Calculation of conformational transitions and barriers in solvated systems: Application to the alanine dipeptide in water

Joannis Apostolakis, Philippe Ferrara, and Amedeo Caflisch[a)]
Department of Biochemistry, University of Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

(Received 31 August 1998; accepted 19 October 1998)

Optimal free energy paths (OFEPs) for conformational transitions are parallel to the mean force at every nonstationary point of the free energy landscape. In contrast to adiabatic paths, which are parallel to the force, OFEPs include the effect of entropy and are relevant even for systems with diffusive degrees of freedom. In this study the OFEPs are computed for the alanine dipeptide in solution. The potential of mean force is calculated and an effective potential is derived that is used to obtain the paths with a minimization based algorithm. The comparison of the calculated paths with the adiabatic paths *in vacuo* shows the influence of the environment on conformational transitions. The dynamics of the alanine dipeptide in water are more complex, since there are more minima and the barriers are lower. Two simpler methods for the calculation of reaction pathways in solution are evaluated by comparing their results with the OFEPs. In the first method the mean electrostatic field of the water is approximated by an analytical continuum model. The obtained paths show qualitative agreement with the OFEPs and the height of the barriers are similar. Targeted molecular dynamics (TMD), the second approach, constrains the distance to a target conformation to accelerate the transition. In the general case, however, it is difficult to assess the physical significance of the obtained paths. Changing the initial conditions by assigning different velocities leads to different solutions for the conformational transition. Furthermore, it is shown that by performing the simulations with different reaction coordinates or in opposite directions different pathways are preferred. This result can be explained by the structure of the free energy landscape around the initial conformations. In a first approximation the physical significance of different pathways is assumed to depend mainly on the free energy at the highest saddle point. In the literature the total energy of the system has often been used to estimate the position and the height of the energy barriers in the path. By comparing the total energy with the calculated free energy it is shown that the former largely overestimates the height of the barriers. Furthermore, the positions of the maxima of the total energy do not coincide with the free energy barriers. Simple approximations to the free energy lead to good quantitative agreement. © *1999 American Institute of Physics.* [S0021-9606(99)50204-3]

I. INTRODUCTION

Transitions between conformational states of molecules are important in a number of biological problems. For example, signal transduction and the regulation of protein properties often depend on such conformational changes. Protein folding and denaturation correspond to the interconversion between the native and the denatured state. Finally, chemical reactions can also be seen as transitions between two conformations (configurations) of the participating nuclei and electrons. The kinetics of a transition depends on the free energy barriers found along the pathway and the transmission coefficient, which takes into account the effect of trajectories recrossing the barrier.[1] Both the free energy barriers and the transmission coefficient depend on the pathway chosen for the transition. Therefore, knowledge of the pathways is necessary for understanding the kinetics. Moreover, it is known that the efficiency of free energy calculations depends on the free energy barriers encountered on the way.[2] This is true for conformational transitions as well as for chemical reactions. Reaction paths are especially important when kinetically trapped states are being considered. They are, however, difficult to observe experimentally. Structure determination methods like nuclear magnetic resonance and x-ray crystallography yield data on average structures of ensembles of many molecules. In most cases it is difficult to synchronize the transitions in such a way that time resolution of the process is possible. Furthermore, it is conceivable that a number of distinct possible pathways exist. Theoretical methods like energy minimization and molecular dynamics offer a convenient possibility to study these issues on a single molecule basis. The price one has to pay for this is the limited time scale. Most large scale conformational changes have to be accelerated in order to be observed in the time scale of a computationally feasible simulation.

In this work three different methods are used to obtain the transition pathways between different conformations of the alanine dipeptide (AD). In the first approach the two-dimensional potential of mean force of the AD is calculated and the free energy is approximated by an analytical func-

[a)]Author to whom all correspondence should be addressed. Electronic mail: caflisch@bioc.unizh.ch

tion. The application of the conjugate peak refinement algorithm (CPR)[3] on this analytical function leads to optimal free energy paths (OFEPs). These are more relevant for kinetics than adiabatic paths, which have been the focus of previous research.[4] Furthermore, the free energy barriers along the pathway can be assessed. The calculated reaction paths prove the dependence of the transition pathway on the environment. A number of pathways that are energetically unfavorable *in vacuo* become accessible in water. Barriers for transitions are generally lower in water.

In the second approach the effect of the solvent is included implicitly through the use of the analytical continuum electrostatics (ACE) solvation model of Schaefer and Karplus.[5] ACE was implemented in the CHARMM23 program and used together with CPR to obtain the adiabatic map and the steepest descent pathways for the AD in solution. The paths qualitatively agree with the ones from the potential of mean force (PMF) calculations. It can be shown that the main electrostatic effect of the water lies in the screening of the intrasolute interactions. This method is, in principle, related to the first, since the implicit model corresponds to a mean force potential of the solvent.

Targeted molecular dynamics (TMD), the third approach, is a simple method to accelerate transitions.[6] It has recently been used independently by Diaz *et al.*[7] and Ma and Karplus[8] to calculate reaction paths for the conformational changes in *ras* p21. In TMD the root mean square distance (RMSD) to a target structure is constrained and continuously decreased during the simulation. The system is slowly forced to find a path from its initial state to the final conformation. This time-dependent constraint introduces a bias which is minimal in the sense that only one degree of freedom is constrained. With TMD the Cartesian distance between conformations is assumed to be an adequate descriptor of the progress along the reaction path, a reaction coordinate in other words. It is easy to imagine paths for which this assumption is not valid. When a number of distinct pathways exist, the choice of the reaction coordinate will influence the results significantly. Furthermore, although in most studies using TMD the forward and reverse direction for the simulation is performed, the question of the stability of the solutions with respect to small perturbations of the initial conditions is still open.

TMD is applied to the conversion of the C7 axial to the C7 equatorial conformation of the AD in solution. The significance of the pathways and the corresponding barriers is investigated through comparison with the results from the PMF calculations. In this work three main differences to the original TMD exist. First, the system is restrained to the conformational subspace corresponding to a particular value of the reaction coordinate (RC) by a harmonic term. This allows control of the stringency, with which the restraint is applied. In the original method the RC is controlled by the application of a holonomic constraint. Second, the RC is discontinuously decreased in small steps to allow equilibration and the calculation of average properties. Last, a different choice for the reaction coordinate is investigated. It is set equal to the ratio η of the RMSD to the initial structure over the sum of the RMSDs to the initial and the final structure. This choice of the reaction coordinate has already been used by Czerminski and Elber in reaction path studies based on energy minimization.[9] It is more general since for any arbitrary conformation its value lies between zero and one. During the simulation η is slowly increased from zero to one. Therefore, this RC is less restrictive. A number of pathways are found in the η simulations that are not found with the original TMD. Multiple simulations with different random seeds for the assignment of the initial velocities show that a number of possible solutions exist for the pathway.

The total energy of the system has been used in the literature to find the energy barriers along the RC and to assess their height.[7,8] The comparison with the PMF calculation shows that the total energy is only weakly correlated with the free energy. Marrone *et al.*[10] have shown that an implicit solvation model, which models the electrostatic effect of the solvent with the help of the Poisson equation, gives a good description of the one-dimensional PMFs of the AD in water. Further, Smart *et al.*[11] have shown that a similar model leads to reasonable results for the free energy minima of the AD in water. The use of the former model in this work leads to an expression for the free energy in solution which shows good agreement with the explicit water simulation for the positions and the heights of the barriers, thus extending previous results. Furthermore, a second free energy function is constructed, based on an approximation to the solvation free energy, which is linear in the interaction energy.[12] This model also shows satisfactory agreement with the PMF calculation.

II. METHODS

All simulations were performed with the CHARMM23 program. The water model and the polar-hydrogen model from the CHARMM parameter set 19 were used, i.e., polar hydrogens are treated explicitly, while nonpolar hydrogens are included in the heavy atom they are bound to.[1,13]

A. Minimal free energy paths

The two-dimensional PMF of the AD in water has been presented previously by other groups.[14,15] In this study it was calculated by an adaptive umbrella potential scheme,[16,17] that was implemented in the CHARMM23 program. The PMF is iteratively approximated by a linear combination of basis set functions:

$$U_i(\lambda_1,\lambda_2)=\sum_{\alpha_1}\sum_{\alpha_2} a_{\alpha_1\alpha_2} f^{(1)}_{\alpha_1}(\lambda_1) f^{(2)}_{\alpha_2}(\lambda_2). \tag{1}$$

U_i is an analytical approximation to the PMF extracted from the ith simulation, $\lambda_{1,2}$ are the degrees of freedom over which the PMF is to be calculated, $f^{(j)}_{\alpha_j}(\lambda_j)$, $j=(1,2)$ are trigonometric functions of the λ variables. The coefficients $a_{\alpha_1\alpha_2}$ are determined by a linear least-squares fit.[17] For the AD in the polar–hydrogen model,[1,13] there are only two soft degrees of freedom, the ϕ and ψ backbone dihedral angles. For each dimension 13 functions were used for the basis set, $\sin(n\lambda)$, $\cos(n\lambda)$ $(n=1–6)$ and a constant.[17] The inverted potential of mean force extracted from simulation i is added to the energy function for simulation $i+1$. Upon convergence

the simulations give uniform sampling in the n degrees of freedom. Two-hundred iterations were performed. Each iteration consisted of 10 ps of equilibration and 40 ps of data collection. The analytical approximation to the PMF that was obtained from the adaptive umbrella potential calculation (U_{200}) was used to create a simplified model of the AD in solution. The model consists of five linearly bonded atoms mimicking the backbone of the AD. The energy of the system is given by:

$$U_{\text{tot}} = U_{\text{bond}} + U_{\text{angle}} + U_{\phi,\psi}, \tag{2}$$

where $U_{\phi,\psi}$ is equal to U_{200}. The bond and bond angle energy terms ($U_{\text{bond}}, U_{\text{angle}}$) are used to keep the geometry similar to that of the AD backbone. Since no nonbonded terms are included and only minimization is applied, the coupling between terms can be neglected. The adiabatic map of this model corresponds to the calculated PMF of the AD in water. The CPR algorithm[3] as implemented in the TRAVEL module of CHARMM23 was used to obtain an approximation to the steepest descent path for the transition from the initial to the final state. CPR starts from the initial and final conformations to find a connecting path. The saddle points on the path are identified and refined. From each side of a refined saddle point a steepest descent minimization is started. Assuming there is a single saddle point on the path, these minimizations should lead to the initial and final conformations. For a more detailed description of the method the reader is referred to the original literature.[3] Application of this algorithm to the simplified AD molecule yielded the optimal free energy paths for the conformational transitions between the different free energy minima in water. Similar to an adiabatic path, which follows the gradient at each nonstationary point, an optimal free energy path is defined to follow the mean force. Given an initial and a final conformation CPR finds exactly one pathway. In order to obtain different pathways for the same transition, an intermediate point close to the alternative saddle point is necessary. Therefore, a fair amount of intuition or, as is the case here, previous knowledge of the energy landscape is necessary.

B. The TMD method

The conformation of a protein can be described as a vector $\mathbf{X}$ in $3N$ dimensional space, where N is the number of atoms in the protein. The length of the difference vector between two conformations is proportional to their RMSD. In the original TMD, simulations are performed with an additional time-dependent holonomic constraint, which acts on the length of the difference vector between the actual conformation and a target structure:

$$\phi(\mathbf{X}) = |\mathbf{X} - \mathbf{X}_f|^2 - \rho(t)^2 = 0. \tag{3}$$

$\mathbf{X}_f$ is the vector corresponding to the target structure. $\rho(t)$ starts at a value equal to the distance between the initial and the final state and is slowly reduced to zero during the simulation. The space of the conformations satisfying this constraint corresponds to a hypersphere of radius $\rho(t)$ around the final conformation. In the present work, a harmonic restraint is employed to control the sampling:

$$\text{RMSD}_f = \frac{|\mathbf{X} - \mathbf{X}_f|}{\sqrt{N}}, \tag{4}$$

$$U_{\text{res}}(\mathbf{X}) = K(\text{RMSD}_f - \rho)^2, \tag{5}$$

where K is the force constant of the harmonic potential, set to 100 kcal/mol Å^2, N is the number of constrained atoms and RMSD_f is the RMSD to the final structure. ρ starts with a value equal to the RMSD between the initial and final structures. It is discontinuously decreased to zero in fifty 2 ps intervals. The RMSD thus corresponds to the RC, describing the progress toward the target conformation. It is important to note that this distance RC is not symmetric with respect to the initial and final structures since the hypersphere on which the system is restrained is centered on the target conformation. This implies that different parts of the conformational space are sampled in the forward and reverse simulations. Also, since the initial value of the RC corresponds to the RMSD between the initial and the target conformation and is monotonically decreased, pathways that lead over structures with a larger RMSD to the target conformation cannot be sampled. It is therefore reasonable to test other, less restrictive, choices for the RC. The ratio RC η is defined as the ratio of the distance to the initial conformation over the sum of the distances to the initial and final conformation:

$$U(\mathbf{X}) = K\left(\frac{|\mathbf{X} - \mathbf{X}_i|}{|\mathbf{X} - \mathbf{X}_i| + |\mathbf{X} - \mathbf{X}_f|} - \eta\right)^2. \tag{6}$$

K was set to 500 kcal/mol Å^2. Performing a simulation with η slowly changing from 0 to 1 pushes the system from its initial to the final state. A value of η between 0 and 1 restrains the system on a sphere at $\mathbf{L}$:[9]

$$\mathbf{L} = \frac{\eta^2 \mathbf{X}_f - (1-\eta)^2 \mathbf{X}_i}{2\eta - 1}, \tag{7}$$

with radius R:

$$R = \left|\frac{\eta(1-\eta)}{2\eta - 1}(\mathbf{X}_i - \mathbf{X}_f)\right|. \tag{8}$$

There are no global restrictions on the pathway. This means that, unlike the distance RC, conformations that differ from the final structure by more than $|\mathbf{X}_i - \mathbf{X}_f|$ can be sampled. Furthermore, this reaction coordinate is symmetric with respect to the initial and final structures. Therefore, there should in principle be no bias toward different paths for the forward and reverse directions of the simulation.

Since rigid body motions make it difficult to restrain the system close to the barriers, sampling of the saddle points becomes inefficient. The Eckart conditions were applied in the form of holonomic constraints to avoid rigid body translation and rotation,[18]

$$\sum_i M_i(\mathbf{r}_i - \mathbf{r}_i^{\text{ref}}) = 0, \tag{9}$$

$$\sum_i M_i \mathbf{r}_i \times \mathbf{r}_i^{\text{ref}} = 0. \tag{10}$$

Here the sum is over the atoms i of the solute, M_i is the atomic mass, $\mathbf{r}_i$ is the atomic coordinate vector and $\mathbf{r}_i^{\mathrm{ref}}$ is the reference coordinate. The reference coordinates are taken from either the initial or final conformation.

C. Simulations in water

Explicit water simulations were performed with periodic boundary conditions and a cubic box of 18.8 Å length. The box contained 206 water molecules. The SHAKE method was used to constrain the bonds in the water molecules.[19] The time integration step was 1 fs. The minimized system was heated during 5 ps by reassigning the velocities every 50 steps according to a Gaussian distribution. The system was then equilibrated for 10 ps at 300 K. For the calculation of the pathway with the distance RC, ρ was initially set to 1.357 Å, corresponding to the RMSD between the reactant and the product. After every 2 ps of simulation time ρ is decreased such that it is equal to zero at 100 ps. For the ratio method η is initially set to zero and increased by 0.02 every 2 ps until it reaches one, leading to a total simulation time of 100 ps.

D. Implicit solvation models

The two implicit solvation models used in this study were based on the following equation:

$$G_{\mathrm{tot}} = E_{\mathrm{vac}} + \Delta G_{\mathrm{solv}}. \tag{11}$$

G_{tot} is the total free energy in solution, E_{vac} the energy of the solute *in vacuo* and ΔG_{solv} the electrostatic free energy of solvation. Nonpolar contributions are usually approximated by a solvent accessible surface-dependent term, which was neglected here, since the accessible surface of the AD changes by 12Å^2 at most. The calculation of the electrostatic solvation term was performed in two different ways: with an approximate model (ACE) and with a finite difference scheme for the solution of the Poisson equation. The ACE force field is an analytical approximation to the semianalytical method of Still *et al.*[20] It is based on the Coulombic approximation for the calculation of the self-energy of solvation and on the generalized Born equation for the screening of solute–solute interactions. ACE is not as accurate as the solution of the Poisson equation, but being analytical it yields the energy gradients which are necessary for minimization and dynamics. It was implemented and used with the parameters described in the original work.[5] The smoothing parameter (α) for the atomic volumes was set to 1.6, the dielectric constant was set to 1.0 and 78.5 for the solute and the solvent, respectively. The Poisson equation was solved by a finite difference scheme with the program UHBD.[21,22] The molecular surface was used to define the dielectric boundary and smoothing was applied.[23] For the dielectric constants the same values were used as in the ACE force field. Two steps of focusing were performed for every structure. The spacing of the final grid was 0.4 Å .

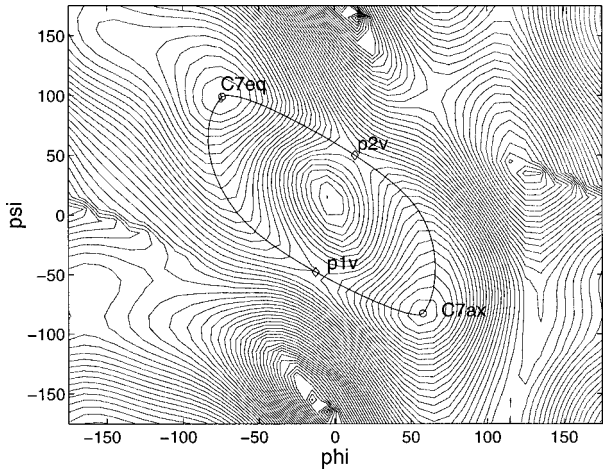

FIG. 1. Adiabatic surface of the AD *in vacuo* in dependence of the ϕ and ψ backbone dihedral angles. Contour lines are drawn at a distance of 0.5 kcal/mol. The adiabatic paths connecting the minima were calculated with CPR.

III. RESULTS

A. Steepest descent paths *in vacuo*

The adiabatic surface of the AD *in vacuo*, shown in Fig. 1, is very similar to the vacuum free energy surface calculated by adaptive umbrella sampling.[17] There are two significant minima ($C_{7\mathrm{eq}}$ and $C_{7\mathrm{ax}}$). Two different steepest descent paths *p1v* and *p2v* were obtained with CPR for the transition between the two minima. The saddle points lie 8.5 and 10.5 kcal/mol higher than the lowest minimum ($C_{7\mathrm{eq}}$), respectively (Table I). Since for every transition path there is exactly one highest saddle point, the paths and the corresponding saddle points are denoted by the same symbol.

B. Optimal free energy paths in water

For the simulations in water it is important to assess the statistical error of the PMF calculation and the error in the analytical approximation to the free energy. The statistical error was estimated as the standard deviation between the results obtained after 100 and 200 updates of the adaptive umbrella potential. This yields an error of 0.3 kcal/mol which lies well below kT (0.6 kcal/mol). The quality of the analytical approximation seems important, since even small deviations in the energy can lead to significant differences in the form of the OFEPs. This is the case particularly in regions where the energy function is flat. There, however, the system is not expected to follow the path faithfully because of thermal fluctuations. For the present purposes it is suffi-

TABLE I. Minima and transition states *in vacuo*. Positions and energies of the minima and highest saddle points along the paths *in vacuo*. Energies are given in kcal/mol.

Name	Coordinates	Energy
$C_{7\mathrm{eq}}$	(−74°,99°)	0.00
$C_{7\mathrm{ax}}$	(57°,−83°)	2.69
p1v	(−12°,−47°)	8.54
p2v	(13°,49°)	10.48

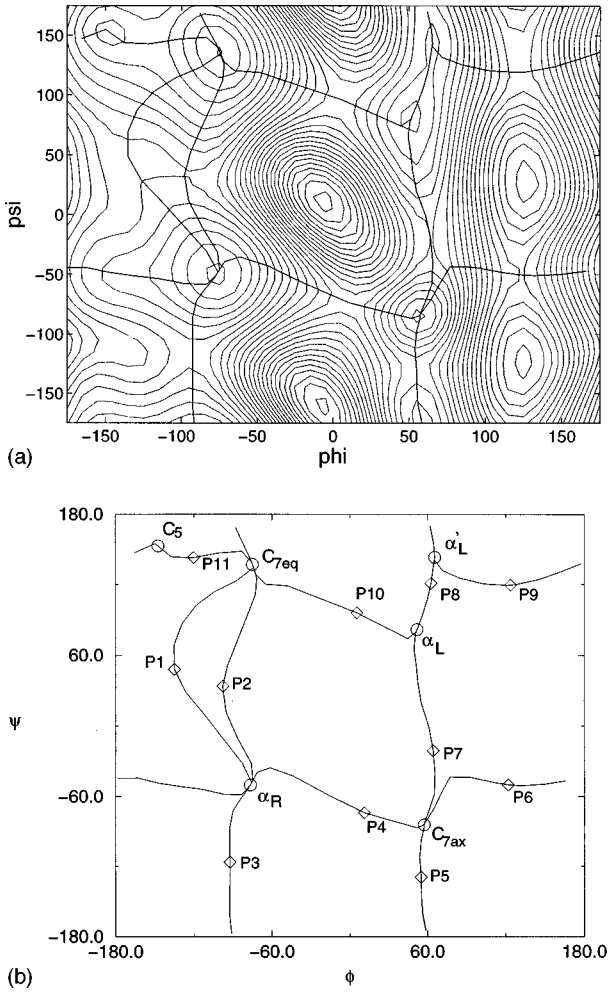

FIG. 2. (a) Free energy surface of the AD in water. Contour lines are drawn every 0.5 kcal/mol. The OFEPs connecting the minima are shown. For clarity the names of the minima and the saddle points are shown in (b). Due to periodicity, paths that end abruptly at the border of the plot are understood to be continued on the other side. (b) OFEPs for the AD in water. Minima are denoted by circles, saddle points by diamonds.

cient that errors in the energy lie below kT. The analytical approximation to the free energy of AD in water shows a standard deviation of 0.2 kcal/mol to the calculated PMF. This agreement is satisfactory and the discussion of the free energy potential will be limited to the analytical approximation. The free energy surface shows five minima already known from the literature[11] $C_{7\mathrm{eq}}, \alpha_R, \alpha_L, C_5$ and $C_{7\mathrm{ax}}$ (Fig. 2). There is an additional minimum (α_L') at position (65°, 143°). This minimum is separated from the α_L minimum by a barrier lower than 0.2 kcal/mol. Therefore, its physical significance is questionable, however it is included here because it influences the form of the transition paths.

In all there are 11 direct paths ($p1$–$p11$) leading from one minimum to another without passing through a third one. Paths $p1$, $p2$ and $p3$ lead from $C_{7\mathrm{eq}}$ to α_R and paths $p4$ and $p6$ lead from α_R to $C_{7\mathrm{ax}}$ via different saddle points. The remaining paths differ in at least one end point from each other. Some minima are not connected directly. For example there are no direct paths leading from $C_{7\mathrm{eq}}$ to $C_{7\mathrm{ax}}$. Instead a number of possible combinations of paths exist that lead over the intermediates α_R or α_L. To form such indirect paths any path leading from the initial conformation to the intermediate can be combined with any path leading from the intermediate to the final conformation. The barrier along such a path is defined as the difference between the highest saddle point and the lowest minimum on the path. The most reasonable paths leading from $C_{7\mathrm{eq}}$ to $C_{7\mathrm{ax}}$ are ($p1$, $p4$), ($p2$, $p4$), ($p3$, $p4$), ($p10$, $p7$), and ($p10$, $p8$, $p5$). A certain degeneracy exists in the sense that the barriers for all of these paths lie within kT, although most of these paths differ significantly from each other. The paths ($p1$, $p4$) and ($p10$, $p7$) can be compared to the $p1v$ and $p2v$ paths in vacuo, respectively. The saddle point $p10$ lies lower than $p4$ by 0.6 kcal/mol. This is in contrast to the situation *in vacuo*, where $p1v$ is more probable than $p2v$ by almost 2 kcal/mol. Furthermore, the barrier at $p10$ is also lower than the barrier at $p1v$ by almost 2 kcal/mol. Hence conformational transitions are facilitated by the solvent.

C. CPR with electrostatic solvation models

The effect of the solvent leads to the stabilization of certain conformations and a general flattening of the energy hypersurface (Figs. 1 and 2). It is known[10] that these effects are mainly electrostatic and can be modeled with the help of implicit solvent models. Here an analytical implicit solvation model is used for the calculation of reaction paths in water. The underlying assumption is that the solvent can be approximated well by the analytical model. In the generalized Born treatment the electrostatic contributions to solvation are divided into a self term and a screened interaction term. The self term corresponds to the direct interaction of solute charges with the dipoles of the solvent, while the screened interaction term describes the effect of the solvent on the solute charge–charge interactions.

The adiabatic map of the AD in the ACE force field is qualitatively very similar to the PMF in water. The agreement is satisfactory with a standard deviation to the PMF in water of 0.9 kcal/mol (Table II). The adiabatic map constructed using the Poisson equation for the electrostatics in the manner of Marrone *et al.*[10] shows a somewhat better qualitative and quantitative agreement with the PMF in water (standard deviation of 0.7 kcal/mol). The calculation of the possible pathways with CPR is straightforward and leads to the paths shown in Fig. 3. The paths are overall quite similar to the ones obtained from the PMF calculations. Therefore, the names of the paths have been maintained. Notable differences are paths $p1$ and $p3$, which with the ACE force field lead from minimum α_R to C_5. This is in contrast to $p1$ and $p3$ in water which lead to $C_{7\mathrm{eq}}$. This difference is more formal than substantial because the energy is relatively flat in that region. Plotting the paths found with the ACE force field in the PMF landscape shows that the energy of the saddle point does not change for $p1$ while it becomes less favorable by 1 kcal/mol for $p3$. In the ACE simulations no paths could be found that correspond directly to $p2$. The barriers are in general somewhat lower than in the explicit water simulation (Table II).

As has been mentioned earlier the AD has been used to show the agreement of numerical continuum electrostatic models based on the Poisson equation or the generalized

TABLE II. Minima and transition states in water. Positions and energies of the minima and highest saddle points along the paths. Energies are given in kcal/mol.

	Explicit water		ACE	
Name	Coordinates	Energy	Coordinates	Energy
C_{7eq}	(−75°,136°)	0.00	(−88°,143°)	0.00
α_R	(−76°,−50°)	1.41	(−84°,−42°)	0.71
C_{7ax}	(57°,−84°)	3.85	(59°,−78°)	4.34
α_L	(51°,81°)	4.38	(55°,73°)	4.35
α_L'	(65°,143°)	4.85	(63°,178°)	4.46
C_5	(−147°,152°)	1.34	(−142°,145°)	0.12
$p1$	(−135°,48°)	4.14	(−140°,42°)	2.61
$p2$	(−97°,33°)	4.07	⋯	⋯
$p3$	(−92°,−116°)	4.19	(−133°,−114°)	3.28
$p4$	(11°,−73°)	7.36	(7°,−75°)	6.88
$p5$	(55°,−129°)	6.34	(56°,−117°)	5.15
$p6$	(121°,−49°)	11.60	(120°,−54°)	9.46
$p7$	(64°,−21°)	6.51	(66°,−16°)	5.98
$p8$	(62°,121°)	4.97	(66°,130°)	4.69
$p9$	(123°,119°)	10.28	(120°,138°)	8.66
$p10$	(5°,96°)	6.75	(7°,98°)	6.47
$p11$	(−120°,142°)	1.83	(−121°,144°)	0.16

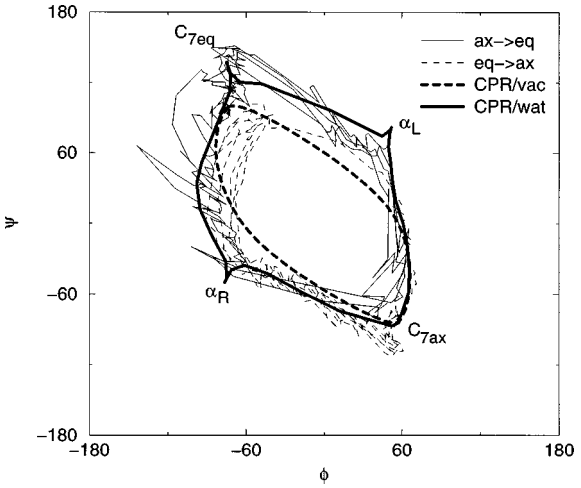

FIG. 4. Distance TMD paths in explicit water. Thin solid lines correspond to the $C_{7ax} \rightarrow C_{7eq}$ transition, thin broken lines correspond to the $C_{7eq} \rightarrow C_{7ax}$ direction. The adiabatic path *in vacuo* is shown in thick broken lines and the closest minimal free energy paths in water are shown in solid thick lines. Due to the initial equilibration and the fact that at the beginning of the simulation the system is constrained on a sphere with a radius of 1.357 Å around the target conformation the TMD trajectories do not start exactly at the initial conformation.

Born approach[24] with explicit water simulations. The results for the ACE force field are satisfactory and show that analytical implicit solvent models can be used to calculate transition pathways for simple systems.

A further point that is of interest, is whether it is the direct electrostatic interactions of the AD with the water dipoles or an indirect screening effect that leads to the flattening of the energy function. The simplest way to address this question is to use electrostatic models that describe only the screening effect of the solvent. The latter can be approximated by assuming a homogeneous dielectric equal to that of water ($\epsilon = 78.5$). In a homogeneous dielectric the self term is zero. In the case of the AD these approximations seem reasonable due to the small size of the solute. The adiabatic map with a constant dielectric permittivity of 78.5 shows qualitative agreement with the PMF in explicit water and the ACE solvation model (standard deviation of 0.9 and 0.7 kcal/mol, respectively). This result indicates that the main effect of water on the AD is the screening of the interactions within the solute. Moreover, more stringent test cases for the validation of implicit solvation models are needed, since it is shown that the homogeneous dielectric model, which is not accurate for more complex systems yields good approximations of the PMF of AD in solution. On the other hand the combination of results for the AD[10,24] with results from the solvation energy of small compounds,[20,25] where the dominant part of the electrostatic solvation is due to the self energy suggests that electrostatic solvation is described satisfactorily by Poisson and numerical generalized Born methods, at least for small systems.

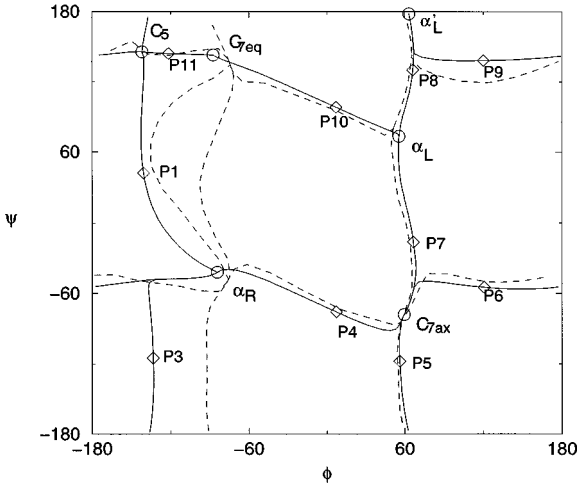

FIG. 3. CPR paths for the AD. Solid lines denote the paths found with the ACE force field, while dashed lines show the results obtained from the explicit water calculations for comparison. ACE minima are denoted by circles, saddle points by diamonds.

D. Distance TMD with explicit water

Six independent trajectories were calculated for the $C_{7ax} \rightarrow C_{7eq}$ and the $C_{7eq} \rightarrow C_{7ax}$ directions. The projections of the snapshots on the $\phi - \psi$ plane are shown in Fig. 4. The trajectories approximately follow the steepest descent paths $p1v$ and $p2v$ and ($p2$, $p4$) and ($p7$, $p10$). Because of the similarity of the paths, it is difficult to determine whether the trajectories follow the vacuum or water paths more closely. Figures 1 and 2 show that some paths in water are significantly different from any of the paths *in vacuo*. These are however not found in the trajectories of the distance TMD simulations, because only direct paths are allowed, i.e., paths for which the RMSD to the final conformation decreases monotonically. Of the six $C_{7ax} \rightarrow C_{7eq}$ trajectories four follow the ($p7$, $p10$) path passing close to the α_L minimum and two follow the path ($p4$, $p2$). In fact one of these two could also be considered to be following the ($p4$, $p1$) path, since $p1$ and $p2$ are quite similar. For the $C_{7eq} \rightarrow C_{7ax}$ direction the

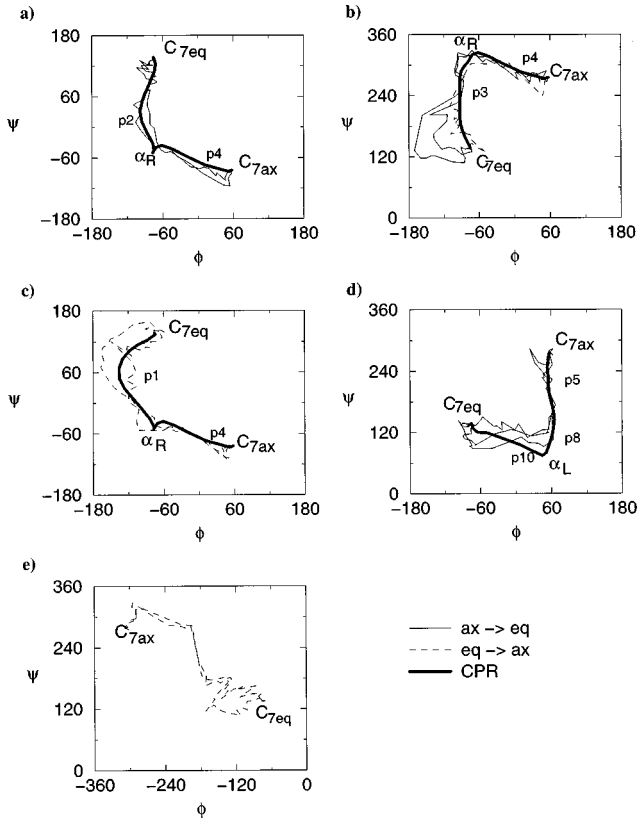

FIG. 5. Ratio TMD paths in explicit water. The results of the 12 trajectories are shown together with the closest minimal free energy paths (thick lines). Thin solid and thin dashed lines correspond to the $C_{7ax} \rightarrow C_{7eq}$ and the $C_{7eq} \rightarrow C_{7ax}$ direction, respectively. To maintain continuity of the paths it was necessary to shift the origin of (b), (d), and (e) by 180° in the ψ direction and by $-180°$ in the ϕ direction in (e). See also the description in the text.

preference is toward the ($p2$, $p4$) path with five trajectories taking approximately this path and only one taking the ($p10$, $p7$) path.

E. Ratio TMD with explicit water

Also for the ratio TMD a total of 12 trajectories were calculated. Five different types of trajectories are obtained. Two trajectory types follow paths that are also found with the distance TMD method [($p2$, $p4$) and ($p1$, $p4$) shown in Figs. 5(a) and 5(c), respectively], whereas two others follow paths that were not found either with CPR *in vacuo* or with the distance TMD [($p3$, $p4$) and ($p10$, $p8$, $p5$) shown in Figs. 5(b) and 5(d)]. For the $C_{7eq} \rightarrow C_{7ax}$ direction two paths are found that do not correspond directly to any of the optimal paths [Fig. 5(e)]. These paths are trapped in high energy regions. The reason for the occurrence of such trajectories is that with TMD the system at every point equilibrates within the confined space that satisfies the restraints. When the value of the restraint is changed the system readapts. This procedure pushes the system to find the optimal way to satisfy the constraints, but is however dependent on its previous position. In other words a local optimization of the single trajectory points takes place that does not take the path as a whole into account. It is therefore possible that single trajectories become trapped. In the case of the transition studied here, this is seen only with the ratio method in the $C_{7eq} \rightarrow C_{7ax}$ direction, due to the energy landscape around the C_{7eq} conformation. With the ratio method the system is pushed out of the reactant minimum in the most favorable way and the trajectories are initially biased toward the direction of minimum C_5. As the value of the RC is increased (>0.5) the system is pushed toward the product conformation. For the C_5 conformation, however, the direct way to the product leads over high barriers [Fig. 2(a)]. This does not occur with the distance TMD, because from the beginning the system is pushed toward the product. For the same reason distance TMD finds only direct routes. It is a peculiarity of the AD free energy landscape that the direct routes lead over reasonable barriers. The trapping is, therefore, not particular to the ratio RC, it can also occur with any other reaction coordinate. The combination of the sets of paths found with the distance and the ratio TMD includes all relevant pathways. This result is likely not to hold for larger systems, since, as will be shown in Sec. III F, the chosen RCs lead to a systematic bias toward particular paths.

F. Influence of the RC and the transition direction

From the TMD simulations in water it seems that different paths are preferred for the two different directions of the conformational change (Figs. 4 and 5). In order to obtain better statistics on this question a large number of additional TMD simulations were performed with the ACE force field, since it was shown that it leads to reasonable pathways with CPR and is significantly faster than the explicit water calculations. The resulting paths are compared to the CPR paths. For each RC and each direction 100 trajectories were calculated. For the distance RC in the $C_{7ax} \rightarrow C_{7eq}$ direction 31 trajectories followed the ($p4$, $p1$) path and 69 the ($p7$, $p10$) path. For the $C_{7eq} \rightarrow C_{7ax}$ direction the preference was reversed with 93 trajectories along the ($p1$, $p4$) path and 7 along the ($p10$, $p7$) path. These preferences agree with the results from the TMD simulations in explicit water. For the ratio method the situation was slightly more complex. For the $C_{7ax} \rightarrow C_{7eq}$ direction there were 59 trajectories corresponding to the ($p7$, $p10$) path, 39 to the ($p5$, $p8$, $p10$) path, one to the ($p4$, $p3$) path and one to the ($p7$, $p8$, $p9$, $p11$) path. For the $C_{7eq} \rightarrow C_{7ax}$ direction 55 trajectories did not correspond directly to any of the low energy paths and looked in principle similar to the paths shown in Fig. 5(e). These are considered to be trapped paths. Of the remaining trajectories, 44 followed the ($p1$, $p4$) path and one the ($p3$, $p4$) path.

The preference for certain paths for the different directions can be explained in a number of ways. In the distance TMD the constraint is not symmetric. In the $C_{7ax} \rightarrow C_{7eq}$ simulation it corresponds to a hypersphere centered on C_{7eq} while in the $C_{7eq} \rightarrow C_{7ax}$ the hypersphere is centered on C_{7ax}. Therefore, the available conformational space depends on the direction of the transition simulation. However, the fact that direction-dependent path preference is observed also with the ratio method, which is symmetric, shows that the asymmetry of the constraint is not the only reason for the direction dependence of the path preferences. Furthermore, the energetic landscape is different around the two confor-

mations. The initial effect of the energy landscape can lead to completely divergent paths. This is a particular problem of methods that calculate continuous trajectories. The path to be followed is chosen according to what appears initially like the lowest energy direction. Assuming there are a number of divergent paths connecting the two conformations, they will, by definition, converge toward the end points and diverge elsewhere. This means that in the beginning it is relatively probable that the trajectory jumps from one path to another. As the system progresses along the reaction coordinate sampling of other pathways is hampered by the increasing distance and the barriers between the paths. At the end points the path is well equilibrated, while later it may easily become trapped. However, the probability for the transition along a certain path depends mainly on the highest barrier encountered, and should be independent of the position of the barrier along the path. These arguments and the results of the TMD simulations stress the importance of performing multiple trajectories when studying conformational changes. Finally, the direction of the change is also important, since the constraint is being changed in small but discrete steps in a finite time. Therefore, the process is by definition irreversible.

In this study the ratio RC was necessary to show that the paths calculated with TMD in water follow the OFEPs in water and not the vacuum steepest descent paths. This is not clear from the distance TMD results. It is important to note that OFEPs are equilibrium paths, whereas with TMD the transition occurs in a relatively short period of time during which equilibration of the solvent may not be sufficient.

G. Free energy estimation

The calculation of multiple trajectories necessitates the use of criteria to assess the quality of the different solutions. In previous studies[7,8] the total energy was used. In solvated systems the total energy contains contributions from degrees of freedom that affect the free energy of the reaction only indirectly. At the same time they lead to significant oscillations in the energy. These may hide more important contributions. In the case of the AD in explicit water it is possible to compare the total energies with the free energies that were obtained by the PMF calculation. For the twelve trajectories performed with the ratio RC in explicit water the energies and conformations were averaged over the 2 ps, during which the RC is kept constant. Averaging was performed in a postprocessing step on conformations that had been saved every 50 steps of dynamics. As can be seen in Fig. 6(a) there is practically no correlation between the total energy and the free energy (the correlation coefficient is equal to 0.2). The statistical error due to the oscillations of the solvent–solvent energy is too large. Furthermore, Fig. 7 shows that the total energy peaks do not coincide with the free energy barriers for one of the twelve runs. Similar results are found for the other runs.

The estimation of the free energy of conformations along the path can be performed by employing an implicit solvation model. The electrostatic solvation energy is calculated by solving the Poisson equation for the conformation of the AD at the end of every 2 ps interval with a finite difference

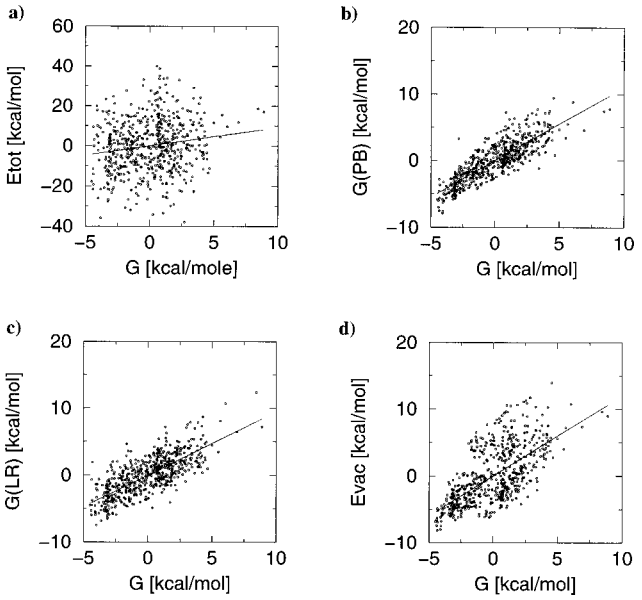

FIG. 6. Plot of different energy terms vs the free energy. (a) Total energy of the solute–solvent system. (b) Approximated free energy based on the Poisson equation. (c) Approximated free energy based on the linear response approximation. (d) Vacuum energy of the AD.

algorithm.[21,22] The free energy is then calculated as the sum of the energy of the AD *in vacuo* and the free energy of electrostatic solvation. The correlation coefficient to the free energy from the PMF calculation is 0.84 [Fig. 6(b)]. As can be seen in Fig. 7 the implicit model free energy is also a good indicator of the position of the barriers along the path. This result extends the results of previous work on the field which has focused on minimal conformations.[11] However, for dynamics it is important that barriers are also reproduced correctly.

Another possibility to estimate solvation energies is given by linear response theory[26] which indicates that the

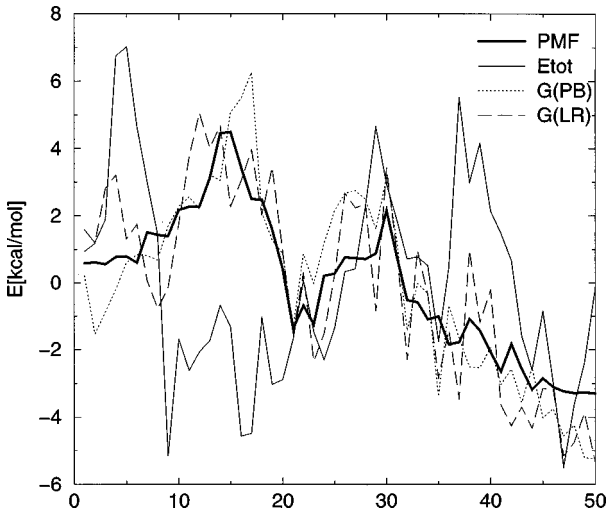

FIG. 7. Plot of the different energy terms for 50 snapshots along a typical trajectory of the ratio TMD. Thick line: Free energy in explicit water. Thin solid line: Total energy of the solute–solvent system. The total energy was scaled down by a factor of 5 for better comparison. Thin dashed line: Approximated free energy based on the Poisson equation. Thin dotted line: Approximated free energy based on the linear response approximation.

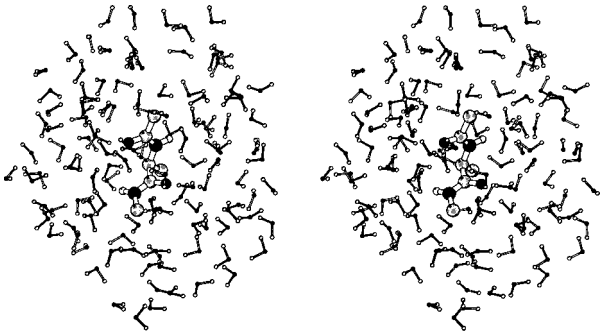

FIG. 8. Snapshot from a simulation with ratio TMD ($\eta = 1.0$) showing the AD and the water molecules that lie within the nonbonded interaction cutoff radius ($R_c = 7.5$ Å). Carbon atoms are light grey, nitrogen atoms black, oxygen atoms dark grey and hydrogen atoms white. The stereo plot was prepared using MOLSCRIPT (Ref. 29).

free energy of electrostatic solvation is proportional to one-half the average solute–solvent electrostatic interaction energy. Figure 8 shows the water molecules that lie within the nonbonded-cutoff distance and thus interact directly with the AD. For the van der Waals contribution a similar approach can be taken, where the factor of one-half is, however, not given and should be determined by fitting.[12] It was found that the van der Waals contribution is more or less constant for the AD. This is in agreement with the finding that the solvent accessible surface does not change significantly. The factor 0.5 for the Lennard–Jones interaction term was used for consistency. This leads to a further model for the free energy approximation as the sum of the average solute–solvent interaction energy with the average vacuum energy of the AD. Also this approximation is significantly better than the total energy both in terms of total correlation with the free energy [Fig. 6(c), $r = 0.75$] and in terms of estimation of barrier height and position (Fig. 7).

In large systems it is generally not possible to calculate the free energy landscape in dependence of the most important degrees of freedom. In such cases it will be of advantage to estimate the free energy with both the implicit and the linear interaction energy model. The extent of agreement between the two is an indirect measure of the validity of the estimation. However, one has to bear in mind that these models only lead to rough estimates of the barrier energy. At the same time both approximations are expected to become more critical for larger systems. Last, even though the two models might seem quite independent, one should bear in mind that implicit solvent models based on the Poisson equation are also essentially linear response models.[26]

IV. CONCLUSIONS

The calculation of optimal transition pathways *in vacuo* has been tackled in many ways. Most of the methods used are based on minimization with additional constraints.[3,4,27,28] However, in solvent the situation is even more complex; for every configuration of the solvent a somewhat different path is obtained.[27] It is obvious that the effect of the solvent has to be averaged in order to obtain relatively well defined optimal paths. At higher temperatures entropic effects become important and should be taken into account. With molecular dynamics the solvent can be equilibrated and its effect on the solute is averaged.

In this study the free energy surface of the AD in solution was calculated in dependence of the two backbone dihedral angles. This leads to the definition of OFEPs as the paths that at every nonstationary point follow the mean force. They can be understood as the mean paths followed in the slow transition limit. The aqueous environment flattens the energetic landscape and makes the dynamics more complex, since there are more minima and the number of probable pathways is significantly increased with respect to the situation *in vacuo*. For the same transition five different pathways with comparable probability exist. It will be interesting to investigate to which extent similar situations arise in more complex conformational changes in peptides or even proteins. Unfortunately, the calculation of OFEPs with the method described here becomes impracticable for all but the simplest systems. The significance of the calculation of the OFEPs for the AD lies in the possibility to use these results to validate simpler methods.

Analytical implicit solvent models based on the generalized Born equation can be used in conjunction with minimization-based schemes to obtain reaction paths. The results that were obtained for the AD with the ACE force field are reasonable. By using a second very simple model for the solvent (Coulomb energy with constant dielectric equal to that of water), it is shown that, for this system, the main effect of the solvent lies mainly in the screening of electrostatic interactions. Both models are very approximate and their validity should be investigated in each case.

In TMD, the system is restrained to changing values of a reaction coordinate during dynamics. The trajectories obtained contain thermal fluctuations that make comparison between pathways more difficult. Even for the simple system AD different possible paths are found by changing the initial conditions. Most of the calculated paths approximately follow the optimal free energy paths. An issue of particular importance with TMD is the bias that the chosen reaction coordinate exerts on the trajectory. A second reaction coordinate was introduced that allows better conformational sampling than when the RMSD is restrained. This second RC leads to more diverse paths and shows that also in TMD the paths depend significantly on the environment. Some of the paths obtained with the ratio RC are trapped in high energy regions. The different RCs bias toward different paths thus indicating that the results obtained depend on the RC choice. The results obtained are partly complementary to each other. The paths obtained with each RC are a subset of the possible paths and even with a large number of simulations certain possible paths are not found. This is particularly important considering the simplicity of the system. However, most of the paths found correspond to the most probable OFEPs. Furthermore, it is found that the direction of the transition has also a significant effect on the paths.

Recently TMD has been used independently by Diaz *et al.*[7] and Ma and Karplus[8] to calculate reaction paths for the conformational changes in *ras* p21. The two studies lead to different results for both the pathways and the height of

the energy barrier. Ma and Karplus suggest that their results are complementary to those of Diaz *et al.* Nevertheless, it is quite surprising that in the first study an energy barrier of approximately 300 kcal/mol is obtained, while in the second no barriers higher than 60 kcal/mol are observed. These results may arise from the different simulation environments used in the two studies. Diaz *et al.* use periodic boundary conditions while Ma and Karplus include only a shell of water around the protein. However, the differences may also be due to intrinsic difficulties in the calculation of the energy barriers. In this work this second possibility could be addressed. The results from the (exact) free energy calculation make it possible to test different models for the estimation of the height and the position of the free energy barrier. The comparison with the calculated free energy shows that the total energy overestimates the height of the barriers and shows maxima at different positions along the path than the free energy. Approximating the free energy with an implicit solvent model based on the Poisson equation is shown to be an efficient and accurate approximation for the AD, also for the high energy regions. Furthermore, the results show that the position and height of the barriers are predicted correctly. The results obtained from the application of a linear response model for this system are not as accurate as the implicit model but much more so than the total energy.

ACKNOWLEDGMENTS

The authors wish to thank Professor J. A. McCammon for providing the UHBD program, which was used for all the finite difference calculations. This work was supported by the Swiss National Science Foundation (National-fonds, Grant No. 31-53604.98), the Basel Chemical Industry Foundation, and the Olga Mayenfisch Foundation.

[1] E. Neria, S. Fischer, and M. Karplus, J. Chem. Phys. **105**, 1902 (1996).
[2] W. F. van Gunsteren, in *Computer Simulation of Biomolecular Systems*, edited by W. F. van Gunsteren and P. K. Weiner (ESCON, Leiden, 1989), pp. 27–53.
[3] S. Fischer and M. Karplus, Chem. Phys. Lett. **194**, 252 (1992).
[4] R. Elber, in *Recent Developments in Theoretical Studies of Proteins*, edited by R. Elber (World Scientific, Singapore, 1996), pp. 65–136.
[5] M. Schaefer and M. Karplus, J. Phys. Chem. **100**, 1578 (1996).
[6] J. Schlitter, M. Engels, P. Krüger, E. Jacoby, and A. Wollmer, Mol. Simul. **10**, 291 (1993).
[7] J. Diaz, B. Wroblowski, J. Schlitter, and Y. Engelborghs, Proteins: Struct., Funct., Genet. **28**, 434 (1997).
[8] J. Ma and M. Karplus, Proc. Natl. Acad. Sci. USA **94**, 119905 (1997).
[9] R. Czerminski and R. Elber, J. Chem. Phys. **92**, 5580 (1990).
[10] T. Marrone, M. Gilson, and J. McCammon, J. Phys. Chem. **100**, 1439 (1996).
[11] J. Smart, T. Marrone, and J. McCammon, J. Comput. Chem. **18**, 1750 (1997).
[12] J. Åqvist, C. Medina, and J.-E. Samuelsson, Protein Eng. **7**, 385 (1994).
[13] B. R. Brooks, R. E. Bruccoleri, B. D. OIlafson, D. J. States, S. Swaminathan, and M. Karplus, J. Comput. Chem. **4**, 187 (1983).
[14] A. G. Anderson and J. Hermans, Proteins: Struct., Funct., Genet. **3**, 262 (1988).
[15] B. M. Pettitt and M. Karplus, Chem. Phys. Lett. **121**, 194 (1985).
[16] R. Hooft, B. vanEijck, and J. Kroon, J. Chem. Phys. **97**, 6690 (1992).
[17] C. Bartels and M. Karplus, J. Comput. Chem. **18**, 1450 (1997).
[18] C. Eckart, Phys. Rev. **47**, 552 (1935).
[19] J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, J. Comput. Phys. **23**, 327 (1977).
[20] W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, J. Am. Chem. Soc. **112**, 6127 (1990).
[21] M. E. Davis and J. A. McCammon, J. Comput. Phys. **10**, 386 (1989).
[22] M. E. Davis and J. A. McCammon, J. Comput. Phys. **11**, 401 (1990).
[23] M. E. Davis and J. A. McCammon, J. Comput. Phys. **12**, 909 (1991).
[24] M. Scarsi, J. Apostolakis, and A. Caflisch, J. Phys. Chem. B **102**, 3637 (1998).
[25] M. Scarsi, J. Apostolakis, and A. Caflisch, J. Phys. Chem. A **101**, 8098 (1997).
[26] A. Dejaegere and M. Karplus, J. Phys. Chem. **100**, 11148 (1996).
[27] R. Elber, J. Chem. Phys. **93**, 4312 (1990).
[28] M. Engels, E. Jacoby, P. Krüger, J. Schlitter, and A. Wollmer, Protein Eng. **5**, 669 (1992).
[29] P. Kraulis, J. Appl. Crystallogr. **24**, 946 (1991).