Coarse molecular dynamics of a peptide fragment: Free energy, kinetics, and long-time dynamics computations

Gerhard Hummer
Laboratory of Chemical Physics, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892

Ioannis G. Kevrekidis
Chemical Engineering, PACM and Mathematics, Princeton University, Princeton, New Jersey 08544

We present a ''coarse molecular dynamics'' approach and apply it to studying the kinetics and thermodynamics of a peptide fragment dissolved in water. Short bursts of appropriately initialized simulations are used to infer the deterministic and stochastic components of the peptide motion parametrized by an appropriate set of coarse variables. Techniques from traditional numerical analysis (Newton–Raphson, coarse projective integration) are thus enabled; these techniques help analyze important features of the free-energy landscape (coarse transition states, eigenvalues and eigenvectors, transition rates, etc.). Reverse integration of coarse variables backward in time can assist escape from free energy minima and trace low-dimensional free energy surfaces. To illustrate the coarse molecular dynamics approach, we combine multiple short (0.5 ps) replica simulations to map the free energy surface of the ''alanine dipeptide'' in water, and to determine the $\sim 1/(1000 \text{ ps})$ rate of interconversion between the two stable configurational basins corresponding to the α-helical and extended minima. © *2003 American Institute of Physics.* [DOI: 10.1063/1.1574777]

I. INTRODUCTION

Molecular dynamics (MD) simulations[1] on classical or quantum energy surfaces provide a unique tool for exploring the phase space of (bio)chemical systems at full atomic resolution. In the biological sciences, realistic MD simulations of protein folding, complex formation and aggregation, enzyme kinetics, channel transport, etc., hold the promise to form not only the basis for new understanding of these fundamental processes but also to accelerate the development of new drugs and treatments for diseases. However, these and many other processes in (bio)materials occur on time scales well beyond the reach of current MD simulations, even if carried out with the most powerful computers available on an approximate classical Born–Oppenheimer energy surface. With femtosecond timesteps required to integrate the fastest atomic motions, classical MD of (bio)molecular systems in condensed phase is only starting to push into the microsecond regime,[2,3] and orders of magnitude less on quantum surfaces. Remarkable progress has been made recently in *overcoming* the time scale limitations of MD;[4–9] *circumventing* them is the goal of this work.

The fundamental difficulty in MD arises from the requirement to integrate the motions of ''all'' nuclear degrees of freedom, not just those of interest. This problem of timescale separation has been formally addressed in the projection-operator formalism of Zwanzig and Mori[10–12] by constructing a generalized Langevin equation that describes the time evolution in a ''slow subspace.'' The coarse dynamics proposed by Kevrekidis and co-workers for equation-free multiscale computations[13–22] (see Ref. 23 for a review) builds on this general framework. In the coarse molecular dynamics (CMD) approach developed here, the existence of

an attracting ''slow manifold'' for the mesoscopic evolution is assumed. However, explicit constructions of the slow manifold and the corresponding evolution equation, memory kernel, and noise function are avoided by using conventional MD to propagate the microscopic system over intermediate times, as discussed below. Here, we implement the CMD method to study arguably the ''simplest biomolecule,'' alanine dipeptide (i.e., N-acetyl alanine N'-methyl amide), dissolved in water. This system was chosen as a fundamental fragment of protein backbones with torsion degrees of freedom φ, ψ, and ω; and polar groups C=O and N–H that interact strongly with each other and the solvent (see Fig. 1). For this reason, the alanine dipeptide has been studied extensively by theory and experiment (see, e.g., Refs. 24–33), and has been used as a model system to analyze the thermodynamics and kinetics of conformational dynamics.[9,34]

The two major goals here are (1) to map the essential features of the (coarse) free energy surface of the dipeptide, in particular to find stable minima and connecting saddle points, and (2) to determine the rates of interconversion between the stable minima.

The paper is organized as follows: In Sec. II A, we outline the concepts of CMD. In Sec. II B, an expression is introduced that relates the drift velocity of an ensemble of trajectories to the underlying free energy surface. In Sec. II C, we discuss how stable and unstable stationary points on that surface (i.e., free energy minima and saddle points) can be found using contraction mapping (e.g., Newton–Raphson) type searches. The concepts of coarse projective forward and reverse integration are introduced in Secs. II D and II E, respectively. In Sec. II F, we show how the Chapman–Kolmogorov identity can be used to estimate rates of slow
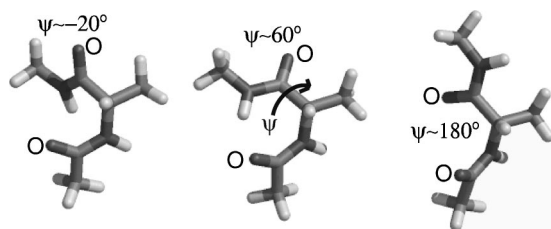
10762

FIG. 1. Structures of the alanine dipeptide in the right-handed α-helical minimum (left), in the extended minimum (right), and at the barrier in between (center). The dihedral angle ψ is indicated in the center structure. Carbonyl oxygen atoms are labeled.
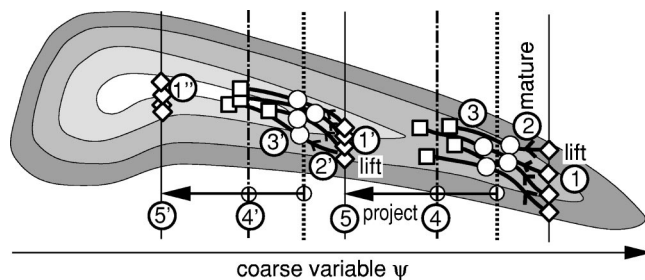

FIG. 2. Schematic of projective forward integration in CMD. The coarse and fast variables are along the horizontal and vertical axis, respectively. Lighter shading indicates lower free energy. (1) *Lifting* creates an ensemble of initial conditions (diamonds) with a prescribed value of the coarse variable ψ (thin vertical line). In short MD *runs* starting from these initial conditions, (2) the fast degrees rapidly *mature* by relaxing on the free energy surface (circles), followed by (3) drift in the coarse variable. At the end of the MD runs, we (4) *restrict* to coarse variables by collapsing the distribution of trajectory endpoints (squares) to their average (dotted–dashed line). From the drift in the average (from dotted to dotted–dashed line), we *estimate* a time derivative and (5) use it to *project* forward in time (horizontal arrow). (1′) Lifting creates a new ensemble of initial conditions (vertical solid line), followed by a second cycle of projective forward integration. In this schematic plot, "convergence" to the free energy minimum is practically achieved in two cycles.

processes from ensembles of short trajectories.

Details of the MD simulations of the solvated alanine dipeptide can be found in Sec. III. Section IV presents the results for the estimate of the free energy surface from drift velocities, with minima and saddles found directly by Newton–Raphson-type searches and by projective forward and reverse integration. The rate of interconversion between the α-helical and extended states of the alanine dipeptide will be estimated using the Chapman–Kolmogorov identity and compared to explicit simulation results.

Finally, Sec. V contains some concluding comments. In particular, we will discuss the relation of CMD to other methods, such as constrained dynamics,[35] the parallel replica method,[7] transition path sampling,[6,36] SWARM-MD,[37] and memory-based dynamics.[9,38] We conclude by outlining future directions.

II. THEORY

A. Coarse molecular dynamics

In CMD, the existence of an attracting "slow manifold" is assumed. We start with a set of "coarse" variables that *parametrize* this manifold: for every combination of values of these variables, a single point exists on the manifold. This means that the expected values of the remaining variables can be conceptually plotted as an (unspecified) function of these coarse variables. We stress that these variables are "observation variables." They *do not span* the subspace on which the coarse dynamic occurs (the "curved" manifold). There exists, however, a one-to-one relation between trajectories *on* this manifold, and *the projections* of these trajectories on the hyperplane spanned by the coarse variables. The number of such coarse (observation) variables should be at least as large (and preferably, for computational economy purposes, the same) as the dimension of the slow manifold. For macromolecular systems, such "coarse" variables will include descriptors of their internal geometry, such as dihedral angles, radii of gyration, end-to-end distances, number, and type of monomer contacts, etc. They may also include solvent coordinates, such as variables describing solvent coordination numbers and structures. In a somewhat different context (e.g., kinetic theory description of fluid flow) the "coarse manifold" will be parametrized by the hydrodynamic variables (density and momentum fields, i.e., low moments of the molecular distribution over velocity space; see, for example, Ref. 17). If necessary, nonlinear couplings be-

tween variables can be introduced, for instance, by using the expectations of products of hydrodynamic variables to parametrize the manifold, as in mode-coupling theories. In the CMD approach to atomistic simulations, the assumption is that microscopically evolving distributions in phase space quickly become effectively low-dimensional.[39]

Along fast directions (such as those describing bond vibrations), the distributions rapidly saturate, while they keep drifting and spreading along slow directions (such as those describing large-scale conformational rearrangements). The properties of motions along the fast degrees of freedom (such as amplitude or characteristic time) depend on the slow degrees of freedom. In a peptide, for instance, the frequency spectrum of fast motions associated with the amide C=O stretch vibration depends on the environment described by the slow variables quantifying, e.g., the type and amount of secondary structure and solvent present. Geometrically, the average of an ensemble of trajectories relaxes onto the attracting "slow manifold" and remains there. This picture invokes an analogy with the so-called "inertial manifolds" for dissipative partial differential equations.[40–42]

The (expected) dynamics of the coarse variables are explored through short multiple replica simulations as follows (see Fig. 2): A coarse initial condition is prescribed, and "lifted" to many microscopic copies consistent with (conditioned on) the coarse variables. This "lifting" step is not unique, since many distributions can be constructed that have the same coarse variables. Lifting can be achieved, for instance, by performing a short MD run with an added potential biasing the coarse variables toward their new target values, as in umbrella sampling.[43] This approach works also for coarse variables consisting of nonlinear combinations of the atomic coordinates, as in the product variables of mode-coupling theories. The initialization will necessarily be at least slightly "off manifold;" that is, the fast degrees will not be initialized to be the "correct" functionals of the coarse

variables. This discrepancy, however, will be quickly "healed," as a relatively short detailed simulation will bring the trajectory down on the manifold (i.e., slave the statistics of the fast degrees of freedom to the coarse variables). After this short healing or "maturing" period we monitor the detailed simulations over short times, and estimate the time derivative of the slow dynamics in the coarse variables. This time-derivative can be used to extrapolate coarse dynamics over relatively long time intervals; repeating the procedure (lift, short run, restrict to coarse variables, estimate coarse variable time derivative, project coarse variables into the future, lift again,...) constitutes the coarse projective integration schemes of Gear and Kevrekidis,[15,19,44] and allows us to extrapolate the long-time coarse behavior from short, repeated bursts of appropriately initialized MD simulation. It is important to stress the relation between these methods and the optimal predictors of Chorin and co-workers.[45–47] In their formalism, a projection-operator approach is used to propagate ensembles of initial conditions in time, with some phase space variables (here, the coarse variables) at constant values in the starting ensemble, and the remaining ones drawn from an equilibrium distribution (as achieved here by "maturing").

Beyond these "accelerated" dynamics, we will see that the CMD approach permits the rapid search of dynamically-relevant features of the free energy surface and the calculation of transition rates, as discussed below. Here, we use the dihedral angle ψ (N–C_α–C–N) of the Ala dipeptide as a first approximation for the coarse coordinates (i.e., the coordinates used to observe the slow dynamics). In the following, we will explain the method in terms of this coordinate. Generalizations to multiple dimensions are straightforward, have been discussed in the literature,[18,22,23] and will be illustrated here by occasionally using the dihedral angle φ (C–N–C_α–C) as a second coarse coordinate.

B. From drift in the coarse variables to free energies

Assume that we have created an ensemble of N configurations i with identical values of $\psi_i(t=0;\psi_0)=\psi_0$, for instance by randomly assigning Maxwell–Boltzmann velocities to all solute and solvent atoms in a given configuration. As these configurations evolve in time, the dynamics of $\psi_i(t;\psi_0)$ will initially be governed by the coupling to fast motions such as bond vibrations and molecular collisions. As a consequence, the $\psi_i(t;\psi_0)$ will "spread" rapidly over a small range. After a sufficiently long time $t>\tau_{\mathrm{mol}}$, however, the average $\overline{\psi(t;\psi_0)}=N^{-1}\Sigma_{i=1}^N\psi_i(t;\psi_0)$ is expected to drift down on the free energy surface (or potential of mean force) $G(\psi)$ towards a stable minimum. "Timestepping" for $\overline{\psi(t;\psi_0)}$ involves lifting, evolving for a time $t>\tau_{\mathrm{mol}}$, and restricting back to coarse variables (i.e., collapsing the distribution of the endpoints of the several trajectories to their average). This "coarse timestepper" (the map resulting from the lift-run-restrict procedure), lies at the heart of the coarse numerical processes (coarse integration, coarse Newton–Raphson, coarse optimization) we perform. Initializing at nearby values of ψ_0 can be used to estimate the partial derivatives of the coarse timestepper with respect to its (coarse)

initial conditions, and from that we can obtain estimates of the local curvature of the free energy surface. From a Langevin approximation to the dynamics of slow variables,[12] we expect that the drift velocity is proportional to the slope of the free-energy surface,

$$\frac{\partial\overline{\psi(t;\psi_0)}}{\partial t}\approx-D\beta\left.\frac{\partial G(\psi)}{\partial\psi}\right|_{\psi=\psi^*}\qquad\text{for }t>\tau_{\mathrm{mol}},\qquad(1)$$

where $\beta^{-1}=k_BT$ (with k_B Boltzmann's constant, T the temperature) and $\psi^*=t^{-1}\int_{t_0}^t\overline{\psi(t';\psi_0)}dt'$. Normally, one would begin averaging after an initial maturation time ($t_0>\tau_{\mathrm{mol}}$); here, we averaged over the whole interval ($t_0=0$). The proportionality constant βD is a mobility, with D a diffusion coefficient that may depend on ψ. We may also expect that the variance in $\psi_i(t;\psi_0)$ of different runs i initialized with the same $\psi_i(0;\psi_0)=\psi_0$ will exhibit diffusive behavior,

$$\frac{\partial}{\partial t}\mathrm{var}[\psi(t;\psi_0)]\approx2D(\psi^*),\qquad(2)$$

where the diffusion coefficient D is related to the mobility in Eq. (1) through the Einstein relation. We expect Eq. (2) to hold for times $\tau_{\mathrm{mol}}<t\ll(\beta D|\partial^2 G/\partial\psi^2|)^{-1}$ long compared to the initial molecular relaxation time τ_{mol} but short with respect to curvature effects of the free energy surface. One can thus use the time-evolution observed through the coarse variables to estimate the deterministic and stochastic components of the motion, and from those estimate the underlying free energy surface, $G(\psi)$, and dynamic properties, such as effective diffusion coefficients $D(\psi)$.

C. Finding stationary points on the coarse free energy surface

In many practical applications, one is interested in finding the stable minima and the (unstable) saddle points on the coarse free energy surface, rather than mapping the complete surface. Protein folding simulations constitute one such example with goals of finding the folded structure and the structures at the "barrier" to folding. From Eq. (1), we expect that the minima and maxima are stable and unstable stationary points for the drift component of the dynamics in the coarse variables. We can thus use the dynamic information to search for these stationary points on the free energy surface, i.e., minima and saddles satisfying

$$\frac{\partial\overline{\psi(t;\psi_0)}}{\partial t}=0\quad\text{or}\quad\psi_0-\overline{\psi(t;\psi_0)}=0.\qquad(3)$$

By estimating $\overline{\dot{\psi(t;\psi_0)}}=\partial\overline{\psi(t;\psi_0)}/\partial t$ or $\psi_0-\overline{\psi(t;\psi_0)}$ for nearby initial points ψ_0, we can estimate $\partial\dot{\bar\psi}/\partial\psi$ and perform a Newton–Raphson type step towards a stationary point,

$$\psi_1=\psi_0-\frac{\dot{\bar\psi}}{\partial\dot{\bar\psi}/\partial\psi_0}\quad\text{or}\quad\psi_1=-a/b,\qquad(4)$$

where a and b are the intercept and slope, respectively, in a straight-line fit of $\psi_0-\overline{\psi(t;\psi_0)}$ versus $[\psi_0+\overline{\psi(t;\psi_0)}]/2$ for nearby starting values of ψ_0. Recursive application of Eq. (4), can be used to converge to a stationary point. This

timestepper approach to coarse steady state computations was introduced for spatially distributed systems,[13] and was illustrated for stochastic systems[18,20,22] (stable and saddle stationary states of kinetic Monte Carlo and Brownian dynamics simulations).

Unlike the thermodynamically determined free energy profiles (e.g., from umbrella sampling[43] or constrained dynamics)[35] the dynamic sampling can provide information about the kinetic relevance of the chosen coarse coordinate. Consider the classic example [see, e.g., Fig. 1(b) in Ref. 48] of a rotated double well system in two dimensions. Projected onto the x axis, the two wells overlap in the barrier region. If the dynamics is monitored only along the x axis but not the "solvent" coordinate y, most of the structures at the barrier in x appear already "committed" to one or the other well. This has important implications. In free energy calculations using umbrella sampling or thermodynamic integration, the system will not be pulled easily over the barrier by a potential acting on x alone; in kinetics calculations, the dynamics along x alone is non-Markovian, making rate calculations difficult. In our dynamic analysis, this situation becomes manifest in a large spread of the estimated drift velocity and direction, $\overline{\dot{\psi}(t;\psi_0)}$, for nearby initial values of ψ_0.

A detailed discussion of practical approaches to determining the need for additional coarse variables is contained in Refs. 18 and 22. The idea is to quantify this need directly, "on-the-fly:" by lifting with more coarse variables ψ_i, we can determine the eigenvalues and corresponding time scales of the coarse local linearization, $J_{ij} = \partial\dot{\psi}_i/\partial\psi_j$. If the time scales introduced by adding new coarse variables are fast compared to the ones already included, as evidenced by a gap in the eigenvalue spectrum of the local linearization, then the current coarse description is still satisfactory. In contrast, the presence of an additional slow time scale (eigenvalue) in the expanded space suggests the need to augment the coarse description, e.g., by including the corresponding eigenvector as a new coarse variable. If additional slow modes are not accounted for, estimates of local derivatives with respect to the coarse variables will become less reliable due to the new "hidden variables." This deterioration can be used as a computational "flag," starting up the direct systematic augmentation procedure. These approaches very much resemble, in spirit, tests for the adaptive mesh refinement or adaptive step size selection in partial-differential-equation simulations: there one compares the results using smaller (or larger) timesteps and finer (or coarser) meshes and decides whether to refine (or coarsen) the mesh. Here, in a completely analogous way, we are pursuing the "adaptive coarse variable augmentation" (and, in some cases, reduction!) for CMD (and other multiscale) codes.

From the CMD analysis, we can also extract immediately properties of the dynamics in the free energy wells. In one dimension, we can estimate the correlation time τ_c for slow (diffusive) motion at the bottom of a well from the slope of the drift velocity with respect to the coarse variable,

$$\tau_c^{-1} = -\left.\frac{\partial\dot{\psi}}{\partial\psi}\right|_{\dot{\psi}=0}. \tag{5}$$

This follows from Eq. (1) and the relation $\tau_c^{-1} = \beta D\kappa$ between the characteristic time τ_c and the effective spring constant $\kappa = \partial^2 G/\partial\psi^2$ for a diffusive harmonic oscillator with diffusion constant D. By using Eq. (5), the local correlation time τ_c in a particular steady state can be determined directly from the Newton–Raphson search upon convergence. The slope of $\dot{\psi}$ with respect to ψ provides the "restoring force" for the coarse variable. If $\dot{\psi}$ remained linear in ψ, then the motion would correspond to a harmonic oscillator in the high-friction (overdamped) limit. In many dimensions, the characteristic frequencies are given by the eigenvalues of the Jacobian matrix $J_{ij} = \partial\dot{\psi}_i/\partial\psi_j$ analogous to the derivative in Eq. (5). The corresponding eigenvectors give the "normal coordinates" for the overdamped motions around a stationary point.[49]

D. Coarse projective integration

The dynamic information in the replica runs $\psi_i(t;\psi_0)$ can also be used to extrapolate toward longer times. Instead of propagating each of the replicas, we extrapolate the *average* position of the slow variable, for instance linearly (exploiting regularity of the expected coarse dynamics with time),

$$\overline{\psi(t';\psi_0)} \approx \overline{\psi(t;\psi_0)} + \frac{t'-t}{t}[\overline{\psi(t;\psi_0)} - \psi_0]. \tag{6}$$

A long "projective" step $t'-t$ is then effected by reinitializing an ensemble at the extrapolated value; this is the simplest "projective forward Euler method." [15] Clearly, instead of simply taking a linear interpolation between the first and last coarse values of a short run, we can record a short "tail" of the coarse evolution after the quick maturing period and use that to construct (linear or higher order) predictors.[15,19,44]

In the Ala dipeptide, we use a harmonic constraint on the torsion potential to initialize ψ at its new target value. With new initial velocities assigned from a Maxwell–Boltzmann distribution, the fast degrees of freedom, such as bonds or solvent positions, re-equilibrate rapidly and the newly initialized state starts drifting again on the free-energy surface. It is interesting that in addition to coarse projective integration and to coarse Newton–Raphson, additional algorithms, like time-stepper based coarse control[21] and coarse optimization,[50] become enabled, and they may be useful in analyzing these surfaces without ever explicitly closing Langevin equations on them (by specifying force, friction, and noise terms).

E. Reverse coarse integration and escaping free energy minima

It is well known that whether one uses MD forward or backward in time (i.e., whether one flips the velocities of the initial configuration or not) one obtains *forward* evolution of the coarse variables. As discussed in Ref. 51, however, it is possible to integrate *the coarse variables* backward in time on the slow manifold (i.e., on the coarse free energy surface) exploiting projective integration as follows: After initializing at the molecular level, and running (whether with flipped or unflipped velocities) long enough for the lifting errors to

heal, we use the estimated coarse *forward* time derivatives $\dot{\overline{\psi}}$ or $\psi_0 - \overline{\psi(t;\psi_0)}$ to perform relatively large *reverse* projections in time. Under appropriate conditions this is a stable algorithm.[52]

For reverse integration from ψ_0 equal to the length of the forward healing/estimation step, we can use a simple Euler integrator,

$$\overline{\psi(-\Delta t;\psi_0)} \approx 2\overline{\psi(0;\psi_0)} - \overline{\psi(\Delta t;\psi_0)}. \tag{7}$$

We then (1) lift again from the projected $\psi(-\Delta t;\psi_0)$, (2) run again MD (whether forward or backward in time) to obtain a new estimate of the local, coarse forward-in-time $\dot{\overline{\psi}}$ or $\psi_0 - \overline{\psi(t;\psi_0)}$; and (3) reverse project again the coarse variable backward in time.

What was discussed here is the simplest "projective forward Euler" method used backward in time; it is clear that more elaborate multistep coarse projective integration methods can be used to accelerate both the forward and reverse coarse integration in our context. Notice that this "reverse integration" is useful not only in a microscopic/stochastic context, but also in the case of stiff deterministic problems in general, and even discretizations of dissipative partial differential equations.[52] Such generalizations of coarse projective integrations (including coarse projective Runge–Kutta, Adams, and even implicit implementations of such algorithms in the "coarse" case) are outlined in Refs. 15 and 44 and are the subject of ongoing research in collaboration with C. W. Gear. This "short step forward in the full space, large step backward approximately on the manifold" can be used to systematically integrate the unavailable equations backward in time on the free energy surface. For this one-dimensional problem, the "saddle" is actually a "source," i.e., an attractor backward in time. Suitably initialized, reverse integration of $\overline{\psi(t;\psi_0)}$ will then converge to the saddle where $\dot{\overline{\psi(t;\psi_0)}} = 0$ and $\partial \overline{\psi(t;\psi_0)}/\partial \psi > 0$. Reverse integration can become an efficient way of both exploring the surface and escaping free energy minima. It appears that, if all but one or two of the coarse backward directions are very stiff (i.e., a separation of time scales prevails for the coarse variable drift), their effect is quickly damped by the short forward integrations. With the appropriate step choices, the reverse integration will probably only "see" the slow backward directions, and use them to "climb back up" the slow backward path(s). Methods for the construction of stable manifolds for low-dimensional dynamical systems (see, e.g., Refs. 53 and 54) can, under favorable time scale separation conditions, be useful in exploring the free energy surface through reverse coarse integration.

F. Calculation of rates

We can also use CMD to estimate rates of interconversion between states. Instead of performing one long run, we determine the short-time dynamics in the space spanned by the coarse variables from many short, appropriately initialized replica runs. The replica runs are used to construct propagators. Assuming some degree of regularity (smoothness) these propagators are then applied recursively to infer the long-time dynamics. From N replica simulations of length t, we can estimate a propagator as a sum of δ functions at the end points of each trajectory,

$$p(\psi,t|\psi_0,0) \approx N^{-1} \sum_{i=1}^{N} \delta[\psi - \psi_i(t;\psi_0)]. \tag{8}$$

Alternatively, propagators could be built by filtering the data or by using cumulants of the distribution at time t. Under the assumption of Markovian dynamics, the distribution at times of $2t$, $3t$, etc., can be determined by recursive application of the Chapman–Kolmogorov identity,

$$p(\psi,2t|\psi_0,0) = h(\psi) \int p(\psi,t|\psi',0)p(\psi',t|\psi_0,0)d\psi'. \tag{9}$$

Without absorbing points [$h(\psi)=1$ for all ψ], we expect to recover the equilibrium distribution by iterating to infinite time, $\beta G(\psi) = -\ln p(\psi,t\to\infty|\psi_0,0)$. Equation (9) provides an alternative route to the free energy surface, in addition to Eq. (1). We note that under the Markovian assumption, the action functional describing the relative probabilities of a dynamic path (here, $\psi_0,\psi_1,\ldots,\psi_N$) in the space of the coarse variables is given by the product of the corresponding propagators,

$$e^{-s(\psi_0,\psi_1,\ldots,\psi_N)} \equiv e^{-\beta G(\psi_0)} \prod_{i=0}^{N-1} p(\psi_{i+1},t|\psi_i,0). \tag{10}$$

We can use Eq. (9) to estimate "reaction rates." To find first-passage time distributions from a "reactant" state to a "product" state, absorbing points can be inserted by multiplying the integral in the Chapman–Kolmogorov relation, Eq. (9), with $h(\psi)=0$ inside the "product" region and $h(\psi)=1$ outside. This assumes that the time t is short relative to the average time for escape from the reactant and product well. Integration of $p(\psi,nt|\psi_0,0)$ over ψ then gives a survival time distribution at time nt starting from $\psi = \psi_0$,

$$S(nt) = \int p(\psi,nt|\psi_0,0)d\psi. \tag{11}$$

Application of Eq. (9) requires propagators at intermediate ψ values. This is where regularity in coarse phase space is assumed: propagators at intermediate values can be estimated by interpolation from replica runs initialized with different values of ψ. For an intermediate value ψ_2, we can use a simple linear interpolation,

$$p(\psi,t|\psi_2,0) \approx (1-\alpha)p(\psi,t|\psi_0,0) + \alpha p(\psi,t|\psi_1,0) \tag{12}$$

or

$$p(\psi,t|\psi_2,0) \approx (1-\alpha)p(\psi+\psi_2-\psi_0,t|\psi_0,0)$$
$$+ \alpha p(\psi+\psi_2-\psi_1,t|\psi_1,0), \tag{13}$$

where $\alpha = (\psi_2-\psi_0)/(\psi_1-\psi_0)$ with propagators for ψ_0 and ψ_1 given by Eq. (8). For diffusion on a linear potential with D constant, the linear interpolation Eq. (13) is exact. If the diffusion coefficient is position dependent and possibly anisotropic, appropriate generalizations can be constructed.

(Notice that our CMD approach *does* provide estimates of the position dependence and the anisotropy of the diffusion coefficient.)

In a certain limit, the concept of an optimal path can be helpful in rate computations. Optimization *over coarse variable paths* can be performed using CMD timesteppers. We can locally approximate both the deterministic *and* the ''diffusive'' part of the evolution through the propagators estimated from replica runs. If we guess a coarse transition path between two coarse minima we can write down the coarse action functional for this path. The coarse action functional can thus be deterministically minimized, using, for example, dynamic programming methods. One can estimate the sensitivity of the functional to (discretized) path variations through nearby short integrations; or, alternatively, ''derivative free'' optimization methods can be used for this goal, and the conversion rate computed (approximated) upon convergence of the optimization.[55] We have, in the past, solved discretized coarse optimization problems for optimal parameter variation policies;[50] we are currently pursuing the application of the same optimization methods for the computation of rates in problems with coarse dimension higher than one.[55]

In many cases, the concept of the optimal path loses its significance in the computation of rates. Instead, one evolves a state density over the coarse free energy surface, solving, in effect, a Fokker–Planck equation for the evolution of this density and its stationary state. For a relatively low coarse-dimensional problem (i.e., if the coarse free energy surface is two- or three-dimensional) it should be possible to solve this Fokker–Planck equation by convolving the propagators above.

In summary, we force the system to sample the dynamics in weakly populated regions of the ''coarse variable space'' through multiple systematic initializations of the coarse variables. This allows us to circumvent the problem that the short-time dynamics ''fully'' samples the fast degrees of motion (as conditioned on the slow ''coarse variables''), but covers only a small range in the slow variables. We are therefore trading multiple initializations for long time dynamics. This allows us to sample even the slow dynamics in the coarse variables. This can be a computationally efficient approach, as can be understood for 1D diffusive barrier crossing. By constructing and interpolating the propagators along the coordinate leading across the barrier, the rate of crossing can be estimated for arbitrarily high barriers. It is important to state, however, that this procedure assumes the dynamics to be ''smooth,'' such that propagators can be interpolated, and that the dynamics in the coarse variables is Markovian at the time scale of the short replica simulations.

Interpolating the propagators to solve a Fokker–Planck equation brings up an analogy with the so-called ''gap-tooth'' methods in Refs. 14 and 23. In this approach, regularity of the solution of a Fokker–Planck equation in space and time can, in principle, be used to accelerate a particle-based solver by evolving particles in ''patches'' of space–time, separated by empty gaps. Communication across the ''gaps'' of the ''teeth'' in which the particle density evolves is, of course, the key to the approach. The construction of

boundary conditions capable of effecting this may significantly enhance the performance of Brownian dynamics type solvers.[23,56,57]

III. MOLECULAR DYNAMICS SIMULATIONS

MD simulations of the hydrated Ala dipeptide are performed with the sander module in the AMBER 6.0 simulation package (University of California at San Francisco) and the parm94 force field.[58] The dipeptide is simulated in a periodically replicated box with 607 TIP3P water molecules.[59] Particle-mesh Ewald summation is used for the long-range electrostatic interactions.[60] The system is simulated at constant volume corresponding to $\sim$1 bar pressure, as determined during an initial equilibration run. The temperature is maintained at 300 K by weak coupling (10 ps time constant) to a Berendsen thermostat.[61] Bond lengths involving hydrogen atoms are constrained using the SHAKE algorithm.[62] After 500 ps of equilibration, we collect data for 7 ns, with configurations saved every 0.5 ps for analysis. All simulations use a time step of 0.001 ps.

Replica runs are initialized by drawing particle velocities from a Maxwell–Boltzmann distribution. For rigid TIP3P water molecules, rigid body translational and rotational velocities are generated in their principal-axes system. Initial velocities along (possibly linked) constrained bonds of the dipeptide are removed recursively.

''Lifting'' to a new target value of the coarse coordinate ψ in CMD is accomplished by a short run with a tight harmonic potential, $k(\psi-\psi_0)^2/2$, acting on the dihedral angle ψ, with $k=100$ kcal mol^{-1} rad^{-1}. With ''lifting'' here occurring not instantaneously but over a finite time interval, the fast variables are given time to adapt to the changes in the coarse variable. This effectively shortens the time needed to mature the system at the new state, as prescribed by the updated coarse variable.

At every iteration of the Newton–Raphson search for the location of stationary points, ten initial configurations are created as structures along a 5-ps run with a harmonic constraint holding ψ near the target value. For each of the ten configurations, ten sets of random initial velocities are drawn from a Maxwell–Boltzmann distribution, followed by 0.5 ps of MD.

IV. RESULTS AND DISCUSSION

A. Equilibrium molecular dynamics

From 7- and 24-ns equilibrium runs with 607 and 265 water molecules, respectively, we estimate the free energy surface $\beta G(\psi)=-\ln p(\psi)$ shown in Fig. 3, where $p(\psi)$ is given by the histogram of the dihedral angle ψ. We find two minima, one corresponding to a right-handed α helix [with $G(\psi\approx-0.3 \text{ rad})\approx0$] and the other to an extended structure [with $G(\psi=\pi)\approx1$ kcal mol^{-1}]. In the two-dimensional $\varphi-\psi$ Ramachandran plot of Fig. 3, we observe a small population in the left-handed α-helical minimum during the 24-ns run, but not the 7-ns run. In the following, we will focus on the equilibrium between the extended and right-handed α helical structures. The lower of the two barriers separating the two minima, with a height of about 3 kcal mol^{-1} near
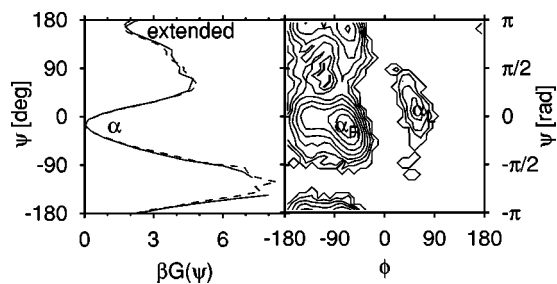
FIG. 3. Free energy surfaces of the alanine dipeptide. (Right) Free energy surface in the φ–ψ plane from the 24-ns run with 265 water molecules ($1 k_B T$ contour lines). (Left) Free energy in units of $k_B T$ (horizontal axis) as a function of the ψ dihedral angle (vertical axis). The solid and dashed lines are the results from the 7- and 24-ns runs with 607 and 265 water molecules, respectively.

$\psi = 1$ rad, shows a considerable amount of structure, with a small dip (7-ns run) or shoulder (24-ns run) near the polyproline P_{II} minimum ($\psi \approx 2\pi/3$).

For the second barrier near $\psi = -2\pi/3$ rad, we estimate a height of about 5 kcal mol^{-1} from umbrella sampling in the barrier region, in agreement with the value obtained by Bolhuis et al.[34] Overall, our equilibrium runs give a very similar $G(\psi)$ to the one obtained by Bolhuis et al.[34] with umbrella sampling.

From the variance $var_\alpha(\psi)$ of ψ in the α-helical minimum, and the decay time τ_c of the corresponding autocorrelation function, we estimate a diffusion coefficient of $D \approx var_\alpha(\psi)/\tau_c \approx 0.15$ rad^2 ps^{-1}.

B. Free energy surface

To map the free energy surfaces with CMD, we assume that the distributions of fast variables (bond lengths, etc.) are quickly slaved to the slow variables, and that the averaged dynamics of coarse variables are governed by the underlying free energy surface. We thus expect that the averaged projections onto the coarse variables drift over a short (here: sub-ps) time scale towards free energy minima. Figure 4 shows $\overline{\psi(t;\psi_0)}$ as a function of time t for runs starting from 13 different initial configurations.

Also shown is the probability distribution $p(\psi)$ of ψ from the 24-ns equilibrium run (with ψ on the vertical axis). We find that $\overline{\psi(t;\psi_0)}$ indeed drifts toward the maxima of $p(\psi)$ and away from minima. Similar behavior is found for the corresponding dynamics projected onto the φ–ψ Ramachandran plane. To get a more quantitative estimate of the free energy surface, we estimate $\dot{\psi}$ throughout the interval $-\pi < \psi < \pi$. From 67 different initial configurations, we run 50 short ($\tau = 0.5$ ps) replica simulations. The $\psi_i(t;\psi_0)$ curves of the replicas are averaged and fitted to straight lines. Figure 5 shows the corresponding slope $\dot{\psi}$, scaled by $D = 0.15$ rad^2 ps^{-1}, as a function of the average angle $\langle \overline{\psi} \rangle = \tau^{-1} \int_0^\tau \psi(t;\psi_0) dt$. As mentioned before, we average here over the whole interval, including the initial maturation time. We also calculated the local diffusion coefficient $D(\psi)$ from Eq. (2). Near the minima of $G(\psi)$, we recover the value for the equilibrium run. Near barriers, however, curvature effects are relevant even at the 0.5 ps time scale because of the
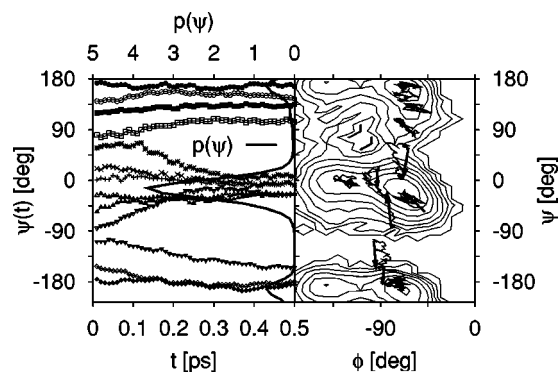

FIG. 4. Drift in the ψ dihedral angle. The panel on the left shows $\overline{\psi(t;\psi_0)}$ as a function of time averaged over 50 runs with different initial velocities. The thick line shows the equilibrium probability (horizontal axis; arbitrary units) of ψ (vertical axis) to demonstrate that $\overline{\psi(t;\psi_0)}$ converges toward the most populated regions of ψ. The panel on the right shows the "drift" of $\overline{\varphi(t;\varphi_0)}$ and $\overline{\psi(t;\psi_0)}$ in a Ramachandran plot (thick lines). Initial configurations of the replica runs were chosen from structures along the 7-ns equilibrium run. Thick arrows point from the initial values of φ and ψ to the final values. Thin lines are free energy contours separated by $1 k_B T$.

relatively rapid relaxation of ψ compared to τ_{mol}. For simplicity, we will in the following use a constant $D = 0.15$ rad^2 ps^{-1} instead of correcting $D(\psi)$ for the curvature effects by using the ψ_0-dependence of the $\overline{\psi(t;\psi_0)}$ data.

As expected from Eq. (1), the $D^{-1}\dot{\psi}$ data scatter around the free energy derivative $-\beta \partial G / \partial \psi$ which is accurately reproduced by a spline approximation to the $D^{-1}\dot{\psi}$ data. Integration with respect to ψ of the $\dot{\psi}$ data (sorted with respect to ψ and linearly interpolated) thus provides an estimate of the free energy $G(\psi)$, as shown in Fig. 6. This is a fundamental result: multiple short (0.5 ps) replica runs projected onto coarse variables can be used to probe the underlying free energy surface.

We also note that the scatter of the derivative data gives direct indication for the presence of additional slow variables. We find the largest scatter near $\psi = 1$ rad where Bolhuis et al.[34] have shown that a slow "solvent" coordinate is relevant for barrier crossing.
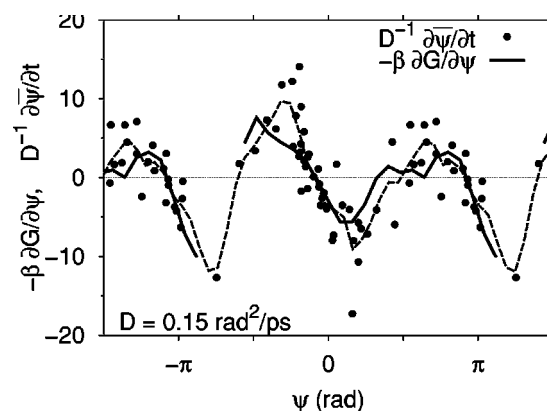

FIG. 5. Drift velocity $\dot{\psi}$ as a function of the ψ dihedral angle. Shown are $D^{-1}\dot{\psi}$ (filled circles) and the mean force $-\beta \partial G(\psi)/\partial \psi$ (solid line) for $D = 0.15$ rad^2 ps^{-1}. The dashed line is a spline approximation to the $D^{-1}\dot{\psi}$ data.

J. Chem. Phys., Vol. 118, No. 23, 15 June 2003

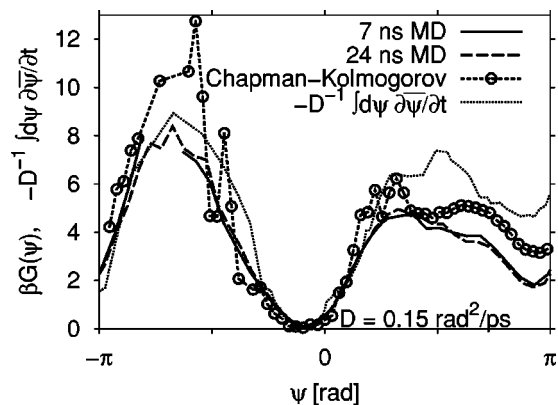Coarse molecular dynamics of a peptide fragment   10769

FIG. 6. Free energy surface as a function of ψ. Results are shown for the two equilibrium runs with 265 (24 ns; long-dashed line) and 607 water molecules (7 ns; solid line), and for integration of $D^{-1}\dot{\overline{\psi}}$ (dotted line). Also shown is the free energy corresponding to the limiting distribution of Chapman–Kolmogorov iterations (short-dashed line with symbols), evaluated at 56 discrete values of ψ.

FIG. 8. Coarse projective integration in the φ–ψ plane. Thin solid lines show the average trajectories ($\overline{\varphi}, \overline{\psi}$). Arrows indicate the projective integration steps (labeled 1, 2, and 3). Dotted lines are contour lines at $1k_BT$ intervals.

C. Free energy minima and saddles as stationary points

From Fig. 5, we expect that Newton–Raphson type, contraction mapping algorithms for finding stationary points should indeed converge to the free energy maxima and minima. We test this for the most unstable saddle near $\psi_{\text{max}} = -2.15$ rad (as determined from umbrella sampling). We save ten configurations along a 5-ps MD trajectory with a tight harmonic constraint holding ψ near $\psi_0 = -\pi/2$. From each of the ten configurations, we start ten $\tau = 0.5$-ps replica simulations with different velocities. Figure 7 shows the difference $\overline{\psi(\tau;\psi_0)} - \psi_0$ as a function of the average $[\overline{\psi(\tau;\psi_0)} + \psi_0]/2$. Despite the scatter in the difference data of round 1, a linear fit already predicts zero drift at $\psi_1 \approx -1.92$ rad close to the free energy maximum ψ_{max}. After a second round initialized near ψ_1, a quadratic fit (as shown in Fig. 7) yields $\psi_2 = -1.98$ rad. Inclusion of data from a third round initiated near ψ_2 does not change the fit and predicts

an unstable ($\partial\dot{\overline{\psi}}/\partial\psi > 0$) stationary point at $\psi = -2$ rad. The rapid convergence of the Newton–Raphson-type search shows how locations of the dominant features of free energy surfaces (i.e., minima and saddles) can be identified dynamically by searching (given good initial guesses!) for stationary points of the coarse dynamics through contraction mappings.

D. Coarse projective integration

We illustrate coarse projective integration in the two-dimensional Ramachandran plane of dihedral angles φ and ψ. In each integration step we start from identical structures but random initial velocities assigned according to a Maxwell–Boltzmann distribution. The forward differences of the dihedral angles, $\overline{\varphi(t;\varphi_0,\psi_0)} - \varphi_0$ and $\overline{\psi(t;\varphi_0,\psi_0)} - \psi_0$, are calculated from 0.5-ps MD simulations of 50 replicas. Linear extrapolation is then used to project forward in time for 0.5 ps. "Lifting" is accomplished through a 0.5-ps MD run with harmonic constraints holding both φ and ψ near their new target values.

Before showing the results, we should point out that this is not a particularly "good" problem for coarse integration; indeed, for these conditions the average solution very quickly (within 2 ps) finds its way to the bottom of the well. Coarse projective integration will probably be beneficial in situations in which the "healing period" is short compared to the slow dynamics (i.e., if there exists a gap in the eigenvalues of the linearization of the drift part of the problem). If that gap is large (i.e., if there exists a long slow transient of the coarse behavior towards the stable minimum), then projective integration has the potential to accelerate the convergence. Here, both processes—healing and drift of the expectation down into the well bottom—are relatively fast, and (for the number of copies we use) quite noisy. Figure 8 shows the resulting dynamics. We confirm that within three forward integration steps, the system reaches the free energy minimum in the φ–ψ plane. Coarse projective integration can thus be exploited to potentially accelerate CMD towards stable stationary points forward in time. The reason for in-
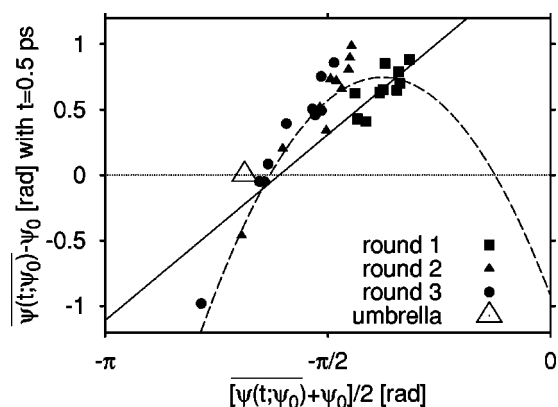

FIG. 7. Newton–Raphson-type search for the highest barrier near $\psi \approx -2.15$ rad ($\psi \approx -123°$). Starting from an initial structure with $\psi \approx -\pi/2$ rad, linear (round 1, squares and solid line) and quadratic fits to $\overline{\psi(t;\psi_0)} - \psi_0$ data (round 2, filled triangles; round 3, circles and dashed parabola) are used to locate the unstable stationary point with $\overline{\psi(t;\psi_0)} - \psi_0$. The open triangle indicates the maximum in the free energy surface, as determined by umbrella sampling.
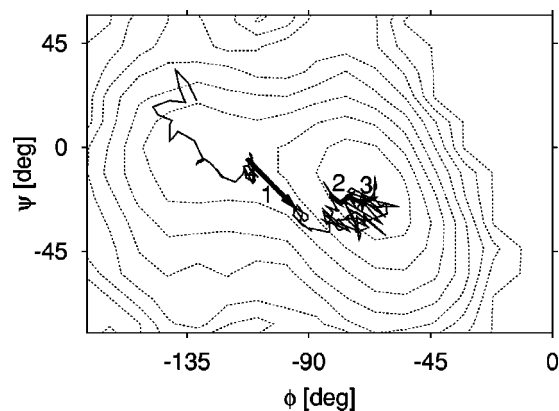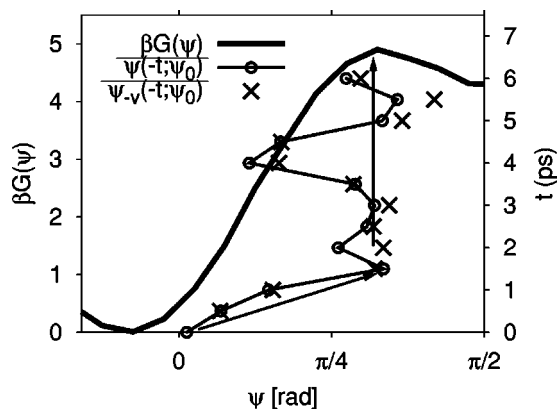
FIG. 9. Reverse integration. The thin line with open circles shows $\overline{\psi}(-t;\psi_0)$ (horizontal axis) as a function of time t (vertical axis, right-hand scale). The thick curve shows the free energy surface (in k_BT units; left-hand scale). The tilted and vertical arrows indicate the initial approach to the unstable stationary point, and the subsequent fluctuations about it, respectively. Shown as crosses are values for $\overline{\psi}(-t;\psi_0)$ estimated from reverse integration using the same starting configuration but initial velocities of opposite sign.

cluding the forward in time coarse projective integration in this particular case, however, is more for completeness, and to motivate the next section.

E. Reverse coarse integration

To escape from a free energy minimum we can also try to use reverse integration. Whereas forward integration of the expected coarse variables [here, $\overline{\psi}(t;\psi_0)$] converges towards *stable* stationary points (i.e., free energy minima), reverse integration goes "up the mountains" towards *source-type, unstable* stationary points. For a one-dimensional coarse problem the "saddle" is indeed such a "source." We start from a configuration with $\psi \approx 0$ near the bottom of the α-helical well and perform 13 reverse integration steps of length $\Delta t = 0.5$ ps. At each step, we initialize 50 replicas with identical structures and random Maxwell–Boltzmann velocities, and run regular MD *forward* in time for $\Delta t = 0.5$ ps. We then use Eq. (7) to estimate $\overline{\psi}(-\Delta t;\psi_0)$ and "lift" the $t=0$ structure by running a 0.5-ps MD with a harmonic constraint on ψ with $\overline{\psi}(-\Delta t;\psi_0)$ as the target value. The final configuration is then used as a starting structure for the forward replica runs in the next reverse integration step.

Figure 9 shows the ψ values (horizontal axis) as a function of time (vertical axis; right-hand scale) for the 13 reverse integration steps. Also shown is the free-energy surface (left-hand scale). We find that the system rapidly escapes from the free-energy minimum and reaches the barrier near $\psi = 1$ rad after three reverse integration steps (-1.5 ps). The subsequent fluctuations of $\overline{\psi}(-t;\psi_0)$ about the free-energy maximum are caused by statistical uncertainties in evaluating $\overline{\psi}(t;\psi_0) - \psi_0$, and by the possible role of an additional slow variable, as inferred from the study of Bolhuis *et al.*[34]

The time reversibility of classical mechanics is the basis for one of the main objections to Boltzmann's kinetic theory, the so-called "irreversibility paradox." Here, one might na-

ively expect that if we reverse each of the replica trajectories by changing the sign of the initial velocities, our *reverse* integration scheme would turn into a *forward* integration. This is not the case. For each of the initial structures along the reverse-integration path found before, we run each of the 50 replicas forward in time with initial velocities of opposite sign. The resulting reverse ψ values are also shown in Fig. 9, and are found to agree well with those obtained for the original initial velocities, illustrating the "irreversible" time evolution of the averaged coarse variables. To qualify this result, we point out that the integrator used in the MD simulations is not fully time reversible because of thermostatting and the bond constraint algorithm.

While reverse integration will take us "up" the free-energy surface, it will eventually search for "mountain tops" rather than saddles. Even so, it can still be used as a tool to help explore the free energy surface. For a coarsely two-dimensional surface, techniques for approximating the stable manifolds of fixed points in dynamical systems can be used to efficiently draw the surface by reverse integration. In effect, a circle of points surrounding a well-bottom gives a one-parameter family of initial conditions for reverse coarse projective integration that can be used to "triangulate" the surface. Computational approaches to these problems for explicit ordinary differential equations are well developed[53,54] and we expect that they can find good use in the case of "coarse stable manifolds." It is also worth mentioning that, if close to a saddle a large separation of time scales exists between a slow unstable mode and many fast stable ones, projective backward integration may indeed approach the saddle.

F. Kinetics of interconversion

To estimate the rate of escape from the α helical minimum into the extended minimum, we apply the Chapman–Kolmogorov relation, Eq. (9). The propagators are constructed from 56 of the runs of 50 replicas used before to estimate the free energy derivative. We use a simple linear interpolation of the propagators, Eq. (12), that ignores a small translational correction for the narrowly spaced starting points ψ_0. Figure 6 compares the free energy surface predicted from Chapman–Kolmogorov iteration, $\beta G(\psi) = -\ln p(\psi,t\to\infty|\psi_0,0)$, to the free energy from the equilibrium run. The agreement is excellent, with the possible exception of the poorly sampled barrier near $\psi = -2.15$ rad. This shows that Chapman–Kolmogorov iterations are indeed applicable to estimate the equilibrium distribution.

To test the applicability to kinetics calculations, we determine the rate of escape from the lowest-free energy well at $\psi_0 = -0.3$ rad into the extended well with 2 rad$< \psi < 2\pi - 2.5$ rad which defines the absorbing region [$h(\psi) = 0$] in the iteration, Eq. (9). By integration over ψ, we find that the survival time distribution rapidly becomes exponential with a time constant of 920 ps (see Fig. 10). As a reference, we also determine the corresponding survival time distribution from the two equilibrium runs. From the 7- and 24-ns runs, we estimate mean-first-passage times of about 400 and 800 ps, respectively. The rate constant for escaping from the alpha-
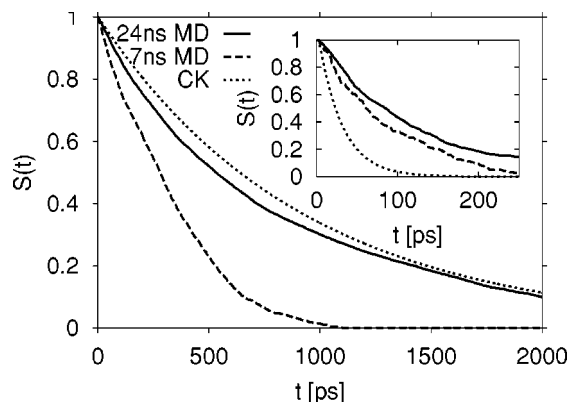
FIG. 10. Survival time distribution to reach the extended state (2 rad$< \psi$ $< 2\pi - 2.5$ rad) from the α-helical well, starting from $\psi_0 = -0.3$. Results are shown for the two equilibrium runs with 265 (24 ns; solid line) and 607 water molecules (7 ns; long-dashed line), and for Chapman–Kolmogorov iterations (dotted line). The inset shows the survival time distributions to reach the α-helical minimum from an extended configuration ($\psi \approx \pi$).

helical well to the extended well calculated by Chapman–Kolmogorov iterations is thus within about a factor of 2 of the value from simulation.

For the backward rate coefficient to go from the extended minimum ($\psi \approx \pi$) to the α-helical minimum (-0.6 rad$< \psi < 0$), the Chapman–Kolmogorov iteration gives a rate coefficient of about $1/(33 \text{ ps})$, consistent with the free energy difference of about 3.3 $k_B T$ (Fig. 3) and the forward rate coefficient of $1/(920 \text{ ps})$ in a two-state model. Both MD simulations give backward rate coefficients of about $1/(100 \text{ ps})$, consistent with the free energy difference of about 2 $k_B T$ and a forward rate coefficient of $1/(800 \text{ ps})$ for the 24-ns run, but too slow for the forward rate of $1/(400 \text{ ps})$ estimated for the 7-ns run. We have confirmed these rate coefficients and the underlying two-state model by determining the decay time of the number correlation function, $\langle \theta[\psi(t)]\theta[\psi(0)]\rangle$ [with $\theta(\psi) = 1$ inside one well and 0 outside] and relative populations in the two wells. In the 24-ns simulation, the number correlation decays exponentially over a broad time range. The decay time is insensitive to the particular choice of dividing lines in the barrier regions, and a two-state model is consistent with the above rates and the observed equilibrium coefficient.

For comparison, we also calculate the rates of interconversion by numerical solution[63] of the Smoluchowski diffusion equation along ψ. With a diffusion coefficient $D = 0.15$ rad^2 ps^{-1} and a free energy surface $G(\psi)$ from the equilibrium simulations, we obtain forward and backward rate coefficients of $1/(600 \text{ ps})$ and $1/(120 \text{ ps})$, respectively. These values are in good agreement with the rate coefficients from the MD simulations and from the Chapman–Kolmogorov iterations.

Our analysis demonstrates how CMD can use subpicosecond dynamics to extrapolate by three orders of magnitude to nanosecond dynamics. One concern may be that in a system with many wells, it will be impossible *a priori* to sample all relevant space in the coarse variables. With reverse integration and Newton–Raphson searches, however, we can create propagators for ensembles of configurations that con-

nect between neighboring wells. Starting from different initial structures in a given well, the search will lead to multiple exit routes from that well. The Chapman–Kolmogorov iteration for this set of local propagators can then be used to project forward in time and carry the system over barriers.

V. CONCLUSIONS

The CMD approach uses methods closely related to those of other approaches aiming at the long time dynamics. In constrained dynamics,[35] a single "coarse" coordinate is held fixed in time and the corresponding mean force is evaluated. In CMD, we evaluate instead the drift in the space of coarse variables resulting from the interactions with the rest of the system. In Voter's parallel replica method,[7] multiple replicas search for exit routes from a metastable state. Here, we use multiple replicas to determine the short-time dynamics in the coarse variables. This allows us to *construct* paths out of free-energy wells by recursive application of the Chapman–Kolmogorov identity, by Newton–Raphson search for saddle points, or, sometimes, by reverse integration. In the former approach, information about rare events is contained in the tails of the propagators constructed from MD, similar to the advancing replica in Voter's parallel replica method.[7]

In the transition path sampling approach developed by Chandler, Dellago and co-workers,[6,36] dynamic paths connecting reactant and product regions[64] are efficiently sampled. In CMD, we can use local short-time propagators to build transition paths in the space of the *coarse* variables. With the Chapman–Kolmogorov approach, we can estimate the kinetics of barrier crossing. In addition, we can sometimes exploit reverse integration to *construct* multiple coarse transition paths. In the case of a single dominant "coarse path," deterministic optimization methods (preferably derivative-free algorithms) may be wrapped around the coarse timestepper to locate the path.

To explore conformation space, Huber and van Gunsteren[37] have developed a method that couples the dynamics of multiple replicas to their average structure. These authors point out that "the average structure of a swarm of molecules converges faster to the structure with lowest energy than individual molecules do." In adapted form, this is an essential element of the CMD approach to mapping the coarse free energy surface. In the Newton–Raphson search and projective forward integration of the coarse variables, we use the time evolution of the averaged coarse variables to converge rapidly to free energy minima. In an earlier paper, Huber et al.[38] have explored the idea of adding memory to molecular dynamics to enhance conformational sampling. In a recent paper, Laio and Parrinello expanded on this approach and ingeniously combined coarse timestepping with "building in" a memory that allows the simulation to explore the free energy surface by, in effect, filling up the wells.[9] In their approach, repulsive markers are left behind along a trajectory projected into the coarse space. Eventually, these markers drive the system over a barrier leading out of the well (and, in the process, effectively mapping the well out). Our "stable manifold through reverse integration" maps out

the wells by essentially reversing the deterministic part of the coarse dynamics.

We have shown here that CMD can be used for rapid searches of the conformation space of flexible molecules in aqueous solution. We demonstrated how the results of multiple short (0.5 ps) replica runs can be combined to determine the free energy surface using two different methods. We have illustrated how projective forward and reverse integration can be used to move toward free energy minima and saddles. We have also shown how dynamics at the subpicosecond time scale can be used to predict the slow (0.5–1 ns) kinetics of barrier crossing.

These encouraging results show that CMD is a robust method. From the work of Bolhuis et al.[34] we know that additional "solvent" coordinates are kinetically relevant. The subspace monitored here does thus not cover the slow dynamics completely. Similar behavior is expected in many practical applications. While CMD is very flexible with respect to the inclusion of additional variables, the construction of slow (or "hydrodynamic") variables is often difficult. Even for the Ala dipeptide, Bolhuis et al.[34] could only identify the second relevant variable in vacuum, but not in water. In some cases, mode-coupling approaches, as discussed by Kostov and Freed[65] for polymer dynamics,[66,67] may be useful in expanding the set of coarse variables. Here, we show that CMD can give promising quantitative results even if only part of the slow dynamics is covered. An explanation for this result is the averaging invoked. By starting from configurations with similar ψ values, but different drift directions (or commitment probabilities), we average over additional slow variables. Moreover, the presence of such variables can be apparent in the data as a large scatter in the drift velocities and directions. Such scatter highlights the need to either add new variables or perform additional simulations.[18]

In summary, the CMD approach provides an integrated framework, individual components of which are closely related to various approaches aimed at overcoming the time-scale problem in MD. Here, we have demonstrated that CMD is a useful approach to extract thermodynamic and kinetic properties of molecular systems, and to extrapolate their long-time dynamics.

One of the ambitions of coarse computation is not only to map "coarse phase space" effectively, but also to search parameter space efficiently. Indeed, coarse bifurcation algorithms can be implemented based on the coarse timestepper approach that converge on parameter values at which qualitative transitions occur for the coarse dynamics (see, for example, Refs. 18, 20, and 22). In the present context, this would correspond to finding the regime of parameters like temperature or solution ionic strength at which coarse free energy wells form. Such coarse timestepper based numerical bifurcation and continuation techniques have been demonstrated for kinetic Monte Carlo, Lattice–Boltzmann, and Brownian dynamics cases; we are currently transferring this computational technology to the CMD context. Finally, the CMD approach is not limited to dynamics on a classical energy surface and can be implemented equivalently if the dynamics occurs on a quantum surface as, for instance, in Car–Parrinello MD.[44,68]

[1] M. P. Allen and D. J. Tildesley, *Computer Simulation of Liquids* (Clarendon, Oxford, 1987).

[2] Y. Duan and P. A. Kollman, Science (Washington, D.C., U.S.) **282**, 740 (1998).

[3] I. C. Yeh and G. Hummer, J. Am. Chem. Soc. **124**, 6563 (2002).

[4] M. Tuckerman, B. J. Berne, and G. J. Martyna, J. Chem. Phys. **97**, 1990 (1992).

[5] A. F. Voter, Phys. Rev. Lett. **78**, 3908 (1997).

[6] C. Dellago, P. G. Bolhuis, F. S. Csajka, and D. Chandler, J. Chem. Phys. **108**, 1964 (1998).

[7] A. F. Voter, Phys. Rev. B **57**, R13985 (1998).

[8] D. Passerone and M. Parrinello, Phys. Rev. Lett. **87**, 108302 (2001).

[9] A. Laio and M. Parrinello, Proc. Natl. Acad. Sci. U.S.A. **99**, 12562 (2002).

[10] R. Zwanzig, Phys. Rev. **124**, 983 (1961).

[11] H. Mori, Prog. Theor. Phys. **33**, 423 (1965).

[12] R. Zwanzig, *Nonequilibrium Statistical Mechanics* (Oxford University Press, New York, 2001).

[13] C. Theodoropoulos, Y. H. Qian, and I. G. Kevrekidis, Proc. Natl. Acad. Sci. U.S.A. **97**, 9840 (2000).

[14] I. G. Kevrekidis, "Coarse bifurcation studies of alternative microscopic/hybrid simulators," Plenary Lecture, CAST Division, AIChE Annual Meeting, Los Angeles, 2000. Slides can be obtained at http://arnold.princeton.edu/~yannis/

[15] C. W. Gear and I. G. Kevrekidis, SIAM J. Sci. Comput. (USA) **24**, 1091 (2003).

[16] O. Runborg, C. Theodoropoulos, and I. G. Kevrekidis, Nonlinearity **15**, 491 (2002).

[17] K. Theodoropoulos, K. Sankaranarayanan, S. Sundaresan, and I. G. Kevrekidis, Phys. Rev. Lett. (to be published); can also be obtained as nlin.PS/0111040 at arXiv.org

[18] A. G. Makeev, D. Maroudas, and I. G. Kevrekidis, J. Chem. Phys. **116**, 10083 (2002).

[19] C. W. Gear, I. G. Kevrekidis, and C. Theodoropoulos, Comput. Chem. Eng. **26**, 941 (2002).

[20] A. G. Makeev, D. Maroudas, A. Z. Panagiotopoulos, and I. G. Kevrekidis, J. Chem. Phys. **117**, 8229 (2002).

[21] C. I. Siettos, A. Armaou, A. G. Makeev, and I. G. Kevrekidis, AIChE J. (to be published); can also be obtained as nlin.cG/0207017 at arXiv.org

[22] C. I. Siettos, M. D. Graham, and I. G. Kevrekidis, J. Chem. Phys. (to be published); can also be obtained as cond-mat/0211455 at arXiv.org

[23] I. G. Kevrekidis, C. W. Gear, J. M. Hyman, P. G. Kevrekidis, O. Runborg, and C. Theodoropoulos, Commun. Math. Sci. (to be published); can also be obtained as physics/0209043 at arXiv.org

[24] P. J. Rossky and M. Karplus, J. Am. Chem. Soc. **101**, 1913 (1979).

[25] B. M. Pettitt and M. Karplus, Chem. Phys. Lett. **121**, 194 (1985).

[26] A. G. Anderson and J. Hermans, Proteins: Struct., Funct., Genet. **3**, 262 (1988).

[27] T. Lazaridis, D. J. Tobias, C. L. Brooks, and M. E. Paulaitis, J. Chem. Phys. **95**, 7612 (1991).

[28] C. L. Brooks III and D. A. Case, Chem. Rev. (Washington, D.C.) **93**, 2487 (1993).

[29] M. Vásquez, G. Némethy, and H. A. Scheraga, Chem. Rev. (Washington, D.C.) **94**, 2183 (1994).

[30] T. J. Marrone, M. K. Gilson, and J. A. McCammon, J. Phys. Chem. **100**, 1439 (1996).

[31] C. Chipot and A. Pohorille, J. Phys. Chem. B **102**, 281 (1998).

[32] J. Apostolakis, P. Ferrara, and A. Caflisch, J. Chem. Phys. **110**, 2099 (1999).

[33] C. D. Poon, E. T. Samulski, C. F. Weise, and J. C. Weisshaar, J. Am. Chem. Soc. **122**, 5642 (2000).

[34] P. G. Bolhuis, C. Dellago, and D. Chandler, Proc. Natl. Acad. Sci. U.S.A. **97**, 5877 (2000).

[35] E. A. Carter, G. Ciccotti, J. T. Hynes, and R. Kapral, Chem. Phys. Lett. **156**, 472 (1989).

[36] C. Dellago, P. G. Bolhuis, and D. Chandler, J. Chem. Phys. **110**, 6617 (1999).

[37] T. Huber and W. F. van Gunsteren, J. Phys. Chem. A **102**, 5937 (1998).

[38] T. Huber, A. E. Torda, and W. F. van Gunsteren, J. Comput.-Aided Mol. Des. **8**, 695 (1994).

[39] N. N. Bogoliubov, *Problems of Dynamic Theory in Statistical Physics* (Gostekhizdat, Moscow, 1946), English Translation: *Problems of a Dynamical Theory in Statistical Physics*, in *Studies in Statistical Mechanics*, edited by J. de Boer and G. E. Uhlenbeck (North-Holland, Amsterdam, 1962), Vol. 1.

[40] P. Constantin, C. Foias, B. Nicolaenko, and R. Temam, *Integral Manifolds and Inertial Manifolds for Dissipative Partial Differential Equations* (Springer Verlag, New York, 1988).

[41] R. Temam, *Infinite-Dimensional Dynamical Systems in Mechanics and Physics*, Vol. 68 in *Applied Mathematical Sciences* (Springer Verlag, New York, 1988).

[42] C. Foias, M. S. Jolly, I. G. Kevrekidis, G. R. Sell, and E. S. Titi, Phys. Lett. A **131**, 433 (1988).

[43] G. M. Torrie and J. P. Valleau, Chem. Phys. Lett. **28**, 578 (1974).

[44] C. W. Gear and I. G. Kevrekidis, J. Comput. Phys. **187**, 95 (2003).

[45] A. J. Chorin, A. P. Kast, and R. Kupferman, Proc. Natl. Acad. Sci. U.S.A. **95**, 4094 (1998).

[46] A. J. Chorin, O. H. Hald, and R. Kupferman, Proc. Natl. Acad. Sci. U.S.A. **97**, 2968 (2000).

[47] A. J. Chorin, O. H. Hald, and R. Kupferman, Physica D **166**, 239 (2002).

[48] P. L. Geissler, C. Dellago, and D. Chandler, J. Phys. Chem. B **103**, 3706 (1999).

[49] G. Lamm and A. Szabo, J. Chem. Phys. **85**, 7334 (1986).

[50] A. Armaou and I. G. Kevrekidis, in Proceedings of the 2003 CDC Conference (2003).

[51] J. Li, P. G. Kevrekidis, C. W. Gear, and I. G. Kevrekidis, SIAM Multiscale Modeling Simul. (to be published).

[52] C. W. Gear and I. G. Kevrekidis, Phys. Lett. A (to be published); can be obtained as nlin.CD/0302055 at arXiv.org

[53] M. E. Johnson, M. S. Jolly, and I. G. Kevrekidis, Numer. Algorithms **14**, 125 (1997).

[54] M. E. Johnson, M. S. Jolly, and I. G. Kevrekidis, Int. J. Bifurcation Chaos Appl. Sci. Eng. **11**, 1 (2001).

[55] D. Barkley, I. G. Kevrekidis, and A. Stuart (unpublished).

[56] C. W. Gear and I. G. Kevrekidis, J. Comput. Phys. (to be published); also NEC Technical Report No. NECI TR 2002-031N, Oct 2002; can be obtained as physics/0211043 in arXiv.org

[57] C. W. Gear, J. Li, and I. G. Kevrekidis, Phys. Lett. A (to be published); can be obtained as physics/0303010 at arXiv.org

[58] W. D. Cornell, P. Cieplak, C. I. Bayly *et al.*, J. Am. Chem. Soc. **117**, 5179 (1995).

[59] W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein, J. Chem. Phys. **79**, 926 (1983).

[60] T. Darden, D. York, and L. Pedersen, J. Chem. Phys. **98**, 10089 (1993).

[61] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. D. Nola, and J. R. Haak, J. Chem. Phys. **81**, 3684 (1984).

[62] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, J. Comput. Phys. **23**, 327 (1977).

[63] D. J. Bicout and A. Szabo, J. Chem. Phys. **106**, 10292 (1997).

[64] L. R. Pratt, J. Chem. Phys. **85**, 5045 (1986).

[65] K. S. Kostov and K. F. Freed, J. Chem. Phys. **106**, 771 (1997).

[66] M. Bixon and R. Zwanzig, J. Chem. Phys. **68**, 1896 (1978).

[67] A. Perico, R. Pratolongo, K. F. Freed, R. W. Pastor, and A. Szabo, J. Chem. Phys. **98**, 564 (1993).

[68] R. Car and M. Parrinello, Phys. Rev. Lett. **55**, 2471 (1985).