

Calculation of the potential of mean force using molecular dynamics with linear constraints: An application to a conformational transition in a solvated dipeptide

Ron Elber

Department of Chemistry M/C 111, The University of Illinois at Chicago, P.O. Box 4348,
Chicago, Illinois 60680

(Received 6 April 1990; accepted 15 June 1990)

An algorithm is proposed for the calculation of mean force potentials using molecular dynamics with linear constraints. The new algorithm is based on free energy perturbation and is especially useful when the analytical form of the reaction coordinate is not known (e.g., for a numerically evaluated steepest descent path). It employs only Cartesian coordinates which makes it easy to use for a variety of systems. In particular, it is useful when the reaction coordinate cannot be represented by a single local coordinate or by a combination of a small number of local (internal) coordinates. The linear constraints are solved explicitly and the computational effort required is similar to that of dynamics with no constraints. The potential of mean force for a conformational transition of a solvated valine dipeptide is calculated. In aqueous solution the barriers for the conformational transitions are reduced with respect to the barriers in vacuo, the helix configuration becomes stabilized.

I INTRODUCTION

The potential of mean force was demonstrated to be a useful tool in the study of the relative stability of conformers and in the study of the kinetic of conformational changes and chemical reactions.¹⁻¹⁸ Thus, given the reaction path from the reactant to the product, the thermal probability of finding the system at different points along this coordinate is calculated. The thermal probability is directly related to the free energy profile and is useful in estimates of the rate based on the transition state theory. If the form of the reaction coordinate is known analytically, a number of different algorithms is available which were demonstrated to give converged results. The first approach which is very useful and widely employed is the use of umbrella sampling.^{1-7,12,13} Here the application of an alternative and a more recent technique, the free energy perturbation,^{8-11,14-17} is considered. Both approaches were employed in the past to calculate the potential of mean force for analytic reaction coordinates, expressed as a combination of a small number of internal coordinates (IC). An example is the rotation around a single^{1,2,6,9} or several^{3,5,13-15} bonds. For more complex systems, analytical reaction paths were suggested using gas phase calculations,^{4,8,10} chemical intuition,^{11,16-18} or exploratory simulations.^{7,12} Unfortunately if one considers the so called "intrinsic reaction coordinate"—the steepest descent path (SDP)—it is very rarely an exact IC, even for cases in which it can be *approximated* by a single internal coordinate.^{1,2,6,9} The SDPs are usually evaluated numerically and are given by a set of structures along the reaction coordinate. In general, it is not possible to represent this set as a motion along a single or a few internal coordinates. A classic example is of the conformational transitions in ring systems.¹⁹ For these molecules a new type of IC (pseudorotation) was invented, since the usual set of ICs (bonds, angles, and torsions) was inconvenient. Furthermore, recent studies²⁰ demonstrated the feasibility of calculating reaction paths in systems with several tens to thousands of atoms.

The representation of such reaction paths in terms of a large number of ICs is no longer advantageous with respect to the use of the considerably simpler Cartesian representation. A method that can employ directly the numerical (Cartesian) form of the SDP and avoids the transformation to internal coordinates is therefore computationally attractive. In this paper such a computational technique is described and applied. The computational example is of a conformational transition in valine dipeptide for which the SDP is a function of at least three ICs (the torsions ϕ , Ψ , χ_1).

Another feature of the present algorithm is that the calculations required are of molecular dynamics subject to linear constraints. In contrast, expressing SDP in terms of ICs requires molecular dynamics simulation subject to nonlinear constraints. We demonstrated in a number of studies^{20(b)-20(d)} that minimizations subject to linear constraints are considerably faster and simpler to implement than minimization subject to nonlinear constraints. It is therefore expected that molecular dynamics simulation will benefit from the simplicity of linear constraints as well. For example, I shall show that the expressions for the Lagrange multipliers are analytical. Hence, no numerical iterative procedure is required in order to find them. As a result the computations are very similar to common molecular dynamics trajectories without constraints.

This manuscript is organized as follows: In Sec. II the computational method is described and the example is presented in Sec. III. Final remarks are in Sec. IV.

II. METHOD

In this section detailed derivation of the computational technique is given. First, the problem to be solved is described more precisely.

A. The problem

The reaction coordinate is denoted by q and the rest of the coordinates which are orthogonal to q are denoted by Q .

q and $\mathbf{Q}$ form a complete set. The atomic level Hamiltonian is denoted by H and the potential of mean force is $W(q)$.

In free energy perturbation the difference in the potential of mean force at two fixed values of q , i.e. q_0 and $q_0 + \delta q$ is given by

$$W(q_0 + \delta q) - W(q_0) = \frac{-1}{\beta} \log [\langle \exp \{ -\beta [H(\mathbf{Q}, \mathbf{P}_{\mathbf{Q}}; q_0 + \delta q) - H(\mathbf{Q}, \mathbf{P}_{\mathbf{Q}}; q_0)] \} \rangle_{\mathbf{Q}, \mathbf{P}_{\mathbf{Q}}}], \quad (1)$$

where β is the inverse temperature ($\beta = 1/K_B T$, T is the absolute temperature, and K_B is the Boltzmann constant). $\mathbf{P}_{\mathbf{Q}}$ is the momentum conjugate to $\mathbf{Q}$. The average $\langle \cdots \rangle_{\mathbf{Q}, \mathbf{P}_{\mathbf{Q}}}$ is an average over the canonical ensemble at a fixed value of q , i.e., q_0 . The averaging can be performed by replacing the integration over momenta and coordinates by a time series and sampling the points from the molecular dynamics trajectory subject to the constraint

$$q(\mathbf{R}) - q_0 = 0. \quad (2)$$

The set of Cartesian coordinates $\mathbf{R}$ is used to represent q . It is not obvious what is the best coordinate set $\mathbf{Q}$ which should be used to calculate the potential of mean force. For example, in rate calculations the coordinate set which maximizes the flux is a reasonable choice. Finding these coordinates is not a simple task for a multidimensional system considered here and I leave this problem for future generations. Instead, the simplest choice which is the most commonly employed¹⁻¹⁷ is adopted here. That is, $\mathbf{Q}$ is set to the cartesian subspace perpendicular to q at q_0 . This choice of $\mathbf{Q}$ is exact for quadratic expansion of the potential around q_0 . Let the cartesian representation of the point q_0 and the step δq be $\mathbf{R}_q$ and $\mathbf{R}_\delta$ respectively. The slope of $q(\mathbf{R})$ (in the Cartesian representation) is denoted by $\mathbf{e}_q$ (Fig. 1)

$$\mathbf{e}_q = \frac{\nabla q(\mathbf{R})}{|\nabla q(\mathbf{R})|}.$$

The subspace of $\mathbf{Q}$ which is the multidimensional plane perpendicular to q is defined by the following scalar product:

$$\mathbf{e}_q \cdot (\mathbf{R} - \mathbf{R}_q) = 0, \quad (3)$$

where $\mathbf{R}$ is the current coordinate set. Note that Eq. (3) defines a linear constraint. The constraint depends only on the first power of the coordinates. This will make life considerably more pleasurable as explained later.

The presentation of q in terms of Cartesian coordinates has the disadvantage that rigid body motions may affect the value of q . It is necessary therefore to eliminate these motions. This is done by adding six constraints (linear again!) on rigid body motions.^{2(b)-2(c)} Three of the constraints are on the translations of the center of mass (4a) and three on rigid body rotations (4b):

$$\sum_{j=1}^N m^j (\mathbf{r}^j - \mathbf{r}_0^j) = 0, \quad (4a)$$

$$\sum_{j=1}^N m^j \mathbf{r}_0^j \times (\mathbf{r}^j - \mathbf{r}_0^j) = 0, \quad (4b)$$

where j is the atom index, m^j is the mass of atom j , $\mathbf{r}^j$ is the Cartesian position vector of the j th atom and $\mathbf{r}_0^j$ is the coordi-

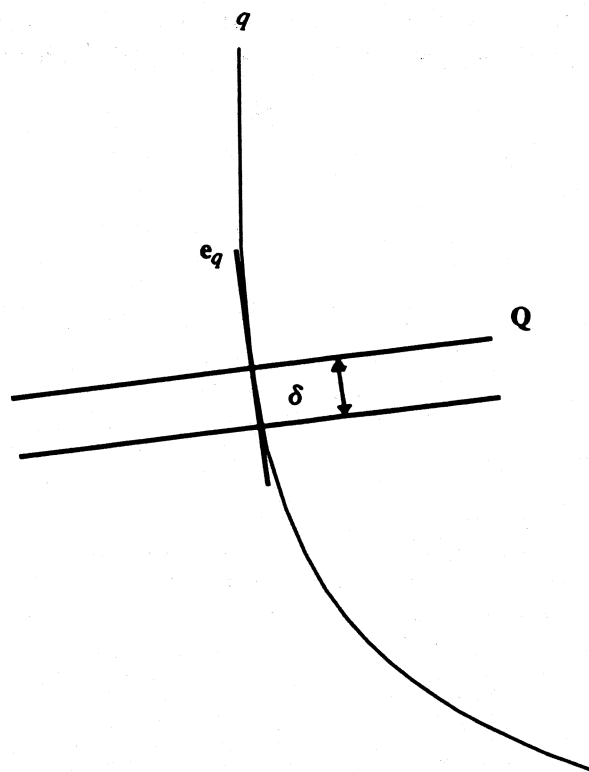


FIG. 1. A schematic representation of a reaction coordinate q , the coordinate perpendicular to it $\mathbf{Q}$, the slope of the path $\mathbf{e}_q$ and the step δ .

nate vector of the j th atom in the reference frame. The reference frame can be (for example) the initial configuration of the trajectory. $\times$ denotes a vector product. N is the number of atoms which "contribute" to q , i.e. on whose positions q is dependent.

The set of the seven linear constraints in Eqs. (3) and (4) which are denoted by $\Delta_{l=1,7}$ can be written in a more compact form:

$$\Delta_l = \sum_{k=1}^{3N} a_l^k X^k = c_l, \quad (5)$$

where the a_l^k and the c_l are constants. k is an index of a cartesian degree of freedom X^k and it is running from one to three times the number of atoms which q depends on. Thus the trajectory which provides the ensemble average for Eq. (1) obeys the following differential equation:

$$\begin{aligned} m^k \frac{\partial^2 X^k}{\partial t^2} &= -\frac{\partial U}{\partial X^k} - \sum_{l=1}^7 \lambda_l \frac{\partial \Delta_l}{\partial X^k} \\ &= -\frac{\partial U}{\partial X^k} - \sum_{l=1}^7 \lambda_l a_l^k, \end{aligned} \quad (6)$$

subject to the constraints of Eq. (5) ($\lambda_{l=1,7}$ are the Lagrange multipliers and U is the microscopic interaction potential). Equation (6) samples configurations with a constant energy (microcanonical ensemble). The canonical averaging [formulated in Eq. (1)] is obtained by velocity scaling. Haile and Gupta proved²¹ that if the trajectory is ergodic then the configurations sampled using the velocity scaling procedure represent the canonical ensemble.

It is important to emphasize that these constraints may

influence the computations in two ways. First, the results of the calculations at a fixed center of mass position or at a fixed rigid body orientation may depend on the choice of the rigid body coordinates (e.g. the results for an impurity in a crystal will depend on the position of the impurity in the unit cell). In these cases it is necessary to repeat the average for all possible values of the rigid body variables. This is not the case in the example in which a *solvated* peptide is considered. Solutions are on the average translationally and rotationally invariant. Second there is a kinetic energy effect. Using classical mechanics and the canonical ensemble it is possible to integrate over all the momenta in Eq. (1) analytically. However, in our case the constraint in rotations complicates the integration over the momenta. The tensor of inertia may be different at q_0 and at $q_0 + \delta q$ and therefore the rotational energy (which we ignore by fixing the body orientation) may be different too. The required correction will be discussed in detail later in a separate subsection.

The next step is to introduce a new set of equivalent and orthogonal constraints. Consider a linear transformation $B_{l,l'}$ to a new set of constraints

$$\sigma_l = \sum_{l'} B_{l,l'} \Delta_{l'} \quad (7a)$$

Obviously

$$\sigma_l = \sum_k A_l^k X^k = C_l, \quad (7b)$$

where A_l^k and C_l are constants too. The gradients of the new constraints which span the same space are required to be orthonormal. That is,

$$\nabla \sigma_l \cdot \nabla \sigma_{l'} = \sum_k A_l^k A_{l'}^k / m^k = \delta_{l,l'}, \quad (8)$$

where $\sigma_{l,l'}$ is the Kronecker delta function. If the velocities (in addition to the coordinates) are constrained, Eq. (7b) can still be used. The only difference for the velocities is that all the C_l are zeros. Note that the scalar product is defined with mass weighting (the reason for that will become clearer later) and that the gradients are constants. This is an advantage of the linear constraints. In contrast to nonlinear constraints the gradients of the linear constraints are calculated only once at the beginning of the simulation [Eq. (6)]. Furthermore, the orthonormalization [Eq. (8)] should be performed only once before the trajectory is initiated and used ever after. For the nonlinear case the orthonormalization should be performed each integration step either by proper matrix manipulation or by an approximate iterative procedure.^{14,22,23}

The problem of how to efficiently run molecular dynamics trajectories with constraints is going back to the important paper of Ryckaert, Ciccotti, and Berendsen.²² There are many extensions on this work, some of them quite recent.^{14,23} The present study is based on the same foundation. In the previous investigations the focus was on the more complex nonlinear constraints. The present problem is simpler (linear constraints) and therefore explicit expressions can be obtained. In the next paragraph a similar procedure to that discussed by Ryckaert, Ciccotti, and Berendsen²² and by Tobias and Brooks¹⁴ for general constraints is

followed to derive an explicit expression for the current problem.

It should be emphasized that the nonlinear constraints that are currently described in the literature and employed in molecular dynamics programs are designed to constrain bonds^{22,23} and other internal coordinates (e.g. torsions) which describe the relative positions of only a few atoms.¹⁴ Since the constraints are coupled by a small number of atoms, the coupling matrix is sparse and the iteration usually converges after a few steps. For SDPs which involve the cooperative motions of a large number of atoms the iterative procedure (SHAKE²²) may converge very slowly. In contrast to that for linear constraints the solution is explicit.

B. The integrator

The solution for the Lagrange multipliers which appear in Eq. (6) can be obtained in a general way. However it is more constructive to consider the specific integration algorithm which will be used in the computations later. That is, the velocity form of the Verlet algorithm.²⁴ Let X_i^k and V_i^k be the value of the k th coordinate and the k th velocity at the i th integration step. The next step ($X_{i+1}^k; V_{i+1}^k$) is given by

$$X_{i+1}^k = X_i^k + V_i^k \Delta t - \frac{\Delta t^2}{2m^k} \left(\frac{\partial U}{\partial X_i^k} + \sum_{l=1}^7 \lambda_{l,i+1} \frac{\partial \sigma_l}{\partial X_i^k} \right), \quad (9a)$$

$$V_{i+1}^k = V_i^k - \frac{\Delta t}{2m^k} \left(\frac{\partial U}{\partial X_i^k} + \frac{\partial U}{\partial X_{i+1}^k} + \sum_{l=1}^7 \lambda_{l,i} \frac{\partial \sigma_l}{\partial X_i^k} + \lambda_{l,i+1} \frac{\partial \sigma_l}{\partial X_{i+1}^k} \right). \quad (9b)$$

It is useful to define an unconstrained move for the coordinate and the velocity $X_{i+1}^{k,u}$ and $V_{i+1}^{k,u}$ which is the step taken in the absence of the constraints:

$$X_{i+1}^{k,u} = X_i^k + V_i^k \Delta t - \frac{\Delta t^2}{2m^k} \frac{\partial U}{\partial X_i^k}, \quad (10a)$$

$$V_{i+1}^{k,u} = V_i^k - \frac{\Delta t}{2m^k} \left(\frac{\partial U}{\partial X_i^k} + \frac{\partial U}{\partial X_{i+1}^k} \right). \quad (10b)$$

Hence Eq. (9) can be rewritten as

$$X_{i+1}^k = X_{i+1}^{k,u} - \frac{\Delta t^2}{2m^k} \sum_{l=1}^7 \lambda_{l,i+1} A_l^k, \quad (11a)$$

$$V_{i+1}^k = V_{i+1}^{k,u} - \frac{\Delta t}{2m^k} \sum_{l=1}^7 (\lambda_{l,i} + \lambda_{l,i+1}) A_l^k, \quad (11b)$$

where the constant value of the gradient of the constraint A_l^k was substituted in the equation. To determine the Lagrange multipliers ($\lambda_{l=1,7}$) it is required that the constraints [Eq. (7b)] will be satisfied for the new set of coordinates and velocities:

$$\sum_{k=1}^{3N} A_l^k X_{i+1}^k = C_l, \quad (12a)$$

$$\sum_{k=1}^{3N} A_l^k V_{i+1}^k = 0, \quad (12b)$$

where N is the number of atoms contributing to q and the summation is running only over these atoms. Substituting

Eq. (11a) in (12a) one obtains

$$\sum_{k=1}^{3N} A_l^k X_{i+1}^{k,u} - \frac{\Delta t^2}{2} \sum_{l'=1}^7 \lambda_{l',i+1} \left(\sum_{k=1}^{3N} A_l^k A_{l'}^k / m_k \right) - C_l = 0. \quad (13)$$

The orthogonality relation defined in Eq. (8) yields

$$\sum_{k=1}^{3N} A_l^k X_{i+1}^{k,u} - \frac{\Delta t^2}{2} \sum_{l'=1}^7 \lambda_{l',i+1} \delta_{l,l'} - C_l = 0, \quad (14)$$

$$\frac{\Delta t^2}{2} \lambda_{l,i+1} = \sum_{k=1}^{3N} A_l^k X_{i+1}^{k,u} - C_l.$$

Similar procedure can be followed for the velocity, to obtain

$$\frac{\Delta t}{2} (\lambda_{l,i} + \lambda_{l,i+1}) = \sum_{k=1}^{3N} A_l^k V_{i+1}^{k,u}. \quad (15)$$

Hence it is possible to write the equations of motion with the constraints explicitly:

$$X_{i+1}^k = X_{i+1}^{k,u} - \frac{1}{m_k} \sum_{l=1}^7 \left(\sum_{k'=1}^{3N} A_l^{k'} X_{i+1}^{k',u} - C_l \right) A_l^k, \quad (16a)$$

$$V_{i+1}^k = V_{i+1}^{k,u} - \frac{1}{m_k} \sum_{l=1}^7 \left(\sum_{k'=1}^{3N} A_l^{k'} V_{i+1}^{k',u} \right) A_l^k. \quad (16b)$$

Eq. (16) is the main result of the present discussion. The computational scheme can be divided into two steps. First, an unconstrained move is generated. That is a move of a straightforward molecular dynamics without constraints [Eq. (10)]. Second, Eq. (16) is employed to correct the "free" move to satisfy the constraints in an explicit way.

Note that the unconstrained move could be generated by any molecular dynamics integrator. Equation (16) will provide the appropriate correction for the constrained dynamics.

This protocol is employed to collect data for the equilibrium average of Eq. (1). The trajectory samples points for the integral at a fixed q , $q = q_0$. A Cartesian coordinate set $\{\mathbf{R}_Q, \mathbf{R}_q\}$ is generated at each step of the constrained molecular dynamics trajectory. The coordinates are then shifted by the constant (Cartesian) vector $\mathbf{R}_\delta$ to the next multidimensional plane along the reaction coordinate $\{\mathbf{R}_Q, \mathbf{R}_q + \mathbf{R}_\delta\}$ [see also Fig. (1)]. The energy difference and the integrand in Eq. (1) are then calculated and stored for statistical averaging. Hence, all the calculations are pursued in the cartesian representation, which is computationally convenient.

Before proceeding to the computational details I still need to describe the correction for rigid body motions.

C. A correction for rigid body motions

An alternative way of writing Eq. (1) is

$$W(q_0 + \delta q) - W(q_0) = -\frac{1}{\beta} \log \left[\frac{\int \exp[-\beta H(\mathbf{Q}, \mathbf{P}_Q; q_0 + \delta q)] d\mathbf{P}_Q d\mathbf{Q}}{\int \exp[-\beta H(\mathbf{Q}, \mathbf{P}_Q; q_0)] d\mathbf{P}_Q d\mathbf{Q}} \right]. \quad (17)$$

It is useful to separate the integration over the momenta and coordinates to three subsets: (i) the "rigid body" variables of the subset of atoms to be constrained ($\mathbf{Q}_r, \mathbf{P}_r$), (ii) the internal coordinates and momenta of the constrained atoms ($\mathbf{Q}_c, \mathbf{P}_c$) and (iii) the rest of the variables ($\mathbf{Q}', \mathbf{P}'$). It is possible to make the following coordinate transformation which sets the origin to $\mathbf{Q}_r$ (i.e., the zero is at the center of mass of the constrained atoms and the directions of the axes follow the principal axes of the tensor of inertia):

$$(\mathbf{Q}'; \mathbf{Q}_c; \mathbf{Q}_r) \rightarrow (\mathbf{Q}' - \mathbf{Q}_r; \mathbf{Q}_c - \mathbf{Q}_r; \mathbf{Q}_r). \quad (18)$$

The following definitions are useful to keep the equations short:

$$\Gamma \equiv (\mathbf{Q}' - \mathbf{Q}_r; \mathbf{Q}_c - \mathbf{Q}_r), \quad \mathbf{Q}'_c \equiv \mathbf{Q}_c - \mathbf{Q}_r.$$

The Hamiltonian is now written as

$$H(\mathbf{Q}', \mathbf{Q}_c, \mathbf{Q}_r, \mathbf{P}_r; q_0) = K_c(\mathbf{P}_c) + K'(\mathbf{P}') + K_r(\mathbf{Q}'_c, \mathbf{Q}_r, \mathbf{P}_r; q_0) + U(\Gamma; q_0), \quad (19)$$

where K_i is the kinetic energy associated with the i th momentum subspace. Choosing $\mathbf{P}'$ and $\mathbf{P}_c$ to be cartesian the

integration over these variables can be carried out immediately. In fact since the results are independent of q the identical terms in the nominator and the denominator of Eq. (17) cancel each other.

The integration over Γ is performed via the time series generated by a trajectory [Eqs. (10) and (16)] and is not discussed in this subsection. Here the integrations over $\mathbf{Q}_r$ and $\mathbf{P}_r$ are completed for a fixed Γ . Note that formally freezing $\mathbf{Q}_r$ should not change the results. The integration is carried out over all space and the choice of the origin and the orientation of the coordinate system (which is one way of viewing $\mathbf{Q}_r$) should not matter. In practice however since the simulations are performed for very finite systems boundary effects may be important and different origins may yield different results. Strictly speaking even simulations of solutions with periodic boundary conditions²⁴ may suffer from problems. In this work, however, it is assumed that the effects of the long range symmetry introduced by the periodicity are not large and that the solvent box itself is large enough to be self-averaging (in the sense that to a good approximation the solvent within the box is translationally and rotationally invariant on the average).

The above discussion is summarized by

$$\exp[-\beta(W(q_0 + \delta q) - W(q_0))] = \frac{\int d\Gamma F(\mathbf{Q}'_c; q_0 + \delta q_0) \exp[-\beta U(\Gamma; q_0 + \delta)]}{\int d\Gamma F(\mathbf{Q}'_c; q_0) \exp[-\beta U(\Gamma; q_0)]} \quad (20)$$

where $F(\mathbf{Q}'_c; q_0)$ is the result of integration over the rigid body variables. It is the partition function for the hypothetical case in which the atoms subject to the seven constraints $\sigma_{I=1,7}$ are placed *in vacuo* and only rigid body motions for the collection of these atoms are possible. Hence none of the distances between the constrained atoms is allowed to vary:

$$F(\mathbf{Q}'_c; q_0) = \int d\mathbf{Q}, d\mathbf{P}, \exp[-\beta K_r(\mathbf{Q}'_c, \mathbf{Q}, \mathbf{P}; q_0)]. \quad (21)$$

Consider the ratio between the two F 's in Eq. (20). The translational component of the partition functions of the two

F 's is the same (the translation partition function does not depend on q). The ratio of the translational partition functions is therefore one. Only the rotational partition function is left. The expression for the asymmetric top (the most common case) is given below:²⁵

$$F(\mathbf{Q}'_c; q_0) = D [I_A(\mathbf{Q}'_c; q_0) I_B(\mathbf{Q}'_c; q_0) I_C(\mathbf{Q}'_c; q_0)]^{1/2}. \quad (22)$$

The D does not depend on the coordinates and it is the same for both F 's. Therefore in the ratio of the two F 's the two D 's cancel each other. $I_{i=A-C}$ are the principal moments of inertia which depend of course on $\mathbf{Q}'_c$ and also on q_0 . The atomic level potential is now redefined,

$$U'(\Gamma, q_0) = U(\Gamma, q_0) + \frac{1}{\beta} \log \{ D [I_A(\mathbf{Q}'_c; q_0) \times I_B(\mathbf{Q}'_c; q_0) I_C(\mathbf{Q}'_c; q_0)]^{1/2} \}, \quad (23)$$

and employed in the formula for thermal averaging [Eq. (1)],

$$\exp\left(-\frac{1}{\beta} [W(q_0 + \delta q) - W(q_0)]\right) = \langle \exp \left\{ -\beta \left[U(\Gamma, q_0 + \delta q) - U(\Gamma, q_0) + \frac{1}{2\beta} \log \left(\frac{I_A(\mathbf{Q}'_c; q_0 + \delta q) I_B(\mathbf{Q}'_c; q_0 + \delta q) I_C(\mathbf{Q}'_c; q_0 + \delta q)}{I_A(\mathbf{Q}'_c; q_0) I_B(\mathbf{Q}'_c; q_0) I_C(\mathbf{Q}'_c; q_0)} \right) \right] \right\} \rangle_{\Gamma}, \quad (24)$$

where the averaging left (over Γ) is done by the constrained simulation.

Hence the rigid body rotations introduced a term corresponding to a thermally averaged centrifugal force. The importance of the rotational correction will depend of course on the system considered. A numerical estimate can be given as follows: For a spherical object with a radius r , a constant mass density ρ , and total mass M the principal moment of inertia is $\frac{2}{5}Mr^2$. For simplicity δq is assumed parallel to r . To the first order in δq the ratio $I(r + \delta q)/I(r)$ is $1 + (2\delta q/r)$. Expanding the log in Eq. (24) gives a centrifugal potential of the order $\delta q/\beta r$. For reasonable molecular values the correction is between $K_B T/10$ to $K_B T/100$ and is quite small. Problems may arise however when there is a significant change between the moments of inertia of the reactants and the products. In this case the small changes may accumulate to a significant error. For example, if one considers alternative conformations of long polymers, the transition from extended chain to compact conformations may change the principal moments of inertia by at least a factor of two. In this case the correction is of order of $K_B T$ and can be important if the energy barrier is not large. To be on the safe size the above correction was included in the simulation program.

To summarize this section: A simple molecular dynamics protocol was developed [Eqs. (10) and (16)] which enables the calculations of the potential of mean force along an arbitrary reaction coordinate. The protocol employs *linear* constraints [Eqs. (3) and (4)] which speed up the calculations and simplify them considerably. The correction for rigid body rotations was considered too and was estimated to be

small in most cases. Nevertheless, since cases with significant corrections are not unheard of, a correction was designed. The exact simulation includes an effective potential with centrifugal terms [Eq. (23)]. In the next section an application of the above protocol will be presented.

III. A COMPUTATIONAL EXAMPLE

The dynamics of conformational transitions in peptides is of interest as a possible model for the kinetics of protein folding. The potential of mean force can be used in estimates of rate based on the transition state theory.²⁶ It is calculated here for a conformational transition in valine dipeptide (Fig. 2) along the steepest descent path (SDP).

The problem of how to calculate the reaction coordinate for the solvated system is not trivial. If the solvent is included in the SDP calculations, the resulting reaction path is probably not unique. A large number of SDPs corresponding to different solvent configurations are likely to be found. To maintain the convenience of a single reaction coordinate a number of past studies^{1-6,8-10,13-15} adopted the single reaction coordinate as obtained from the gas phase. If the interaction with the solvent molecules is not too strong, the approximation should be sound. Unfortunately water molecules are expected to have a significant effect on conformational transitions in hydrogen bonding systems. This makes the choice of SDP for the solvated dipeptide system even more difficult. However, one of the CHARMM force fields (the force field with the distance dependent dielectric "constant," ϵ_{rdie})²⁷ was parametrized to model the dielectric screening of water. The gas phase SDP for valine dipeptide

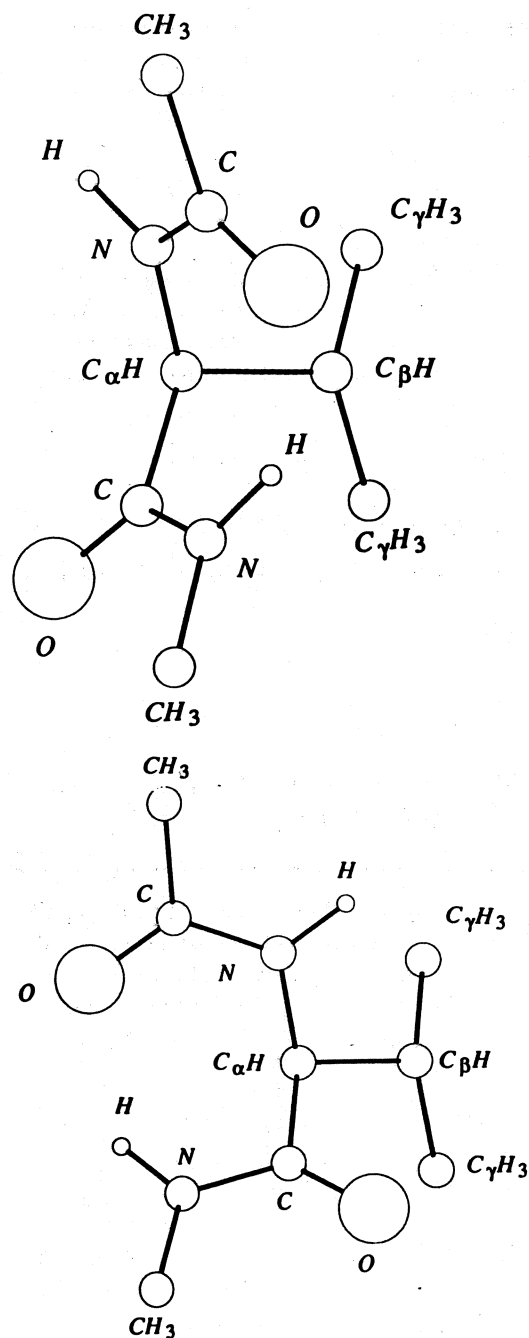


FIG. 2. Sticks and balls model for valine dipeptide: (a) The configuration at ($\phi = 61.4^\circ$, $\Psi = -64.0^\circ$), (b) the configuration at ($\phi = -77.3^\circ$, $\Psi = 77.9^\circ$).

using the rdie force field is therefore expected to be a reasonable choice for the reaction coordinate in the solvated system.

The free energy calculation for the solvated system in which the water molecules were included explicitly were pursued using (of course) Coulomb electrostatic with no phenomenological terms. This treatment of electrostatic is called in the text "cdie" (constant dielectric) to be contrasted with rdie.

It is interesting to note that if the dielectric screening is not taken into account (i.e., the SDP is calculated in the gas

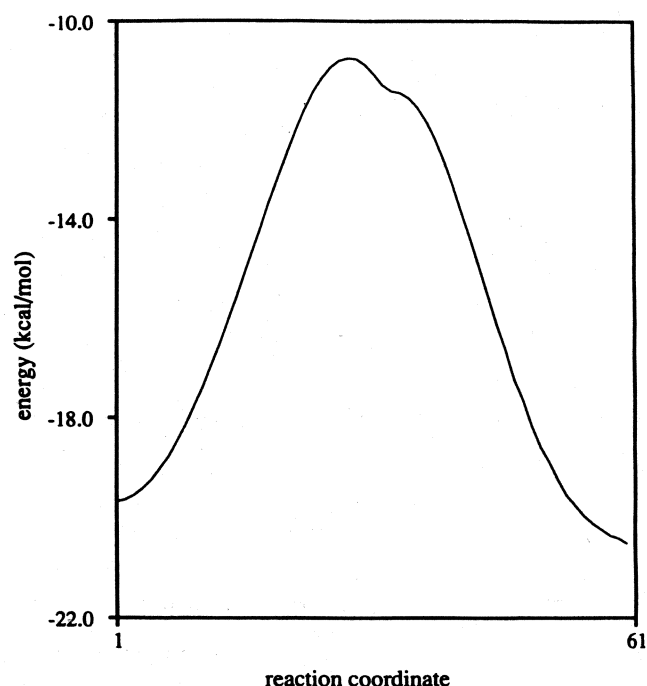


FIG. 3. The minimum energy path for conformational changes in valine dipeptide *in vacuo* using the cdie option of the CHARMM force field (Ref. 27) as a function of the structure number. The energy is in kcal/mol. See text for more details.

phase with Coulomb law) the SDP that is obtained is considerably different. There is no minimum at the helix configuration and the barrier height is significantly larger than for the solvated system (Figs. 3 and 6). The SDP using the rdie has a minimum at the helix and the barriers are much lower. In the calculations, the SDP from the rdie was employed. The resulting free energy profile has even deeper minimum at the helix configuration which suggests that the rdie path is closer to the reaction path in solution. To prove that, however, a second calculation along the cdie SDP will be required. This calculation will be performed in a future study.

Structural changes in these biologically important molecules (peptides) are usually interpreted as changes in the ϕ , Ψ dihedral angles. In Fig. 4 an adiabatic energy map (with rdie option) for these variables is shown (i.e., the energy function was minimized for a fixed value of the above dihedral pair). Also shown on the adiabatic map is a projection of the calculated steepest descent path (SDP) on the ϕ , Ψ variables. The algorithm of Ulitsky and Elber^{20(c)} was employed to calculate the SDP. This algorithm was added as a separate module in a modified version of the molecular dynamics program CHARMM.²⁷ The SDP was calculated between the two minima which are shown on the map at (61.4° , -64.0°) and (-77.3° , 77.9°) (see also a plot of the "reactant" and the "product" in Fig. 2). 60 grid points were used in the *gas phase* calculation of the SDP.

It should be emphasized that the SDP includes changes in **all** internal coordinates. Maximal changes in bond lengths and bond angles along the SDP were of order of 0.02 Å and 2° . The amide planes were distorted by up to 4° . Larger changes along the SDP ($\approx 10^\circ$) can be found in the dihedral angle χ_1 (Fig. 5). Thus even for this relatively simple sys-

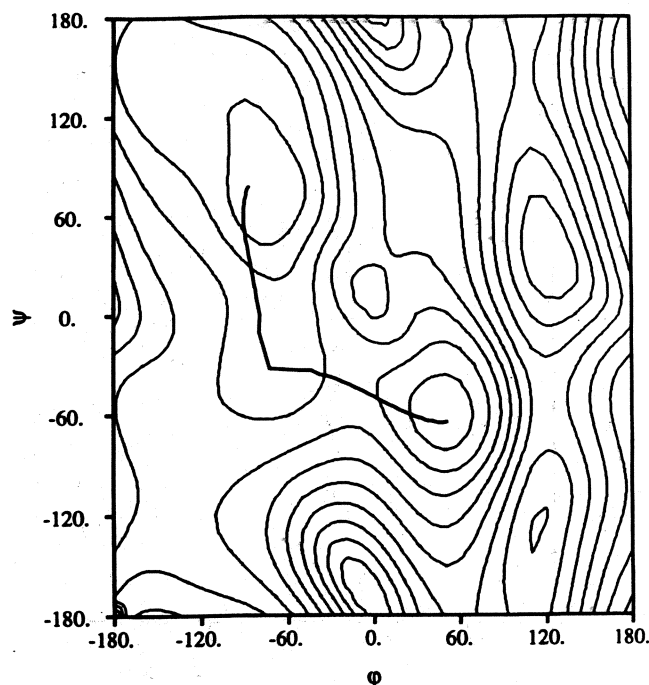


FIG. 4. Adiabatic energy gap for the ϕ , Ψ torsions of valine dipeptide using the rdie option of CHARMM (Ref. 27). Also shown on the map the steepest descent path calculated using the Ulitsky and Elber algorithm [Ref. 20(c)].

tem, the reaction coordinate deviates significantly from what can be assigned to a single internal coordinate. Note that the small variation in bond length, bond angles, and χ_1 is important for the determination of barrier heights. Keeping these variables fixed increases the barrier by approximately ten kcal/mol.

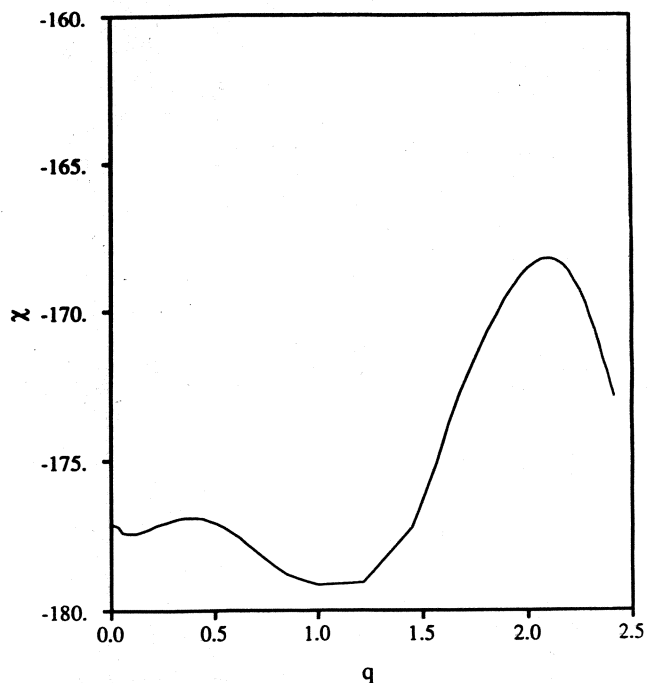


FIG. 5. The projection of the steepest descent path on the χ_1 torsion (in degrees) of valine dipeptide. The changes in χ_1 are given as a function of the length of the path in angstroms starting from the configuration of the reactants (Fig. 4).

The accuracy of the path calculations was measured by the average energy gradient in the direction perpendicular to the path and it was of order of 0.01 kcal/mol Å. The root mean square (rms) distance between the structures along the SDP was of order of 0.02 Å and the resulting energy profile is shown in Fig. 6. Note the shallow minimum at the helical profile is shown in Fig. 6. Note the shallow minimum at the helical conformation $\sim (-70^\circ, -30^\circ)$. The helix conformation is not stabilized by a hydrogen bond in dipeptides and only van der Waals attractions and polar group interactions contribute to the stability. As is quite clear from the projection of the SDP on the ϕ , Ψ map, the SDP changes its direction sharply at the minimum of the helix. At this configuration the path is continuous but not differentiable, corresponding to a switch in low energy eigenvectors of the second derivative matrix.

To pursue the free energy calculations the valine dipeptide molecule was placed in a 19.5 Å cubic box with 207 water molecules. Periodic boundary conditions were employed and the water model was TIP3P of Jorgensen.²⁸ The constant dielectric option of the CHARMM force field with a cutoff distance of 8.0 Å was used. The scaling factor for 1–4 interactions was one and the extended atom model (all the CH_n groups modeled by spheres) was employed. The water structure was equilibrated around the peptide using a molecular dynamics simulation at 300 K for 50 ps. During the equilibration the structure of the dipeptide was held fixed.

The constrained dynamics routines which use the velocity form of the Verlet integrator [Eqs. (10) and (16)] were implemented as a separate head in a modified version of the CHARMM program. In the simulation the linear con-

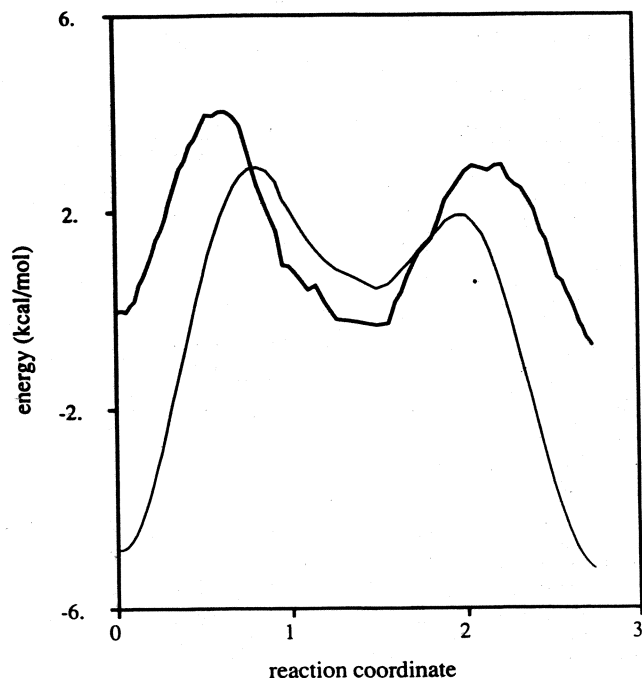


FIG. 6. The minimum energy path (kcal/mol) as calculated with the distance dependent dielectric as a function of the length of the steepest descent path (in angstrom). Also shown (thick line) the potential of mean force (kcal/mol). See text for more details.

straints for the free energy calculations were imposed only on the peptide atoms (14 in all). The time step was 1 femtosecond and the temperature was maintained at 300 K by velocity scaling every 0.2 ps. The velocities were further reassigned at random from Boltzmann distribution of 300 K every 1 ps (the maximum fluctuations in the temperature during the simulation were of order of 10 degrees). The slope of the path [$\mathbf{e}_q$ —Eq. (3)] was estimated numerically by a finite difference. That is,

$$\mathbf{e}_q = \frac{\nabla_{\mathbf{R}} q(\mathbf{R})}{|\nabla_{\mathbf{R}} q(\mathbf{R})|} \cong \frac{\mathbf{R}_{q+\delta q} - \mathbf{R}_{q-\delta q}}{|\mathbf{R}_{q+\delta q} - \mathbf{R}_{q-\delta q}|}. \quad (25)$$

Note that the potential of mean force for dipeptides was calculated in the past by fixing both soft torsions (ϕ, ψ) and averaging over the rest of the degrees of freedom.^{5,13,14} The calculations were repeated for numerous values of this pair of coordinates. The disadvantage of calculating the complete map for soft variables is that the number of free energy differences which one needs to calculate is growing exponentially with dimension. Then if there are D dimensions and each dimension is represented by M points, M^D free energy calculations are required. In the present algorithm for reaction path calculation only one dimension is considered which reduces the number of free energy differences required to approximately M . (It is assumed of course that the identified SDP is indeed the most important pathway.) Note however that the computer time required for calculating a *single free energy difference* is generally longer in the proposed approach. Hence the expected saving in computational effort is smaller than the factor M^{D-1} . The reason for the slower convergence *per point* is as follows: It is expected that some portions of the potential will be quite flat (corresponding to soft degrees of freedom other than q), and averages over the flat portion of the potential requires relatively longer simulations. In Fig. 7(a) the range in the (ϕ, ψ) map which was sampled in the direction perpendicular to q during the calculation of a single free energy difference is shown. This range is of ≈ 200 degrees and corresponds to large structural fluctuations. The possibility of such large structural changes makes the calculations more time consuming since the convergence rate is slower. Nevertheless the computational effort per point is growing considerably slower than the *number* of points resulting still in considerable savings in computer time. Furthermore not all averages are evaluated at flat portions of the potential energy. In Fig. 7(b) another sample is shown with spread of order of $\approx 50^\circ$. The difficulties in free energy calculations in flexible systems were pointed out by Straatsma and McCammon¹⁵ in their study of 1,2 dimethoxy ethane.

Another possible application of the present code is in the calculation of the relative free energy of different conformers. In such calculation one needs to find a thermodynamic path which connects the two conformers. For maximum efficiency the thermodynamic path should have the smallest possible variations in energy along the path. This will make a large number of points "acceptable" since the exponent in Eq. (1) is small:

$$\beta |H(\mathbf{Q}, q_0) - H(\mathbf{Q}, q_0 + \delta q)| \approx 0.$$

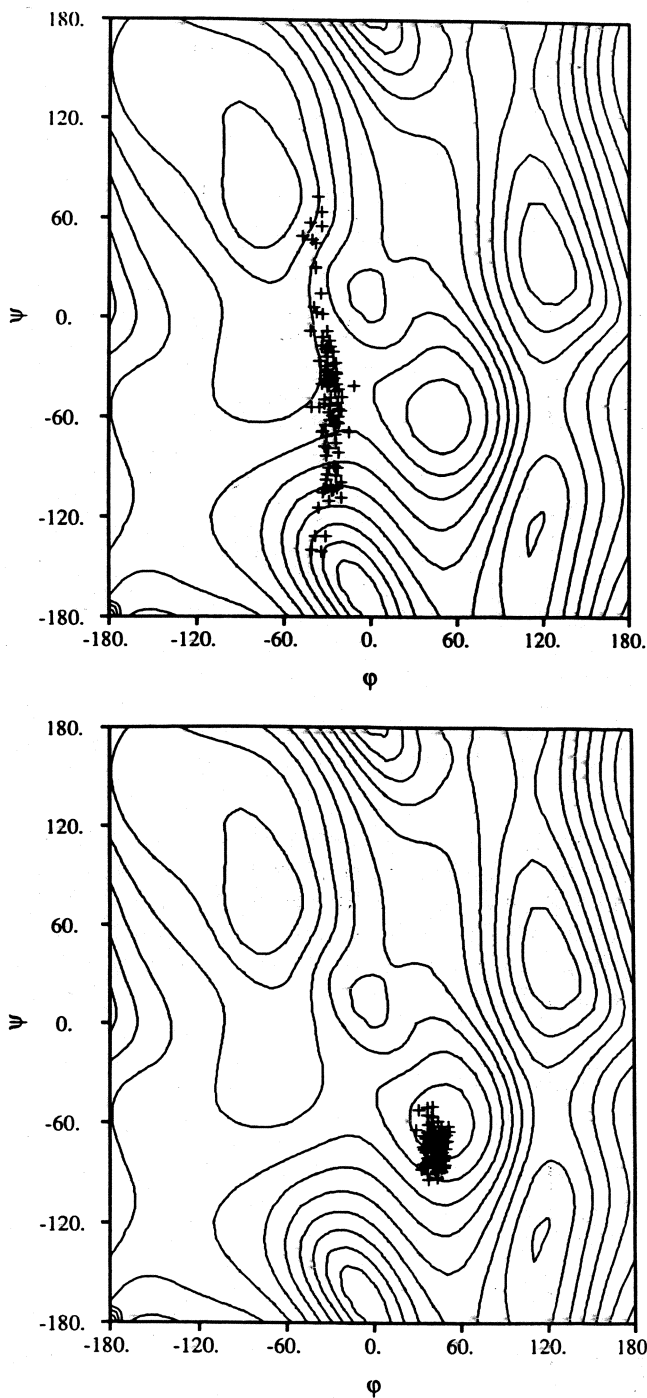


FIG. 7. The range in the (ϕ, ψ) map as sampled in a single trajectory used in the calculation of free energy difference along the reaction coordinate: (a) A trajectory with large structural fluctuations ($\approx 200^\circ$); (b) a trajectory with relatively small fluctuations ($\approx 50^\circ$).

It will enable the use of a larger δq value for a step. The steepest descent path fulfills this requirement.

In the present calculation 5 ps for equilibration and additional 25 ps for data collection were employed for each free energy difference. In the neighborhood of the helix configuration the convergence was slower and 45 ps trajectories were employed for data collection. All the calculations were done on a TITAN minisupercomputer. The required CPU

time was approximately one month. The standard deviations of the free energy differences were of order of one percent, the errors however were significantly larger. They were estimated by calculating each point in the forward and the backward directions for all points and also by doubling the time for the simulations of 10 points equally spaced along the reaction coordinate, and checking for the deviation from the shorter simulation. Most errors were of the order of 10 to 20 percent but two (coming to and getting out from the helix) were about 30 percent. I invested a considerable effort trying to reduce these errors with limited success. The direction of the path changes by 90° at the helical configuration which makes the convergence difficult to obtain.

In Fig. 6 the calculated free energy profile is shown as a function of the reaction coordinate. The reaction coordinate is given in angstroms and it is expressed as the length of the path starting from the reactants [the minimum at $(61.4^\circ, -64.0^\circ)$]. Also shown is the profile of the energy along the minimum energy path (SDP). Two differences between the energy and the free energy profiles are evident. The first difference is the significant decrease in the barrier height for the conformational transition from 7.7 kcal/mol in the minimum energy path to 4.0 ± 0.5 kcal/mol in the potential of mean force. The second clear difference is the considerable stabilization of the helix. The reduction of the barrier is associated with the possibility of exchanging internal hydrogen bonds with hydrogen bonds to water. This option is not available in the gas phase. The relative energy of the two conformers changes only slightly between the "gas phase" and the solution phase calculations which supports the rdie model which was designed to take into account implicitly the solvent. However it seems that the rdie parametrization overestimates the barrier heights.

One may expect that the helix configuration will be more stable when water molecules are included explicitly, since the N-H and the carbonyl groups can form hydrogen bonds (HB) to water molecules. It is not surprising therefore that the calculation with constant dielectric *in vacuo* yields no minimum at the helical configuration (Fig. 3). To examine the solvation effects the number of hydrogen bonds was estimated. The hydrogen bonds were defined by a cutoff distance of 3 Å between the heavy atoms and an angle of less than 90° . The number of hydrogen bonds between a water molecule and an N-H group (averaged over the simulation at the helix configuration) was 0.98 and the number of hydrogen bonds to the carbonyl group was 2.4. For the initial configuration ($\phi = 61.4^\circ$, $\Psi = -64.0^\circ$) the average number of hydrogen bonds between water molecules and the carbonyl groups was 1.91 and the number of hydrogen bonds between water molecules and an N-H groups was 0.69. The N-H and the carbonyl group also form (in the initial configuration) an internal hydrogen bonds (0.44). The total sum of hydrogen bonds in the helix configuration is therefore $2 \cdot (2.44 + 0.98) = 6.84$, to be contrasted with the values for the initial configuration $2 \cdot (1.91 + 0.69) + 0.44 = 5.66$. This explains why the helix conformation is stabilized in water. Mezei *et al.*⁵ found a similar effect for alanine dipeptide—the stabilization of the helix conformation by hydrogen bonding primarily to the carbonyl groups.

IV. CONCLUSIONS

A simple and efficient algorithm to calculate free energy profiles along a discretized reaction path was introduced. The technique samples configurations in a restricted space (at a fixed value of the reaction coordinate) using molecular dynamics subject to linear constraints. The solution for the Lagrange multipliers for linear constraints is analytic and the computations involved are therefore similar to dynamics without constraints. The proposed algorithm is quite general since any reaction coordinate can be easily discretized using cartesian coordinates.

A computational example is given in which the conformational transitions in valine dipeptide are considered. It is demonstrated that water molecules reduce significantly the barrier height for the transitions and that the helix conformation is stabilized significantly in the presence of water molecules.

ACKNOWLEDGMENTS

This research was supported by NIH Grant No. GM40698. R.E. is a Camille and Henry Dreyfus New Faculty. The calculations were performed on a TITAN mini-supercomputer purchased by NIH shared instrumentation Grant No. RR04884 to the University of Illinois at Chicago, Department of Chemistry.

- ¹R. O. Rosenberg, R. Mikkilineni, and B. J. Berne, *J. Am. Chem. Soc.* **104**, 7647 (1982).
- ²W. L. Jorgensen, *J. Chem. Phys.* **77**, 5757 (1982).
- ³S. H. Northrup, M. R. Pear, C. Y. Lee, J. A. McCammon, and M. Karplus, *Proc. Natl. Acad. Sci. US* **79**, 4035 (1982).
- ⁴J. Chandrasekhar, S. F. Smith, and W. L. Jorgensen, *J. Am. Chem. Soc.* **107**, 154 (1985); J. Chandrasekhar and W. L. Jorgensen, *ibid.* **107**, 2974 (1985).
- ⁵M. Mezei, P. K. Mehrotra, and D. L. Beveridge, *J. Am. Chem. Soc.* **107**, 2239 (1985).
- ⁶C. D. Bell and S. C. Harvey, *J. Phys. Chem.* **90**, 6595 (1986).
- ⁷D. A. Case and J. A. McCammon, *Proc. N. Y. Acad. Sci.* **482**, 222 (1986).
- ⁸W. L. Jorgensen, J. K. Buckner, S. E. Huston, and P. J. Rossky, *J. Am. Chem. Soc.* **109**, 1891 (1987).
- ⁹W. L. Jorgensen and J. K. Buckner, *J. Phys. Chem.* **91**, 6083 (1987).
- ¹⁰P. A. Bash, M. J. Field, and M. Karplus, *J. Am. Chem. Soc.* **109**, 5531 (1987).
- ¹¹B. Owenson, R. D. MacElory, and A. Pohorille, *J. Am. Chem. Soc.* **110**, 6992 (1988).
- ¹²J. Kottalam and D. A. Case, *J. Am. Chem. Soc.* **110**, 7690 (1988).
- ¹³A. G. Anderson and J. Hermans, *Proteins* **3**, 362 (1988).
- ¹⁴(a) D. J. Tobias and C. L. Brooks III, *Chem. Phys. Lett.* **142**, 472 (1987); (b) *J. Chem. Phys.* **89**, 5115 (1988).
- ¹⁵T. P. Straatsma and J. A. McCammon, *J. Chem. Phys.* **90**, 3300 (1989).
- ¹⁶J. van Eerden, W. J. Briels, S. Harkema, and D. Feil, *Chem. Phys. Lett.* **164**, 370 (1989).
- ¹⁷L. X. Dang and P. A. Kollman, *J. Am. Chem. Soc.* **112**, 503 (1990).
- ¹⁸D. Rojewska and R. Elber, *Proteins* **7**, 265 (1990).
- ¹⁹U. Burkert and N. L. Allinger, *Molecular Mechanics* (ACS, Washington, D.C., 1982).
- ²⁰(a) R. Elber and M. Karplus, *Chem. Phys. Lett.* **139**, 375 (1987); (b) R. Czerminski and R. Elber, *Proc. Natl. Acad. Sci. US* **86**, 6963 (1989); (c) A. Ulitsky and R. Elber, *J. Chem. Phys.* **92**, 1510 (1990); (d) R. Czerminski and R. Elber, *J. Chem. Phys.* **92**, 5580 (1990); (e) R. Czerminski and R. Elber, *Int. J. Quant. Chem.* (in press); (f) C. Choi and R. Elber (in progress); W. Nowak and R. Elber (in progress).
- ²¹J. M. Haile and S. Gupta, *J. Chem. Phys.* **79**, 3067 (1983).
- ²²J. P. Ryckaert, G. Cicciotti, and H. J. C. Berendsen, *J. Comput. Phys.* **23**, 327 (1977).

- ²³S. G. Lambrakos, J. P. Boris, E. S. Oran, I. Chandrasekhar, and M. Magumo, *J. Comput. Phys.* **85**, 473 (1989).
- ²⁴D. W. Heermann, *Computer Simulation Methods in Theoretical Physics* (Springer, Berlin, 1986).
- ²⁵D. A. McQuarrie *Statistical Mechanics* (Univ. Science, 1974).
- ²⁶J. D. Doll, *J. Chem. Phys.* **73**, 2760 (1980); **74**, 1074 (1981).
- ²⁷B. R. Brooks, R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus, *J. Comput. Chem.* **4**, 187 (1983).
- ²⁸W. L. Jorgensen, J. Chandrasekhar, and J. D. Madura, *J. Chem. Phys.* **79**, 926 (1983).