



Published in final edited form as:

J Phys Chem B. 2016 July 14; 120(27): 6612–6620. doi:10.1021/acs.jpcc.6b05119.

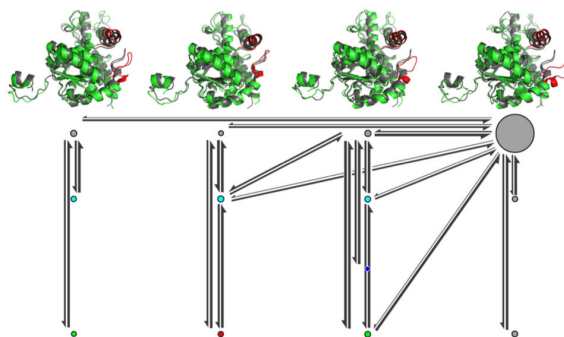
Conformational Heterogeneity in the Michaelis Complex of Lactate Dehydrogenase: An Analysis of Vibrational Spectroscopy Using Markov and Hidden Markov Models

Xiaoliang Pan and Steven D. Schwartz*

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

Abstract

Lactate dehydrogenase (LDH) catalyzes the interconversion of pyruvate and lactate. Recent isotope-edited IR spectroscopy suggests that conformational heterogeneity exists within the Michaelis complex of LDH, and this heterogeneity affects the propensity toward the on-enzyme chemical step for each Michaelis substrate. By combining molecular dynamics simulations with Markov and hidden Markov models, we obtained a detailed kinetic network of the substates of the Michaelis complex of LDH. The ensemble-average electric fields exerted onto the vibrational probe were calculated to provide a direct comparison with the vibrational spectroscopy. Structural features of the Michaelis substates were also analyzed on atomistic scales. Our work not only clearly demonstrates the conformational heterogeneity in the Michaelis complex of LDH and its coupling to the reactivities of the substates, but it also suggests a methodology to simultaneously resolve kinetics and structures on atomistic scales, which can be directly compared with the vibrational spectroscopy.



*Corresponding Author, sschwartz@email.arizona.edu. Phone: 520-621-6363. .

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jpcc.6b05119.

Box plots of the distributions of the minimum distances between carbonyl oxygen (Figure S1)/carboxyl oxygens (Figure S2)/methyl carbon (Figure S3) of pyruvate, and active site residues in different substrate binding states; implied relaxation time scales as a function of lag time (Figure S4); and output probabilities from hidden to observable discrete states (Figure S5) (PDF)

Notes

The authors declare no competing financial interest.

INTRODUCTION

It has been long recognized that proteins are highly dynamic entities and their structures are best described as an ensemble of interconverting conformations on different time scales. The same holds true for the enzyme–substrate (Michaelis) complex which contains an ensemble of interconverting conformations, each with its own reactivity (k_{cat}) as shown by recent single-molecule experiments,^{1–3} despite that it is usually assumed as a single active conformation in virtually all the textbooks.

Lactate dehydrogenase catalyzes the interconversion of pyruvate and lactate, using the cofactors NADH and NAD⁺, respectively, and it has been studied to characterize the conformational heterogeneity in the Michaelis complex and the coupling of this heterogeneity to enzymatic catalysis.^{4–10} Recently, the conformational heterogeneity of the ternary complex of LDH (LDH-NADH-pyruvate) was revealed through isotope-edited IR spectroscopy.⁹ The C = O stretching vibration of pyruvate was used as the vibrational probe, which is a direct measure of the strength of the electric field on the bond in a particular conformation and it is also highly correlated with the propensity toward the on-enzyme chemical reaction of that conformation.¹¹ For local high-frequency vibration such as the C = O stretch, the vibration is assumed to behave as a one-dimensional oscillator. Vibrational Stark spectroscopy¹² and theoretical computations¹³ have shown that the C = O stretching vibration responds to electric fields in a linear fashion, as suggested by linear Stark effect theory,¹⁴ which is believed to be an intrinsic property of the oscillator. Recently, it has been shown that a good linear correlation can be captured between the C = O frequencies and the ensemble-average electrostatic fields calculated by using the classical molecular dynamics (MD) with fixed-charge force fields.¹⁵

Isotope-edited IR spectroscopy has been proven to be a valuable tool to probe the conformational heterogeneity of the Michaelis complex and measure the reactivity of each Michaelis substate. Temperature-jump experiments combined with a kinetics model has been proposed for the Michaelis complex of LDH,¹⁰ in which four distinct Michaelis substates were resolved. However, structural information about degrees of freedom other than the probing carbonyl is mostly unknown. This information might shed light on important mechanisms such as how the enzyme interacts with the substrate in each Michaelis substate. It is a challenging problem to obtain a detailed kinetic network on atomistic scales for straightforward MD simulations, since the time scales of the involved motions range from nanosecond to microsecond, even millisecond. Markov state models (MSMs) have been shown to be able to reproduce long-time conformational dynamics of biomolecules using data from MD simulations that are individually much shorter. MSMs have been successfully applied to study the coupling of conformational dynamics of enzyme and substrate bindings in the lysine-, arginine-, ornithine-binding,¹⁶ choline-binding,¹⁷ and trypsin¹⁸ proteins.

Previously, we studied the catalytically relevant loop motions in purine nucleoside phosphorylase¹⁹ and lactate dehydrogenase^{20,21} using straightforward MD simulations. Recently, we estimated the free energy surface of the Michaelis complex of LDH using the cut-based free energy profile (FEP) method.²² Even though different substrate binding

modes were resolved, the basins on the FEP and the substates resolved by vibrational spectroscopy were only compared indirectly through hydrogen bond analysis. Due to the limitation of the method, the relaxation time scales of the system were underestimated by about 4 orders of magnitude. In this work, 80 μs trajectories of MD simulations have been analyzed with a Markov model which is further coarse-grained into a hidden Markov model (HMM) to obtain the metastable states. A total of 13 substates which differ in protein conformations as well as the substrate binding modes were resolved. The structural features of the substates were analyzed, and a kinetic network was established between them. The relaxation time scales are closer to the experimental values. Ensemble-average electric fields were calculated for each substate and compared with the vibrational spectroscopy to assign the IR bands to the Michaelis substates quantitatively. Our results not only support the experiment-observed conformational heterogeneity in the Michaelis complex of LDH but also provide a detailed view of the interactions between LDH and pyruvate in each observed Michaelis substate.

METHODS

Molecular Dynamics Simulations

The MD simulation and force field parameters are identical to the settings used in ref 22. The initial structure was built on the X-ray crystal structure of human heart LDH complexed with the cofactor NADH and the substrate mimic oxamate (PDB ID: 110Z). Two asymmetric units were included to form the tetramer model, and the substrate pyruvate was created by manually substituting the oxamate nitrogen atom with a carbon atom. The proteins and the cofactors were modeled using the CHARMM36 force field,^{23,24} and the force field parameters for pyruvate were obtained using the CHARMM general force field.²⁵ The system was solvated in a cubic TIP3P²⁶ water box, and neutralized with sodium ions. To prevent the substrate from diffusing too far away from the active site, a soft wall potential was added to the hydride donor–acceptor distance at 15 Å. A total of 25 independent simulations of 0.8 μs were conducted with different initial velocities using the NAMD program.²⁷ The coordinate trajectories of each subunit in the tetramer were treated as independent trajectories and saved every 100 ps for analysis, resulting in a total simulation time of 80 μs (8×10^5 snapshots).

Markov State Models

The first step of estimating an MSM is to transform the trajectories in Cartesian coordinate space into trajectories in a preselected feature space which can capture the conformational changes of the protein. After various descriptors were tested, we found that methods only including the backbone atoms ($C\alpha$) of the protein were not sufficient to resolve different conformations of the protein. Thus, we defined a feature set to describe the contacts between a set of 28 active site residues which includes the active site loop (residues 98–110 and 112), important active site residues (residues 138–139, 168, 193–195), and the contacting helix (residues 237–239, 242–243, and 246–248). First, the minimum inter-residue distance r_{\min} , which is defined as the distance of the closest pair of heavy atoms between two residues, is calculated for each pair of the selected residues which are at least two residues apart from each other. The minimum inter-residue distance r_{\min} can capture the conformational changes

involving the backbones as well as the side chains of the protein. Then, instead of using a cutoff value (step function) to turn r_{\min} into a binary contact indicator, we used a continuous function

$$x = \frac{1}{2} \left(\tanh(r_{1/2} - r_{\min}) + 1 \right)$$

where \tanh is the hyperbolic tangent function and $r_{1/2} = 5 \text{ \AA}$, to transform r_{\min} into a contact value x between 0 and 1 which captures the degree of contact of two residues. Next, we exclude the pairs of residues which have no contact with each other ($x < 0.02$) through the whole trajectories. In this way, each snapshot of protein can be represented by a 261-dimensional vector \vec{x}_{prot} . Similar procedures were also applied to get the contact values between the carbonyl O of pyruvate and the selected active site residues, resulting in a 28-dimensional vector \vec{x}_{sub} which was used as the input features to build the MSM to resolve different substrate binding states.

For the next step, a time-lagged independent component analysis (TICA)^{28,29} was performed on the featurized trajectories to find the slow linear subspace of the input features, and each TICA component was scaled according to its corresponding eigenvalue to obtain a kinetic map³⁰ in which Euclidean distances are proportional to kinetic distances, providing an optimal space to perform clustering. When projecting the featurized trajectories, instead of keeping a selected number of TICA components, we chose to keep the TICA components which have eigenvectors that can account for 90% of the total variation in kinetic distance. Then, the k -means clustering method was employed to group the snapshots into microstates. Finally, the reversible transition matrix was estimated using the maximum likelihood estimator.³¹ All the estimations and analysis of MSM were done by using the PyEMMA 2 program.³²

We first analyzed the microstates obtained by using \vec{x}_{prot} combined with \vec{x}_{sub} as the input features, which contains the protein conformations as well as the substrate binding. While the microstates resolved different protein conformations, they did not distinguish different binding states of the substrate, because the protein motions are on much slower time scales than the substrate motions in the current simulation system. Thus, we obtained another set of 500 microstates by using only \vec{x}_{prot} as the input features, and then manually dividing each microstate by giving each snapshot an additional label which incorporated the substrate binding states. We determined the substrate binding states of each snapshot by estimating a MSM with 500 microstates using only \vec{x}_{sub} as the input features, and lumping the microstates into four metastable states with the Perron cluster cluster analysis (PCCA) method.^{33,34} The metastable states were used as the label describing the substrate binding states. Thus, each microstate obtained by using \vec{x}_{prot} as the input features was further divided into up to four possible new microstates, which yielded a new set of 758 microstates. We calculated the implied relaxation time scales as a function of the lag time τ which showed that the time scales become constant starting at a lag time of ~ 30 ns (Figure S4a). Therefore, $\tau = 30$ ns was used as the lag time to estimate the final MSM.

Coarse-Grained Kinetic Model

By using the same set of 758 microstates, we also estimated a coarse-grained kinetic model using a HMM.³⁵ First, the original MSM was coarse-grained into four metastable states using the PCCA+ method³⁶ to obtain an initial guess for the hidden transition matrix $\tilde{\mathbf{P}}$ the output probability matrix χ .³⁵ Then, maximum likelihood HMM was estimated using the Baum–Welch method^{37,38} as implemented in the PyEMMA 2 program. Implied relaxation time scale as a function of lag time for the HMM was also calculated (Figure S4c), which shows reasonably converged time scales except for the time scales at a lag time of 10 ns where the reversible transition matrix estimation did not converge. Because the HMM was estimated by coarse-graining the MSM, the lag time of the HMM is the same as the lag time of the MSM (30 ns).

The expected value and variance of observable A in hidden state i can be estimated as

$$\langle A \rangle_I = \sum_i \chi_{Ii} \langle A \rangle_i \quad (1)$$

$$\text{Var}(A)_I = \sum_i \chi_{Ii} (\langle A \rangle_i - \langle A \rangle_I)^2 \quad (2)$$

where χ_{Ii} is the output probability of hidden state I to observed state (microstate) i and $\langle A \rangle_i$ is the expected value of A in the observed state i , which can be calculated as the mean of A for all the members of i . When the metastable state I was further divided into substates I_n on the basis of the substrate binding states n , the output probability of I_n to observed state i is

$$\chi_{I_n i} = \begin{cases} \chi_{Ii} & \text{if } i \in n \\ 0 & \text{otherwise} \end{cases}$$

Then, the expected value and variance of observable A in substate I_n can be estimated using eqs 1 and 2. The transition probability between the substates is

$$\tilde{\mathbf{P}}' = \chi_{I,I_n}^T \tilde{\mathbf{\Pi}}_I \tilde{\mathbf{P}} \chi_{I,I_n} \quad (3)$$

where $\tilde{\mathbf{\Pi}}_I = \text{diag}(\pi_I)$, and π_I is the stationary probability of I . χ_{I,I_n} is the output probability of I to I_n , which can be calculated as

$$\chi_{I,I_n} = \sum_{i \in n} \chi_{I,i}$$

The stationary probability of I_n is

$$\pi_{I_n} = \sum_{i \in n} \pi_i \quad (4)$$

Bayes Factor Analysis

Bayes factor analysis³⁹ was used to identify the key contacts defining each metastable state. The Bayes factor for a specific contact between residues i and j in state k can be calculated as

$$BF_k(c_{ij}) = \frac{P(k|c_{ij}=1)}{P(k|c_{ij}=0)} = \frac{P(c_{ij}=1|k) P(c_{ij}=0)}{P(c_{ij}=0|k) P(c_{ij}=1)} \quad (5)$$

where c_{ij} is the contact indicator; i.e., $c_{ij} = 1$ if a contact is present between residues i and j and $c_{ij} = 0$ otherwise. Thus, for a specific contact in a specific state, the Bayes factor is the ratio of the probability of finding the system in this state given that the contact exists and the probability of finding the system in this state given that the contact does not exist. The Bayes factor can be seen as the degree of uniqueness for a contact to define a state. In this work, Bayes factors were calculated for the contacts between the same set of 28 active site residues used to define r_{\min} , and c_{ij} was calculated as a step function of $r_{\min}(i, j)$

$$c_{ij} = \begin{cases} 1 & \text{if } r_{\min}(i, j) < 4 \text{ \AA} \\ 0 & \text{otherwise} \end{cases}$$

In practice, the probability of observing $c_{ij} = 1$ in state k , $P(c_{ij} = 1|k)$, was estimated as the expected value of c_{ij} in state k using eq 1. The unconditional probability of observing $c_{ij} = 1$, $P(c_{ij} = 1)$, was estimated as the average of $P(c_{ij} = 1|k)$ over all k 's, weighted by the stationary distribution of k . Then, $P(c_{ij} = 0|k) = 1 - P(c_{ij} = 1|k)$ and $P(c_{ij} = 0) = 1 - P(c_{ij} = 1)$. To avoid the contacts with small $P(c_{ij} = 1|k)$ from showing large Bayes factors, $P(c_{ij} = 1|k)$ was regularized by 10^{-4} , which will suppress Bayes factors of rare contacts without affecting the others.

Electric Field Calculation

The electric field exerted onto the C = O stretching vibration of pyruvate by the environment (the whole system except for pyruvate) can be calculated from the electrostatic forces on the carbonyl atoms of pyruvate. Because the electrostatic forces are not readily available through the Tcl interface of the NAMD program, we adopted a similar strategy as in ref 15 to get the electrostatic forces on the carbonyl atoms for each snapshot using eqs 6–8.

$$\vec{f}_{\text{elec}}^i = \vec{f}_{\text{tot}}^i - \vec{f}_{\text{nonelec}}^i \quad (6)$$

$$\vec{F}^i = \vec{f}_{\text{elec}}^i / q_i \quad (7)$$

$$|F_{\text{vib}}| = \frac{1}{2} \left(\vec{F}^{\text{C}} \hat{u}_{\text{CO}} + \vec{F}^{\text{O}} \hat{u}_{\text{CO}} \right) \quad (8)$$

First, the snapshots were postprocessed using the same simulation parameters as in the MD simulations to extract the total forces on the C and O atoms of the carbonyl in pyruvate,

\vec{f}_{tot}^i , where i denotes the C or O atom. Then, the snapshots were postprocessed again with the partial charges of all atoms of the environment set to zero, and the total forces on the C and O atoms, $\vec{f}_{\text{nonelec}}^i$, which did not include the electrostatic forces exerted by the environment due to the lack of partial charges of the environment atoms, were also extracted. By taking the difference between the total forces from the two rounds of postprocessing calculations for the C or O atom (eq 6), we can get the electrostatic forces on the C or O atom exerted by the environment, \vec{f}_{elec}^i . The electric field at the position of the C or O atom, \vec{F}^i , can be calculated as the electrostatic force on the corresponding atom dividing by the partial charge q_i of the same atom (eq 7). Finally, the electric field experienced by the C = O stretching vibration, $|F_{\text{vib}}|$, is calculated by projecting the electric field at the C or O atom onto the unit vector along the C = O bond, \hat{u}_{CO} , and then averaging the two projections between the two atoms (eq 8).

RESULTS AND DISCUSSION

Multiple Metastable Substrate Binding States

Multiple substates of the Michaelis complex of LDH have been shown in experiment by vibrational spectroscopy. The C = O stretching vibration of the substrate pyruvate has been used as the vibrational probe in these experiments. The frequency shifts of the probe in different substates can show their propensities toward the chemical step, and can provide a direct measurement of the local electrostatic environments of the probe, which is ultimately determined by the interactions between the substrate and the protein/solvent such as hydrogen bonds. Thus, it is important to resolve the metastable substrate binding states in the MSM in order to compare with the experiments.

In this work, we first built an MSM on the basis of the contacts between the protein and the substrate only (the information about protein conformations not included explicitly; see Methods for details) and then lumped the microstates of the MSM into four metastable states. The first two metastable states are the same two binding states we found previously,²² the substrate binding state in the X-ray structure S_{xray} in which the carbonyl O of pyruvate forms hydrogen bonds with Arg106, Asn138, and His193 (Figures 1a and S1a) and the carboxyl O forms hydrogen bonds with Arg106, Arg169, and His193 (Figures 1a and S2a) and an alternative binding state S_{flip} in which pyruvate flips around its main chain by 180°,

the carbonyl O forms a hydrogen bond with Arg169 (Figures 1b and S1b), and the carboxyl O forms hydrogen bonds with Arg106, Asn138, Arg169, and His193 (Figures 1b and S2b). The third metastable state is a newly resolved binding state S_{rotate} in which the substrate rotates in the molecular plane by $\sim 90^\circ$ and shifts away from Arg169, the carbonyl O forms hydrogen bonds with Asn138 and a loop residue Gln100 (Figures 1c and S1c), and the carboxyl O forms hydrogen bonds with Arg106, Asn138, and His193 (Figures 1c and S2c). Furthermore, the methyl group of pyruvate also forms a nonpolar contact with Thr248 (Figures 1c and S3c) which could help in stabilizing the binding. In the last metastable state $S_{\text{a/u}}$, the substrate cannot form a metastable binding to the protein, and is only weakly associated with or unbound from it, so this state is very heterogeneous (Figures S1d, S2d, and S3d) in terms of substrate binding, and should not have significant contributions to the vibrational spectra of the probe.

The slowest time scales of the MSM are at sub-microseconds (Figure S4b). However, it should be noted that the MSM built upon the contacts between the enzyme and the substrate should only be considered to be qualitative, since the time scales of the slowest motions in the system actually correspond to the protein motions which is neglected deliberately in this model, so we did not draw any quantitative conclusions from this model and only used the metastable states as an additional label for the microstates built upon the enzyme conformations to resolve different binding states of the substrate (see Methods).

Microsecond Time Scale Protein Motions

By manually incorporating the substrate binding states into the discretization of the input feature space, we obtained a set of 758 microstates which were used to estimate a MSM. The MSM was further coarse-grained into a four-state HMM because there is a gap after the third relaxation time scales (Figure S4a) which indicates there are four metastable states in the model. Figure 2 shows the metastable states of the protein conformations. Representative structures for the metastable protein states are shown with the active site loop and its contacting helix highlighted and the crystal structure 1I0Z (in black) superimposed for comparison. By visual inspection, we found the overall structural differences between the metastable states mainly came from the active site loop region. At equilibrium, the vast majority (over 95%) of the system is in the open state S^{open} , while each of the remaining states only accounts for 1–2% of the whole population. The slowest conformational change in the system is governed by a relaxation time scale of $\sim 1 \mu\text{s}$, which corresponds to the transition between the open state S^{open} and the remaining (closed and half-open) states. This is in agreement with the previous experiments which showed the rate-limiting step in the turnover of LDH is the closure of the active site loop over the substrate binding pocket.⁴⁰ At the currently used lag time (30 ns), S^{closed} and S^{half} can interconvert with each other, while $S^{\text{half*}}$ cannot interconvert with S^{closed} or S^{half} directly without going through S^{open} . S^{half} and $S^{\text{half*}}$ have similar conformations of the active site loop in terms of the loop–helix distances, but the inter-residue contacts defining each state are different (Figure 3; see below).

To compare the overall structures of the metastable states, we calculated the HMM-weighted distributions of the minimum distances between the active site loop residues and the contacting helix in each metastable state (Figure 3a–d). In S^{closed} , the active site loop

fluctuates around its position in the crystal structure, and remains in the closed position, while, in S^{open} , the active site loop is in a wide open position compared with its position in the crystal structure, especially toward the end of the loop where the average minimum distances exceed 15 Å for Leu107 and Asn108. In both of the states S^{half} and $S^{\text{half*}}$, the active site loop is slightly more open than in the crystal structure.

To compare the structural features of the metastable states on a more detailed level, we calculated the Bayes factors³⁹ (using eq 5; see Methods) for all of the inter-residue contacts used to define the protein conformation features (261 dimensions) to find what set of contacts uniquely define each metastable state. Parts e–h of Figure 3 show the results of the Bayes factor analysis, in which the magnitudes of the Bayes factors for the contacts in each metastable state reflect the degrees of uniqueness of the contacts with respect to other metastable states. There are two sets of contacts uniquely defining S^{closed} (Figure 3e). One consists of the contact between Arg106 and Asn138 and the contact between Arg106 and His193, which also exist in the crystal structure. These residues are also important substrate binding residues. The other set of contacts that uniquely define S^{closed} is formed by the helix residue Tyr239 and the loop residues 103–106, which suggests that these inter-residue contacts play an important role in stabilizing the closed conformation of the protein. In S^{half} , the overall Bayes factors are quite similar to those in S^{closed} (Figure 3f), though the values are much lower for S^{half} , which suggests the similarities of the contacts in the two metastable states and is consistent with the fact that S^{closed} and S^{half} are kinetically close to each other (Figure 2). For $S^{\text{half*}}$ (Figure 3g), the contact with the largest Bayes factor is between two loop residues Gln100 and Leu107. There are also two contacts with relatively large Bayes factors: the contact between Ser105 and Asp195 and the contact between the loop residue Glu102 and the helix residue Gly246. There are no contacts with significant Bayes factors in S^{open} (Figure 3h), which suggests that the conformations in S^{open} are mostly unstructured, so there are no contacts that can uniquely define this metastable state.

Substates of the Michaelis Complex

Due to the separation of time scales between protein and substrate motions, the HMM mainly captured the protein motions; i.e., the substrate binding states are still mixed in each of the metastable protein states. To compare with the substates resolved in experiment, each of the four states in the HMM were further divided into up to four substates on the basis of the additional labels (substrate binding states) of the microstates, which resulted in 13 substates because some combinations of the substrate binding states and protein states were not observed in this model. Figure 4 shows the kinetic network of the substates, where the transition probabilities and stationary distribution are calculated using eqs 3 and 4. The majority of the population is in $S_{a/u}^{\text{open}}$ (~94%), which suggests that LDH is not a strong binder for pyruvate, whereas the populations of the remaining substates range from 0.1 to 1%. The substrate binding state S_{rotate} only exists in the metastable state S^{half} . The intermetastable transitions happen mainly when the substrate is in $S_{a/u}$ or S_{flip} , which are not reactive binding states, and mainly through $S_{a/u}^{\text{open}}$, which works like a central hub in the kinetic network. The reactive substates ($S_{\text{xray}}^{\text{closed}}$, $S_{\text{xray}}^{\text{half}}$, $S_{\text{xray}}^{\text{half*}}$, and $S_{\text{rotate}}^{\text{half}}$) cannot interconvert

directly without going through a nonreactive substate (see below), which supports the experimental observations.¹⁰

Direct Comparison with Substates Resolved by Vibrational Spectroscopy

To have a direct comparison with vibrational spectroscopy, the electric field exerted onto the vibrational probe (the C = O stretching vibration of pyruvate) for each snapshot was calculated using eqs 6–8, and then the expected value and variance of the electric field for each substate were estimated using eqs 1 and 2. By excluding the substates with the protein in the open state or the substrate in the associated/unbound state, the remaining states can be grouped into four clusters on the basis of their expected values of the electric field. If we assign the four clusters to the four peaks of the IR spectrum of the LDH·NADH·pyruvate complex (1673, 1679, 1686, and 1699 cm⁻¹),⁹ we can find a reasonably good linear correlation ($R^2 = 0.96$) between the peak vibrational frequencies and the ensemble-average electric field (Figure 5), which suggested that the substates resolved in our model could be assigned to the observed substates of the Michaelis complex. It should be noted that the correlation between the standard deviation of the electric field distribution and the IR band's line width is rather poor, as can be seen in Figure 5. One potential source of such large errors is that, even though the fixed-charge force field can describe the average properties of the electric field quite well, it is not adequate to capture the field heterogeneity. It has been suggested that reliable estimations of the IR line width for the C = O bond can be achieved by incorporating the polarizable force field.⁴¹ Another potential source is the limitation of the Stark effect theory itself due to its rather primitive form. More rigorous and well-parametrized models like the amide frequency maps,⁴² the solvatochromic charge model,⁴³ or the more recent SolEFP method⁴⁴ have been proposed to model the vibrational solvatochromism more versatily and reliably.

It has been suggested that each C = O band could potentially represent more than one substate,⁹ which is indeed the case as has been seen in our model. The C = O band with the largest frequency shift (1673 cm⁻¹) corresponds to $S_{\text{xray}}^{\text{closed}}$ in which the active site is highly structured with the loop closed and C = O of pyruvate is highly polarized by the three hydrogen bonds formed with Arg106, Asn138, and His193. The C = O band at 1679 cm⁻¹ corresponds to two substates $S_{\text{xray}}^{\text{half}}$ and $S_{\text{xray}}^{\text{half*}}$ in which the substrate is still bound in a similar fashion as in the crystal structure but the active site loop is more open due to the loss of some key contacts in the closed state. The hydrogen bonds between the carbonyl O of pyruvate and the residues Arg106, Asn138, and His193 were weakened because of the higher flexibility of the active site, so the C = O band in $S_{\text{xray}}^{\text{half}}$ and $S_{\text{xray}}^{\text{half*}}$ is less shifted than that in $S_{\text{xray}}^{\text{closed}}$. The C = O band at 1686 cm⁻¹ corresponds to $S_{\text{rotate}}^{\text{half}}$ in which the substrate rotates in the molecular plane and can only form two hydrogen bonds with Gln100 and Asn138, so the C = O is less polarized. The fitting of the C = O band at 1686 cm⁻¹ is slightly worse than those of the other three bands, which could be caused by the fact that the band near 1686 cm⁻¹ also contains the out-of-phase C=C stretch motions of the reduced nicotinamide moiety of NADH.⁹ The C = O band at 1699 cm⁻¹ can be assigned to the substates with the substrate flipped regardless of the protein state ($S_{\text{flip}}^{\text{closed}}$, $S_{\text{flip}}^{\text{half}}$, and $S_{\text{flip}}^{\text{half*}}$) in which the substrate flips around its main chain, and can only form one hydrogen bond with

Arg169, so it has the least frequency shift. It has been proposed that the observed substates at 1673, 1679, and 1686 cm^{-1} are all reactive, whereas the observed substate at 1699 cm^{-1} is nonreactive.¹⁰ On the contrary to our previous proposal in which the substates with the flipped substrate were assigned to one of the three reactive observed substates,²² by comparing the calculated electric field and IR bands directly in this work, we found that the substates with the flipped substrate actually corresponded to the nonreactive substate at the IR band of 1699 cm^{-1} .

CONCLUSIONS

The Markov and hidden Markov models have been shown to be a powerful tool to obtain a detailed kinetic network of the protein conformations and the coupled substrate bindings. When combining with the electric field calculations, they can provide a way to compare with the vibrational spectroscopy directly, which can help in interpreting and complementing the experiments on atomistic scales. In this work, we explored the conformational dynamics of LDH and its coupling to the interactions between LDH and pyruvate by combining MD simulations and Markov/hidden Markov models. Besides the associated/unbound state, three substrate binding states are resolved, which differ in the hydrogen bond network formed with the protein, including a previously unresolved binding mode in which the substrate rotates in the molecular plane by $\sim 90^\circ$ and shifts away from Arg169 compared to the crystal structure. For protein conformations, a hidden Markov model resolved four metastable states including closed state, two half-open states, and open state. The closed state resembles the crystal structure in terms of loop–helix distances and residue contacts in the active site, and the contacts between the helix residue Tyr239 and the loop residues 103–106 help in stabilizing the closed conformation. For the two half-open states, the loop–helix distances are slightly larger than the closed state. One half-open state is kinetically closer to the closed state, and has similar inter-residue contacts, while the other one has different contacts with the closed state. For the open state, the active site loop is fully open in terms of the loop–helix distances, and the conformations of the active site loop are unstructured.

To have a direct comparison with the vibrational spectroscopy, each of the metastable states was further divided into up to four substates according to the substrate binding states of the microstates, and the expected value of the electric field exerted onto the vibrational probe was calculated for each substate. We found a good linear correlation between the calculated electric field and the peaks of the infrared spectra, so we can assign the substates resolved in our model to the IR bands, and the structural features of substates corresponding to each IR band can be obtained. The IR bands with the largest three frequency shifts correspond to the binding state as in the crystal structure while the protein is in the closed state (1673 cm^{-1}), or in either of the two half-open states (1679 cm^{-1}), and the binding state with the rotated substrate while the protein is in one of the half-open states (1686 cm^{-1}). These substates are all reactive but differ in the propensities toward the chemical reaction according to experiments. The IR band with the smallest frequency shift (1699 cm^{-1}) corresponds to the binding state with the flipped substrate while the protein is not in the open states. The transition probabilities between the substates support the proposal that the reactive substates cannot interconvert directly without going through the nonreactive substates.

Our work supports the dynamical point of view of enzymatic catalysis⁴⁵ and suggests a methodology to determine the detailed kinetic network of the Michaelis complex on atomistic scales which can be directly compared with the vibrational spectroscopy. For LDH, there is heterogeneity in the Michaelis complex in terms of protein conformations and substrate bindings, and the femtosecond on-enzyme chemical step⁴⁶ is modulated by the microsecond protein conformational changes and the coupled nanosecond substrate binding mode changes. Thus, the conventional picture of enzymatic catalysis in which the on-enzyme chemical step is initiated from a single species, presumably the one resembling the crystal structures of the enzyme–substrate complex, might not provide the whole picture of the reaction. The methodology used in this work can be used to resolve all the distinct reactive substates which can be used as the starting structures in the study of the on-enzyme chemical step. Furthermore, the methodology used in this work can also be used in studying evolutionary adaptation of function such as adaptation to varying thermal environments, and allosteric regulation of enzyme, which are believed to be through the modulation of the Michaelis complex ground state ensemble distribution.⁴⁷

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

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

REFERENCES

- (1). Min W, English BP, Luo G, Cherayil BJ, Kou SC, Xie XS. Fluctuating Enzymes: Lessons from Single-Molecule Studies. *Acc. Chem. Res.* 2005; 38:923–931. [PubMed: 16359164]
- (2). Min W, Gopich IV, English BP, Kou SC, Xie XS, Szabo A. When Does the Michaelis-Menten Equation Hold for Fluctuating Enzymes? *J. Phys. Chem. B.* 2006; 110:20093–20097. [PubMed: 17034179]
- (3). Xie XS. Enzyme Kinetics, Past and Present. *Science.* 2013; 342:1457–1459. [PubMed: 24357307]
- (4). Gulotta M, Deng H, Deng H, Dyer RB, Callender RH. Toward an Understanding of the Role of Dynamics on Enzymatic Catalysis in Lactate Dehydrogenase. *Biochemistry.* 2002; 41:3353–3363. [PubMed: 11876643]
- (5). McClendon S, Vu DM, Clinch K, Callender R, Brian Dyer R. Structural Transformations in the Dynamics of Michaelis Complex Formation in Lactate Dehydrogenase. *Biophys. J.* 2005; 89:L07–L09. [PubMed: 15879476]
- (6). Deng H, Brewer S, Vu DM, Clinch K, Callender R, Dyer RB. On the Pathway of Forming Enzymatically Productive Ligand-Protein Complexes in Lactate Dehydrogenase. *Biophys. J.* 2008; 95:804–813. [PubMed: 18390601]
- (7). Zhadin N, Gulotta M, Callender R. Probing the Role of Dynamics in Hydride Transfer Catalyzed by Lactate Dehydrogenase. *Biophys. J.* 2008; 95:1974–1984. [PubMed: 18487309]
- (8). Deng H, Vu DV, Clinch K, Desamero R, Dyer RB, Callender R. Conformational Heterogeneity within the Michaelis Complex of Lactate Dehydrogenase. *J. Phys. Chem. B.* 2011; 115:7670–7678. [PubMed: 21568287]
- (9). Peng H-L, Deng H, Dyer RB, Callender R. Energy Landscape of the Michaelis Complex of Lactate Dehydrogenase: Relationship to Catalytic Mechanism. *Biochemistry.* 2014; 53:1849–1857. [PubMed: 24576110]

- (10). Reddish MJ, Peng H-L, Deng H, Panwar KS, Callender R, Dyer RB. Direct Evidence of Catalytic Heterogeneity in Lactate Dehydrogenase by Temperature Jump Infrared Spectroscopy. *J. Phys. Chem. B.* 2014; 118:10854–10862. [PubMed: 25149276]
- (11). Deng H, Zheng J, Clarke A, Holbrook JJ, Callender R, Burgner JW. Source of Catalysis in the Lactate Dehydrogenase System. Ground-State Interactions in the Enzyme-Substrate Complex. *Biochemistry.* 1994; 33:2297–2305. [PubMed: 8117687]
- (12). Park ES, Andrews SS, Hu RB, Boxer SG. Vibrational Stark Spectroscopy in Proteins: A Probe and Calibration for Electrostatic Fields. *J. Phys. Chem. B.* 1999; 103:9813–9817.
- (13). Choi J-H, Cho M. Vibrational Solvatochromism and Electrochromism of Infrared Probe Molecules Containing C≡O, C≡N, C = O, or C-F Vibrational Chromophore. *J. Chem. Phys.* 2011; 134:154513. [PubMed: 21513401]
- (14). Bublitz GU, Boxer SG. Stark Spectroscopy: Applications in Chemistry, Biology, and Materials Science. *Annu. Rev. Phys. Chem.* 1997; 48:213–242. [PubMed: 9348658]
- (15). Fried SD, Bagchi S, Boxer SG. Measuring Electrostatic Fields in Both Hydrogen-Bonding and Non-Hydrogen-Bonding Environments Using Carbonyl Vibrational Probes. *J. Am. Chem. Soc.* 2013; 135:11181–11192. [PubMed: 23808481]
- (16). Silva D-A, Bowman GR, Sosa-Peinado A, Huang X. A Role for Both Conformational Selection and Induced Fit in Ligand Binding by the LAO Protein. *PLoS Comput. Biol.* 2011; 7:e1002054. [PubMed: 21637799]
- (17). Gu S, Silva D-A, Meng L, Yue A, Huang X. Quantitatively Characterizing the Ligand Binding Mechanisms of Choline Binding Protein Using Markov State Model Analysis. *PLoS Comput. Biol.* 2014; 10:e1003767. [PubMed: 25101697]
- (18). Plattner N, Noé F. Protein Conformational Plasticity and Complex Ligand-binding Kinetics Explored by Atomistic Simulations and Markov Models. *Nat. Commun.* 2015; 6:7653. [PubMed: 26134632]
- (19). Núñez S, Wing C, Antoniou D, Schramm VL, Schwartz SD. Insight into Catalytically Relevant Correlated Motions in Human Purine Nucleoside Phosphorylase. *J. Phys. Chem. A.* 2006; 110:463–472. [PubMed: 16405318]
- (20). Pineda JRET, Callender R, Schwartz SD. Ligand Binding and Protein Dynamics in Lactate Dehydrogenase. *Biophys. J.* 2007; 93:1474–1483. [PubMed: 17483170]
- (21). Pineda JRET, Antoniou D, Schwartz SD. Slow Conformational Motions That Favor Sub-picosecond Motions Important for Catalysis. *J. Phys. Chem. B.* 2010; 114:15985–15990. [PubMed: 21077591]
- (22). Pan X, Schwartz SD. Free Energy Surface of the Michaelis Complex of Lactate Dehydrogenase: A Network Analysis of Microsecond Simulations. *J. Phys. Chem. B.* 2015; 119:5430–5436. [PubMed: 25831215]
- (23). Best RB, Zhu X, Shim J, Lopes PEM, Mittal J, Feig M, MacKerell AD. 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.* 2012; 8:3257–3273. [PubMed: 23341755]
- (24). MacKerell AD, Bashford D, Bellott M, Dunbrack RL, Evanseck JD, Field MJ, 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. [PubMed: 24889800]
- (25). 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. [PubMed: 19575467]
- (26). Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* 1983; 79:926–935.
- (27). Phillips JC, Braun R, Wang W, Gumbart J, Tajkhorshid E, Villa E, Chipot C, Skeel RD, Kale L, Schulten K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* 2005; 26:1781–1802. [PubMed: 16222654]

- (28). Pérez-Hernández G, Paul F, Giorgino T, Fabritiis GD, Noé F. Identification of Slow Molecular Order Parameters for Markov Model Construction. *J. Chem. Phys.* 2013; 139:015102. [PubMed: 23822324]
- (29). Schwantes CR, Pande VS. Improvements in Markov State Model Construction Reveal Many Non-Native Interactions in the Folding of NTL9. *J. Chem. Theory Comput.* 2013; 9:2000–2009. [PubMed: 23750122]
- (30). Noé F, Clementi C. Kinetic Distance and Kinetic Maps from Molecular Dynamics Simulation. *J. Chem. Theory Comput.* 2015; 11:5002–5011. [PubMed: 26574285]
- (31). Prinz J-H, Wu H, Sarich M, Keller B, Senne M, Held M, Chodera JD, Schütte C, Noé F. Markov Models of Molecular Kinetics: Generation and Validation. *J. Chem. Phys.* 2011; 134:174105. [PubMed: 21548671]
- (32). Scherer MK, Trendelkamp-Schroer B, Paul F, Pérez-Hernández G, Hoffmann M, Plattner N, Wehmeyer C, Prinz J-H, Noé F. PyEMMA 2: A Software Package for Estimation, Validation, and Analysis of Markov Models. *J. Chem. Theory Comput.* 2015; 11:5525–5542. [PubMed: 26574340]
- (33). Schütte C, Fischer A, Huisinga W, Deuffhard P. A Direct Approach to Conformational Dynamics Based on Hybrid Monte Carlo. *J. Comput. Phys.* 1999; 151:146–168.
- (34). Weber M. Improved Perron Cluster Analysis. ZIB-Report. 2003; 03–04
- (35). Noé F, Wu H, Prinz J-H, Plattner N. Projected and Hidden Markov Models for Calculating Kinetics and Metastable States of Complex Molecules. *J. Chem. Phys.* 2013; 139:184114. [PubMed: 24320261]
- (36). Deuffhard P, Weber M. Robust Perron Cluster Analysis in Conformation Dynamics. *Linear Algebra Appl.* 2005; 398:161–184.
- (37). Baum LE, Petrie T, Soules G, Weiss N. A Maximization Technique Occurring in the Statistical Analysis of Probabilistic Functions of Markov Chains. *Ann. Math. Stat.* 1970; 41:164–171.
- (38). Welch LR. Hidden Markov Models and the Baum-Welch Algorithm. *IEEE Inf. Theory Soc. Newsl.* 2003; 53:1–13.
- (39). Zhou G, Voelz VA. Using Kinetic Network Models To Probe Non-Native Salt-Bridge Effects on α -Helix Folding. *J. Phys. Chem. B.* 2016; 120:926–935. [PubMed: 26769494]
- (40). Dunn CR, Wilks HM, Halsall DJ, Atkinson T, Clarke AR, Muirhead H, Holbrook JJ. Design and Synthesis of New Enzymes Based on the Lactate Dehydrogenase Framework. *Philos. Trans. R. Soc., B.* 1991; 332:177–184.
- (41). Fried SD, Wang L-P, Boxer SG, Ren P, Pande VS. Calculations of the Electric Fields in Liquid Solutions. *J. Phys. Chem. B.* 2013; 117:16236–16248. [PubMed: 24304155]
- (42). la Cour Jansen T, Dijkstra AG, Watson TM, Hirst JD, Knoester J. Modeling the Amide I Bands of Small Peptides. *J. Chem. Phys.* 2006; 125:044312.
- (43). Lee H, Choi J-H, Cho M. Vibrational Solvatochromism and Electrochromism. II. Multipole Analysis. *J. Chem. Phys.* 2012; 137:114307. [PubMed: 22998262]
- (44). Błasiak B, Cho M. Vibrational Solvatochromism. III. Rigorous Treatment of the Dispersion Interaction Contribution. *J. Chem. Phys.* 2015; 143:164111. [PubMed: 26520502]
- (45). Callender R, Dyer RB. The Dynamical Nature of Enzymatic Catalysis. *Acc. Chem. Res.* 2015; 48:407–413. [PubMed: 25539144]
- (46). Basner JE, Schwartz SD. How Enzyme Dynamics Helps Catalyze a Reaction in Atomic Detail: A Transition Path Sampling Study. *J. Am. Chem. Soc.* 2005; 127:13822–13831. [PubMed: 16201803]
- (47). Hochachka, PW.; Somero, GN. *Biochemical Adaptation: Mechanism and Process in Physiological Evolution.* Oxford University Press; Oxford, U.K.: 2002.

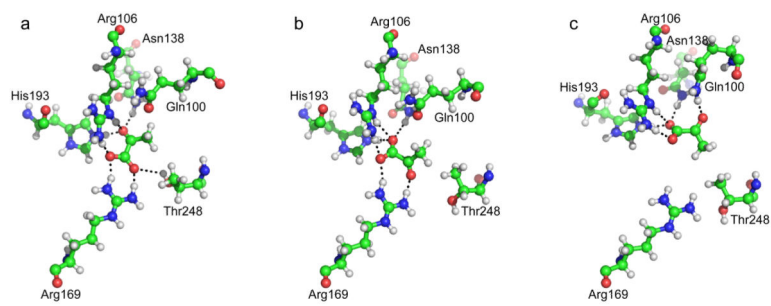


Figure 1. Representative structures from (a) the state of X-ray structure S_{xray} , (b) the flipped state S_{flip} , and (c) the rotated state S_{rotate} .

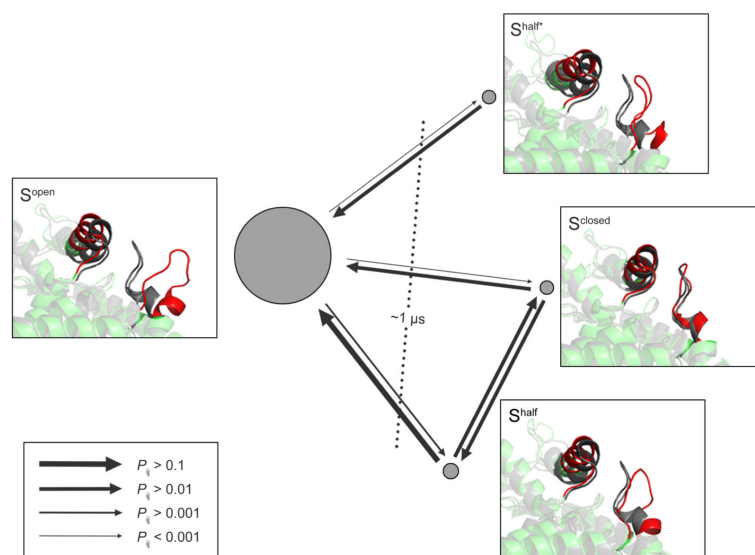


Figure 2. Representative structures, equilibrium distribution, and kinetics of the four metastable protein states. The arrowed lines represent the transition probabilities between the metastable states. The areas of the discs are proportional to the probabilities of the corresponding metastable states at equilibrium. The dashed line represents the slowest relaxation time scale and its corresponding transition process.

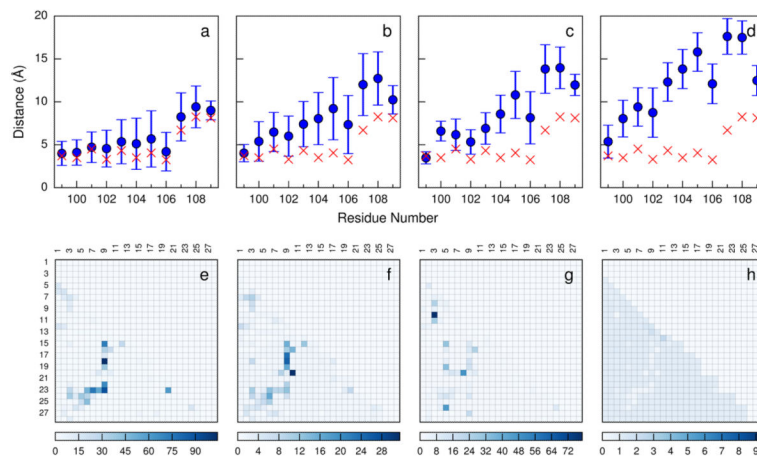


Figure 3. Structural features of the protein metastable states. HMM-weighted distributions of the minimum distances between the active site loop residues and the contacting helix in (a) S^{closed} , (b) S^{half} , (c) S^{half^*} , and (d) S^{open} . The red markers represent the minimum distances calculated from the crystal structure; Bayes factors for the contacts used to define the protein conformation features for (e) S^{closed} , (f) S^{half} , (g) S^{half^*} , and (h) S^{open} .

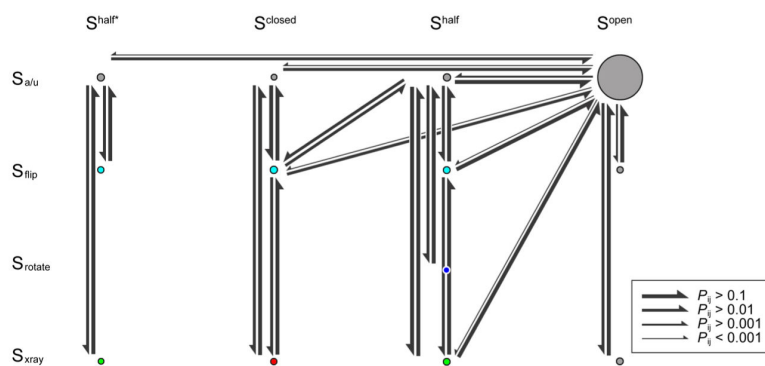


Figure 4.
Kinetic network of substates of the metastable states.

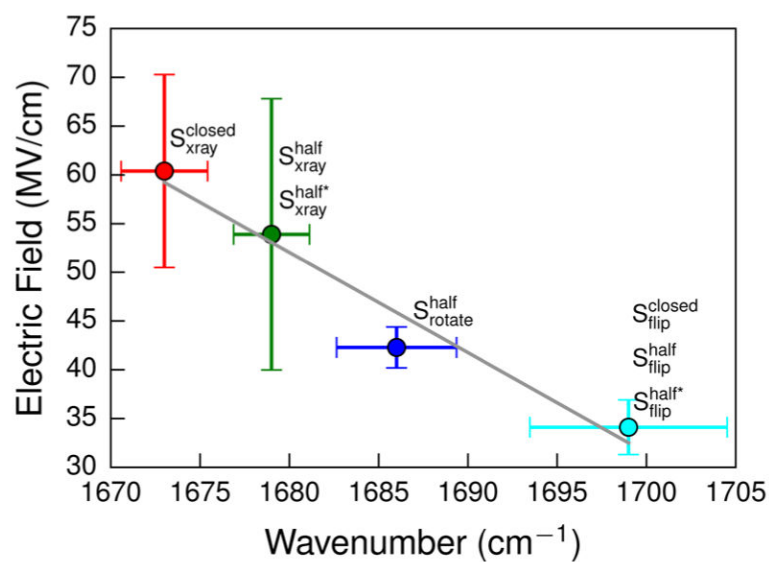


Figure 5. Correlation between calculated electric fields and the C = O stretching vibration of pyruvate for the substates of the metastable states. Error bars along the x - and y -axes correspond to the standard deviations of the IR spectrum and the electric field, respectively.