Deng, H.; Vu, D. V.; Clinch, K.; Desamero, R.; Dyer, R. B.; Callender, R. Conformational Heterogeneity within the Michaelis Complex of Lactate Dehydrogenase. *J. Phys. Chem. B* **2011**, *115*, 7670–7678.
- (2) Peng, H.-L.; Deng, H.; Dyer, R. B.; Callender, R. Energy Landscape of the Michaelis Complex of Lactate Dehydrogenase: Relationship to Catalytic Mechanism. *Biochemistry* **2014**, *53*, 1849–1857.
- (3) Reddish, M. J.; Peng, H.-L.; Deng, H.; Panwar, K. S.; Callender, R.; Dyer, R. B. Direct Evidence of Catalytic Heterogeneity in Lactate Dehydrogenase by Temperature Jump Infrared Spectroscopy. *J. Phys. Chem. B* **2014**, *118*, 10854–10862.
- (4) Schmidt, R. K.; Gready, J. E. Molecular Dynamics Simulations of L-Lactate Dehydrogenase: Conformation of a Mobile Loop and Influence of the Tetrameric Protein Environment. *J. Mol. Model.* **1999**, *5*, 153–168.
- (5) *Maestro*, version 9.6; Schrödinger, LLC, New York, NY, 2013.
- (6) Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.
- (7) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926–935.
- (8) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
- (9) Best, R. B.; Zhu, X.; Shim, J.; Lopes, P. E. M.; Mittal, J.; Feig, M.; MacKerell, A. D. Optimization of the Additive CHARMM All-Atom Protein Force Field Targeting Improved Sampling of the Backbone  $\phi$ ,  $\psi$  and Side-Chain  $\chi_1$  and  $\chi_2$  Dihedral Angles. *J. Chem. Theory Comput.* **2012**, *8*, 3257–3273.
- (10) MacKerell, A. D.; Bashford, D.; Bellott, M.; Dunbrack, R. L.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; et al. All-Atom Empirical Potential for Molecular Modeling and Dynamics Studies of Proteins. *J. Phys. Chem. B* **1998**, *102*, 3586–3616.
- (11) Vanommeslaeghe, K.; Hatcher, E.; Acharya, C.; Kundu, S.; Zhong, S.; Shim, J.; Darian, E.; Guvench, O.; Lopes, P.; Vorobyov, I.; et al. CHARMM General Force Field: A Force Field for Drug-like Molecules Compatible with the CHARMM All-Atom Additive Biological Force Fields. *J. Comput. Chem.* **2010**, *31*, 671–690.
- (12) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An  $N \log(N)$  Method for Ewald Sums in Large Systems. *J. Chem. Phys.* **1993**, *98*, 10089–10092.
- (13) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of  $n$ -Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.
- (14) Martyna, G. J.; Tobias, D. J.; Klein, M. L. Constant Pressure Molecular Dynamics Algorithms. *J. Chem. Phys.* **1994**, *101*, 4177–4189.
- (15) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant Pressure Molecular Dynamics Simulation: The Langevin Piston Method. *J. Chem. Phys.* **1995**, *103*, 4613–4621.
- (16) Zhou, T.; Caflisch, A. Distribution of Reciprocal of Interatomic Distances: A Fast Structural Metric. *J. Chem. Theory Comput.* **2012**, *8*, 2930–2937.
- (17) Rao, F.; Caflisch, A. The Protein Folding Network. *J. Mol. Biol.* **2004**, *342*, 299–306.
- (18) Krivov, S. V.; Muff, S.; Caflisch, A.; Karplus, M. One-Dimensional Barrier-Preserving Free-Energy Projections of a  $\beta$ -Sheet Miniprotein: New Insights into the Folding Process. *J. Phys. Chem. B* **2008**, *112*, 8701–8714.
- (19) Krivov, S. V.; Karplus, M. J. One-Dimensional Free-Energy Profiles of Complex Systems: Progress Variables that Preserve the Barriers. *Phys. Chem. B* **2006**, *110*, 12689–12698.
- (20) Krivov, S. V.; Karplus, M. Diffusive Reaction Dynamics on Invariant Free Energy Profiles. *Proc. Natl. Acad. Sci. U. S. A* **2008**, *105*, 13841–13846.
- (21) Seeber, M.; Cecchini, M.; Rao, F.; Settanni, G.; Caflisch, A. Wordom: A Program for Efficient Analysis of Molecular Dynamics Simulations. *Bioinformatics* **2007**, *23*, 2625–2627.
- (22) Seeber, M.; Felling, A.; Raimondi, F.; Muff, S.; Friedman, R.; Rao, F.; Caflisch, A.; Fanelli, F. Wordom: A User-Friendly Program for the Analysis of Molecular Structures, Trajectories, and Free Energy Surfaces. *J. Comput. Chem.* **2011**, *32*, 1183–1194.
- (23) Deng, H.; Zheng, J.; Clarke, A.; Holbrook, J. J.; Callender, R.; Burgner, J. W. Source of Catalysis in the Lactate Dehydrogenase System. Ground-State Interactions in the Enzyme-Substrate Complex. *Biochemistry* **1994**, *33*, 2297–2305.
- (24) Pineda, J. R. E. T.; Antoniou, D.; Schwartz, S. D. Slow Conformational Motions That Favor Sub-Picosecond Motions Important for Catalysis. *J. Phys. Chem. B* **2010**, *114*, 15985–15990.
- (25) Pineda, J. R. E. T.; Callender, R.; Schwartz, S. D. Ligand Binding and Protein Dynamics in Lactate Dehydrogenase. *Biophys. J.* **2007**, *93*, 1474–1483.