Tryptophan-47 rotational isomerization in variant-3 scorpion neurotoxin

A combination thermodynamic perturbation and umbrella sampling study

Christopher Haydock, Joseph C. Sharp, and Franklyn G. Prendergast
Department of Biochemistry and Molecular Biology, Mayo Foundation, Rochester, Minnesota 55905 USA

ABSTRACT A combination thermodynamic perturbation and umbrella sampling study predicts two free energy wells for the rotational isomerization of the variant-3 scorpion neurotoxin tryptophan-47 indole side chain. One well has the indole side chain in the crystallographic orientation; the other has the indole rotated ~220 degrees to form a new conformation with a relative free energy of 3 ± 2 kcal/mol. The activation barrier is 8.5 kcal/mol from the crystallographic well, from which transition state theory predicts a rate of escape of 2×10^5 s^{-1}. Correlations in the displacements of side chains neighboring tryptophan-47 and the isomerization reaction coordinate last up to 20 ps. Favorable conditions of experimental verification are discussed.

INTRODUCTION

The fluorescence intensity decay of single tryptophan proteins is frequently multiexponential (1). This lifetime heterogeneity has been ascribed to the existence of multiple rotational isomers of the indole side chain (2). Rotational isomerization can be modeled over a broad range of time scales by activated molecular dynamics (3, 4). Ring flips have been shown to make an important contribution to the order parameter for fluorescence depolarization probes of tyrosine motions in proteins (5). In general, the interpretation of fluorescence intensity and polarization spectra of proteins requires a knowledge of the dynamics of the fluorophore and surrounding protein on the nanosecond time scale (3, 6, 7). Here we focus on the rotational isomerization of the indole side chain of tryptophan-47 in variant-3 scorpion neurotoxin. In favorable cases activated barrier crossing events can also be detected with fluorescence spectroscopy. For example, the fluorescence lifetime is sensitive to the interactions of the tryptophan fluorophore with juxtaposing amino acid residues and solvent. If the interactions of tryptophan and neighboring residues differ significantly for each rotational isomer, we expect to observe corresponding exponential decay components in the fluorescence intensity.

In a many-bodied system as complex as a protein it is not practical to survey every bump and valley in the energy throughout the entire phase space. In this paper we model the relative free energy of rotational isomers and the height of the barrier separating these as a function of a simple one-dimensional reaction coordinate. The question of what makes a good reaction coordinate has been explored in considerable detail (8–11). Roughly speaking the reaction coordinate should correspond to the arc length along a typical barrier crossing trajectory. Its magnitude should increase steadily as this trajectory emerges from the reactant region of configuration space, ascends the activation barrier, passes through the transition state bottleneck, descends the barrier and finally is lost in the product region of configuration space. With the introduction of a one-dimensional reaction coordinate we need not deal explicitly with most of the details of the complex potential energy surface; instead we compute the free energy, which amounts to averaging over all the other degrees of freedom. The ideal reaction coordinate will faithfully describe the average behavior of the otherwise extraordinarily complex energy surface and accurately give the free energy of the rotational isomers of interest and interconversion rate between these.

Previous studies of rotational isomerization provide important initial guidance in the selection of a reaction coordinate. The first detailed theoretical study of an activated process in a globular protein examined $\chi^1 \times \chi^2$ torsion space rotations of tyrosines in the interior of bovine pancreatic trypsin inhibitor (12). In the case of the free amino acid tryptophan the χ^2 dihedral angle is evidently a suitable rotational isomerization reaction coordinate (13). Indeed for tryptophan and simple derivatives χ^1 and χ^2 are probably suitable reaction coordinates for rotations about the C^α—C^β and C^β—C^γ bonds, respectively, and these rotational motions are fairly independent of each other. However, when a tryptophan residue is introduced into a protein the free energies for rotations about χ^1 and χ^2 are coupled and simple separation of the tryptophan $\chi^1 \times \chi^2$ torsion space into independent χ^1 and χ^2 rotamers and reaction coordinates is probably no longer possible, even for a surface tryptophan with a highly exposed side chain, not to mention a case such as

0006-3495/90/06/1269/11 $2.00

neurotoxin tryptophan-47, which is partially buried by the surrounding protein matrix. This coupling of χ^1 and χ^2 is suggested by the observation that in proteins the rotations about χ^1 and χ^2 are cross-correlated. For example, in a 60-ps vacuum simulation of lysozyme the six tryptophans exhibited anticorrelations of these rotations ranging from -0.2 to -0.8 (14). Below we shall explore several alternative $\chi^1 \times \chi^2$ torsion space reaction coordinates for the isomerization of tryptophan-47 in variant-3 scorpion neurotoxin.

THEORY

A rotational isomerization model requires the free energy as a function of the reaction coordinate. In this section we develop an expression for the Helmholtz free energy as a function of a reaction coordinate by combining thermodynamic perturbation and umbrella sampling (3, 4, 15). We have selected the Helmholtz free energy because it is simple to sample the canonical ensemble when a stochastic boundary is present. We assume that the reaction coordinate is a configuration coordinate, in which case the Helmholtz free energy along this coordinate is a simple configuration space integral over the remaining coordinates and is typically referred to as the potential of mean force

$$A(q) = -k_B T \ln P(q), \tag{1}$$

where q is the reaction coordinate and $P(q)$ is the reaction coordinate probability density. In the canonical ensemble with variables N,V,T the probability density

$$P(q_1) = Z_N^{-1} \int \cdots \int e^{-E_N(q_1,\ldots,q_{3N})/k_BT}\, dq_2 \ldots dq_{3N}$$
$$Z_N = \int \cdots \int e^{-E_N(q_1,\ldots,q_{3N})/k_BT}\, dq_1 \ldots dq_{3N}, \tag{2}$$

where the configuration coordinates have been labeled so that the first coordinate q_1 is the reaction coordinate. Here $E_N(q_1, \ldots, q_{3N})$ is the N particle configurational energy, expressed as a function of the reaction coordinate and $3N - 1$ remaining configuration coordinates $q_2, \ldots, q_{3N}$, and Z_N is the configurational partition function. Eq. 2, together with Eq. 1, specifies the Helmholtz free energy at a given reaction coordinate in terms of the canonical ensemble probability of observing that value of the reaction coordinate. The configurational partition function in Eq. 2 normalizes the probability integral over the entire reaction coordinate to unity. Strictly speaking, both integrals in Eq. 2 and all integrals over $dq_1 \ldots dq_{3N}$ and $dq_2 \ldots dq_{3N}$ leading up to Eq. 4 should contain the factor $|\det \mathbf{T}(q_1, \ldots, q_{3N})|^{1/2}$, where $\mathbf{T}(q_1, \ldots, q_{3N})$ is the metric tensor for the configuration space (16, 17) and the square root of the metric determinant comes from the integration over the canonical momenta (18). Because this has no effect on the derivation of Eq. 4 or on our final result, it is omitted.

Eq. 2 for the probability density is inadequate in, for instance, the case of a macromolecular system with a reaction coordinate which specifies two stable conformations and a reaction pathway over a barrier separating these conformations. If the free energy height of the barrier is large compared to k_BT, then the corresponding conformations are exceedingly rare in the canonical ensemble. In application only a reasonable number of samples from the ensemble can be generated. Thus it is common practice to modify the configurational energy with so called umbrella potentials which bias the canonical ensemble towards otherwise unlikely conformations. We require an expression analogous to Eq. 2 which gives the probability density for the reaction coordinate in the absence of the perturbing umbrella potential in terms of integrals over the perturbed configurational partition functions. The essence of our method is to compute a total probability correction relative to the unperturbed ensemble as a product of corrections across a series of overlapping umbrellas. This amounts to determining all the constants needed to match the free energies of adjacent umbrella samples. Each matching constant turns out to be given by the perturbation method difference in the perturbed energies of adjacent umbrellas.

To derive an expression for the unperturbed probability density along a reaction coordinate in terms of perturbed densities we first derive an auxiliary expression relating the probability densities of two special systems. These systems, designated a and b, both contain N particles and have configuration energies which are identical except for a difference which is a function of the reaction coordinate,

$$\Delta_{ab}(q_1) = E_b(q_1, \ldots, q_{3N}) - E_a(q_1, \ldots, q_{3N}). \tag{3}$$

Eq. 2 applied to system a gives the probability density along q_1,

$$P_a(q_1) = \frac{\int \cdots \int e^{-E_a/k_BT}\, dq_2 \ldots dq_{3N}}{\int \cdots \int e^{-E_a/k_BT}\, dq_1 \ldots dq_{3N}}.$$

Replacing the subscript a with b gives the analogous probability density for system b. We require an expression for P_a in terms of P_b. This can be accomplished by multiplying by unity in a form such that an exponential of $-\Delta_{ab}/k_BT$ is inserted into the integral in the numerator and the normalization for P_b appears in the denominator:

$$P_a(q_1) = e^{\Delta_{ab}(q_1)/k_BT} \frac{\int \cdots \int e^{-[E_a+\Delta_{ab}(q_1)]/k_BT}\, dq_2 \ldots dq_{3N}}{\int \cdots \int e^{-E_b/k_BT}\, dq_1 \ldots dq_{3N}} \cdot \frac{\int \cdots \int e^{-E_b/k_BT}\, dq_1 \ldots dq_{3N}}{\int \cdots \int e^{-E_a/k_BT}\, dq_1 \ldots dq_{3N}}.$$

Recalling Eq. 3, the second factor is P_b and the numerator of the third may be integrated over $dq_2 \ldots dq_{3N}$. Thus we have

$$P_a(q_1) = e^{\Delta_{ab}(q_1)/k_BT} P_b(q_1) \int e^{-\Delta_{ab}(q_1)/k_BT} P_a(q_1) dq_1$$
$$= e^{\Delta_{ab}(q_1)/k_BT} P_b(q_1) \langle e^{-\Delta_{ab}(q_1)/k_BT} \rangle_a, \quad (4)$$

where the angle brackets with subscript a indicate the average over the system a ensemble. After interchanging the labels a and b, Eq. 4 may be immediately rearranged to give a second expression for P_a in terms of P_b:

$$P_a(q_1) = e^{\Delta_{ab}(q_1)/k_BT} P_b(q_1) \langle e^{\Delta_{ab}(q_1)/k_BT} \rangle_b^{-1}. \quad (5)$$

It is shown below that the accuracy of the ensemble average estimate in Eqs. 4 and 5 is inversely proportional to the magnitude of the derivative of Δ_{ab} with respect to q_1.

Now suppose that a sequence of umbrellas is specified, each of which constrains a reaction coordinate q to some region of interest. In each of the constraint regions the Helmholtz free energy is given by Eq. 1 and the probability density along the reaction coordinate corrected for umbrella bias

$$P_0(q) = P_j(q) e^{U_j(q)/k_BT} c_0 \prod_{i=1}^{j-1} c_i, \quad j = 1, 2, 3, \ldots, M, \quad (6)$$

where $P_j(q)$ is the probability density from the jth umbrella sample, $U_j(q)$ is the jth umbrella constraint potential and there are a total of M umbrellas. The perturbation constants

$$c_0 = \langle e^{-U_1(q)/k_BT} \rangle_0 = \langle e^{U_1(q)/k_BT} \rangle_1^{-1},$$
$$c_i = \langle e^{-[U_{i+1}(q) - U_i(q)]/k_BT} \rangle_i = \langle e^{[U_{i+1}(q) - U_i(q)]/k_BT} \rangle_{i+1}^{-1},$$
$$i = 1, 2, 3, \ldots, M - 1, \quad (7)$$

where the subscript on the angle brackets gives the number of the umbrella perturbing the ensemble average, except for the subscript 0 on the angle brackets, which indicates that the ensemble average is taken over the unperturbed system. To prove Eq. 6 by induction, let E_0 be the configurational energy in the absence of any perturbing umbrellas. Setting $E_a = E_0$ and $E_b = E_0 + U_1(q)$ in Eq. 4 and 5 proves the case $j = 1$. Then setting $E_a = E_0 + U_j(q)$ and $E_b = E_0 + U_{j+1}(q)$ and assuming Eq. 6 is true for j finishes the proof.

The computation of free energy along a reaction coordinate based on Eq. 6 involves the selection of umbrella potentials and sampling of the canonical ensemble for each of these perturbations. The accuracy of the result depends critically on the degree of umbrella overlap. Eq. 7 gives each perturbation constant as the expectation of the exponential of the difference between neighboring umbrella potentials. For example, if the umbrella potentials are harmonic in the reaction coordinate, then this difference is linear in the reaction coordinate with a proportionality constant which increases with the separation of the two umbrella minima. Thus, the sampling error incurred in computing the perturbation constants increases dramatically if the umbrellas are spaced too widely. As with the other perturbation constants c_i, the accuracy of c_0 depends upon the choice of umbrellas. Because c_0 serves only as an overall normalization constant for $P_0(q)$, a number of alternatives are available to calculate c_0. Eq. 7 is accurate only when the first umbrella overlaps the unperturbed reaction coordinate minimum. Furthermore the second expression for c_0 in Eq. 7 is only accurate when the width of the first umbrella and the unperturbed distribution are about the same. If instead of the first umbrella, umbrella R coincides with the unperturbed minimum, then the normalization constant

$$c_0 = c_R \prod_{i=1}^{R-1} c_i^{-1}, \quad c_R = \langle e^{-U_R(q)/k_BT} \rangle_0. \quad (8)$$

Of course, if a sample of the unperturbed ensemble is not available, then c_0 may be determined by normalizing the composite $P(q)$. In the present work the normalization is not necessary, instead we choose to set $c_R = 1$ in Eq. 8. This has the effect of setting the free energy approximately to zero near the unperturbed minimum in the reaction coordinate.

METHODS

The molecular dynamics simulations, which were required both for the testing of reaction coordinates and for the combination thermodynamic perturbation and umbrella sampling studies, were computed with version 20 of the program CHARMM (19). Our simulations treated all nonpolar hydrogens as part of extended heavy atoms. Topology and parameter files corresponding to this approximation were obtained from version 19 of CHARMM. The TIP3P water model in these files is essentially identical to Jorgensen's (20). The protein structure was modified by removing the two tryptophan-47 bond angles C^{γ}—$C^{\delta 2}$—$C^{\epsilon 3}$ and $N^{\epsilon 1}$—$C^{\epsilon 2}$—$C^{\zeta 2}$. Without this modification the version 19 parameter file generates a distorted butterfly mode of tryptophan. All bonds involving a hydrogen were constrained with the SHAKE algorithm (21). The simulation time step was 2 fs and all nonbonded interactions were cutoff between 7.5 and 8.0 Å. The complete protein-solvent system was treated according to the stochastic boundary method (22). The specifications and a brief outline of the construction procedures for the boundary follow. First the energy of the crystallographic neurotoxin structure (23) plus polar hydrogens was minimized in 1,000 iterations of the Powell method (24, 25) to adjust van der Waals contacts and optimize bond lengths. The reaction and buffer regions were centered on the indole side chain and the outer radii were set to 10 and 12 Å, respectively. Amino acid residues beyond the 12-Å cutoff were deleted. The reaction region included the side chains of residues partially within the 10-Å cutoff. Each backbone atom was assigned to the reaction or buffer regions strictly based upon whether its distance from the origin was less than or greater than 10 Å. The energy minimized structure also provided the reference positions for the buffer region harmonic con-

straints. The constraint force constants were derived from the crystallographic temperature coefficients averaged over all residues in groups defined by remoteness from the backbone. A bulk TIP3P water sphere, which had previously been equilibrated with the stochastic boundary, was then overlaid; every water molecule whose oxygen was within 2.8 Å of an existing heavy atom was deleted. To insure complete protein solvation this overlay procedure was repeated four times, each time with the bulk overlay water sphere randomly rotated to a new orientation. This was followed by 8.0 ps of solvent molecular dynamics in the presence of the stochastic boundary and fixed protein field and then four more bulk water overlays. The final reaction zone included 335 protein atoms and 115 water molecules. In subsequent simulations all atoms were free to move about, subject to the empirical potential and boundary forces specified above as well as a stochastic force on all buffer region nonhydrogen atoms. For the selected protein atoms the stochastic friction constant was set to 125 ps^{-1}. Both this friction constant and the harmonic constraint constants were scaled by a switching function that took the value 0.5 beyond 12 Å and 0.0 within 10 Å. The stochastic friction constant for TIP3P water oxygens in the buffer region was set to 62 ps^{-1}.

We investigated the suitability of χ^1, χ^2, $\chi^1 + \chi^2$, and $\chi^2 - \chi^1$ as reaction coordinates for tryptophan-47 side chain rotational isomerization. The torsion angles χ^1 and χ^2 are equal to χ^1_{47} and $\chi^{2,2}_{47}$, respectively, where χ^1_{47} is along atoms N_{47}—C^{α}_{47}—C^{β}_{47}—C^{γ}_{47} and $\chi^{2,2}_{47}$ is along atoms C^{α}_{47}—C^{β}_{47}—C^{γ}_{47}—$C^{\delta 2}_{47}$ (26). These reaction coordinates were constrained by 2π periodic harmonic constraint functions

$$U_j(\xi) = k(\xi - \xi_j)^2, \; -\pi < \xi - \xi_j \leq \pi, \tag{9}$$

where ξ is any one of the candidate reaction coordinates, k = 200 kcal · mol^{-1} · rad^{-2} is the energy constant and ξ_j is the position of the energy minimum of the jth constraint potential. Because this definition lacks the factor of ½, k is referred to as the energy constant rather than the force constant. The constraint potentials preserve the 2π periodicity of the $\chi^1 \times \chi^2$ torsion space, i.e., for every χ^1 and χ^2, $U(\chi^1, \chi^2) = U(\chi^1 + 2\pi, \chi^2) = U(\chi^1, \chi^2 + 2\pi)$. We conducted a series of $2M + 1$ 1-ps simulations with constraint function indices $j = -M, \ldots, -1, 0, 1, \ldots, M$. In each series the first simulation has an index of $j = 0$ and corresponds approximately to the crystallographic structure and subsequent simulations have constraint function energy minima at positions given by $\xi_j = \xi_0 + j\Delta\xi$, where $\Delta\xi$ = 10 degrees. The $2M + 1$ resulting molecular configurations were merged to form a coordinate trajectory, which was studied with the HYDRA molecular graphics program, and the corresponding $2M + 1$ values for χ^1 and χ^2 were plotted to form a $\chi^1 \times \chi^2$ torsion space trajectory. The final M values, based on continuous monitoring of these trajectories, were 70, 70, 105, and 140 for the χ^1, χ^2, $\chi^2 - \chi^1$, and $\chi^1 + \chi^2$ reaction coordinates, respectively.

The reaction coordinate for the combination thermodynamic perturbation and umbrella sampling studies was $\xi = \chi^1 + \chi^2$. As described above, during the reaction coordinate test studies ξ was scanned for 1,400 degrees in the negative direction starting from the crystallographic value of $\xi = -90$ degrees. We used the final equilibration coordinate sets from the portion of the negative ξ scan between -160 and -510 as the initial coordinates for 36 umbrella samples. Since ξ is 2π periodic, the umbrella sampling results can be presented with ξ in the range from $-150 \equiv -510 \pmod{360}$ to $200 \equiv -160 \pmod{360}$ degrees. The umbrella sampling constraint potentials were 2π periodic harmonic functions that were identical to the scan constraint potentials given by Eq. 9, except for a softer constraint constant of k = 20 kcal · mol^{-1} · rad^{-2}. After 11 ps of equilibration, a 12-ps umbrella sample was collected for each umbrella position. The $\chi^1 \times \chi^2$ torsion space coordinates were saved at each time step and the full reaction zone coordinates every 0.04 ps. The torsion space time series were converted into reaction coordinate time series and analyzed in accordance with Eq. 6, which was derived in the Theory section.

All molecular simulations were at a temperature of 300 degrees Kelvin and were computed on a Cray X-MP/48 computer system or DEC VAX 3600.

RESULTS AND DISCUSSION

By several criteria $\xi = \chi^1 + \chi^2$ emerged as the most suitable reaction coordinate for studying rotational isomerization of the scorpion neurotoxin tryptophan-47 side chain, primarily because it generates relatively uniform rotation of the indole side chain. This is visible as fairly smooth indole rotation as a function of coordinate set number in the display of the scan coordinate trajectory. Furthermore, the indole rotation angle is a 2π periodic function of the $\chi^1 + \chi^2$ reaction coordinate. This is visible in the coordinate trajectory and in the $\chi^1 \times \chi^2$ torsion space trajectory. These uniformity and periodicity characteristics imply that a plus one radian increment in $\chi^1 + \chi^2$ very nearly corresponds to a one radian counterclockwise rotation of indole. A minor departure from strict $\chi^1 + \chi^2$ periodicity is apparent by comparing the positive and negative portions of the $\chi^1 + \chi^2$ scan. Although each portion is 2π periodic, the positive and negative scans are not quite identical. The average positive scan position is shifted several angstroms toward tyrosine-4, 40 and 42 compared to the average negative scan. The $\chi^1 \times \chi^2$ torsion space trajectory of the positive and negative scans also differ. Although both start at the crystallographic position, only the negative scan returns to this position after 2π radians. The positive scan generates a periodic loop which only passes within about one radian of the starting position. For this reason, the umbrella sampling is based on the negative scan.

The only other candidate reaction coordinate that even comes close to meeting the minimum requirements is χ^2; however, by both the periodicity and uniformity criteria, χ^2 is of limited usefulness. The main difficulty is that negative scanning grossly fails to generate complete rotations of the indole side chain. Instead the indole rocks back and forth through a total angle of about one radian and with a period of 2π. In contrast the positive χ^2 scan rotates the indole side chain in the counterclockwise direction with a period of 2π along the scan. This suggests that χ^2 might serve as a unidirectional reaction coordinate. However, even the positive χ^2 scan by itself has the problem that the rotational motion is not very uniform. For example, in the 30-degree interval of reaction coordinate approximately given by $-150 \leq \chi^2 \leq -120$ degrees the indole ring rotates through an angle of almost 120 degrees. Finally, scanning $\chi^2 - \chi^1$ and χ^1 in either direction produces indole side chain rocking, waving and slicing motions with periods of 4π and 2π, respectively. Though these motions may be of spectroscopic interest,

the reaction coordinates are not suitable for studying rotation.

The free energy for rotational isomerization of tryptophan-47 is presented in Fig. 1. This is computed by repeated application of Eq. 1 and 6 in each of the 36 umbrellas. Also plotted in Fig. 1 are the cumulative perturbation constants

$$A_j = -k_B T(\ln c_0 + \sum_{i=1}^{j-1} \ln c_i), j = 1, 2, 3, \ldots M, \quad (10)$$

where c_0 and c_i are given by Eq. 8 and 7, respectively, and $M = 36$ is the total number of umbrellas. In comparing the free energies by umbrella with the cumulative perturbation constants it is evident that the latter contain the essential information about the free energy as a function of the ξ reaction coordinate. In both cases the free energy functions reveal that tryptophan-47 can rotate counterclockwise from the crystallographic orientation through an angle of ~220 degrees into a new rotational isomer. Both the crystallographic and new isomers are stable energy minima. The individual umbrella free energies provide important additional information. Umbrellas 22, 25, and 29 are located on the right-hand side of the barrier separating the new isomer from the crystallographic isomer, yet the free energies derived from each of these umbrellas have positive slopes that are incongruent with the negative slope of this side of the barrier. The anomalous slopes of these three umbrellas are responsible for the more ragged appearance of the free energy in the new rotational isomer well as compared with the crystallographic free energy well. This suggests that either the free energy of the new rotational isomer has not completely converged or that the reaction coordinate is missing important degrees of freedom. We emphasize that the individual umbrella free energies are not simply visually aligned in their regions of overlap, rather the required offsets are given by the thermodynamic perturbation constants as specified in Eqs. 6–8.

To assess further the convergence of our results, we extracted a series of 1-ps snapshots from the umbrella simulations. The results of an analysis of three of these shapshots are presented in Fig. 2. Note that a 23-ps simulation was computed in each umbrella and that Fig. 1 is the result of the analysis of the simulation interval from 11.0 to 23.0 ps. Thus the second and third snapshots in Fig. 2 represent the beginning and end of the interval analyzed in Fig. 1. The initial free energy snapshot contains a steep residual energy gradient generated during the initial reaction coordinate scanning in the negative direction. The scan residual greatly increases the barrier

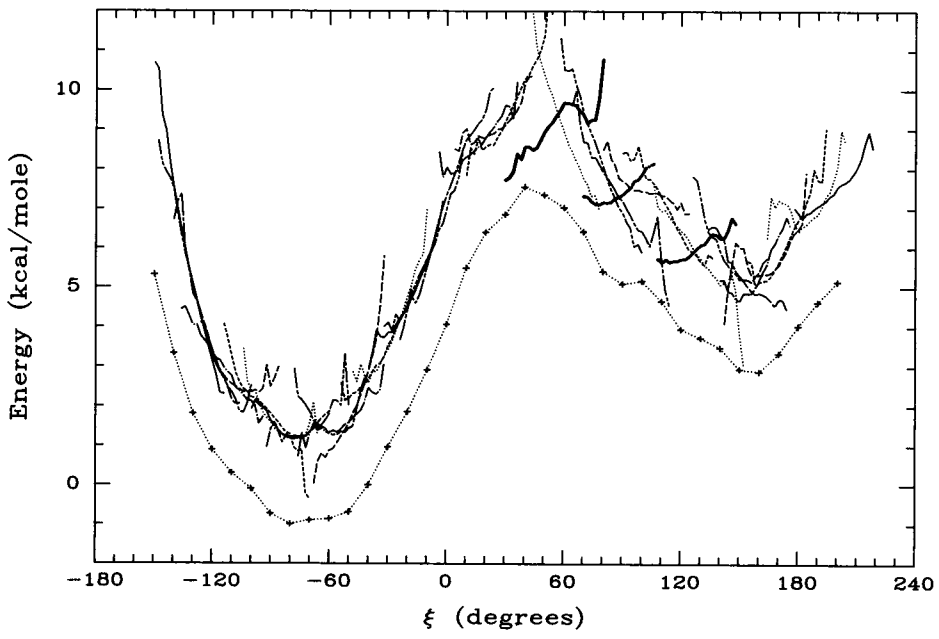

FIGURE 1 Helmholtz free energy by umbrella and cumulative perturbation constants along the ξ reaction coordinate. A separate curve is plotted for the free energy derived from each umbrella. The line styles are alternated over the umbrella numbers as follows: *solid*, 1, 8, 15, 22, 29, 36; *long, short, short*, 2, 9, 16, 23, 30; *long, short*, 3, 10, 17, 24, 31; *long, dot*, 4, 11, 18, 25, 32; *long*, 5, 12, 19, 26, 33; *medium*, 6, 13, 20, 27, 34; *dot*, 7, 14, 21, 28, 35. The lines for umbrellas 22, 25, and 29 are widened for emphasis. The cumulative perturbation constants for the 36 umbrellas are plotted as small crosses with an interconnecting dotted line. The perturbation constant of umbrella 8 is arbitrarily set to −1.0 kcal/mol. The average position and orientation of the indole side chain in the left-hand minimum is very close to that in the crystallographic structure.

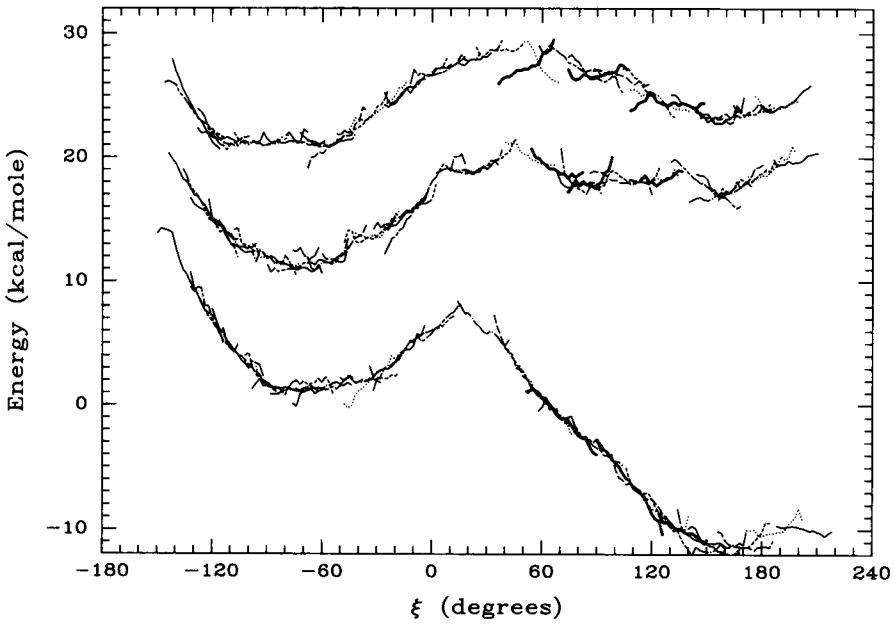

FIGURE 2 Free energies by umbrella for 1-ps snapshots of the 23-ps simulations within each umbrella. The snapshots are taken at 0.0–1.0, 11.0–12.0, and 22.0–23.0 ps. The line styles and line widths are as in Fig. 1. The second and third snapshots are offset by 10.0 and 20.0 kcal/mol, respectively, so that they can be easily distinguished.

to escape from the new isomer and also shows up as positive offset in the position of the crystallographic minimum. The scan residual is not present in the second or third snapshots. In a snapshot taken from 4.0 to 5.0 ps, which is not shown in Fig. 2, the scan residual is still clearly present, but has diminished in amplitude by at least 50%. This suggests that the residual relaxes with a time constant of ~5.0 ps, which corresponds to the time it takes to scan about one radian at the rate of 10.0 degrees per picosecond. Another interesting feature in Fig. 2 is that the potential well of the new isomer deepens by 4.0 kcal/mol during the time interval between the second and third snapshots. The time constant for this process must be comparable to the duration of the individual umbrella simulations or ~20 ps. In view of this long time constant for well deepening we estimate the minimum free energy of the new isomer is 3.0 ± 2.0 kcal/mol relative to the crystallographic. This relative energy is from an analysis of the final 6.0 ps of the umbrella simulations, which is the interval from 17.0 to 23.0 ps. The error estimate is simply the difference in well depths from analysis of the last two 6.0-ps intervals.

Both scan gradient relaxation and potential well deepening are apparent in the snapshot time series. In Fig. 2 these effects are easiest to identify in the overall dependence of the free energy on the reaction coordinate. As discussed in connection with Fig. 1, this overall dependence is contained in the thermodynamic perturbation constants. However, the scan relaxation and well deepening are also apparent in the time development of the individual umbrella free energies. For example, consider the right-hand side of the barrier separating the crystallographic and new rotational isomers and corresponding to the reaction coordinate range ξ = 30–150 degrees. In the initial snapshot the individual slopes are all negative and very steep, confirming the presence of the scan gradient. In the second snapshot there are a random-looking assortment of positive and negative gently sloping free energies. In the third snapshot more systematic negative slopes reappear, again confirming the well deepening process. Comparison of individual umbrella free energies in the last two snapshots and in Fig. 1 reveals the time evolution of the free energy during the 11.0–23.0-ps simulation interval. It is particularly interesting to examine the free energies in umbrellas 22, 25, and 29, because these have anomalous positive slopes in Fig. 1. In both umbrellas 25 and 29 the positive slope is observed to flatten out and presumably in a longer simulation would converge to the negative slope of the barrier's right-hand side. In contrast the umbrella 22 free energy has a negative slope in snapshot 2 and a positive slope in snapshot 3. In this case a transition from the new isomer back into the crystallographic has occurred. Evidently the ξ reaction coordinate cannot hold the indole side chain exactly at the top of the rotational activation barrier. This suggests that additional degrees of freedom must be included in the reaction coordinate before actual barrier crossing trajectories can be efficiently simulated.

It is interesting to examine the correlations of the translational fluctuations of residues surrounding tryptophan-47 with the displacements along the ξ reaction coordinate and how the amplitudes of these correlations diminish with distance from tryptophan-47. The following procedure for measuring these cross-correlations exploits the systematic scanning of the entire reaction coordinate that is inherent in an umbrella sampling simulation. The essential idea is to test for interumbrella fluctuations in the average positions of individual residues. Because the system is approximately in a state of equilibrium in each umbrella, the interumbrella displacement of these residues should be viewed as a zero-time cross-correlation with the average reaction coordinate in that umbrella. However, in light of the 5–20-ps relaxation time constants seen in the snapshot analysis it is likely that these cross-correlations persist for times >10 ps. Perhaps the simplest statistics describing the interumbrella fluctuations are the root-mean-square displacements of the umbrella average structures from the grand average over all umbrellas. Where the umbrella average structures and grand average structure are computed by atom and then the mean-square by atom fluctuations are mass weighted and averaged by residue, giving the interumbrella root-mean-square fluctuations on a per-residue basis. Although these statistics are very useful, two shortcomings are overcome by a slightly more complex analysis that follows. First, the interumbrella fluctuations should be compared to the fluctuations within each umbrella. And second, in addition to the overall average interumbrella differences, it is important to look at differences between groups of neighboring umbrellas. This amounts to asking if smaller displacements of the reaction coordinate are correlated with smaller displacements of neighboring residues. These shortcomings of looking only at the interumbrella fluctuations are overcome by averaging the total fluctuations over groups of umbrellas, where the group size ranges from one umbrella up to all 36 umbrellas and each grouping contains groups of approximately twice the size of the previous. We use groups of size 1, 2, 4, 9, 18, and 36 umbrellas. For each group size we compute by atom the mean-square fluctuations about the average structure in each group. The mean-square fluctuations are then averaged over all groups of a given size and these group averages are mass weighted and averaged over each residue. For a group size of 1 this procedure gives the average intraumbrella fluctuations. A group size of 36 gives the fluctuations about the grand average structure. These total fluctuations thus include both the intraumbrella and interumbrella contributions.

The results of this cross-correlation analysis are presented in Fig. 3. As might be expected the tryptophan-47 fluctuations are the most strongly correlated with rotations of the ξ reaction coordinate. The other residues exhibiting large fluctuations are glutamates 2 and 49, tyrosines 4, 38, 40, 42, and 58, glycine-39, and leucine-51. With the exception of tyrosines 38 and 58 all of these residues are direct neighbors of tryptophan-47. The boundary construction procedures described in the Methods section placed both of these tyrosine's side chains in the reaction region without any harmonic position constraints, even though the benzyl ring centers have average radial positions of 12.9 and 11.2 Å, respectively. In fact the large fluctuations of both these tyrosines are entirely due to these side chains, which more or less stick out into the reservoir region. In evaluating Fig. 3 it is important to compare the fluctuations for the various averaging group sizes. Aside from tryptophan-47 itself, tyrosine-40 and 42 and glutamate-2 show the largest fluctuation increases with each increasing step in the group size. The strong cross-correlations of the displacements of these residues and the reaction coordinate suggest that significant changes in their interactions with tryptophan-47 are also occurring. In spite of smaller cross-correlations, significant changes in tryptophan-47 interactions with tyrosine-4 and glutamate-49 are also likely because their side chains are bulky and charged, respectively, and the indole translates from a position nearer glutamate-2 and 49 to a position nearer tyrosine-4, 40, and 42 as it rotates from the crystallographic to the new isomer. In Fig. 3 the residues constrained in the buffer region consistently show the smallest fluctuations and these are fairly independent of averaging group size. Thus the rotational isomerization of the tryptophan-47 side chain is a spatially localized event, only the displacements of residues that neighbor tryptophan-47 on the antiparallel beta sheet show strong cross-correlations with the reaction coordinate displacements. Because our simulations use the stochastic boundary method we cannot exclude the possibility that nonlocal conformational changes accompany rotational isomerization. The reaction zone radius and thus the number of particles simulated are simply adjusted to achieve adequate convergence in a reasonable number of computing hours and accurately model the localized changes.

As discussed above, ξ is 2π periodic. The free energy $A(\xi)$, which is the average given in Eqs. 1 and 2 over the remaining $3N - 1$ degrees of freedom, must have the same periodicity as the ξ reaction coordinate. For example, in Fig. 1 the free energy at $\xi = -150$ degrees and $\xi = 210 \equiv -150 \pmod{2\pi}$ should be identical. Extrapolating the free energy dependence on ξ at both ends of the computed region suggests that this 2π periodicity is poorly satisfied. The origin of this problem is clear if we examine the $\chi^1 \times \chi^2$ torsion space probability density in Fig. 4. Note that all umbrellas are weighted equally in this plot. Thus this figure shows the overall pattern of torsion space sampling and is not directly related to the

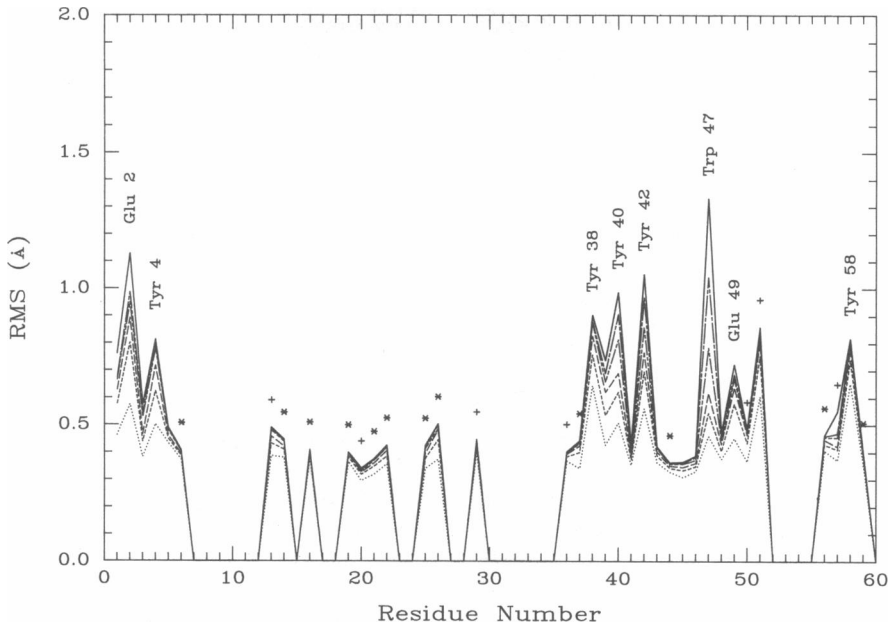

FIGURE 3 Root-mean-square fluctuations per residue averaged over umbrella groupings of various sizes. The line styles correspond to group sizes as follows: *dot*, 1; *medium*, 2; *long*, 4; *long, short, short*, 9; *long, short*, 18; *solid*, 36. This figure is based on the 11.0–23.0-ps simulation intervals of the 36 consecutive umbrellas. Those residues showing zero fluctuations are in the reservoir region, which also includes residues 61–65. The presence of harmonic constraints on the entire residue or backbone only is indicated by an asterisk or plus sign above the peak fluctuations, respectively. These markers were placed if cutoffs of 0.20 and 0.0 kcal · mol^{-1} · Å^{-2} on the backbone and side chain root-mean-square constraint constant were exceeded, respectively. The residues with asterisks have root-mean-square constraint energy constants ranging from 0.91 to 1.38 on the backbone and 0.26 to 1.05 on the side chains and in the case of the plus markers 0.58 to 1.38 on the backbone.

free energy. From the typical values of ξ in each umbrella given in Fig. 1 we can guess how individual umbrellas contribute to the overall probability density plotted in Fig. 4. Indeed, inspection of the individual umbrella densities on torsion space confirms that they do trace out a fairly smooth diagonal path starting with umbrella 1 at $(\chi^1, \chi^2) \approx (170, 60)$ or $\xi = -130$ and ending with umbrella 36 at $(\chi^1, \chi^2) \approx (-70, -90)$ or $\xi = -160 \equiv 200 \pmod{2\pi}$. The questionable periodicity of $A(\xi)$ at -150 degrees shows up in torsion space as a large separation between the regions where umbrella 1 and 36 contribute to the probability density. It is not the separation itself but rather the lack of probability mixing that spoils the periodicity. Because the free energies in umbrella 1 and 36 are about the same according to Fig. 1, the umbrella 36 probability density should exhibit a subsidiary maximum in probability near the maximum of umbrella 1 density at $(\chi^1, \chi^2) \approx (160, 40)$. As long as each umbrella gives the correct probability density over the entire torsion space consistent with the constraint potential on the reaction coordinate ξ, the free energy will show the correct periodicity. The average indole side chain orientation during the simulations in umbrella 1 and 36 are very similar; however, the average position in the two umbrellas differs by a translation of 2 or 3 Å. In fact, the torsion space path connecting umbrella 1 and 36 along a line of constant $\xi \approx -145$ degrees is a translational isomerization, which we describe as a waving motion of the indole side chain. The lack of probability mixing between umbrella 1 and 36 is evidently due to a barrier along the line $\chi^2 \approx 0$, which is high enough to prevent any waving transitions during a simulation of reasonable length. Thus it is necessary to average along lines of constant ξ in $\chi^1 \times \chi^2$ torsion space to get the 2π periodicity of the free energy. We prefer to leave the results as they appear in Fig. 1. Averaging would leave a false impression of a small barrier at $\xi \approx -160$ degrees between the crystallographic and new rotational isomers.

Transition state theory provides a simple estimate of the transition rate between the crystallographic and new rotational isomers. It is clear in Fig. 4 that the choice of $\xi = \chi^1 + \chi^2$ as reaction coordinate does not imply that the reaction path follows a diagonal of constant $\chi^2 - \chi^1$ in $\chi^1 \times \chi^2$ torsion space. Indeed, the vertical orientation of the transition state saddle at $(\chi^1, \chi^2) \equiv (-155, -160)$ suggests that near the transition state the isomerization reaction path follows a line of constant $\chi^1 \approx -155$ and that fluctuations in the χ^1 dihedral angle can be ignored.

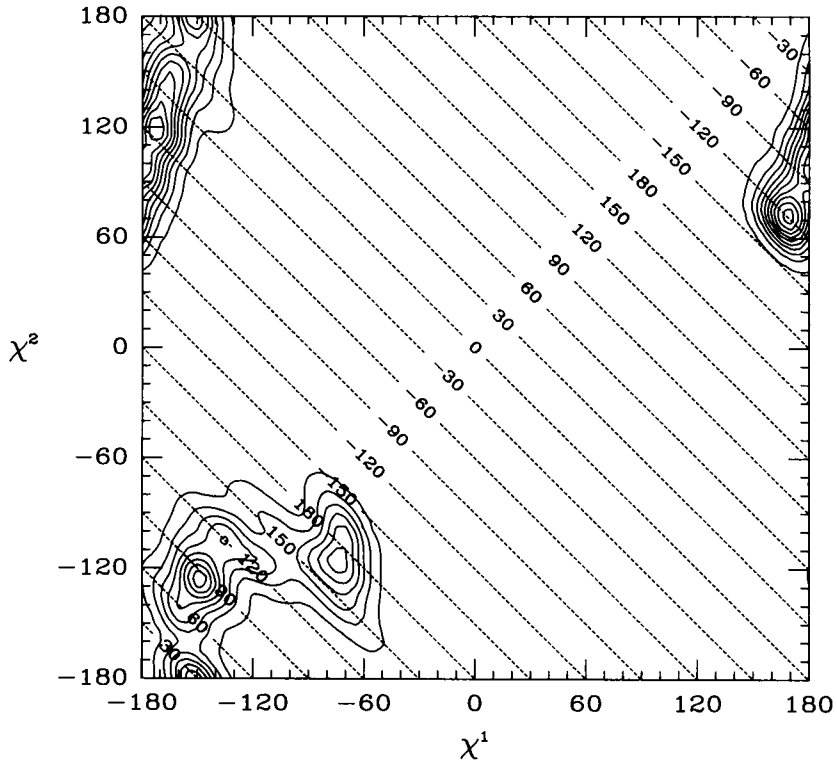

FIGURE 4 Probability density in $\chi^1 \times \chi^2$ torsion space averaged over the 11.0–23.0-ps simulation intervals of the 36 umbrellas. The ξ reaction coordinate is constant along the diagonal dashed lines and equal to the labeled value. The minimum in Fig. 1 at $\xi \approx -75$ degrees corresponds in torsion space to $(\chi^1, \chi^2) \approx (180, 105)$. For comparison, the crystallographic structure is at the point $(\chi^1, \chi^2) = (175.2, 94.5)$.

Thus it is appropriate to model the transition state as a torsional rotor that fluctuates only about the χ^2 dihedral angle. The reaction rate (3, 27) is

$$k_{TST} = \frac{1}{2} \langle |\dot{q}| \rangle \frac{P(q^\ddagger)}{\int_R P(q)\,dq},$$

where $\langle |\dot{q}| \rangle$ is the average absolute value of the Maxwell velocity distribution for the reaction coordinate q in the transition state, $P(q)$ is the reaction coordinate probability density as in Eq. 1, $P(q^\ddagger)$ is the probability per unit q in the transition state, and the subscript R on the integral indicates that the integration is over the entire reactant well. We evaluate the integral by quadratic approximation of the free energy.

$$A(q) = A(q_0) + \tfrac{1}{2} k_0 (q - q_0)^2 + \cdots, \qquad (11)$$

where q_0 is the position of the reactant minimum and k_0 is a force constant. For accuracy, the quadratic term must dominate the higher order terms until $A(q)$ is several times $k_B T$ above the minimum. With this approximation,

$$\int_R P(q)\,dq = \left[2\pi \cdot \frac{k_B T}{k_0}\right]^{1/2} e^{-A(q_0)/k_B T},$$

and the rate

$$k_{TST} = \nu_0 e^{-\Delta A/k_B T}, \qquad (12)$$

where $\Delta A = A(q^\ddagger) - A(q_0)$ is the free energy height of the activation barrier, and ν_0 is the harmonic oscillator frequency. The above expressions apply to any configurational reaction coordinate. When q is a torsional reaction coordinate it must simply be kept in mind that q is expressed in radians and that the Maxwell velocity distribution also describes torsional angular velocities if the particle mass is replaced with the moment of inertia, which gives the average absolute value for the angular velocity

$$\langle |\dot{q}| \rangle = \left[\frac{2}{\pi} \cdot \frac{k_B T}{I}\right]^{1/2}, \qquad (13)$$

where I is the moment of inertia. Also $P(q)$ must be expressed as probability density per radian and the torsional oscillator frequency

$$\nu_0 = \frac{1}{2\pi} \sqrt{\frac{k_0}{I}}, \qquad (14)$$

where the force constant k_0 for the quadratic approximation is expressed per radian squared and I is the moment of inertia.

By a least-squares fit to the cumulative perturbation constants from umbrellas 3–14 the force constant for quadratic approximation of the crystallographic well is 6.1 kcal · mol^{-1} · rad^{-2}. The cumulative perturbation constant for umbrella 20, which is the top of the activation barrier, is 8.5 kcal/mol above the minimum. The moment of inertia for rotation of indole about the χ^2 dihedral angle is 4.3×10^{-14} g · cm^2 · mol^{-1}. This value is derived from the crystallographic structure with $H^{\epsilon 1}$ positioned by CHARMM and the remaining nonpolar hydrogens included as extended heavy atoms. At 300 degrees K Eq. 12 gives 2×10^5 s^{-1} for the escape rate from the crystallographic well.

CONCLUSION

We have computed the relative free energy of the crystallographic rotational isomer of tryptophan-47 and predict a new orientation of the indole side chain in addition to the crystallographic one. Because the transition rate is 2×10^5 s^{-1}, these rotational isomers do not interconvert on the nanosecond time scale. Experimental verification of the existence of the new rotational isomer by fluorescence spectroscopy depends upon the relative population of each isomer and differences in interactions with neighboring amino acids that generate distinct fluorescence lifetimes for each rotational isomer. In view of the wide range of the relative free energy estimate of 1–5 kcal/mol, which corresponds to probabilities of 0.02–20%, the new isomer would probably be detectable by a fluorescence lifetime measurement as well as other experimental techniques. This would hold provided the relative free energy is less than ~3 kcal/mol, which corresponds to a >1% probability of being in the new isomer.

The authors wish to thank Martin Karplus for suggesting the tryptophan rotational isomerization problem and providing access to CHARMM 20 source code. Charles L. Brooks III provided crucial assistance in the application of the stochastic boundary method. Roderick E. Hubbard provided the HYDRA molecular graphics program. We thank John E. Aldag of Cray Research, Inc., for processor time on a Cray X-MP/48 computer system and Sara K. Graffunder and Martin T. Storma for their assistance with software conversion.

This work is supported by grant GM34847 from the United States Public Health Service.

Received for publication 23 October 1989 and in final form 29 January 1990.

REFERENCES

1. Alcala, J. R., E. Gratton, and F. G. Prendergast. 1987. Interpretation of fluorescence decays in proteins using continuous lifetime distributions. *Biophys. J.* 51:925–936.
2. Beechem, J. M., and L. Brand. 1985. Time-resolved fluorescence of proteins. *Annu. Rev. Biochem.* 54:43–71.
3. Brooks, C. L., III, M. Karplus, and B. M. Pettitt. 1988. Proteins: a theoretical perspective of dynamics, structure, and thermodynamics. *Adv. Chem. Phys.* 71:1–259.
4. McCammon, J. A., and S. C. Harvey. 1987. Dynamics of Proteins and Nucleic Acids. Cambridge University Press, Cambridge. 234 pp.
5. Levy, R. M., and R. P. Sheridan. 1983. Combined effect of restricted rotational diffusion plus jumps on nuclear magnetic resonance and fluorescence probes of aromatic ring motions in proteins. *Biophys. J.* 41:217–221.
6. Tanaka, F., and N. Mataga. 1987. Fluorescence quenching dynamics of tryptophan in proteins. *Biophys. J.* 51:487–495.
7. Szabo, A. 1984. Theory of fluorescence depolarization in macromolecules and membranes. *J. Chem. Phys.* 81:150–167.
8. Bennett, C. H. 1977. Molecular dynamics and transition state theory: the simulation of infrequent events. *In* Algorithms for Chemical Computations. R. E. Christofferson, editor. American Chemical Society, Washington, DC. 63–97.
9. Northrup, S. H., M. R. Pear, C. Y. Lee, J. A. McCammon, and M. Karplus. 1982. Dynamical theory of activated processes in globular proteins. *Proc. Natl. Acad. Sci. USA.* 79:4035–4039.
10. McCammon, J. A., C. Y. Lee, and S. H. Northrup. 1983. Sidechain rotational isomerization in proteins: a mechanism involving gating and transient packing defects. *J. Am. Chem. Soc.* 105:2232–2237.
11. Carter, E. A., G. Ciccotti, J. T. Hynes, and R. Kapral. 1989. Constrained reaction coordinate dynamics for the simulation of rear events. *Chem. Phys. Lett.* 156:472–477.
12. Gelin, B. R., and M. Karplus. 1975. Sidechain torsional potentials and motion of amino acids in proteins. *Proc. Natl. Acad. Sci. USA.* 72:2002–2006.
13. Engh, R. A., L. X.-Q. Chen, and G. R. Fleming. 1986. Conformational dynamics of tryptophan: a proposal for the origin of the non-exponential fluorescence decay. *Chem. Phys. Lett.* 126:365–372.
14. Ichiye, T., and M. Karplus. 1983. Fluorescence depolarization of tryptophan residues in proteins: a molecular dynamics study. *Biochemistry.* 22:2884–2893.
15. Mezei, M., and D. L. Beveridge. 1986. Free energy simulations. *Ann. NY Acad. Sci.* 482:1–23.
16. Goldstein, H. 1980. Small oscillations. *In* Classical Mechanics. 2nd ed. Addison-Wesley Publishing Co., Reading, MA. 243–274.
17. Wilson, E. B., Jr., J. C. Decius, and P. C. Cross. 1980. Connection between G matrix and kinetic energy. *In* Molecular Vibrations. Dover Publications, New York. 307–308.
18. H. Cramér. 1946. Matrices, determinants and quadratic forms. *In* Mathematical Methods of Statistics. Princeton University Press, Princeton, NJ. 103–121.
19. Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Computational Chem.* 4:187–217.

20. Jorgensen, W. L., J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

21. van Gunsteren, W. F., and H. J. C. Berendsen. 1977. Algorithms for macromolecular dynamics and constraint dynamics. *Mol. Phys.* 34:1311–1327.

22. Brooks, C. L., III, A. Brünger, and M. Karplus. 1985. Active site dynamics in protein molecules: a stochastic boundary molecular-dynamics approach. *Biopolymers.* 24:843–865.

23. Almassy, R. J., J. C. Fontecilla-Camps, F. L. Suddath, and C. E. Bugg. 1983. Structure of variant-3 neurotoxin from *Centruroides sculpturatus* Ewing, refined at 1·8 Å resolution. *J. Mol. Biol.* 170:497–527.

24. Powell, M. J. D. 1977. Restart procedures for the conjugate gradient method. *Mathematical Programming.* 12:241–254.

25. van Gunsteren, W. F., and M. Karplus. 1980. A method for constrained energy minimization of macromolecules. *J. Computational Chem.* 1:266–274.

26. IUPAC-IUB Commission on Biochemical Nomenclature. 1970. Abbreviations and symbols for the description of the conformation of polypeptide chains. Tentative rules (1969). *Biochemistry.* 9:3471–3479.

27. Glasstone, S., K. J. Laidler, and H. Eyring. 1941. Statistical treatment of reaction rates. *In* The Theory of Rate Processes. McGraw-Hill Book Co., New York. 153–201.