Proc. Natl. Acad. Sci. USA
Vol. 86, pp. 6963–6967, September 1989
Biophysics

Reaction path study of conformational transitions and helix formation in a tetrapeptide

(ϕ,Ψ torsions/barriers/melting/constrained optimization)

RYSZARD CZERMINSKI* AND RON ELBER

Department of Chemistry, The University of Illinois at Chicago, Box 4348, Chicago, IL 60680

Communicated by Hans Frauenfelder, June 15, 1989 (received for review March 22, 1989)

ABSTRACT **Conformational transitions between the 112 stable states of the tetrapeptide isobutyryl-Ala$_3$-NH-methyl (IAN) are studied theoretically. The objective of the investigation is to advance the understanding of helix formation and of conformational transitions in polypeptides. The possible reaction paths between extended chain and helical configurations are examined in detail. The study of the multiple reaction paths in this 48-atom molecule became possible due to development of a new computational algorithm. It is shown that the helix-coil transitions in IAN follow a sequence of local dihedral flips and that the number of the available routes for the transition is significantly lower than in a random search. A quasi-melting point is obtained at 5 ± 1 kcal (1 cal = 4.18 J)/mol above the lowest energy minimum. Below this point the molecule is trapped in one or very few minima, and above it the molecule hops between a large number of configurations.**

One of the unsolved problems in the formation of secondary structure elements (e.g., helices) is how the polypeptide chain finds the correct configuration on a reasonable time scale. It is believed that random search through all possible conformations is too slow and that the actual folding mechanism is biased by certain stable intermediates (1). The structural properties and the energies of these intermediates and the reaction paths connecting them are not known.

Detailed microscopic calculations of the dynamics of secondary structure formation are not as yet feasible. However, studies of simpler model systems have been performed in the past, aiming at qualitative insights into the conformational changes in polypeptides. In particular, a great body of research was devoted to the study of the conformations in a dipeptide and the transitions between them (2–4). However, the dipeptide does not have the 1–4 stabilizing hydrogen bond of the helix configuration and therefore cannot model the formation of the last. Longer peptides were not studied in the past, since the available techniques were limited to systems (*i*) with a relatively small number of degrees of freedom and (*ii*) (more importantly) with a small number of stable conformations. The number of energy minima in peptides grows very rapidly with the molecular size. For example, only three major energy minima and four reaction paths exist in a dipeptide, while in the tetrapeptide there are at least 112 energy minima interconnected by 257 reaction paths. To address this sharp increase in complexity, a new computational technique had to be introduced.

Here we present a detailed calculation on a larger model system—an analog of Ala$_4$ [isobutyryl-Ala$_3$-NH-methyl (IAN)]. IAN is a flexible tetrapeptide that can form a full helical turn (Fig. 1). In addition to the two configurations resembling respectively a helical turn and an extended chain, IAN has a large number of additional minimum energy configurations. We present a detailed theoretical study of the transitions between the different minima, including the transitions between the particularly interesting helix and extended chain configurations. We show that the conformational transitions in IAN are selective (each minimum is connected to a small number of neighbors) and localized on a small part of the molecule—typically a single peptide unit. Since the transitions between the minima proceed via localized conformational changes (localized already on a size scale of a tetrapeptide), similar localized transitions are likely to prevail in larger polypeptides as well. In *Methods* we introduce a computational technique designed for finding minimum energy paths in molecules with multiple minima. In *Results and Discussion* a detailed description of a reaction path between an extended chain and a helix is presented. We also present the statistical analysis of the transitions between all of the different minima. We conclude with a general remark on possible mechanisms of helix formation in polypeptides.

METHODS

The reaction path approach and the transition-state theory are well-established methods to calculate rates of chemical reactions or conformational changes (5–8). To apply these techniques, the coordinates of a transition state between the reactants and the products are required. The transition state is usually identified as a saddle point of the potential energy surface. In practice, computational searches of saddle points are difficult (9). Most of the available techniques are limited to systems with a relatively small number of atoms (<30) and with few energy minima and transition states (9–12). Recently Elber and Karplus (13) introduced a new algorithm that made possible for the first time calculations of reaction coordinates in large molecular systems (≤1000 atoms). However, the convergence of the new algorithm was still relatively slow. Here we present an alternative approach that is between 9 and 20 times faster depending on the type of constraints and on the tolerance imposed on the gradient. This approach is the only one practical for the current problem in which thousands of reaction paths need to be calculated for a 48-atom molecule. Detailed comparison of the different computational approaches will be published elsewhere. The principles of the new technique are outlined below.

We denote by $V(\mathbf{R})$ the potential energy and by $\mathbf{R}_1$ and $\mathbf{R}_2$ the coordinate vectors of two different potential energy minima between which we search for the connecting minimum energy path. $\mathbf{R}_1$ and $\mathbf{R}_2$ are input required by the algorithm. We impose on $\mathbf{R}$ the following holonomic constraint which depends on a parameter α

$$(\mathbf{R} - \mathbf{R}_\alpha)\cdot(\mathbf{R}_1 - \mathbf{R}_2) = 0, \qquad [1]$$

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: IAN, isobutyryl-Ala$_3$-NH-methyl.

*Permanent address: Institute of Physics, Polish Academy of Sciences, Al. Lotnikow 32/46, 02-668 Warsaw, Poland.

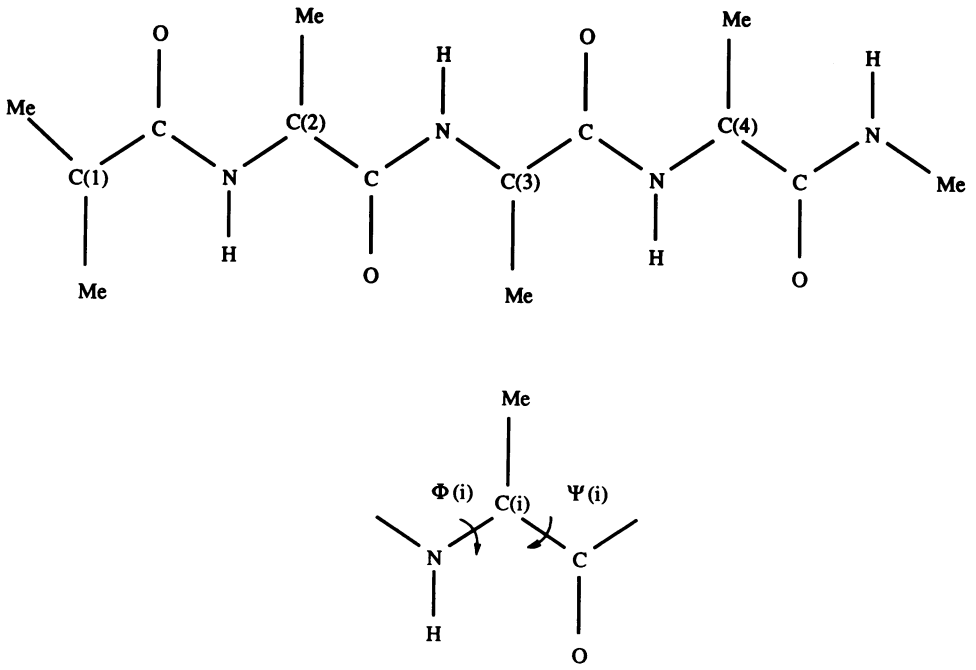

FIG. 1. Schematic representation of IAN. The calculation used the united-atom model in which the CH_n are approximated by a single sphere. All other degrees of freedom were included explicitly in the calculation. We also show the soft degrees of freedom of a single unit. The ϕ,Ψ dominate the structural changes, though the inclusion of the rest of the degrees of freedom is essential to obtain reasonable barrier values.

where

$$\mathbf{R}_\alpha = [(1 - \alpha)\mathbf{R}_1 + \alpha\mathbf{R}_2].$$

The set of points $\mathbf{R}$ that satisfy Eq. 1 for a given α form a multidimensional plane between $\mathbf{R}_1$ and $\mathbf{R}_2$. α parameterizes a zero-order reaction coordinate and varies from zero (reactants) to one (products) in small steps. At each step (different α), the potential energy is minimized while forcing the system to remain in the plane (1). The optimized structures for different α values give a discrete representation of the minimum energy path between the two minima $\mathbf{R}_1$ and $\mathbf{R}_2$.

The cartesian coordinates used are convenient for implementation on computers. A difficulty in their use is that rigid body motions may affect the "α" constraint. (These motions cannot change the value of the potential energy, which is a function of the relative coordinates only.) It is necessary therefore to project these motions out. This is done by adding six additional linear constraints given below—three on the translation and three on the rotation:

$$\sum_i M_i(\mathbf{r}_i - \mathbf{r}_i^{\text{ref}}) = 0, \qquad [2]$$

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

where the sum is over the atoms, M_i is the atomic mass, $\mathbf{r}_i$ is the atomic coordinate vector to be optimized, and $\mathbf{r}_i^{\text{ref}}$ is the reference coordinate ($\mathbf{R}_1$ or $\mathbf{R}_2$).

Thus, we compute at each point of the reaction coordinate a minimum of the potential energy $V(\mathbf{R})$ subject to seven linear constraints, which we denote by $\sigma_i (i = 1, 7)$.

The general idea of locating transition states by using constrained optimization is not new (9, 10, 13). However, the current approach is different in two respects: first, the calculated path satisfies the following criterion in the definition of a reaction coordinate; at any point along the reaction coordinate only one direction may have negative energy curvature. This criterion is satisfied since the energy is minimized for all degrees of freedom except one—the direction perpendicular to the plane. In that direction the potential energy curvature may be negative. Second, the optimization in the present formulation is subject only to linear constraints, which speeds up the calculation to the extent of making it feasible.

The minimization subject to linear constraints is performed as follows. Consider a coordinate set $\mathbf{R}^0$ that satisfies $\sigma_i(\mathbf{R}^0) = 0$ ($i = 1, 7$). A displacement $\boldsymbol{\delta}$ that is added to $\mathbf{R}$ does not violate the linear constraints if $\sigma(\mathbf{R}^0 + \boldsymbol{\delta}) = \nabla\sigma\cdot\boldsymbol{\delta} = 0$ (that is, if $\boldsymbol{\delta}$ is orthogonal to the gradient of the constraints). Such orthogonalization is straightforward to perform in our case since the gradients of the constraints are constant and do not vary during the optimization. Therefore, we project the gradient of the potential to the subspace of the constraints during each evaluation of the forces. This assures that the step $\boldsymbol{\delta}$, which is given by a linear combination of the current and the older gradients, lies in the subspace defined by the constraints [see Powell conjugate gradient algorithm (14)]. The numerical errors of the constraints are smaller than 10^{-13} amu·Å for the translations, 10^{-10} amu·Å^2 for the rotations and 10^{-13} Å^2 for the α constraint (amu = atomic mass unit).

To ensure the accuracy of the results, the calculated transition states were further refined by the method of Banarjee *et al.* (12). [We incorporated in our program the subroutine of Baker (15), which is based on the formalism of ref. 12.] The energy was maximized along the eigenvector of the second-derivative matrix with a negative eigenvalue and minimized for all the other directions. The initial structure was the barrier configuration located by our technique. The refinement decreased considerably the value of the energy gradient from $\approx 10^{-1}$ to $\approx 10^{-6}$ kcal/mol·Å. However, in most of the cases the changes in the structures and the energies of the transition states due to the refinement were small. The energy changes were of the order of 0.01–0.1 kcal/mol, and the changes in the dihedral angles (the soft degrees of freedom of the system) were ≈5 degrees.

The above computational procedure was used in the study of transitions between different minima of IAN. We focused initially on the transition from the extended chain to the helix. Idealized α-helix and β-sheet configurations $(\phi,\Psi) = (-60, -60)$ for α and $(-60, +120)$ for β were minimized by using the

CHARMM potential energy (16). We note that although the solvent (water) was not included explicitly in the study, the distance-dependent dielectric option in the CHARMM force field that we used is parameterized to recover some of the solvent effects. The resulting α-helix "reactant" and β-sheet "product" were used in the reaction-path calculation. The calculated minimum-energy path passed through several intermediate minima and maxima. To explore the possibilities of alternative transition routes, the newly discovered minima were further investigated. The reaction paths between all of the pairs of the minima were calculated. This secondary calculation resulted in a number of additional minima and was repeated until converged. Convergence was assumed when all of the $N(N-1) = 12{,}544$ reaction path calculations for $N = 112$ different minima did not result in new minima. (We comment that some additional minima may be left undetected by using the above protocol.) The calculated paths between the 112 minima include paths from the "reactant" to the "product" and from the "product" to the "reactant" (for additional reliability). Finally we obtained only 257 direct paths (with no intermediates) out of $N(N-1)/2 = 6216$ possible paths examined.

The torsional angle distribution of the resulting 112 minima is similar to that found in a dipeptide (2–4) and includes three major configurations at $\phi\Psi \sim (-60, -60)$ (28%); $(-60, 120)$ (42%); $(60, -60)$ (30%). The percentile given is average over all the different minima, and all the $\phi\Psi$ pairs in each stable structure.

RESULTS AND DISCUSSION

In Fig. 2 the lowest energy path between the helix and the extended chain conformation is presented. Only the changes in the Ψ angles are shown (Fig. 2*a*), since the variations in the rest of the degrees of freedom were small. (The changes in ϕ along the reaction coordinate were less than 30 degrees.) The path is composed of sequential, weakly overlapping Ψ transitions. Along the reaction coordinate, each Ψ transition is close to completion before the next transition is initiated. The first Ψ flip replaces the 1–4 hydrogen bond by a β turn. The second flip results in two γ turns, and finally the extended chain conformation includes three γ turns. Hydrogen bonds contribute a substantial fraction of the electrostatic and the van der Waals energies (Fig. 2*b*), and the energy barriers are associated with their flips. In Fig. 2 the lowest energy path is presented (the highest barrier is of 4.9 kcal/mol); additional 1068 indirect paths can be constructed between the helix and the extended chain with up to 10 intermediates and a barrier below 6 kcal/mol.

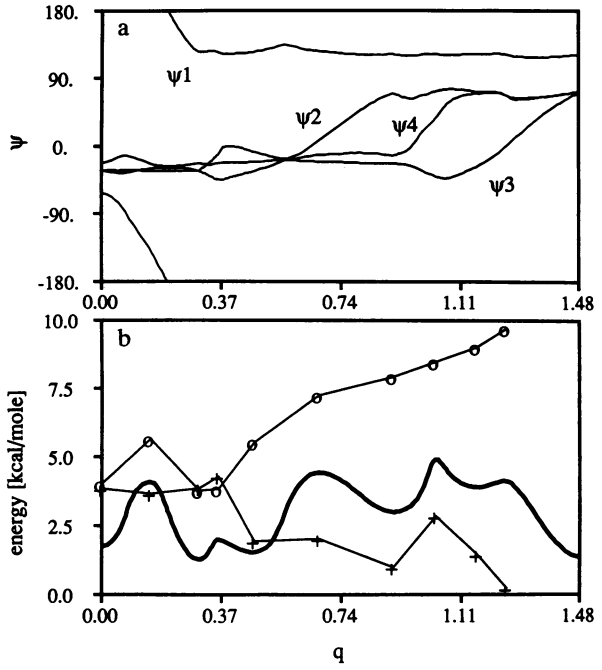

FIG. 2. (*a*) The changes in the Ψ dihedral angles along the reaction coordinate q (in Å) from the helix to the extended chain conformation. The changes in all ϕ values are less than 30 degrees. (*b*) The lowest energy path for the transition from the helix to the extended chain. The helix is on the left, and the extended chain conformation is on the right. The reaction coordinate includes four intermediates related by hydrogen bond flips. The energy profile is dominated by the electrostatic terms (+) and by the van der Waals interactions (○).

We consider next the properties of the 257 direct transitions (without intermediate minima) between the 112 minima of the IAN. We should like to understand the mechanisms of the conformational changes in this tetrapeptide and to indicate the possible relevance of these mechanisms to formation of secondary structure elements (not necessarily helices) in polypeptides in general. Therefore, it is of interest to consider statistical properties of the calculated paths and the resulting network between them. All of the possible routes between the different minima are fully characterized by the direct paths—i.e., the paths with a single maximum. This is analogous to the traveling salesman problem (17) in which the information on direct flights is sufficient to construct all possible connections.

In Fig. 3 we show a connectivity tree in which the minima are placed at the tips of the tree branches (●), and the barriers are located at the junctions of the branches. For clarity only the dominant branch of the tree is shown. At the highest energy of this branch, efficient hopping between a large number of minima is possible. To further investigate the transitions between minima we plot in Fig. 4 the distribution

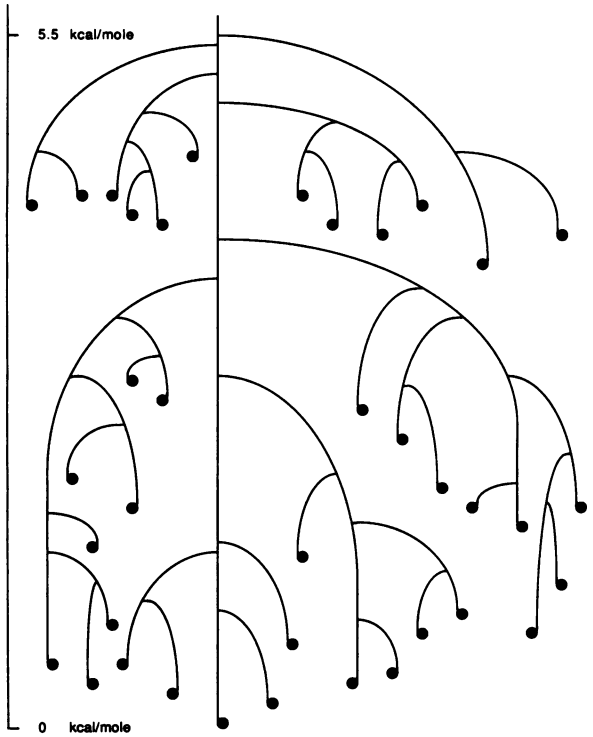

FIG. 3. A schematic plot of the dominant branch in the connectivity tree of IAN. The energy range considered is from zero (the lowest energy minimum) to 5.5 kcal/mol. A minimum is described by a dark full circle and a barrier as junction between different branches. For clarity, only the lowest energy path between any two minima or clusters is shown.

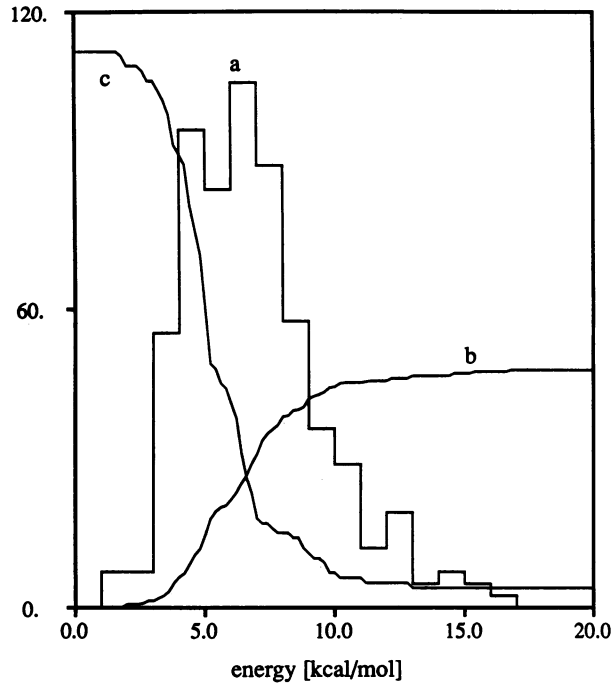

FIG. 4. Graph a shows the distribution of the barrier heights as a function of the energy in kcal/mol (the zero is set at the lowest energy minimum). The number of barriers at the maximum is 22. Curve b shows the average number of the directly connected neighbors to a given minimum (the maximum of this plot is 4.0). Curve c shows the number of disjoint clusters of interconnected minima.

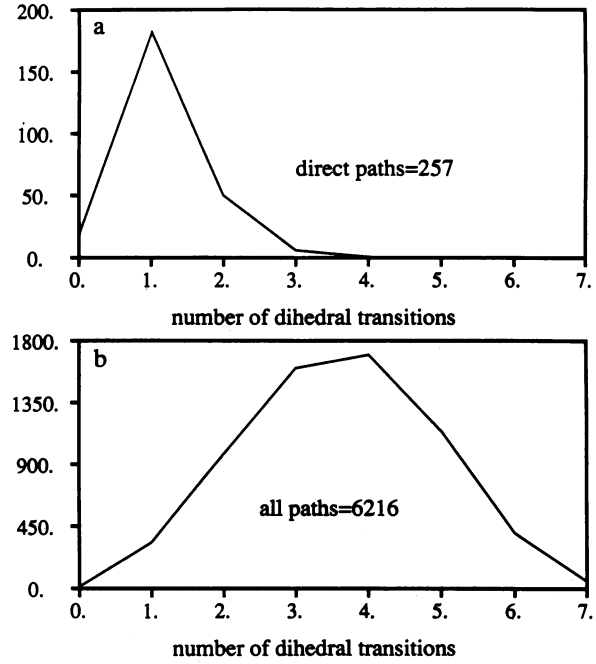

FIG. 5. (*a*) A histogram plot of the number of direct paths with a given number of dihedral transitions. The x axis is the number of dihedral flips, and the y axis is the number of calculated paths. A flip is defined as a change of at least 60 degrees. (*b*) A histogram plot of the number of paths with a given number of dihedral transitions for the hypothetical case in which all the minima are directly connected to each other. That is, the 112 minima are directly connected to 111 neighbors (maximum connectivity).

of the barrier heights as a function of the energy (graph a), the average number of neighbors (i.e., the average number of minima that are connected to a given minimum by direct paths) (curve b), and the number of clusters of minima with energy barriers (between the clusters) which are higher than a given value (curve c). The barrier distribution rises sharply at ≈5 kcal/mol. Below this energy the minima are mostly disconnected from each other, and the average number of neighbors is zero. Above this energy a large number of hopping channels become open, resulting in a sharp decrease in the number of disjoint clusters of minima and an increase in the average number of neighbors to 4.0. It is thus apparent that the system has two distinct dynamical domains. In the short timescale limit (or in the low temperature limit), the system is trapped in a single minimum or in very few interconnected minima, whereas on longer timescales (or at higher temperatures), efficient hopping is expected between numerous minima. While the detailed kinetics of the system was not addressed in this study, one can make a rough estimate for the hopping rate as $K = \omega\exp(-\beta E)$ with $\omega \approx 10^{12}$ s^{-1}, and $E \approx 6$ kcal/mol. Taking into consideration a possible error of some 2 orders of magnitude in this rough estimate, we can still suggest for 25°C efficient hopping on a timescale of $t > 10^{-7}$ s and trapping in a single minimum on a timescale of $t < 10^{-9}$ s. We note that similar hierarchy of timescales in conformational changes was measured in the past for the more complex "polypeptides"—proteins (18). Possible experiments to probe this transition between the two dynamical domains in IAN may be line-shape analysis (to detect the transition from homogenous to inhomogenous broadening as the temperature of the experiment decreases) or incoherent and inelastic neutron scattering to probe the hydrogen fluctuations.

The question of "local" versus "extended" reaction coordinate is addressed next. That is, we examine the spatial extent of the conformational changes occurring in the direct reaction paths. The major conformational changes occur in the "soft" degrees of freedom—i.e., the dihedral angles. In Fig. 5*a* the number of direct paths is shown as a function of the number of dihedral flips occurring during a given direct transition. Clearly, most of the direct paths are associated with a single dihedral flip. For comparison, we show in Fig. 5*b* a "maximum connectivity" distribution, which would be obtained if all of the minima were directly connected to each other (each of the 112 minima would then have 111 neighbors). The "maximum connectivity" distribution peaks at three and four flips per path. Hence, the actual connectivity is not only lower than the maximal value but also has a strong preference for localized events; 70% of the localized flips are in Ψ and only 30% are in ϕ. The small number of double flips found (19%) are in neighboring dihedral angles either on the same residue or on two sequential peptide units.

We finally consider a possible relevance of our results to conformational changes in polypeptides in general. In our view the important results in this respect are the dilute connectivity between the minima and the localization of the conformational transitions. This localization can simplify considerably construction of models for folding, since it is easier to parameterize local properties than the extended ones. The dilute connectivity results in channeling to specific folding paths, which may limit the conformational space and accelerate the search for the "right" conformation. Since the channeling and the localization are apparent already in a small peptide such as IAN, these properties are likely to survive in larger polypeptides.

R.E. thanks Prof. P. G. Wolynes for useful discussions and Prof. J. Kuriyan for a careful reading of the manuscript and many helpful suggestions. R.E. is a Camille and Henry Dreyfus new faculty scholar. R.C. wishes to acknowledge Grant CPBP 01.12 from the Polish Academy of Sciences for partial support during this research. This research was supported by National Institutes of Health Grant GM40698-01.

1. Cantor, C. R. & Schimmel, P. R. (1980) in *Biophysical Chemistry* (Freeman, San Francisco), Vol. 3, pp. 1075–1107.
2. Mezei, M., Mehrotra, P. K. & Beveridge, D. L. (1985) *J. Am. Chem. Soc.* **107**, 2239–2245.
3. Pettitt, B. M. & Karplus, M. (1985) *Chem. Phys. Lett.* **121**, 194–201.
4. Anderson, A. G. & Hermans, J. (1988) *Proteins* **3**, 362–365.
5. Depristo, A. D. & Rabitz, H. (1978) *J. Chem. Phys.* **68**, 4017–4021.
6. Miller, W. H., Handy, N. C. & Adams, J. E. (1980) *J. Chem. Phys.* **72**, 99–112.
7. Truhlar, D. G. & Garret, B. C. (1980) *Acc. Chem. Res.* **13**, 440–448.
8. Pechukas, P. (1981) *Annu. Rev. Phys. Chem.* **32**, 159–177.
9. Bell, S. & Crighton, J. S. (1984) *J. Chem. Phys.* **80**, 2464–2475.
10. Halgren, T. A. & Lipscomb, W. N. (1977) *Chem. Phys. Lett.* **49**, 225–232.
11. Nguyen, D. T. & Case, D. A. (1985) *J. Phys. Chem.* **89**, 4020–4026.
12. Banarjee, A., Adams, N., Simons, J. & Shepard, R. (1985) *J. Phys. Chem.* **89**, 52–57.
13. Elber, R. & Karplus, M. (1987) *Chem. Phys. Lett.* **139**, 375–380.
14. Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1986) *Numerical Recipes* (Cambridge Univ. Press, Cambridge, U.K.), pp. 294–307.
15. Baker, J. (1986) *J. Comput. Chem.* **7**, 385–395.
16. Brooks, B. R., Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. & Karplus, M. (1983) *J. Comput. Chem.* **4**, 187–217.
17. Press, W. H., Flannery, B. P., Teukolsky, S. A. & Vetterling, W. T. (1986) *Numerical Recipes* (Cambridge Univ. Press, Cambridge, U.K.), pp. 328–334.
18. Austin, R. H., Beeson, K. W., Eisenstein, L., Frauenfelder, H. & Gunsalus, I. C. (1975) *Biochemistry* **14**, 5355–5373.