Dynamic coupling between coordinates in a model for biomolecular isomerization

Ao Ma, Ambarish Nag, and Aaron R. Dinner[a)]
Department of Chemistry, James Franck Institute, and Institute for Biophysical Dynamics, The University of Chicago, 5640 S. Ellis Avenue, Chicago, Illinois 60637

(Received 19 December 2005; accepted 13 February 2006; published online 14 April 2006)

To understand a complex reaction, it is necessary to project the dynamics of the system onto a low-dimensional subspace of physically meaningful coordinates. We recently introduced an automatic method for identifying coordinates that relate closely to stable-state commitment probabilities and successfully applied it to a model for biomolecular isomerization, the $C_{7\mathrm{eq}} \rightarrow \alpha_R$ transition of the alanine dipeptide [A. Ma and A. R. Dinner, J. Phys. Chem. B **109**, 6769 (2005)]. Here, we explore approximate means for estimating diffusion tensors for systems subject to restraints in one and two dimensions and then use the results together with an extension of Kramers theory for unimolecular reaction rates [A. Berezhkovskii and A. Szabo, J. Chem. Phys. **122**, 014503 (2005)] to show explicitly that both the potential of mean force and the diffusion tensor are essential for describing the dynamics of the alanine dipeptide quantitatively. In particular, the signficance of off-diagonal elements of the diffusion tensor suggests that the coordinates of interest are coupled by the hydrodynamic-like response of the bath of remaining degrees of freedom. © *2006 American Institute of Physics*. [DOI: 10.1063/1.2183768]

I. INTRODUCTION

The notion of a reaction mechanism inherently involves dynamics. The functions of biomolecules, in particular, depend on specific motions, as revealed by pioneering spectroscopic[1] and simulation[2] studies of the dissociation of diatomic ligands from myoglobin. For that protein and others, unprecedented levels of information about motions and their coupling are now becoming accessible experimentally due to technical advances in both single-molecule[3,4] and time-resolved ensemble (x-ray crystallographic[5] and multidimensional infrared[6,7]) measurements. Although molecular dynamics simulations already play an important role in interpreting such data,[5,6] continued improvements in the systematic analysis of trajectories are needed to identify correlation functions that enable direct connection between the calculated and observed dynamics.

The macroscopic parameter that traditionally provides the most direct information about dynamics is the rate. For many reactions, it is dominated by the behavior in a relatively small region of the phase space accessible to the system, the transition state ensemble. In principle, the transition state ensemble is defined by both coordinates and momenta, but, in practice, many biomolecular reactions are sufficiently diffusive that the latter can be ignored. The transition state ensemble thus reduces to the set of structures that lead with equal likelihood to the reactants and products in trajectories initiated with momenta drawn isotropically from a Maxwell-Boltzmann distribution (the so-called "stochastic separatrix"). The probabilities of commitment to the stable states can be calculated for structures off the stochastic separatrix as well and can be used as a measure of reaction progress. For a two-state reaction from basin A to basin B, we denote the probability of commitment to the latter basin by p_B ($p_A = 1 - p_B$); $0 \leqslant p_B \leqslant 1$ with structures on the stochastic separatrix having $p_B = 0.5$ by definition. In that it derives directly from dynamic information, p_B is a perfect reaction coordinate. At the same time, it is not in itself useful because it provides no direct insight into a mechanism.

Consequently, it is necessary to relate p_B to physically meaningful variables. This task can be quite challenging even for seemingly simple systems.[8–11] However, we recently made progress toward automating it by adapting means for obtaining quantitative structure-activity relationships (QSARs).[12] We used artificial neural networks to determine the functional dependence of p_B on sets of up to four coordinates from a database of candidate system descriptors and a genetic algorithm to select the combinations that gave the best fits. The fact that commitment probabilities for a representative set of structures need to be calculated only once permits the genetic neural network (GNN) method to test many combinations of coordinates efficiently, and the restriction of the database to variables chosen manually facilitates both the search and its subsequent interpretation. Application of the method to a model for biomolecular isomerization enabled us to identify a collective solvent coordinate that eluded other investigators.[9,10,12]

When the physically meaningful variables identified are adequate for describing the dynamics of interest, we expect the transition state ensemble to coincide with the saddle point in the projection of the free energy onto those coordinates. Satisfying this criterion is necessary but not sufficient for making quantitative connection with macroscopic parameters. The rate depends not only on the likelihood of reaching the barrier top (typically taken to be an exponential function

[a)]Fax: (773) 834-5250. Electronic mail: dinner@uchicago.edu

of the difference in free energy between transition states and reactants) but also the velocity along the reaction coordinate in the saddle point region (which enters as a prefactor to the exponential). Although this partitioning serves as the basis for many kinetic theories,[13–15] it has been explored explicitly for remarkably few complex systems.[16–18]

Consequently, in the present study, we introduce a general means for estimating the diffusion tensor in the reduced space of physically meaningful variables and use it together with a recent extension of Kramers theory[19] to estimate the rate for the $C_{7eq} \rightarrow \alpha_R$ isomerization of the alanine dipeptide in implicit solvent. In this "toy" system, there are two sequential backbone dihedral angles that are of interest, and the remaining 28 degrees of freedom (along with the friction term in the Langevin integrator) constitute the "bath." Good agreement is obtained with the rate calculated directly from simulations. The success of the procedure further validates the functional dependence of p_B on the coordinates obtained with the GNN method.[12] Much more importantly, we show for the first time in biomolecular simulations that off-diagonal elements of the diffusion tensor are essential for understanding the orientation of the stochastic separatrix in the low-dimensional space of interest, which makes clear the importance of dynamic coupling between the coordinates.

II. METHODS

The goal of the present paper is to explore how physical coordinates that relate closely to the commitment probability of a reaction can be used quantitatively and in so doing to gain insight into the nature of the dynamics of biomolecular isomerization. The basis for the procedure is a recent extension of Kramers theory for noise-activated dynamics.[19] Here, we review essential features of this reaction-rate theory and introduce means for calculating associated quantities: the diffusion tensor and the potential of mean force in the space of selected coordinates. Simulation details are given at the end of each subsection.

A. Kramers theory

The calculation of rate constants for condensed phase chemical reactions has a long history, most of which builds on the seminal work of Kramers[14] (reviewed in Hanggi *et al.*[15]). In such approaches, the basic form of the rate constant for a unimolecular process subject to sufficiently high friction is taken to be

$$k = \left[\left(\int_{\cup} e^{-\beta W(q)} dq \right) \left(\int_{\cap} \frac{e^{\beta W(q)}}{D(q)} dq \right) \right]^{-1}. \tag{1}$$

Here, q is a generalized (reaction) coordinate, $W(q)$ is the potential of mean force along that variable, $D(q)$ is a position-dependent diffusion constant, β is the inverse temperature in energy units, and $\cap$ and $\cup$ indicate that the integration is performed over the barrier and reactant basin regions in q, respectively. This expression is predicated on the idea that the rate is determined by diffusion of the system across a relatively high free energy barrier that separates the reactants and products. The physical assumption at the heart of such an approach is that the system can be meaningfully divided into a slowly varying reaction coordinate (q) and a rapidly varying and thus equilibrated bath [that enters Eq. (1) through D]. Our objective in the present study is to make this separation in a quantitative way for a biomolecular system based on systematic analysis of atomic-resolution dynamic data.

The Kramers theory was generalized to diffusive barrier crossing in a multidimensional space by Langer,[20] but Berezhkovskii and Szabo[19] showed recently that it is always possible to reduce such a representation to a one-dimensional one by linear combination of the pertinent coordinates when the barrier is parabolic. The direction of the resulting reaction coordinate is obtained by solving the equation

$$\mathbf{VDe} = \lambda \mathbf{e} \tag{2}$$

for the eigenvector $\mathbf{e}$ with negative eigenvalue λ. Above, $\mathbf{D}$ is the diffusion tensor in the multidimensional space, and $\mathbf{V}$ is the second-derivative matrix (Hessian) for the corresponding potential of mean force scaled by the temperature. In other words, $\beta W(\mathbf{q}) = \mathbf{q}^T \mathbf{V} \mathbf{q}/2$ in the saddle point region.[21] Rhee and Pande extended this work to consider systems without parabolic barriers and to emphasize the connection to commitment probabilities.[22] Berezhkovskii and Szabo also show that $\mathbf{e}$ is perpendicular to the stochastic separatrix ($p_B = 0.5$), which provides a bridge between approaches based on Kramers theory[15] and recent computational work.[23,24] Of particular importance for the present study, Eq. (2) makes clear that p_B is determined by both the potential of mean force and $\mathbf{D}$.

Given the eigenvector that defines the reaction coordinate, Eq. (1) can be recast for the multidimensional case,[19]

$$k(\mathbf{e}) = \frac{1}{2\pi} \left(\frac{\det \mathbf{V}_R}{|\det \mathbf{V}|} \right)^{1/2} \frac{\mathbf{e}^T \mathbf{D} \mathbf{e}}{|\mathbf{e}^T \mathbf{V}^{-1} \mathbf{e}|} e^{-\beta \Delta}. \tag{3}$$

Here, Δ is the difference in free energy between the saddle point and the bottom of the reactant basin and $\mathbf{V}_R$ is the Hessian at the position of the latter in the reduced space. To obtain the reaction coordinate and rate, we thus need to calculate the diffusion tensor and the potential of mean force in the space of slowly varying coordinates. Each of these is discussed in turn below.

B. Diffusion in the presence of a restraining potential

In simulations of homogenous liquids, the diffusion coefficient (or the diagonal diffusion tensor for an anisotropic bulk medium) is typically calculated from the Einstein relation,

$$\lim_{t \rightarrow \infty} \langle [\mathbf{q}(t) - \mathbf{q}(0)]^2 \rangle = 2dDt, \tag{4}$$

or the physically equivalent Green-Kubo formula,

$$D = \frac{1}{d} \int_0^{\infty} dt \langle \dot{\mathbf{q}}(t) \cdot \mathbf{q}(0) \rangle. \tag{5}$$

Above, d is the spatial dimension and t is the time. The use of these relations to determine the elements of the diffusion tensor for the Kramers theory is inappropriate in the present case because they do not account for the fact that the dynam-

ics are also governed by a potential of mean force.

Berne and co-workers[25–27] introduced a method (extended by Woolf and Roux[28]) in which the time-dependent friction is obtained as a perturbation to an analytically tractable (linear) model (see Appendix A of Hummer[29] for discussion). In this approach, a stiff harmonic restraint is applied to the reaction coordinate. Although the results do not depend on the specific choice of force constant for the restraint, care must be used to avoid changing the relative time scales of the dynamics of the system and the bath. Of greater concern for the present study, in which we limit ourselves to static friction by assumption, it is clearly desirable to obtain **D** by means that are more direct than integrating the time-dependent friction to avoid unnecessary numeric uncertainty (due to both the inherent noise of the integrand and the choice of integration range[29]).

The essence of a number of recent such methods is that **D** is chosen to match selected statistics calculated from Langevin dynamics and simulations in which the bath degrees of freedom are treated explicitly. In the most direct application of this idea, Berne and co-workers obtained the diagonal elements of the diffusion tensor for water molecules in the direction orthogonal to an air-water interface by comparing the survival probability for remaining in a predefined layer of the system in Langevin and molecular (Newtonian) dynamics simulations.[30] Those authors further showed that, although it required some correction, the static friction Langevin equation was adequate in that case. Hummer introduced an approach that is similar in spirit but employs Bayesian statistics to determine the position-dependent diffusion coefficient that gives the best fit between the observed dynamics and a diffusive dynamics defined by region-to-region hopping transitions for systems subject to strong friction.[29]

Either of the above methods can be generalized to obtain off-diagonal elements in a position-dependent manner because the diffusive dynamics are in effect simulated directly. However, they require determination of survival probabilities for multiple **D**, which can be computationally costly. Thus it is of interest to bound **D** as well as possible prior to resorting to such dual-simulation procedures. In this regard, the simple approach put forward by Im and Roux[31] is attractive. They assume that displacement arises only from Brownian forces and solve for the (homogeneous) diffusion coefficient,

$$D = \frac{\langle [\Delta q(t) - \langle \Delta q(t) \rangle]^2 \rangle}{2t}. \tag{6}$$

This expression requires that the time be sufficiently long that memory effects are negligible. In practice, t is selected empirically. In the case that the system is bounded, however, it cannot be too long because D goes to zero as the denominator approaches infinity and the numerator remains finite.

Since an appropriate choice of t in Eq. (6) is not obvious, it is instructive to consider an exactly solvable model to assess how varying this parameter impacts the estimated diffusion constant; we explore a one-dimensional harmonic oscillator governed by Langevin dynamics in Appendix A. Motivated by the fact that the potential of mean force in the space of slow variables influences their velocities over finite time intervals, we also study an alternative correlation function,

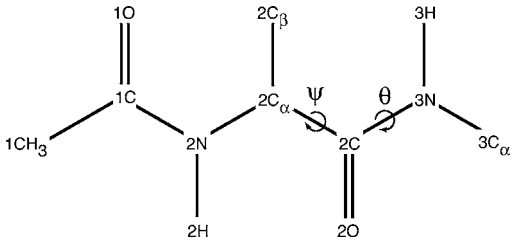

FIG. 1. Structure of the alanine dipeptide with the coordinates identified by the GNN method indicated. We used a polar-hydrogen representation (Ref. 41) and truncated the nonbonded interactions in a manner consistent with the parametrization of the implicit solvent term (Ref. 32). Unless otherwise noted, in all the dynamics calculations, the positions of the atoms were updated with the Langevin integrator in CHARMM (Ref. 42) with friction coefficients from Smith and Pettitt (Ref. 43) and a 1 fs time step.

$$C_q(t) = \langle \dot{q}(0)[q(t) - q(0)] \rangle. \tag{7}$$

Both Eqs. (6) and (7) yield lower bounds for D, but the latter reaches a higher maximum more quickly for the exactly solvable model. In general, we expect the use of the maximum of $C_q(t)$ to be the preferable method for estimating the diffusion constant because many potentials of mean force are approximated well by harmonic functions. In Appendix B, we show that, for sufficiently high friction, an analogous expression is appropriate for two-dimensional cases. In other words,

$$D_{ij} = \max \langle \dot{q}_i(0)[q_j(t) - q_j(0)] \rangle \tag{8}$$

for coordinates q_i and q_j with $i,j \in \{1,2\}$. Because the differences between Eqs. (6) and (7) diminish as the friction increases, it is likely that a similar extension of the former yields comparable accuracy, but that approximation is much more cumbersome to explore analytically.

As mentioned in the Introduction, we study the $C_{7\mathrm{eq}} \rightarrow \alpha_R$ isomerization of the alanine dipeptide with the solvent represented implicitly by the analytical continuum solvent (ACS) term[32] in CHARMM (Ref. 33) (Fig. 1). Here, we take the α_R or A basin to include structures with $-110° \leqslant \psi \leqslant 0°$ and the $C_{7\mathrm{eq}}$ or B basin to include structures with $60° \leqslant \psi \leqslant 180°$; for both the reactant and product, $-180° \leqslant \phi \leqslant 0°$. We modified the definitions of the basins slightly from those used in Ma and Dinner[12] to better ensure that paths entering a basin were committed to it. With the definitions given above, new reactive paths were harvested with transition path sampling,[24,34] commitment probabilities were evaluated for representative structures, and the genetic neural network procedure was reapplied exactly as in Ma and Dinner.[12] The two coordinates that in combination yield the best prediction for p_B are again the dihedral angles ψ (2N-2C$_\alpha$-2C-3N) and θ (2O-2C-3N-3C$_\alpha$).[12]

We calculate the diffusion tensor for $\mathbf{q} = (\psi, \theta)$ by the following procedure. For each configuration in the anticipated saddle point region [$20° < \psi < 40°$ and $-10° < \theta < 10°$ (Ref. 12)], we initiate 400 000 trajectories of 80-fs length from which we calculate over all possible starting points for a moving window of length $t < 2$ fs the averages

$\langle\dot{\psi}(0)[\psi(t)-\psi(0)]\rangle$, $\langle\dot{\psi}(0)[\theta(t)-\theta(0)]\rangle$, and $\langle\dot{\theta}(0)[\theta(t)-\theta(0)]\rangle$ for each $\psi\theta$ pair. For these averages, it is necessary to bin data into finite width (2°) intervals in ψ and θ; in this regard, it is important to note that the ψ and θ values to which the data from a trajectory segment are assigned are $\psi=[\psi(0)+\psi(t)]/2$ and $\theta=[\theta(0)+\theta(t)]/2$, not $\psi(0)$ and $\theta(0)$. Given the averages, the maximum values are identified and used as the elements of **D** in Eqs. (2) and (3). In principle, even better estimates could be obtained by generalizing procedures like those of Hummer[29] and Liu *et al.*[30] to treat off-diagonal elements of the diffusion tensor in two (or more) dimensions, but the results presented below suggest that our simple approach is adequate for the system of interest in the present study.

C. Potential of mean force

In contrast to the diffusion tensor, there are many well-established means for calculating the free energy as a function of a relatively small number of coordinates. We sample configurations using Monte Carlo (MC) because we showed recently that it permits much more rapid convergence of equilibrium averages for peptides than does molecular dynamics.[35] MC has the further advantage that energy terms and acceptance criteria based on arbitrary coordinates can be more readily incorporated because derivatives are not needed in general, although this limitation of molecular dynamics is not an issue for the dihedral angle coordinates that we consider in the present study.

The specific MC algorithm that we use is the Wang-Landau extension of the multicanonical approach.[36–38] The essential idea is that one adds to the energy function (E) a biasing term that is equal to the opposite of the free energy $[U(\mathbf{q})=-W(\mathbf{q})]$. Then, a simulation governed by the acceptance criterion,

$$P_{\mathrm{acc}}(\mathbf{x}\rightarrow\mathbf{x}')=\min\left\{1,\frac{\exp[-\beta(E(\mathbf{x}')+U(\mathbf{x}';\mathbf{q}'))]}{\exp[-\beta(E(\mathbf{x})+U(\mathbf{x};\mathbf{q}))]}\right\}, \tag{9}$$

samples different positions in the low-dimensional space of interest to the same extent. In other words, a histogram of the **q** values visited is flat.

Of course, $W(\mathbf{q})$ is not known *a priori*. As a result, it is necessary to guess a form for $U(\mathbf{q})$ and improve it iteratively. In practice, two arrays are carried during the simulation: one that contains $U(\mathbf{q})$ in a tabulated form and one that monitors the sampling at any given phase of the simulation, $N(\mathbf{q})$. At each step, a random change to the system configuration is made, the criterion in Eq. (9) is applied; then, regardless of whether the step is accepted or rejected, for the value of **q** kept at the end of the step, $U(\mathbf{q})$ is increased by an amount f, and $N(\mathbf{q})$ is incremented by one. Periodically, the standard deviation over all the $N(\mathbf{q})$ bins is calculated and, if it is less than a set fraction of the mean of $N(\mathbf{q})$, f is decreased by a constant factor until a desired tolerance is reached.

We implemented this procedure in the MC module[39,35] of the program CHARMM (Version 31a1).[33] At each step, isotropic single atom displacements of up to 0.075 Å were selected to be applied with a probability of 0.75, and otherwise, an individual dihedral angle was rotated by up to 30°. The corresponding acceptance rates were 0.42 and 0.56, respectively. The system was constrained to remain in the rectangular region $-7°<\psi<64°$ and $-18°<\theta<24°$, which was expected to contain the saddle point based on the predicted commitment probabilities obtained from the trained neural network and the values of ψ and θ sampled by typical transition paths.[12] Initially, $U(\mathbf{q})=0$ and $f=0.5k_BT$; k_B is the Boltzmann constant and T is the temperature, 300 K. Every 2×10^5 MC steps, the standard deviation of $N(\mathbf{q})$ was calculated as described above, and, if it was less than 1% of the mean, f was reduced by a factor of 2. We terminated the simulation when $f=1\times10^{-7}k_BT$. The parts of the resulting potential of mean force within 1 kcal/mol of the saddle point were fit by a quadratic function in ψ and θ and the coefficients were used to evaluate the Hessian.

D. Direct calculation of the rate

By definition, we can quantify the extent of transition from one basin (A) to another (B) with the correlation function,

$$C(t)=\frac{\langle h_A(\mathbf{x}_0)h_B(\mathbf{x}_t)\rangle}{\langle h_A(\mathbf{x}_0)\rangle}. \tag{10}$$

Here, $h_X(\mathbf{x}_t)=1$ for $\mathbf{x}$ in basin X ($X=A$ or B) at time t, and $h_X(\mathbf{x}_t)=0$ otherwise. The reaction rate constant is the plateau of the time derivative of this function.[13,24,34] Elegant means for evaluating $C(t)$ through umbrella sampling in path space are available,[24,34] but straightforward molecular dynamics simulations are sufficient for the present system. Specifically, we initiated 1500 independent trajectories of 5-ns length from structures taken from a 300 K canonical MC simulation in basin A. Momenta were drawn from the Maxwell-Boltzmann distribution at 300 K and equilibrated for 10 ps. Then, we evaluated $h_A(\mathbf{x}_0)h_B(\mathbf{x}_t)$ and $h_A(\mathbf{x}_0)$ for $\mathbf{x}_0$ distributed every 50 fs and $\mathbf{x}_t$ with $0\leqslant t\leqslant5$ ps. The ratio of these two quantities was computed, and the slope of $C(t)$ was obtained from linear regression over points with $t>2.5$ ps.

III. RESULTS AND DISCUSSION

As discussed above, the goal of the present study is to connect the commitment probability (p_B) with Kramers theory[14,15,19,20] explicitly through a specific system and, in so doing, to gain insight into the nature of the dynamics in a reduced space defined by coordinates identified by automatic means.[12] To this end, we study the $C_{7\mathrm{eq}}\rightarrow\alpha_R$ isomerization of the alanine dipeptide with the solvent treated implicitly by the ACS potential[32] (Fig. 1). As discussed in Ma and Dinner[12] and in Sec. II B, we used a GNN procedure to identify coordinates that in combination predict p_B with good accuracy and found the best two to be the dihedral angles ψ and θ (Fig. 1). Below, we denote the numerically determined observed commitment probability by p_B and the analytical function that corresponds to the trained neural network by p_B^{GNN}.

In the Kramers theory, the rate is dominated by the dynamics in the transition state region, which we take here to be the rectangular region of points with $15°\leqslant\psi\leqslant45°$ and

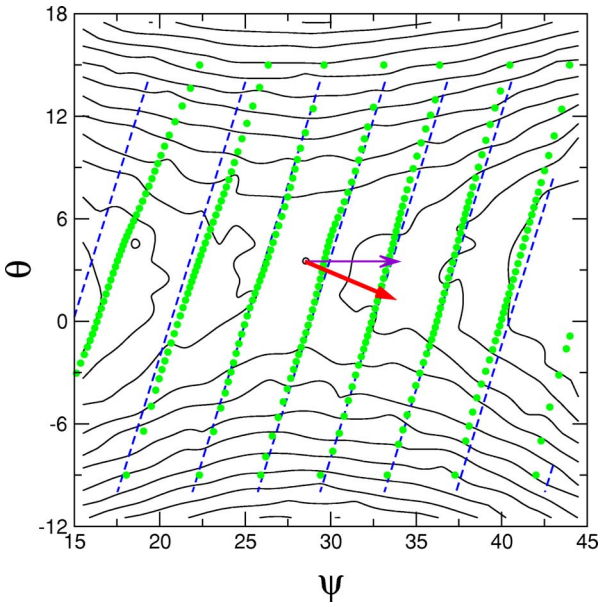

FIG. 2. Comparison of dynamics with the potential of mean force. Contours of $0.1k_BT$ in the potential of mean force are indicated by solid lines. Dashed lines are predicted isocommitor contours, which range from $p_B=0.1$ on the left to $p_B=0.8$ on the right in intervals of 0.1. Exact isocommitors obtained as described in the text are indicated by filled circles. The directions of the eigenvectors obtained by diagonalizing either **V** or **VD** are indicated by the thin and thick arrows, respectively. Angles are measured in degrees.

$-12° \leqslant \theta \leqslant 18°$. In this region, the predicted commitment probability varies over the range of $0.1 < p_B < 0.8$. To verify that this region did indeed contain low free energy points close to the stochastic separatrix, we evaluated p_B for 180 000 configurations in this region randomly selected from a 300 K canonical MC simulation of 18 000 000 MC steps using the move set described in Sec. II C. To determine the lines of constant commitment probability (isocommitors), we partitioned the region into a grid with 2° spacing in each direction and averaged the p_B values for the structures at each site. Consistent with the low error in p_B^{GNN} demonstrated for this system,[12] the agreement between the observed and predicted commitment probabilities is excellent (Fig. 2).

We obtained the potential of mean force in the transition state region as described in Sec. II C (Fig. 2). There is indeed a saddle point in this region at $\psi=27.6°$ and $\theta=1.7°$, and the dependence on the coordinates is almost perfectly quadratic (the rms error for the quadratic fit is about $0.01k_BT$). However, the eigenvector with imaginary frequency obtained by diagonalizing the Hessian at this point is essentially in the ψ direction,

$$\mathbf{e}_V = 0.999\,98\hat{\psi} - 0.006\,11\hat{\theta}. \tag{11}$$

In other words, examination of the potential of mean force by itself obscures the importance of the second coordinate (θ) in determining the dynamics.

To make clear the role played by θ, it is necessary to incorporate the diffusion tensor, as suggested by Eq. (2), which follows from the Kramers-Langer-Berezhkovskii-Szabo theory.[19] We thus used the procedure that we introduced in Sec. II B to calculate **D** in a position-dependent manner. Because each of the elements of the diffusion tensor

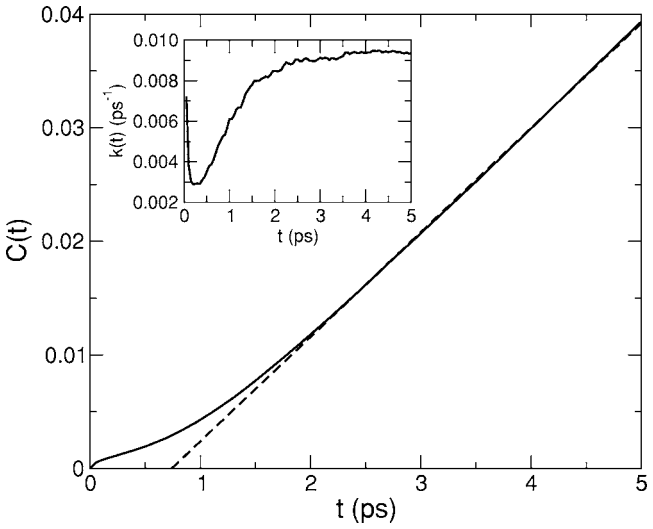

FIG. 3. Correlation function $C(t)$ [Eq. (10)] (solid line) and linear regression over points with $t>2.5$ ps (dashed line). (Inset) Numerical derivative of $C(t)$ to obtain the time-dependent rate coefficient $k(t)$.

was found to vary by only 5.6% at most from its mean over the region examined and paths are expected to deviate from the calculated saddle point, we averaged over positions to obtain

$$D_{\psi\psi} = 0.34 \text{ rad}^2 \text{ ps}^{-1}, \quad D_{\theta\theta} = 0.65 \text{ rad}^2 \text{ ps}^{-1},$$
$$D_{\psi\theta} = D_{\theta\psi} = 0.30 \text{ rad}^2 \text{ ps}^{-1}. \tag{12}$$

The value of $D_{\psi\psi}$ is comparable to but slightly higher than a published estimate for diffusion along the ψ coordinate of the alanine dipeptide in the presence of explicit water molecules.[29] More interestingly, in the two-dimensional case considered here, there is a significant off-diagonal element. As a result, the reactive eigenvector of the product **VD** is rotated in relation to $\mathbf{e}_V$ in Eq. (11),

$$\mathbf{e} = 0.923\hat{\psi} - 0.384\hat{\theta}. \tag{13}$$

This vector deviates by only 5.5° from one orthogonal to the stochastic separatrix ($p_B=0.5$).

To compare estimates of the rate obtained with and without θ to that obtained by more direct means (Sec. II D), we used the Wang-Landau method (Sec. II C) to calculate one-dimensional potentials of mean force along the lines defined by **e** and $\hat{\psi}$. The respective dimensionless barrier heights are $\beta\Delta_{\mathbf{e}}=3.49$ and $\beta\Delta_{\psi}=3.40$; substitution of these values into Eq. (3) gives $k=0.0110$ ps^{-1} and $k=0.0106$ ps^{-1}. Thus both estimates are in good agreement with $k=0.0095$ ps^{-1} obtained by direct means (Fig. 3). In considering these results, it is necessary to keep in mind that accurate rates can be obtained using arbitrary order parameters, so long as the reactive flux correlation function for trajectories initiated from the barrier in the corresponding projection of the free energy converges in a tractable amount of computational time.[13,40] Because ψ by itself can be used to predict p_B with some accuracy (see Fig. 9 of Ma and Dinner[12]), that variable provides a reasonable estimate of the rate even without a transmission coefficient. Thus, in this system at least, the direction of the reactive eigenvector (which is perpendicular to

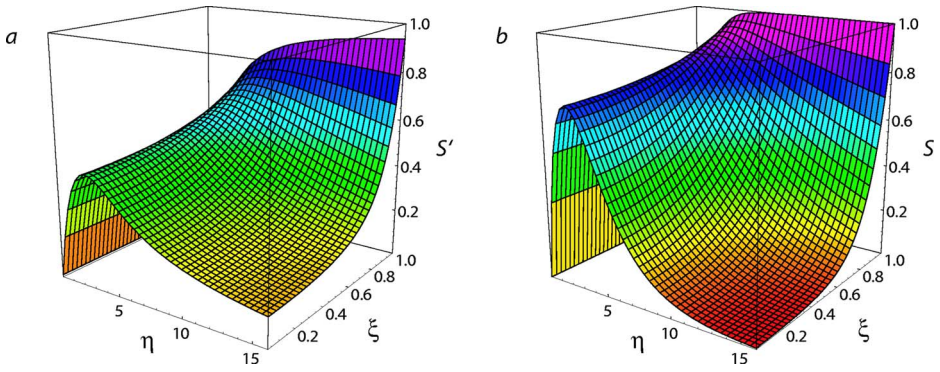

FIG. 4. Comparison of (a) $S'(\xi,\eta)$ defined by Eq. (A3) and (b) $S(\xi,\eta)$ defined by Eq. (A4) for an overdamped Langevin harmonic oscillator [Eq. (A1)] for dimensionless oscillator parameters ξ and η.

the stochastic separatrix) provides much more sensitive discrimination between dynamic models than does the rate.

Rate theories that allow for anisotropic diffusion[19] typically consider only the case of a diagonal tensor with nonequal nonzero elements. Here, detailed analysis of an atomic-resolution model of a molecular system demonstrates that the diffusion tensor for coordinates that best describe the dynamics has significant off-diagonal elements. Given their importance for obtaining a reactive eigenvector (**e**) that is roughly orthogonal to the observed stochastic separatrix, it is of interest to consider a "hydrodynamic" physical interpretation. In the system studied, degrees of freedom other than ψ and θ are considered part of the bath. The value of $D_{\psi\theta}$ suggests that when one of the coordinates of interest moves, it creates a drag on the other through the remaining degrees of freedom. To the best of our knowledge, this effect has not been quantitatively documented previously in the context of chemical reaction dynamics.

IV. CONCLUSION

In summary, in this paper we used a recent extension by Berezhkovskii and Szabo[19] of the Kramers-Langer theory[14,15,20] to investigate the meaning of a secondary reaction coordinate and its relationship to the primary one for the isomerization reaction of the alanine dipeptide in implicit solvent. We observed that, although there is no coupling between the two coordinates in the potential of mean force in this basis, the secondary coordinate motion significantly influences the primary one through the perturbation to the bath of remaining degrees of freedom. Although the current model bears only limited relevance to systems of biological interest due its small size and the use of implicit solvent, we emphasize that it was not constructed to produce such a dynamic effect. Rather, it arises naturally from an energy function and an integrator representative of those used to treat larger molecules. It will clearly be of interest to determine the extent of similar behavior in such systems.

ACKNOWLEDGMENTS

The authors wish to thank Karl Freed, Richard Stratt, Benoît Roux, and Gerhard Hummer for helpful suggestions regarding the work and Albert Pan for a critical reading of the manuscript.

APPENDIX A: COMPARISON OF APPROXIMATIONS FOR *D*

Here, we exploit the fact that we can analytically calculate correlation functions for a one-dimensional harmonic oscillator subject to Langevin dynamics to compare different methods for estimating the diffusion constant along the reaction coordinate. The equation of motion is

$$\ddot{q}(t) = -\omega^2 q(t) - \gamma\dot{q}(t) + \frac{1}{m}R(t), \tag{A1}$$

where ω is the harmonic frequency, γ is a collision frequency (related to the diffusion constant through $D = k_B T/m\gamma$), and $R(t)$ is a random force. This differential equation can be solved using Laplace transforms and gives

$$q(t) = q(0)e^{-\gamma t/2}\left[\cos(\Omega t) + \frac{\gamma/2}{\Omega}\sin(\Omega t)\right] + \frac{\dot{q}(0)}{\Omega}e^{-\gamma t/2}\sin(\Omega t) + \frac{1}{m\Omega}\int_0^t R(t-z)e^{-\gamma z/2}\sin(\Omega z)dz. \tag{A2}$$

Above, the parameter $\Omega = \sqrt{\omega^2 - \gamma^2/4}$ is a renormalized frequency. Because we are interested in motion along slow coordinates in the presence of fast bath degrees of freedom, we study the case that $\gamma/2 > \omega$. As a result, Ω is imaginary; the trigonometric functions become hyperbolic and do not oscillate. In other words, the system is overdamped.

Substituting the solution into Eq. (6),

$$\frac{\langle[\Delta q(t) - \langle\Delta q(t)\rangle]^2\rangle}{2t} = D\frac{1}{\xi^2\eta}\left[e^{-\eta}\cosh\xi\eta - (2 - e^{-\eta}) + \frac{1}{1-\xi}(1 - e^{-(1-\xi)\eta}) + \frac{1}{1+\xi}(1 - e^{-(1+\xi)\eta})\right] = DS'(\xi,\eta). \tag{A3}$$

where S' is defined by the above equation. Here, $\xi = -2i\Omega/\gamma$ $(0 < \xi \leqslant 1)$ characterizes the difference between the system and bath motion time scales, and $\eta = \gamma t$ measures the time in units defined by the bath frequency. These two dimensionless variables enable us to plot S' independent of

the specific choices of ω and γ [Fig. 4(a)]. The closer S' is to one, the closer the calculated diffusion coefficient is to the input value. For $\xi<1$ and finite η, $S'<1$. Along the $\xi=1$ axis ($\omega \ll \gamma$), $S' \to 1$ as $\eta \to \infty$ ($t \to \infty$), and along the $\xi=0$ axis ($\omega \simeq \gamma/2$) $S' \to 0$ as $\eta \to \infty$ as argued above. For all ξ, the best value of D is obtained from the maximum in Eq. (A3) as a function of time.

Alternatively, we substitute the solution in Eq. (A2) into Eq. (7). Doing so yields terms with static coefficients involving $\langle \dot{q}(0) \rangle$, $\langle q(0)\dot{q}(0) \rangle$, and $\langle \dot{q}(0)\dot{q}(0) \rangle$, but only the last of these is nonzero for isotropically distributed $\dot{q}(0)$. Thus

$$C_q(t) = D\left[\frac{2}{\xi} e^{-\eta/2} \sinh(\xi\eta/2)\right] = DS(\xi,\eta). \tag{A4}$$

In analogy to Eq. (A3), S [Fig. 4(b)] is defined by the above equation. For $\xi=1$ ($\omega \ll \gamma$), the maximum of $S(\xi,\eta)$ occurs as $\eta \to \infty$. Along that axis,

$$C_q(t) = D(1 - e^{-\eta/2}), \tag{A5}$$

which becomes D as $\eta \to \infty$. On the other hand, for $\xi=\epsilon \to 0$ ($\omega \simeq \gamma/2$),

$$C_q(t) = \frac{2D}{e} \simeq 0.735D, \tag{A6}$$

which provides a lower limit for the diffusion constant. In practice, D is obtained from the maximum of S or S' for a given ξ (Fig. 4). Because the maximum in S is higher and occurs at an earlier time than that in S', Eq. (7) is preferable to Eq. (6) for the exactly solvable model considered.

APPENDIX B: TWO-DIMENSIONAL DIFFUSION TENSOR

In this appendix, we extend the approach introduced in Sec. II B and Appendix A to two dimensions. To this end, we consider a system governed by the potential of mean force,

$$W(\mathbf{q}) = \frac{1}{2}H_1' q_1^2 + \frac{1}{2}H_2' q_2^2 + H_c' q_1 q_2, \tag{B1}$$

where the components of $\mathbf{q}=(q_1,q_2)$ are mass-weighted coordinates and primes denote the original basis. Neglecting the random force, which does not contribute to relevant time correlation functions, the equation of motion is

$$\ddot{\mathbf{q}}(t) = -\mathbf{H}' \cdot \mathbf{q}(t) - \mathbf{\Gamma}' \cdot \dot{\mathbf{q}}(t), \tag{B2}$$

with

$$\mathbf{H}' = \nabla W(\mathbf{q}) = \begin{pmatrix} H_1' & H_c' \\ H_c' & H_2' \end{pmatrix}, \tag{B3}$$

and

$$\mathbf{\Gamma}' = \begin{pmatrix} \gamma_1' & \gamma_c' \\ \gamma_c' & \gamma_2' \end{pmatrix}. \tag{B4}$$

The friction tensor $\mathbf{\Gamma}'$ is related to the diffusion tensor $\mathbf{D}$ through the generalized Einstein relation $\mathbf{D}=k_BT(\mathbf{\Gamma}')^{-1}$.

We now define $\mathbf{U}$ such that it diagonalizes $\mathbf{\Gamma}'$. Then,

$$\mathbf{U} \cdot \mathbf{\Gamma}' \cdot \mathbf{U}^{-1} = \mathbf{\Gamma} = \begin{pmatrix} \gamma_1 & 0 \\ 0 & \gamma_2 \end{pmatrix}, \tag{B5}$$

and

$$\mathbf{U} \cdot \mathbf{H}' \cdot \mathbf{U}^{-1} = \mathbf{H} = \begin{pmatrix} H_1 & H_c \\ H_c & H_2 \end{pmatrix}. \tag{B6}$$

Left multiplying both sides of Eq. (B2) by $\mathbf{U}$, we obtain the equation of motion in the transformed coordinates described by $\mathbf{y}=\mathbf{U}\cdot\mathbf{q}$:

$$\ddot{\mathbf{y}} = -\mathbf{H} \cdot \mathbf{y} - \mathbf{\Gamma} \cdot \dot{\mathbf{y}}. \tag{B7}$$

As in the one-dimensional case, the formal solution to this equation can be obtained using Laplace transforms. Up to terms proportional to $y_i(0)$, which vanish upon averaging, it is

$$y_i(t) = \frac{\dot{y}_i(0)}{\Omega_i} e^{-\gamma_i t/2} \sin(\Omega_i t) - H_c \int_0^t d\tau \frac{1}{\Omega_i} e^{-\gamma_i (t-\tau)} \times \sin[\Omega_i (t-\tau)] y_j(\tau), \tag{B8}$$

where $\Omega_i = \sqrt{H_i - \gamma_i^2/4} \simeq i\gamma_i/2$ for $H_i \ll \gamma_i^2$. In other words, we assume that the bath relaxes quickly relative to the system dynamics. The resulting correlation functions of intererest are

$$\langle y_i(t)\dot{y}_i(0) \rangle = \frac{k_BT}{\Omega_i} e^{-\gamma_i t/2} \sin(\Omega_i t) - H_c \int_0^t d\tau \frac{1}{\Omega_i} e^{-\gamma_i (t-\tau)/2} \sin[\Omega_i (t-\tau)] \times \langle y_j(\tau) y_i(0) \rangle, \tag{B9}$$

$$\langle y_i(t)\dot{y}_j(0) \rangle = -H_c \int_0^t d\tau \frac{1}{\Omega_i} e^{-\gamma_i (t-\tau)/2} \sin[\Omega_i (t-\tau)] \times \langle y_j(\tau) y_j(0) \rangle. \tag{B10}$$

We now show that the cross-correlation functions can be neglected in the high friction limit. To this end, we note that the integrand of a cross-correlation function is a product of two functions that have the same behavior as $S(\xi,\eta)$ up to the point that $S(\xi,\eta)$ reaches its maximum. Since the two functions are out of phase (at $\tau=0$ or $\tau=t$ one term will be at its maximum, while the other term is zero at the same time), their product remains very small in the interval $[0,t]$. Consequently, if we invoke the mean value theorem to represent the integral as $\bar{S}t$, where $\bar{S}$ is the mean value of the integrand over the whole interval, we arrive at

$$\langle y_i(t)\dot{y}_j(0) \rangle = -(H_c t)\bar{S}. \tag{B11}$$

Since the scale of t is determined by γ_i, we have $H_c t \ll 1$ in the high friction limit, which leads to the conclusion that the cross-correlation functions are of negligible magnitude compared with the autocorrelation functions. Therefore, we have

$$\langle y_i(t)\dot{y}_i(0)\rangle \simeq k_B T \gamma_i^{-1} \frac{\gamma_i}{\Omega_i} e^{-\gamma_i t/2} \sin(\Omega_i t)$$
$$= k_B T \gamma_i^{-1} S_i(\gamma_i, H_i, t), \tag{B12}$$

$$\langle y_i(t)\dot{y}_j(0)\rangle \simeq 0, \tag{B13}$$

where $S_i(\gamma_i, H_i, t)$ is analogous to $S(\xi, \eta)$ in the one-dimensional case [Eq. (A4)]. At a time t^* when $S_i(\gamma_i, H_i, t^*)$ reaches its maximum value (S_i^*), we have

$$\begin{pmatrix} \langle y_1(t)\dot{y}_1(0)\rangle & \langle y_1(t)\dot{y}_2(0)\rangle \\ \langle y_2(t)\dot{y}_1(0)\rangle & \langle y_2(t)\dot{y}_2(0)\rangle \end{pmatrix}$$
$$= k_B T \begin{pmatrix} \gamma_1^{-1} S_1^* & 0 \\ 0 & \gamma_2^{-1} S_2^* \end{pmatrix}$$
$$\simeq k_B T S^* \begin{pmatrix} \gamma_1^{-1} & 0 \\ 0 & \gamma_2^{-1} \end{pmatrix} = k_B T S^* \boldsymbol{\Gamma}^{-1}, \tag{B14}$$

where the second step follows from the fact that in the high friction limit we have $S_1^* \simeq S_2^* \equiv S^*$. Applying a similarity transformation based on $\mathbf{U}$ to both sides of Eq. (B14), we have

$$\begin{pmatrix} \langle q_1(t)\dot{q}_1(0)\rangle & \langle q_1(t)\dot{q}_2(0)\rangle \\ \langle q_2(t)\dot{q}_1(0)\rangle & \langle q_2(t)\dot{q}_2(0)\rangle \end{pmatrix} = k_B T S^* \mathbf{U}^{-1} \cdot \boldsymbol{\Gamma}^{-1} \cdot \mathbf{U} = S^* \mathbf{D}, \tag{B15}$$

which shows that the maximum values of the corresponding correlation functions provide reasonable estimates for the diffusion tensor elements in the high friction limit. Given the restriction to this friction regime in the two-dimensional case, it is likely that our approach yields very similar results to those that one would obtain using a generalized version of Eq. (6), but exploration of that quantity is more mathematically involved.

[1] R. H. Austin, K. W. Beeson, L. Eisenstein, H. Frauenfelder, and I. C. Gunsalus, Biochemistry **14**, 5355 (1975).
[2] D. A. Case and M. Karplus, J. Mol. Biol. **132**, 343 (1979).
[3] X. S. Xie and R. C. Dunn, Science **265**, 361 (1994).
[4] H. P. Lu, L. Y. Xun, and X. S. Xie, Science **282**, 1877 (1998).
[5] G. Hummer, F. Schotte, and P. A. Anfinrud, Proc. Natl. Acad. Sci. U.S.A. **101**, 15330 (2004).
[6] M. T. Zanni and R. M. Hochstrasser, Curr. Opin. Struct. Biol. **11**, 516 (2001).
[7] Y. S. Kim, J. P. Wang, and R. M. Hochstrasser, J. Phys. Chem. B **109**, 7511 (2005).
[8] A. R. Dinner and M. Karplus, J. Phys. Chem. B **103**, 7976 (1999).
[9] P. G. Bolhuis, C. Dellago, and D. Chandler, Proc. Natl. Acad. Sci. U.S.A. **97**, 5877 (2000).
[10] T. A. McCormick and D. Chandler, J. Phys. Chem. B **107**, 2796 (2003).
[11] M. F. Hagan, A. R. Dinner, D. Chandler, and A. K. Chakraborty, Proc. Natl. Acad. Sci. U.S.A. **100**, 13922 (2003).
[12] A. Ma and A. R. Dinner, J. Phys. Chem. B **109**, 6769 (2005).
[13] D. Chandler, J. Chem. Phys. **68**, 2959 (1978).
[14] H. A. Kramers, Physica (Amsterdam) **7**, 284 (1940).
[15] P. Hanggi, P. Talkner, and M. Borkovec, Rev. Mod. Phys. **62**, 251 (1990).
[16] N. D. Socci, J. N. Onuchic, and P. G. Wolynes, J. Chem. Phys. **104**, 5860 (1996).
[17] E. Neria and M. Karplus, J. Chem. Phys. **105**, 10812 (1996).
[18] G. Hummer, A. E. Garcia, and S. Garde, Phys. Rev. Lett. **85**, 2637 (2000).
[19] A. Berezhkovskii and A. Szabo, J. Chem. Phys. **122**, 014503 (2005).
[20] J. S. Langer, Ann. Phys. (N.Y.) **54**, 258 (1969).
[21] Although the second derivative should be a covariant one in a nonlinear space, it reduces to the normal one due to the fact that the first derivative at the saddle point vanishes.
[22] Y. M. Rhee and V. S. Pande, J. Phys. Chem. B **109**, 6780 (2005).
[23] R. Du, V. S. Pande, A. Y. Grosberg, T. Tanaka, and E. S. Shakhnovich, J. Chem. Phys. **108**, 334 (1998).
[24] C. Dellago, P. G. Bolhuis, and P. L. Geissler, Adv. Chem. Phys. **123**, 1 (2002).
[25] J. E. Straub, M. Borkovec, and B. J. Berne, J. Phys. Chem. **91**, 4995 (1987).
[26] B. J. Berne, M. E. Tuckerman, J. E. Straub, and A. L. R. Bug, J. Chem. Phys. **93**, 5084 (1990).
[27] J. E. Straub, B. J. Berne, and B. Roux, J. Chem. Phys. **93**, 6804 (1990).
[28] T. B. Woolf and B. Roux, J. Am. Chem. Soc. **116**, 5916 (1994).
[29] G. Hummer, New J. Phys. **7**, 1 (2005).
[30] P. Liu, E. Harder, and B. J. Berne, J. Phys. Chem. B **108**, 6595 (2004).
[31] W. Im and B. Roux, J. Mol. Biol. **319**, 1177 (2002).
[32] M. Schaefer and M. Karplus, J. Phys. Chem. **100**, 1578 (1996).
[33] B. R. Brooks *et al.*, J. Comput. Chem. **4**, 187 (1983).
[34] C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, J. Chem. Phys. **108**, 1964 (1998).
[35] J. Hu, A. Ma, and A. R. Dinner, J. Comput. Chem. **27**, 203 (2006).
[36] B. A. Berg and T. Neuhaus, Phys. Lett. B **267**, 249 (1991).
[37] F. Wang and D. P. Landau, Phys. Rev. Lett. **86**, 2050 (2001).
[38] F. Calvo, Mol. Phys. **100**, 3421 (2002).
[39] A. R. Dinner, Ph.D. thesis, Harvard University, 1999.
[40] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications* (Academic, London, 2002).
[41] E. Neria, S. Fischer, and M. Karplus, J. Chem. Phys. **105**, 1902 (1996).
[42] R. W. Pastor, B. R. Brooks, and A. Szabo, Mol. Phys. **65**, 1409 (1988).
[43] P. E. Smith and B. M. Pettitt, J. Phys. Chem. **97**, 6907 (1993).