Equilibrium states of rigid bodies with multiple interaction sites: Application to protein helices

B. Erman

Polymer Research Center, Bogazici University and TUBITAK Advanced Polymeric Materials Research Center, Bebek 80815, Istanbul, Turkey

I. Bahar

Polymer Research Center, Bogazici University and TUBITAK Advanced Polymeric Materials Research Center, Bebek 80815, Istanbul, Turkey and Molecular Structure Section, Laboratory of Experimental and Computational Biology Division of Basic Sciences, National Cancer Institute, National Institutes of Health, MSC 5677, Room B-116, Bldg. 12B, Bethesda, Maryland 20892-5677

R. L. Jernigan

Molecular Structure Section, Laboratory of Experimental and Computational Biology Division of Basic Sciences, National Cancer Institute, National Institutes of Health, MSC 5677, Room B-116, Bldg. 12B, Bethesda, Maryland 20892-5677

Equilibrium configurations of rigid building blocks with multiple embedded interaction sites are investigated, as a coarse-grained approach for conformational sampling of protein structures with known secondary structure. First, hypothetical structures of asymmetric shapes, and pairs of rods composed of multiple interaction sites are considered. The rods are either disconnected or joined by a flexible loop. The sites are assumed to interact with a classical 6-12 Lennard-Jones potential. Subsequently, the investigation is extended to the study of two disconnected α helices composed of homogeneous interaction sites and to the ROP monomer, a small protein consisting of two heterogeneous α helices connected by a loop. Residue-specific long-range and short-range potentials extracted from a protein database are used. A Monte Carlo procedure combined with an energy minimization algorithm, originally developed by Li and Scheraga [Proc. Natl. Acad. Sci. USA **84**, 6611 (1987)] is used to generate a set of low energy conformations over the full conformational space. Results show that: (i) The potential of mean force between two rods as a whole exhibits an inverse linear dependence on the separation between rods despite the individual sites interacting via a 6-12 Lennard-Jones potential. (ii) As the length of the rods (or helices) increases, they tend to align parallel to one other. (iii) This tendency to become parallel is enhanced when the density of interaction sites is higher. (iv) The angle between the principal axes of the rods is found to scale as $n^{-5/3}$ with the number n of sites. (v) The native conformation of the ROP monomer, including the detailed rotational states of the virtual bonds located in the loop connecting the α helices is correctly predicted. This lends support to the adoption of such a coarse-grained model and its parameters for future simulations. © *1997 American Institute of Physics.* [S0021-9606(97)50430-2]

I. INTRODUCTION

The determination of the equilibrium configuration of structural elements containing rigidly embedded interaction sites is a problem of interest in several disciplines. For example, in the investigation of complex biomolecular processes such as protein folding, ligand binding, protein–protein interactions, etc., a simple approach is to assume a rigid-body approximation for some structural elements which possess sufficient internal stability, and simulate the statistics and dynamics of these bodies. Likewise, in polymer simulations monomeric repeat units are conveniently approximated by unified groups.[1,2] In the case of proteins, depending on the degree of resolution, individual amino acids as a whole or secondary structural elements such as α helices, may be approximated as rigid blocks.[3–7] At a more coarse-grained level, structural domains, or individual molecules forming a complex or an aggregate may be viewed as rigid building blocks whose mechanics and dynamics are of interest.

Here, we will concentrate on the spatial organization of such rigid bodies. For a system of n interaction sites, the determination of the global minimum in the $3n$-dimensional energy landscape is known as an NP-hard problem, elusive to conventional energy minimization searches on computer, i.e., the complexity of the problem increases exponentially, and not via a polynomial with the increasing size of a molecule. Embedding these sites into m rigid blocks, $m \ll n$, on the other hand, reduces the total number of degrees of freedom to $6m$, or to $6(m-1)$ if the absolute location and orientation in space are neglected. This brings a significant reduction in the complexity of the problem.

The key, general strategy for finding the minimum en-

ergy configuration of such models is the following: First, the relevant variables of the coarse-grained structure are chosen. Second, the total free energy is minimized with respect to these variables. As shown in the Appendix, the equations of motion controlling the linear and angular momentum of the system are highly nonlinear and possess large numbers of local minima. One may easily be trapped in a local minimum, which depends on the starting point in the configuration space. In order to avoid such events, we adopt here a heuristic approach and generate a large number of "low energy conformations" starting from random initial points over the complete phase space. The method closely resembles the Monte Carlo (MC) procedure with minimization originally developed by Li and Scheraga,[8] and shown by Abagyan and Maiorov to be an efficient conformational search technique.[9] We then analyze the ensemble of low energy conformations to determine the most probable state. A key strategy adopted at this step is to determine the loci of the lowest energy states as a function of a single geometric variable ξ, all other variables being optimized to achieve the lowest energy configuration subject to the fixed ξ value.

The paper is organized as follows: In the next section, the basic theoretical approach is presented with illustrative calculations for a few hypothetical collections of rigid building blocks, including two disconnected rods and two rods connected by a flexible loop. In the Sec. II B, these concepts are extended to helical protein fragments. Both disconnected pairs of helices composed of homogeneous interaction sites and helices connected by a loop, comprising specific interaction sites, are considered. ROP monomer[10] is considered as a specific protein for illustrating the application of the method. In addition to the correct registration of the two helices, the detailed conformation of the loop connecting the helices is predicted by an efficient energy minimization algorithm.

II. RESULTS

A. Illustrative calculations with simple models

1. Asymmetric shapes in two dimensions

In Fig. 1(a), two rigid bodies M_1 and M_2 in a two-dimensional space are shown. O_1 and O_2 are two arbitrary points fixed in the two respective bodies, to which the coordinate systems $x_1y_1z_1$ and $x_2y_2z_2$ are affixed. $\mathbf{a}_i$ and $\mathbf{b}_j$ are the position vectors of the ith and jth sites in M_1 and M_2. M_1 and M_2 are assumed to have m and n rigidly embedded interaction sites, respectively. In the absence of the rigid body assumption, the determination of the minimum energy configuration would require the simultaneous solution of $2(m+n)$ equations. Here, the problem reduces to the simultaneous solution of three equations, only, two linear momentum and one angular momentum balance, for three unknowns, say R_x, R_y, and θ. Here, R_x and R_y are the components of the position vector $\mathbf{R}$ pointing from O_1 to O_2, and θ is the angle between the axes y_1 and y_2, which define the rotational state of M_2 with respect to M_1. The set of equations and their solution are outlined in Appendix A.

Figure 2 illustrates some equilibrium configurations obtained from the solution of Eqs (A8) for various collections
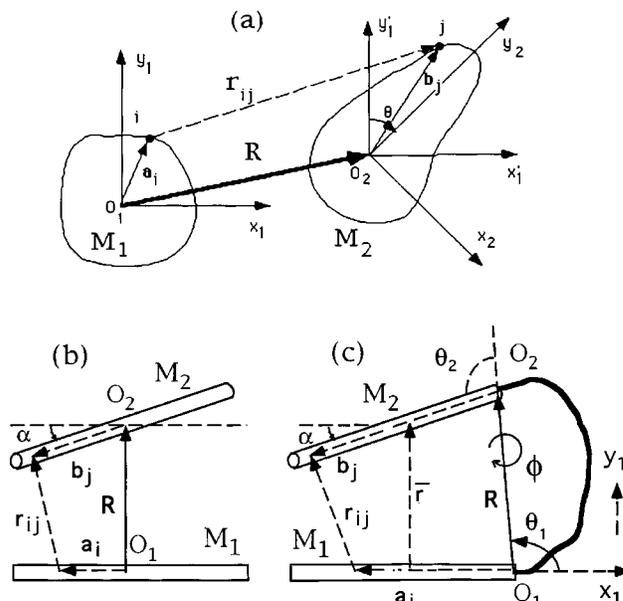

FIG. 1. Model structures used for illustrative calculations. (a) Two rigid bodies M_1 and M_2 in two-dimensions. Interaction sites rigidly embedded in M_1 and M_2 are expressed by the respective position vectors $\mathbf{a}_i$ and $\mathbf{b}_j$, with respect to the molecule-embedded coordinate systems $O_1x_1y_1z_1$ and $O_2x_2y_2z_2$. $\mathbf{R}$ is the vector pointing from O_1 to O_2. It is conveniently expressed in terms of its magnitude R and the angle θ between the axes y_1 and y_2. $\mathbf{r}_{ij}$ is the separation vector between sites i and j. (b) Two disconnected rods. α is the angle between the axes of the rods. (c) Two rods connected by a flexible spacer. The origins O_1 and O_2 are chosen here at the connections of the spacer. $\bar{\mathbf{r}}$ denotes the separation between the centers of the rods. Two supplemental angles θ_1 and θ_2, and one torsional angle ϕ define the orientation of M_2 with respect to M_1, provided that the rods are cylindrically symmetric.

of rigid bodies. The interaction sites are located at the corners of the blocks, in each case. The interaction energy between sites i and j is assumed to obey a Lennard-Jones (LJ) potential, with attractive and repulsive coefficients equal to
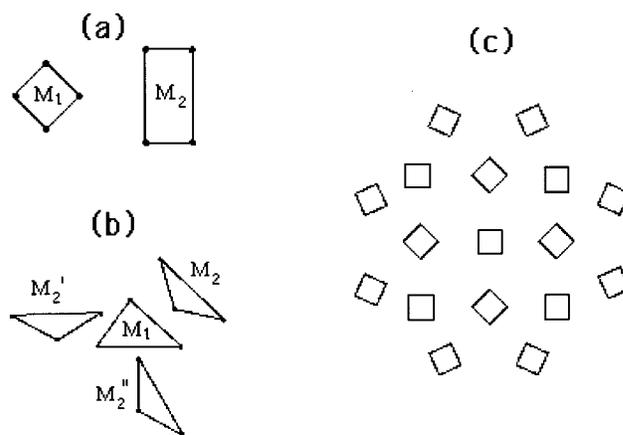

FIG. 2. Equilibrium states of collections of rigid bodies in two dimensions. This illustrates some equilibrium configurations obtained from the solution of Eqs. (A8) for various collections of rigid bodies. The interaction sites are located at the corners of the blocks, in each case. The interaction energy between sites i and j is assumed to obey a Lennard-Jones (LJ) potential, with attractive and repulsive coefficients equal to $A_{ij} = 1$ kcal Å^6/mol and $B_{ij} = 1088$ kcal Å^{12}/mol, respectively.

$A_{ij} = 1$ kcal Å^6/mol and $B_{ij} = 1088$ kcal Å^{12}/mol, respectively. This choice of parameters yields an equilibrium distance of 3.6 Å between two single interacting points in space.

The solutions displayed in Fig. 2 are obtained sequentially. First, Eqs. (A8) are solved for two bodies only. Their equilibrium positions obtained in this manner are then held fixed, and the third body is placed at its equilibrium configuration by solving Eqs. (A8), and so on. The configurations shown in Fig. 2 are unique and exactly reproducible, irrespective of the starting configurations. In order to test the validity of keeping all previously placed $n-1$ bodies fixed while placing the nth body, relaxation experiments are performed. For example, for Fig. 2(b), the triangles were initially placed in the order, M_1, M_2, M_2', M_2''. After placing the fourth triangle M_2'', the triangle M_2 is relaxed and its new equilibrium configuration relative to the other three fixed triangles is recalculated. The result did not differ noticeably from the original equilibrium configuration of M_2 relative to M_1. This is due to the presence of M_1 between M_2, M_2' and M_2''. The distances between the interaction sites of M_2 and those of M_2' and M_2'' are much larger than the effective range of the LJ potentials so that the location of M_2 is predominantly determined by its interaction with M_1.

2. Pair of rods composed of multiple interaction sites

In Fig. 1(b), two rods with their centers at a separation of $\mathbf{R}$ are displayed. When the interaction sites are located symmetrically with respect to the center of each rod, $\mathbf{R}$ becomes mutually perpendicular to the two rods and the solution for the equilibrium configuration of the two rods simplifies considerably. Then, the two unknowns to be evaluated are the magnitude of $\mathbf{R}$ and the angle α between the two rods. Results of calculations with the same potential and coefficients as those of Sec. II A 1 are presented in Fig. 3. The ordinate is the angle α and the abscissa is the number of interaction sites on each rod. The three curves are obtained for different values of the parameter η, where $\eta \equiv r_{eq}/l$ is the ratio of the equilibrium distance between two sites to the separation between consecutive sites on a given rod. Defined in this manner η is a measure of the density of interaction sites on a given rod. The range $2 \leqslant \eta \leqslant 3$ is typical of atomic sites, real bond lengths being about two or three times shorter than the sum of the van der Waals radii of the atoms. For C–C bonds, for example, $\eta = 2.35$, taking $l = 1.53$ Å and $r_{eq} = 3.6$ Å.

The most probable angle α between the rods decreases from 90° to 0° with an increasing number n of interaction sites (Fig. 3), the decrease being sharper in the case of more distant interaction sites (smaller η) along the chain. The decay curves are plotted on a logarithmic scale in the inset. The straight lines in the latter, which are obtained by least squares fits, have approximately equal slopes of $-5/3$. Thus α may be expressed by a power relation $\alpha \sim n^{-5/3}$. In general, the distance $\bar{r}$ between the centers of the two rods at equilibrium is observed to remain slightly below the equilibrium distance between two free sites.
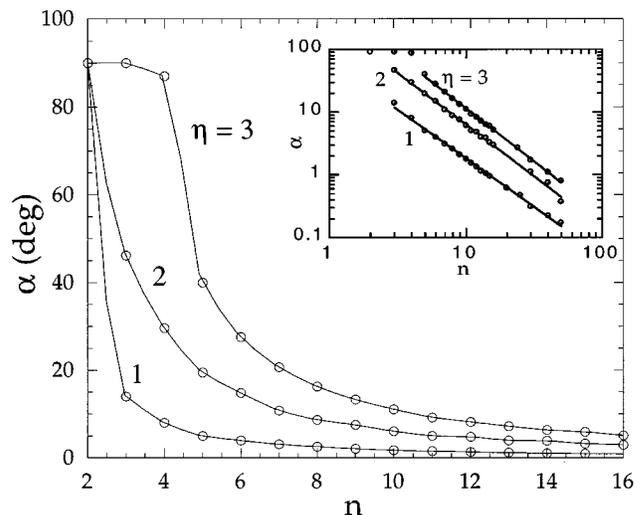

FIG. 3. Dependence of the angle α between the axes of the rods on the number of interaction sites, n, rigidly embedded in the rods. The curves are drawn for various choices of the density parameter $\eta \equiv r_{eq}/l$, where r_{eq} is the equilibrium distance between two nonbonded interaction sites and l is the separation between two consecutive sites on a given rod. The logarithmic plot in the inset indicates a power law of the form $\alpha \sim n^{-5/3}$.

3. Pair of rods composed of multiple interaction sites, joined by a loop

The presence of a flexible connector at the ends of the rods modifies the problem of the previous section. The total free energy of the system is now found from the sum of intermolecular interactions and the free energy of the loop. The latter may conveniently be chosen as the elastic free energy change of entropic origin $\Delta A_{el} = -kT \ln W(R)$ for a flexible chain whose end-to-end separation obeys the Gaussian distribution

$$W(R) = K \exp\left\{ -\frac{3(R-\bar{R})^2}{2\langle (R-\bar{R})^2 \rangle_0} \right\}. \tag{1}$$

Here, K is the normalization constant and $\bar{R}$ is the equilibrium end-to-end separation of the connector in the absence of the rods. The loop is therefore approximated as a spring with a force constant equal to $3kT/2\langle (R-\bar{R})^2 \rangle > 0$. A suitable choice of geometric variables for determining the minimum energy configuration(s) may be the magnitude R of the vector $\mathbf{R}$ joining the two ends of the loop, the angle θ_1 between $\mathbf{R}$ and the first rod, the torsion angle ϕ by which M_2 has rotated about $\mathbf{R}$, and the angle θ_2 between $\mathbf{R}$ and M_2. The problem of obtaining the equilibrium configuration is more complicated relative to that of Fig. 1(b). This is because the previous symmetry no longer holds due to the constraints imposed by the loop, unless the loop were sufficiently flexible so as to permit the optimal placement of the two disconnected rods.

A set of low energy conformations is determined by the solution of Eq. (A8), as outlined in Appendix A. The configurational energy values associated with these most likely states are presented in Fig. 4 as a function of the distance between the midpoints of the rods. The curves are obtained
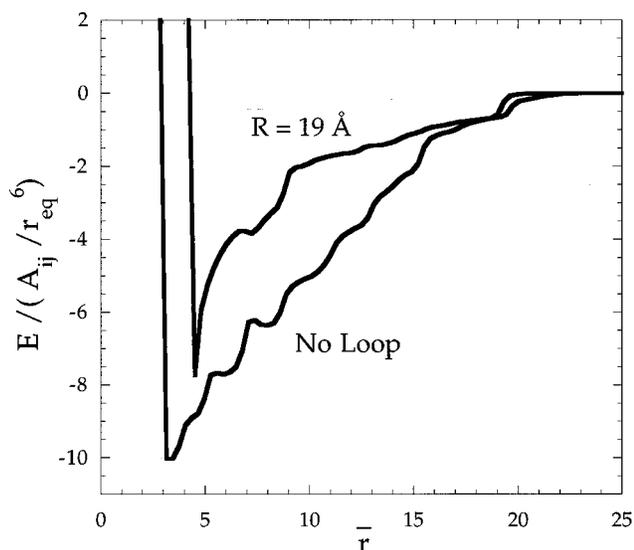
FIG. 4. Energies of the lowest energy states of the pair of rods shown in Fig. 1 as a function of the separation $\bar{r}$ between the centers of the rods. The upper curve represents the loci of local minima for the case of rods connected by a flexible spacer whose equilibrium end-to-end separation is $\bar{R} = 19$ Å with a force constant of $3 kT/2\langle(R-\bar{R})^2\rangle_0 = 0.5$ kcal mol^{-1} Å^{-2}. The lower curve is obtained for the same pair of rods in the absence of the loop. The lowest energy configuration is obtained at $\bar{r}_{min} = 3.5$ Å. In the lowest energy configuration, the ends of the two rods are positioned at $R = 16.5$ Å. Here, the loop constrains the rods to assume a relatively higher energy configuration, and the most probable separation between the rods is shifted from 3.5 to 4.4 Å. The ordinate values are normalized with respect to the LJ attractive energy parameter A_{ij} and the equilibrium separation r_{eq} of nonbonded sites, by dividing with $10^3 A_{ij}/r_{eq}^6$. A very sharp increase in energy occurs at shorter separations due to strong repulsive interactions. The energy approaches zero at about 20 Å, indicating that the attractive potential between the rods is highly nonlocal.

for $n = 20$ and $\eta = 2.35$. The energies are normalized by dividing by $10^3 A_{ij}/r_{eq}^6$. The lower curve is obtained in the absence of the loop. The lowest energy configuration occurs at $\bar{r}_{min} = 3.5$ Å. In the lowest energy configuration, the ends of the two rods are positioned at $R = 16.5$ Å. The upper curve is obtained in the presence of a loop which obeys the distribution of Eq. (1) with $\bar{R} = 19$ Å and $3kT/2\langle(R-\bar{R})^2\rangle_0 = 0.5$ kcal mol^{-1} Å^{-2}. These values cause the energy of the spring to compete with the intermolecular energy of the rods and $\bar{r}$ is shifted to a larger value, $\bar{r}_{min} = 4.4$ Å. An extremely sharp increase in the energy takes place at shorter separations, i.e., the repulsive part of the interaction is very strong. The curve levels off at values of $\bar{r}$ as high as 20 Å, indicating that the attractive potential between the centers of the two rods is highly nonlocal.

In Fig. 5, the data of Fig. 4 are plotted in double logarithmic coordinates after removing the leftmost, rising portion of the curves. Presented in this form, the slopes of the curves provide an estimate of the power law $E(\bar{r}) \sim -(1/\bar{r})^p$ governing the attractive part of the potential of mean force between the rods in the absence and in the presence of a connector. The two curves exhibit a qualitatively similar character, regardless of the presence of a connector. The upper straight line, drawn for comparison, has a slope of -1, and approximately matches the portion of the two en-
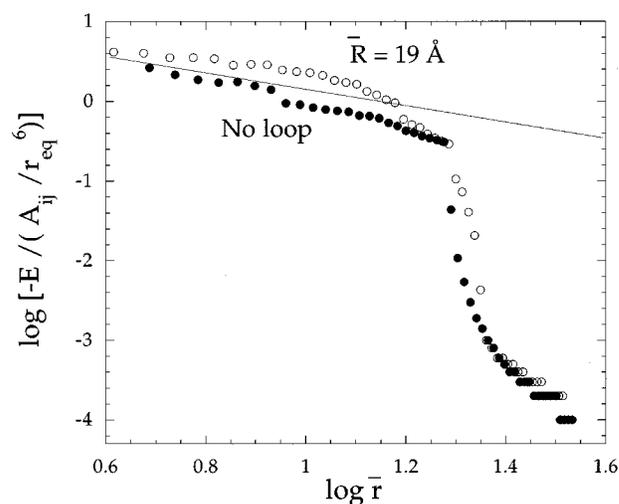

FIG. 5. Another representation of the results displayed in Fig. 4. The slope of the upper line drawn through the data is equal to -1, showing that the energy of the most probable configurations decreases linearly with the separation $\bar{r}$ between the centers of the rods. Thus the potential of mean force between the rods obeys a relationship of the form $E(\bar{r}) \sim -(1/\bar{r})^p$ with $p = 1$ in the range $\bar{r} \lesssim 20$ Å, approximately, although the individual sites interact via a 6–12 LJ potential.

ergy curves up to about 20 Å. This indicates that although an inverse sixth power (LJ) relation is used for the attractive potential between all the individual interaction sites embedded in the rods, the potential of mean force between the rods as a whole obeys a $1/\bar{r}$ relation, over a wide range of $\bar{r}$ values.

Finally, in Fig. 6, the energy of interaction is plotted as a function of the angle α between the rods. In parallel with Figs. 4 and 5, the ordinate displays the total potential of
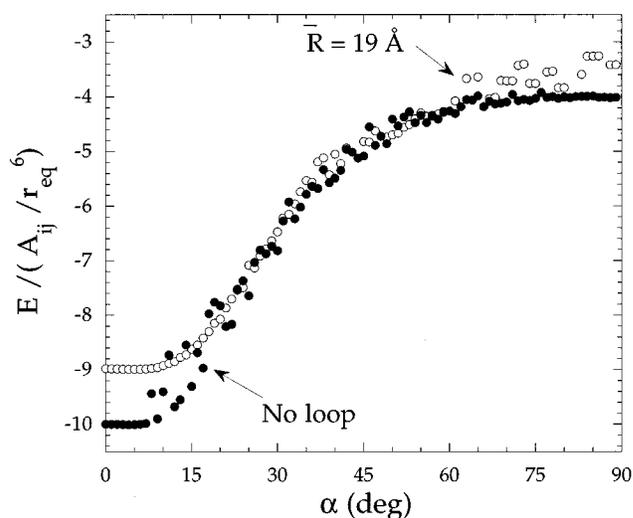
FIG. 6. Energies of the lowest energy states of the pair of rods shown in Figs. 1(a) and 1(b), as a function of the angle α between their axes. In parallel with Figs. 4 and 5, the filled and open circles refer to the most probable states of disconnected and connected pairs of rods, respectively. As in the two figures, the ordinate displays the total potential of mean force between the rods at the lowest energy configuration found for each fixed value of the abscissa. Again, the two curves exhibit a similar dependence on α.

FIG. 7. Schematic representation of two α helices M_1 and M_2 connected by a flexible spacer. The angles ϑ and ψ define the orientation of $\mathbf{R}$ with respect to the frame $O_1 x_1 y_1 z_1$. $\mathbf{R}$ is the separation vector between the centroids of the helic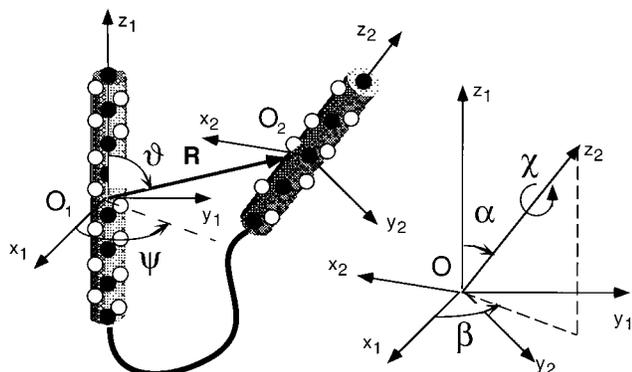es. The orientation of the second helix is defined in terms of the three Euler angles α, β, and χ, which are illustrated on the right panel. In the case of an *idealized* pair of α-helices, in which the heterogeneity of the interaction sites is neglected and thereby the cylindrical symmetry assumption holds, the variables ψ and χ are eliminated.
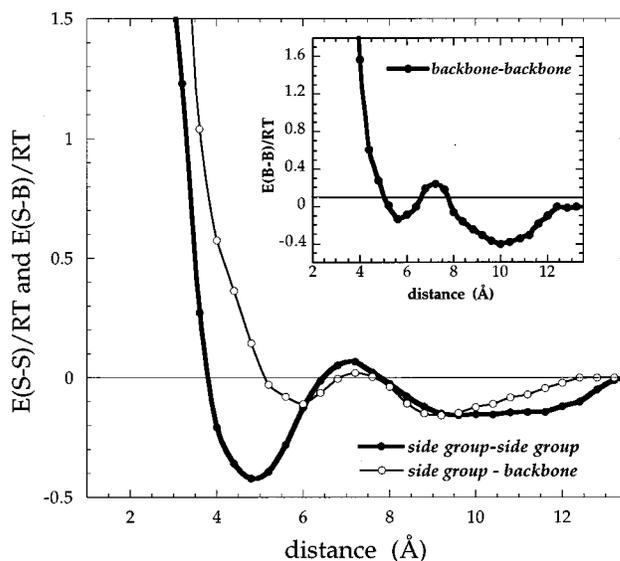

FIG. 8. Potentials of mean force between sidechain (S) and backbone (B) sites of amino acids in globular proteins, averaged over all types of residues. $E(S-S)$, $E(S-B)$, and $E(B-B)$ refer to sidechain–sidechain, sidechain–backbone, and backbone–backbone interaction energies, respectively, extracted from known protein structures using all pairs of sites separated by five or more virtual bonds (Ref. 14). Here, these potentials account for the interactions between the homogeneous sites affixed to the *idealized* pair of α-helices.

mean force between the rods at the lowest energy configuration found for each fixed value of the abscissa. The filled and empty circles show results for disconnected and connected pair of rods, respectively. Again, the two curves exhibit a similar dependence on α. A shallow minimum is found around $\alpha = 5°$ in both cases. The similarity of the curves for connected rods to those for disconnected rods in Figs. 4–6 indicates that the interactions between the sites embedded in the two rods dominate the observed qualitative behavior, rather than the elastic characteristics of the loop. The constraints imposed by the loop have a secondary effect, in the sense that they affect the equilibrium energies without significantly altering the dependence of the lowest energy states on the geometric variables R and α.

B. Extension to protein helices

In this section, M_1 and M_2 are modeled so as to mimic the behavior of α-helices in proteins. Each residue is represented by two interaction sites, one on the backbone (B) and the other on the sidechain (S). These sites may be conveniently identified with the α and β carbons of amino acids. The number of unknowns required to characterize a configuration $\{\Phi\}$ of the two bodies is equal to four provided that the helices are cylindrically symmetric, i.e., the interaction sites are all equivalent and uniformly distributed along the helices. These four variables may conveniently be taken as those described in Fig. 1(c). One has to add to this set two more variables, say the torsion angles about the principal axes of M_1 and M_2, if the helices are not cylindrically symmetric. This choice of variables is not unique, certainly. Another set of variables, also suitable for calculation, is illustrated in Fig. 7. This consists of the polar (ϑ) and azimuthal (ψ) angles of the vector $\mathbf{R}$ connecting the centroids of the two helices, expressed in the system $O x_1 y_1 z_1$, the magnitude of R, and three Euler angles α, β, and χ describing the orientation of

M_2 with respect to $O x_1 y_1 z_1$ (see Appendix B). These six variables $\{R, \vartheta, \psi, \alpha, \beta, \chi\}$ fully describe a given configuration $\{\Phi\}$ of the pair of helices.

1. An idealization: Two disconnected α helices composed of homogeneous interaction sites

Here, the local geometric characteristics and the energy functions and parameters associated with the $B-B$, $B-S$, and $S-S$ interactions are taken to be those of an average amino acid, derived from known protein structures (Fig. 8). This simplification, referred to as an *idealized* pair of α helices, permits us to eliminate two geometric variables, say ψ and χ, in the search for the most probable states. Residue-specific characteristics will be taken into consideration in the next subsection, when analyzing a real pair of α helices. Details about the model and parameters are presented in Appendix B.

The results obtained for helices of 30 residues are shown in Fig. 9. The most probable states are selected from a large number (of the order of 10^5) of local minima located with a MC–Metropolis procedure coupled with a simplex[11] energy minimization algorithm, which closely conforms with a previously developed conformational search technique.[8,9] In parallel with the basic approach outlined above, the curves display the energies of the most probable states as a function of a given geometric variable, R and α in parts (a) and (b) of Fig. 9, respectively. $E(R)$ in part (a) represents, for example, the energy $E\{R\} \equiv E\{R, \vartheta, \alpha, \beta\}$ minimized with respect to $\{\vartheta, \alpha, \beta\}$ within the full range accessible to these angles, and with respect to R within each successive $R \pm 0.25$ Å. Likewise, $E(\alpha)$ represents the energy minimized with respect to
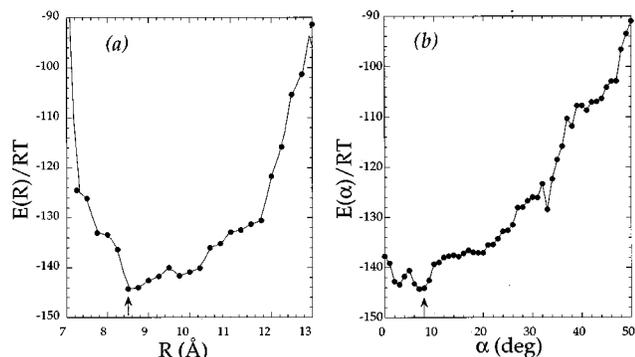
FIG. 9. Energies of the most probable configurations of a pair of idealized helices of $n = 30$ residues as a function of (a) the separation R between the centroids of the helices, and (b) the angle between the principal axes of the helices. The ordinate values represent the loci of the local minima located by the simplex algorithm for each fixed value of the geometric variable R in (a) or α in (b). A global minimum is observed at $R = 8.5$ Å and $\alpha = 7.5°$.

FIG. 10. Loci of the lowest energy configurations as a function of R and α, for the pair of idealized helices of $n = 30$ residues. The energy values are obtained after minimization of each $E\{\Phi\} \equiv E\{R, \vartheta, \alpha, \beta\}$ value with respect to β and ϑ within square bins of size $\Delta R = 0.25$ Å and $\Delta \alpha = 1°$ in the range $0 \leq R \leq 20$ Å and $0 \leq \alpha < 180°$. The surface is smoothed out over segments of 40 grids, which permits a clearer visualization of a global minimum. The general roughness can be estimated from Fig. 9.

the remaining variables. The lowest energy state ($\sim -140\,RT$) is indicated by an arrow in each case. Both curves reveal that the optimal distance R between the centers of helices and the angle α between their principal axes are uniquely determined by the present analysis. These minima appear furthermore to be easily accessible in view of the shallow shape of the energy profile in their neighborhood.

Results of similar calculations for energy as a function of ϑ and β, not shown for brevity, exhibit multiple minima of comparable depths. The most probable polar angle is found to be confined to the region $60° \leq \vartheta \leq 120°$. $E(\beta)$, on the other hand, displays three minima differing by $\leq 0.5RT$ in depth with the most favorable value of $\beta \approx 0°$. Inasmuch as the two helices are almost parallel to each other in the most stable configuration, the exact choice of the variable β is relatively unimportant.

Examination of two-dimensional energy surfaces permits one to distinguish coupling between the different degrees of freedom. Loci of approximately 4×10^4 minima, $E(R, \alpha)$, are displayed as a function of R and α in Fig. 10. These are obtained after minimization of each $E\{\Phi\}$ $\equiv E\{R, \vartheta, \alpha, \beta\}$ value with respect to β and ϑ within square bins of size $\Delta R = 0.25$ Å and $\Delta \alpha = 1°$ centered around (R, α) values in the range $0 \leq R \leq 20$ Å and $0 \leq \alpha < 180°$. The surface is drawn by smoothing out the energies over 40 grids along the horizontal axes. This permits a clearer visualization of the global minimum, although the depth of the well is somewhat reduced.

Calculations were repeated for pairs of helices of different lengths. A distinctive feature in shorter helices is that the low energy ($\leq 3RT$ above the most stable state) configurations are obtainable over a wide range ($0 \leq \alpha \leq 50°$) of tilting angle α between helices, although in the case of longer helices configurations of comparable stability were confined to the range $2 \leq \alpha \leq 9°$, only. The decrease in the angle between the principal axes with increasing asymmetry or length of the rigid building blocks emerges as a general property, irrespective of the detailed structural and energetic characteristics.
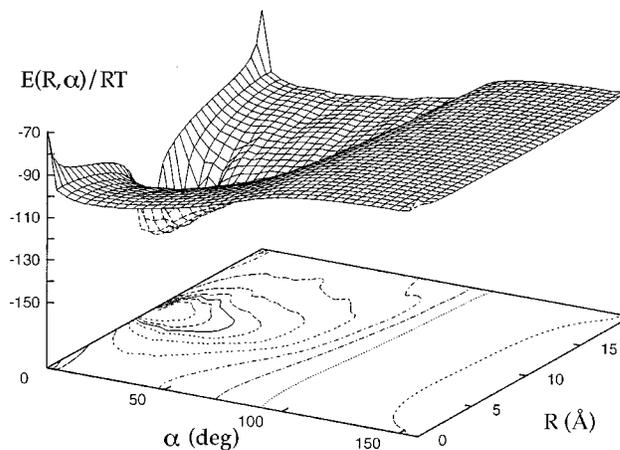
The following two points observed for the pair of rods may provide guidelines for estimating the lowest energy configuration of a pair of helices:

(i) For two identical rods, the angle θ between the vector $\bar{\mathbf{r}}$ and the rod axis was 90°. In the present section the corresponding angle is found to be $\vartheta = 90° \pm 12°$. If one accepts the value, $\theta = 90°$, obtained for the pair of rods as a reference, the deviations of ϑ from 90° may be regarded as perturbations resulting from (i) inhomogeneities at the ends of the helices, (ii) the three-dimensional geometry of the helical structure as opposed to linear structure of the rods, and (iii) the shape of the non-bonded interaction potential displayed in Fig. 8.

(ii) The distance between the centers of the helices is found to be about 8.5 Å. This value approximates the weighted average of the two minima of the potential energy curves displayed in Fig. 8, and may be viewed as an equilibrium separation, r_{eq}, between collective sites representative of α-helical turns. Each helical turn may thus be viewed as an entity equivalent to the interaction sites of the rods treated in the previous section, such that $l = 5.41$ Å and the parameter $\eta \equiv r_{eq}/l$ becomes 1.6, approximately. The corresponding α values for the presently investigated pairs of helices comprising eight turns, is estimated upon extrapolation from Fig. 3 to be 6°. This may be compared to the value $\alpha = 7.5°$ at the lowest energy computed in the present section.

2. A protein consisting of two heterogeneous α-helices connected by a loop: ROP monomer

ROP is a bacterial protein involved in regulating DNA replication.[12] This is a dimeric four-helix bundle, the monomers being identical but oppositely oriented helical

hairpins[10] of 63 residues, each. Thus each monomer may be modeled as two rigid α helices M_1 and M_2 comprising m and n sites, respectively, and connected by a loop. Three essential features differing from the above idealized pair of helices are considered in this case:

(i) The interaction sites are heterogeneous. Sidechain–sidechain ($S_i - S_j$), and sidechain–backbone ($S_i - B_j$) interaction potentials $E_{AB}(S_i - S_j)$ and $E_A(S_i - B_j)$, characteristic of each particular type (A, B, etc.) of residue are used, as opposed to the homogeneous potentials of Fig. 8. These potentials were recently extracted from 302 known protein structures.[13,14]

(ii) As a natural consequence of the heterogeneity and specificity of the interaction sites, the cylindrical symmetry approximation is no longer valid, and the angles ψ and χ are included in the analysis. These angles are expected to be related to the hydrophobicity moments of the helices,[15] inasmuch as the helices have a tendency to assume torsional angles that minimize the solvent exposure of their hydrophobic residues. Accordingly, the helix-embedded coordinate systems $O_1 x_1 y_1 z_1$ and $O_2 x_2 y_2 z_2$ displayed in Fig. 7 are defined with reference to the hydrophobicity moment vectors μ_1 and μ_2 of the helices in addition to their principal axes. The y_1 axis is chosen in the plane spanned by $\mathbf{z}_1$ and μ_1. The x_1 axis completes a right-handed coordinate system. Its angular deviation from the projection of the z_2 axis on the $O_1 x_1 y_1$ plane defines the angle ψ. The same recipe is adopted for defining the axes y_2 and x_2 of the system $O_2 x_2 y_2 z_2$. The rotation χ is accounted for by the angle between the hydrophobicity moments of the respective helices. In particular, $\chi = 180°$ refers to the face-to-face placement of hydrophobic moments μ_1 and μ_2, i.e., the optimal configuration from the point of view of burial of hydrophobic residues at the interface. The other extreme case ($\chi = -180°$) corresponds to the full exposure of hydrophobic surface to solvent.

(iii) The connectivity of the helices, and the constraints imposed by the finite size and specific energetics of the residues in the loop region are rigorously considered here. Let the flexible spacer comprise s virtual bonds. The bond lengths and angles in the spacer are held fixed at values characteristic of the particular residues in the native primary structure (Table I). The flexibility of the loop is ensured by the torsional mobility of the virtual bonds of the loop. The corresponding dihedral angles are indicated as φ_i, with $1 \leq i \leq s$. The overall configurational potential $E\{\Phi\} \equiv E\{R, \theta, \psi, \alpha, \beta, \chi\}$ becomes

$$E\{\Phi\} = \sum_i \sum_j [E(S_i - S_j) + E(S_i - B_j) + E(B_i - S_j)$$

$$+ E(B_i - B_j)] + \sum_{k=1}^{s} E(\varphi_k) + \sum_{k=1}^{s+1} \Delta E(\varphi_{k-1}, \varphi_k).$$

$$(2)$$

Here, the subscripts A, B, etc., are omitted for brevity, although each term depends on the specific type of interacting residue i or j. The first summation in Eq. (2) is performed

TABLE I. Backbone geometry of ROP monomer A in the virtual bond approximation.[a]

i	φ_i	Θ_i	i	φ_i	Θ_i
4	-126.5	89.3	30	53.8	89.7
5	-131.6	89.0	31	-21.1	84.1
6	-126.2	90.4	32	2.72	75.6
7	-130.0	88.1	33	-137.3	87.9
8	-132.0	88.2	34	-130.6	87.1
9	-131.0	88.5	35	-129.4	90.4
10	-130.4	90.1	36	-127.9	90.9
11	-128.9	87.5	37	-133.6	89.5
12	-130.6	91.1	38	-129.5	85.8
13	-130.3	89.6	39	-125.2	91.5
14	-129.3	86.9	40	-133.3	91.2
15	-132.1	89.7	41	-126.7	87.4
16	-126.8	90.8	42	-129.9	90.4
17	-135.2	87.1	43	-129.7	90.8
18	-126.8	85.7	44	-131.4	90.7
19	-125.9	91.6	45	-128.5	89.2
20	-129.1	89.6	46	-129.4	90.3
21	-129.8	89.4	47	-125.4	89.8
22	-133.4	87.8	48	-133.5	86.6
23	-121.6	91.2	49	-129.4	87.9
24	-132.1	88.6	50	-133.5	89.4
25	-128.2	89.0	51	-126.9	89.8
26	-132.1	90.1	52	-132.8	88.2
27	-126.6	91.2	53	-127.1	88.6
28	-125.7	89.8	54	-132.7	86.3
29	-134.0	88.0	55	-138.4	79.9

[a]The virtual bonds of the monomer are indexed from 2 to 63, in conformity with the residue indices in the PDB coordinates; the terminal bonds, which are not determined by x ray, are not present in the table. The ith bond connects the $(i+1)$st and ith alpha carbons. The dihedral angles refer to torsions relative to the *trans* conformer, such that $\varphi = 0°$, $-120°$ and $+120°$ for the *trans*, *gauche$-$* and *gauche$+$* states. Bond bending angle Θ_i refers to the angle between the virtual bonds i and $i+1$. The virtual bonds have torsional angles of $129° \pm 8°$ and bending angles of $88.5° \pm 3.5°$, which are typical of α-helices in the virtual bond approximation, except for the central bonds 29–31 which fold into a tight turn.

over all sites $1 \leq i \leq m$ embedded in M_1. The second summation of the first sum is performed over all sites $1 \leq j \leq s + n$ on the flexible spacer and on M_2, provided that the sites i and j are separated by at least five bonds along the backbone. The last two summations are performed over the bonds of the loop region, using residue-specific short-range conformational potentials.[16] φ_0 and φ_{s+1} are adopted for the dihedral angles of the bonds flanking the spacer, as the torsional states of these two bonds also affect the overall potential $E\{\Phi\}$. In general, the contribution of the dihedral angle torsions (last two summations) in Eq. (2) is negligibly small compared to that of the nonbonded interactions (first double summation).

The lowest energy configurations among the 60 000 local minima determined by the simplex algorithm are presented in Table II. Only those configurations within $\leq 8RT$ of the lowest energy configuration are listed in the table. These are classified in order of increasing energy or decreasing stability. The dihedral angles $\varphi_1 - \varphi_7$ of the fragment Asn27–Gln34 are listed in columns 2–9. The succeeding six columns list the variables $\{R, \vartheta, \psi, \alpha, \beta, \chi\}$ characterizing the relative position of the two helices. The terminal column is the energy of the particular configuration. The structural

TABLE II. Characteristics of the lowest energy configurations of the fragment Asn27–Gln34 in ROP monomer.[a]

{Φ}	φ_1	φ_2	φ_3	φ_4	φ_5	φ_6	φ_7	$R(\text{Å})$	ϑ	ψ	α	β	χ	$E\{\Phi\}/RT$
1	−134.1	−132.5	49.2	−34.9	2.1	−125.0	−126.2	8.95	81.2	151.0	19.5	42.9	100.8	−12.0
2	−127.7	−155.8	26.5	−8.5	35.7	−130.6	−121.5	8.46	79.3	137.1	19.5	42.8	124.8	−12.0
3	−111.0	−164.2	53.3	−4.8	15.6	−122.8	−131.2	8.58	83.3	138.8	19.3	38.1	119.5	−12.0
4	−130.3	−128.2	45.3	−29.8	16.9	−134.1	−118.7	8.79	74.8	150.8	18.0	48.5	104.0	−11.7
5	−129.7	−129.3	54.9	−33.6	4.4	−130.1	−118.4	8.90	80.8	149.2	15.1	44.0	99.2	−11.2
6	−130.6	−131.7	41.3	−26.3	21.7	−134.0	−113.4	8.80	77.9	143.5	15.7	45.3	111.7	−11.1
7	−134.3	−132.9	46.1	−34.3	6.7	−125.8	−122.9	8.68	82.1	146.8	17.9	43.4	105.0	−10.6
8	−124.1	−151.9	46.9	−17.4	15.2	−120.7	−121.5	8.65	87.2	142.9	18.8	39.5	120.8	−9.1
9	−124.4	−133.8	48.9	−22.1	29.8	−128.6	−78.0	9.02	86.3	120.0	1.6	−163.8	132.4	−8.7
10	−108.5	−161.4	51.2	12.1	25.3	−145.4	−124.8	8.99	74.9	137.9	17.6	43.8	114.0	−8.5
11	−121.8	−152.3	30.0	−11.7	32.9	−129.3	−126.8	9.00	75.1	140.7	20.0	41.9	117.9	−8.4
12	−123.3	−138.4	34.2	−39.3	27.4	−111.6	−123.9	8.36	79.2	142.2	17.1	42.9	118.7	−7.7
13	−148.6	−133.5	31.3	−28.2	12.2	−132.9	−122.1	8.42	84.6	138.6	20.1	40.4	116.2	−7.1
14	−133.9	−125.1	49.3	−37.9	25.6	−114.3	−66.1	8.97	90.0	115.5	4.9	−149.7	141.4	−7.0
15	−122.7	−155.8	−11.7	−14.6	74.1	−131.7	−120.1	9.84	66.6	124.4	12.7	33.8	130.6	−5.4
16	−115.4	−133.1	53.4	−20.8	32.2	−125.1	−77.9	9.16	83.7	127.3	1.1	−129.3	130.7	−4.9
17	−122.0	−154.1	−10.1	−12.3	78.2	−134.2	−111.3	9.72	66.9	124.4	10.9	37.0	132.9	−4.6
18	−115.3	−145.5	14.3	−20.3	56.0	−129.4	−119.0	9.33	67.2	134.3	14.0	42.9	123.3	−4.2
19	−138.7	−133.8	23.1	−32.1	29.4	−126.0	−118.6	8.87	78.9	135.7	18.1	37.7	123.1	−4.0
[b]	−122.1	−134.0	53.8	−21.1	2.7	−137.4	−118.6	8.95	74.2	151.5	18.9	49.6	94.0	0.0[c]

[a]All angles are in degrees, R is in Ångstroms.
[b]Native structure.
[c]Native energy taken as reference point.

characteristics of the native structure are presented in the last row. We note that the most stable states reached by the simplex algorithm have energies lower than that of the native state.

Specification of the variables $\{R, \vartheta, \psi, \alpha, \beta, \chi\}$ is sufficient for defining a given configuration $\{\Phi\}$ of the pair of helices, and there may be several sets of dihedral angles $\{\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \varphi_6, \varphi_7\}$ compatible with a given configuration $\{\Phi\} = \{R, \vartheta, \psi, \alpha, \beta, \chi\}$. In a sense the latter may be viewed as a macrostate achieved by several microstates. We will concentrate on the most stable macrostates, although the results in Table II indicate that even the microstates described by the set $\{\varphi_1, \varphi_2, \varphi_3, \varphi_4, \varphi_5, \varphi_6, \varphi_7\}$ exhibit some well defined characteristic; whereas two stable macrostates, shortly referred to as $\{\Phi\}_1$ and $\{\Phi\}_2$, are distinguishable: These are characterized by $\{\Phi\}_1 = \{8.7 \pm 0.3$ Å, $81°\pm 7°$, $144 \pm 7°$, $17.6 \pm 2.5°$, $44 \pm 6°$, $112 \pm 13°\}$ and $\{\Phi\}_2 = \{9.0 \pm 0.03$ Å, $88 \pm 2°$, $118 \pm 3°$, $3.3 \pm 1.7°$, $-157 \pm 7°$, $137 \pm 5°\}$. The former is energetically more favorable than the latter by an energy difference of about 3 RT, and is in close agreement with the native structure. Furthermore, the fact that 16 out of the 19 lowest energy microstates belong to macrostate $\{\Phi\}_1$ suggests that the latter is also favored by kinetic effects. In particular, we note that the lowest energy microstate shown in the first row of the table, exhibits structural characteristics quite close to those of the native configuration: The separation $R = 8.95$ Å between the centroids of the helices is exactly reproduced, the angle α between the two principal axes and the azimuthal angle ψ agree within less than 1°. The remaining three variables θ, β, and χ differ by $\leq 7°$ compared to the corresponding values in the native protein.

A more critical analysis of the loci of the lowest energy

states may be carried out by examining the dependence of $E\{\Phi\}$ on each of the geometric variables R, ϑ, ψ, α, β, and χ. The results are displayed in Figs. 11(a)–11(f). These curves provide an estimate of the depth and width of the lowest energy states along the six variables of the multidimensional energy surface. In parallel with the curves of Figs. 4–6 and 9, the dots represent the lowest energy states obtained upon minimization of the total energy with respect to all variables other than the one shown on the abscissa. $E(R)$ represents, for example, the energy of the lowest energy state achievable at the given R, by optimizing the variables θ, ψ, α, and β over their full range, and repeating this for successive R at 0.25 Å intervals. Likewise $E(\alpha)$, $E(\vartheta)$, $E(\beta)$, $E(\psi)$, and $E(\chi)$ shown in Figs. 11(b)–11(f) are the loci of the most stable microstates obtained by confining the arguments to fixed, small intervals along the abscissa, while allowing the remaining five variables to take the energetically most favorable combination. The native value is written on each figure, for comparison with the predicted lowest energy value.

The following properties are extracted from Fig. 11. Figure 11(a) demonstrates that a unique minimum, which coincides exactly with the native value is predicted for the spatial separation R of the helices. Figure 11(b) shows that the angular inclination of $\alpha = 19.5°$ between the principal axes of the helices is preferred over the approximately parallel alignment of the helices taking place in the macrostate $\{\Phi\}_2$ not only by its lower energy, but also by the shape of the potential energy surface. The value $\theta = 74.17°$ of the native structure is located within the lowest energy well, as may be verified from Fig. 11(c). This implies that the centroids of the helices are not necessarily at the same elevation but the second helix is located slightly upwards, which is correctly
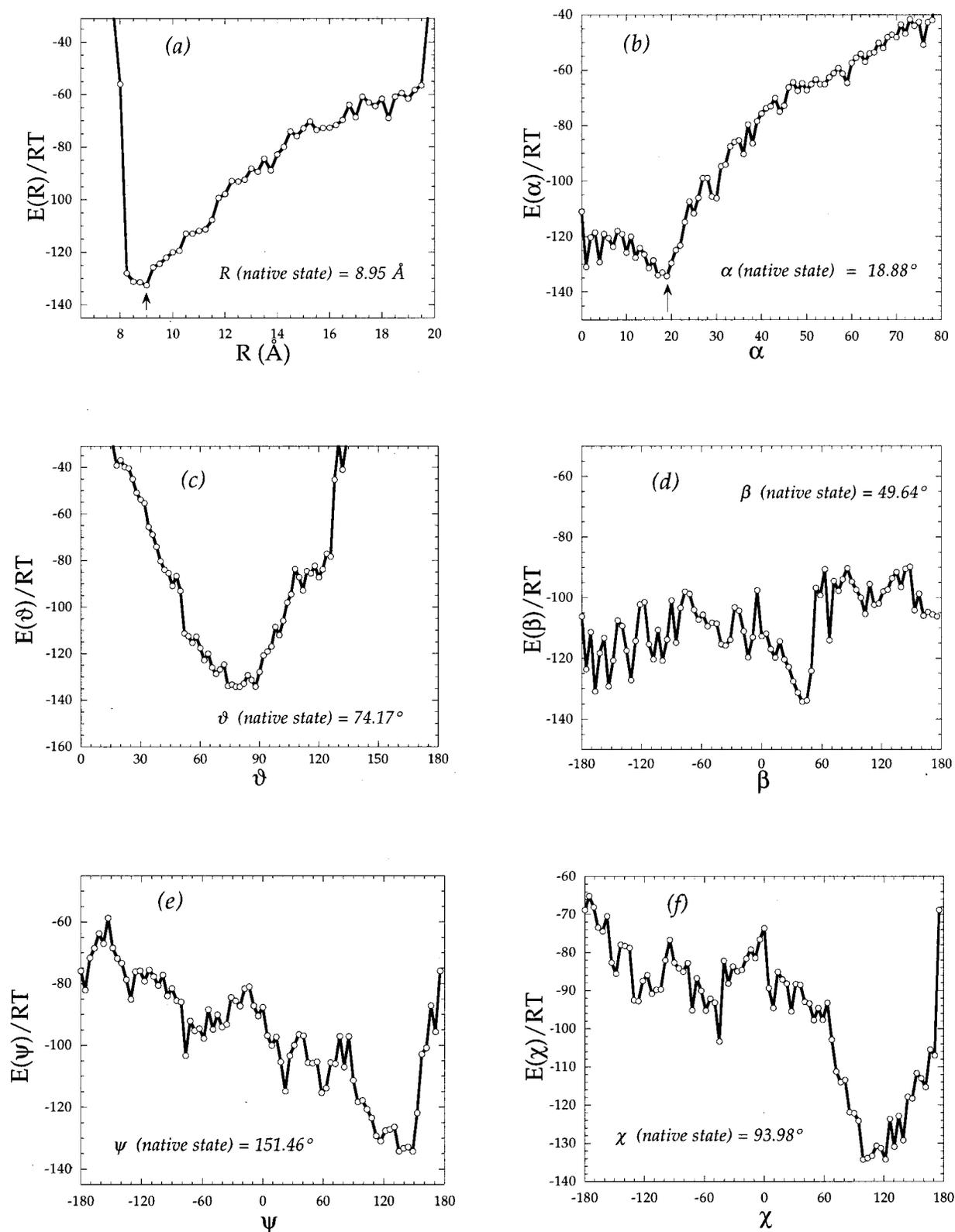
FIG. 11. Results obtained for the ROP monomer. Energies of the most probable conformations are displayed as a function of the geometric variables (a) R, (b) α, (c) ϑ, (d) β, (e) ψ, and (f) χ. The geometric data of the crystal structure are indicated in each case, demonstrating the close agreement between the native structure and the predicted lowest energy configuration. Here, an enlarged set of local minima (60 000 of them) is generated in parallel with Figs. 5–7 and 9, and the lowest energy configurations are plotted as a function of each geometric variable after minimization with respect to the other five.

predicted. As to Fig. 11(d), in parallel with the results obtained for idealized helices, β assumes quite distinct values in different macrostates. This is explained by the fact that the two helices are almost parallel to each other and therefore the choice of a given azimuthal angle has a weak effect on the overall energy. Nevertheless, the broadest and most stable minimum is in satisfactory accord with the native structure. The most stable ψ value in Fig. 11(e) is easily recognized to overlap with the native value. Finally, a deep well at $\chi = 90-125°$ is observed in Fig. 11(f). Here, negative χ values are quite unfavorable. As described above, $\chi \leqslant 0°$ refer to configurations in which the hydrophobic residues are exposed to solvent. Such configurations are disfavored by a significant energy difference (~ 50 RT). It is interesting to note that the rotation $\chi = 180°$, which should allow for maximal interaction of the hydrophobic groups is not energetically favorable here, presumably due to a steric overlap. Instead the angle $\chi \approx 105°$ is preferred. This is slightly ($\sim 10°$) distorted in favor of the association of the two hydrophobic surfaces compared to the native structure. The deviation of about $10°$ may be the result of a fine tuning upon dimerization of the ROP monomer, inasmuch as a smaller χ angle would allow for a better association of the two monomers at the interface. By the same reasoning, one can anticipate that the macrostate $\{\Phi\}_2$ which involves an even tighter interaction of the hydrophobic groups between the two helices of a given monomer, thus lacking enough hydrophobicity at the interface of the monomers, which could be unfavorable for the dimer.

Finally, the most probable energy surface $E(R,\alpha)$ obtained for ROP monomer, not shown for brevity, is found to exhibit the same characteristics as the loci of energy minima displayed in Fig. 10, i.e., a broad global minimum. Its location is slightly shifted towards a larger α value ($19.5°$). Likewise the separation between the centers of the helices is increased by 0.5 Å compared to the idealized helices, which is understandable in view of the bulkier side chains existing in ROP monomer. However, the essential features, i.e., a funnel-like distribution in favor of an entropically favorable energy minimum is again observed with the present coarse-grained approach. The general roughness of the surface can be estimated from the graphs given in Fig. 11.

III. DISCUSSION AND CONCLUSION

Here, a Monte Carlo procedure combined with a simple energy minimization algorithm is used to generate a stack of low energy conformations. Full coverage of the conformational space is accomplished, inasmuch as the loci of energy minima are obtainable as a function of the geometric variables or generalized coordinates in the form of smooth, discontinuous curves. Such a combination of MC procedure with energy minimization goes back to the original work of Li and Scheraga.[8] As pointed out in recent studies,[17,18] visiting a maximum number of different local energy minima using a minimal number of function evaluations, and creating a stack of low energy conformations within a certain energy range are essential for identifying the most probable

state in such multiple minima problem. Low energy conformations are used here to determine the effective energy profile, or the potential of mean force as a function of each of the generalized coordinates.

Exploratory calculations with simple models such as pairs of rods comprising uniformly distributed interaction sites reveal some characteristics that may have important implications insofar as real molecular structures are concerned.

First, the potential of mean force between the rods as a whole exhibits a linear dependence on the separation between rods (Fig. 5), despite the individual sites interacting via a classical 6–12 Lennard-Jones potential. This implies that the range of the effective potential between the rods is much longer than may be inferred from its individual embedded sites. And the existence of such a long-range smooth potential can explain the recognition of the lowest energy conformation of the two bodies from a relatively distant initial position. That the native supersecondary and tertiary structure is spontaneously found, in spite of the enormous number of accessible conformations, i.e., the Levinthal paradox, may be partly explained by the fact that the range of the operating potentials of mean force between secondary structural elements is sufficiently long.

Second, as the length of rods (or helices) increases, they tend to align more parallel to each other. Also, as the density of interaction sites increases they exhibit a stronger tendency to become parallel. This conforms with observations that helices of four or more turns pack optimally into elongated bundles; whereas shorter helices can pack into a number of other geometries.[19] Such criteria are pointed out to be important in designing protein structures.[20] Here, the angle α between the principal axes of the rods is found to scale as $n^{-5/3}$ with the number n of sites, the absolute value depending on the density η of sites. For α helices using $\eta = 1.6$, an estimate of the angle between two adjacent helices—in the absence of the perturbations due to specific effects—may be made directly from Fig. 3.

Calculations for the ROP monomer depart from previous simulations[5,21] in both the model and the method of calculation. Here, a recently developed coarse-grained model with energy parameters extracted from known structures has been employed.[14,16,22] Loci of energy minima are plotted with respect to one variable, after minimizing with respect to all others, which is helpful in locating the most probable state. In fact, good conformations could be built by choosing these independent average values for each variable. This confirms our previous experiences applying this approach to DNA.[23] The fact that the native conformation of the monomer is correctly reproduced, and even the detailed rotational states of the virtual bonds located in the loop connecting the α helices are satisfactorily accounted for, lends support to the adoption of the present coarse-grained model and parameters for future simulations. A major advantage of such low resolution models and empirical potentials is certainly the elimination of several degrees of freedom which might complicate the energy surface.

An interesting observation is that the most probable state is also entropically favorable, i.e., the energy surface exhibits

a funnel-like shape, in conformity with recent theoretical arguments,[24,25] provided that the loci of local minima are examined as a function of two generalized coordinates, after minimization with respect to all others. Such a sufficiently broad global minimum may indeed be a prerequisite for the correct folding of proteins, as suggested by lattice simulations.[26,27]

If folding nuclei could be predicted, then the present method could be utilized directly for the construction of the remainder of the protein, insofar as it were within computational limits. Another possible application of our approach would be to use it together with secondary structure predictions. Usually the most certain parts of these predictions are the centers of helices. These central helix parts were used here in the ROP example, and the ends of the helices next to the loop were permitted to be flexible. For larger helical proteins, sequentially adjacent helices are not always the strongest interacting pairs, but presumably the present method is sufficiently robust to be able to detect the strongest interacting pairs.

ACKNOWLEDGMENTS

APPENDIX A

Here, the method of calculation for the minimum energy configuration of two rods joined by a loop is described. The application to the case of disconnected rods or asymmetric bodies in two-dimensions is straightforward.

The rods contain m and n interaction sites, respectively. The origins of the molecule-embedded coordinate systems O_1 and O_2 are assumed to be located at the end of the rods, such that $\mathbf{R}$ defines the end-to-end separation vector of the loop between the rods. One of the rods, M_1, is assumed to be fixed in space. The x_1 axis is taken along the axis of rod M_1. The y_1 axis is chosen such that the x_1y_1 plane contains the vector $\mathbf{R}$. The z_1 axis completes a right-handed system. θ_1 is the angle between the x_1 axis and $\mathbf{R}$. The position of the second rod in space is determined uniquely by θ_1, the magnitude R, the torsion angle ϕ about $\mathbf{R}$, and the angle θ_2 between the axis of rod M_2 and $\mathbf{R}$, provided that the two rods are cylindrically symmetric.

The four equations required for the solution of the four unknowns R, θ_1, θ_2, and ϕ are obtained by minimizing the Helmholtz free energy ΔA of the system with respect to four variables

$$\frac{\partial \Delta A}{\partial \xi} = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{\partial \Delta A}{\partial r_{ij}^2} \frac{\partial r_{ij}^2}{\partial \xi} = 0, \quad \xi = R, \theta_1, \theta_2, \phi, \quad \text{(A1)}$$

where r_{ij} is the magnitude of the separation vector $\mathbf{r}_{ij} = \mathbf{r}_j - \mathbf{r}_i$ between the ith and jth interaction sites belonging to M_1 and M_2, $1 \leq i \leq m$ and $1 \leq j \leq n$. The Helmholtz free energy is taken to be the sum of the potentials of mean force

associated with the interaction of the m sites in M_1 with the n sites in M_2, and the conformational energy of the loop. By assuming a Gaussian end-to-end distribution for the loop and a Mie type potential for the interaction between all sites, the total free energy change may be written as

$$\Delta A = \frac{1}{2} \sum_i \sum_j \left[-\frac{A_{ij}}{r_{ij}^p} + \frac{B_{ij}}{r_{ij}^q} \right] + \frac{3kT}{2\langle (R-\bar{R})^2 \rangle_0} (R-\bar{R})^2 \quad \text{(A2)}$$

Here, A_{ij} and B_{ij} are the parameters associated with the attractive and repulsive parts of the potential, respectively, p and q are the exponents. $\langle (R-\bar{R})^2 \rangle_0$ is the mean square difference between R and its free state value $\bar{R}$, i.e., in the absence of the two rods. The Mie potential reduces to the Lennard-Jones potential when $p=6$ and $q=12$. The equilibrium separation between two independent interaction sites is $r_{\text{eq}} = (qB_{ij}/pA_{ij})^{1/(q-p)}$. The vector between sites i and j is conveniently expressed as

$$\mathbf{r}_{ij} = \mathbf{R} + \mathbf{T}_1 \mathbf{T}_2 \mathbf{b}_j - \mathbf{a}_i = \mathbf{R} + \mathbf{b}_j' - \mathbf{a}_i, \quad \text{(A3)}$$

where $\mathbf{a}_i$ and $\mathbf{b}_j$ are the fixed position vectors of the respective sites i and j, with respect to the molecule-embedded coordinate systems affixed to O_1 and O_2. $\mathbf{b}_j'$ is the representation of $\mathbf{b}_j$ in the coordinate system $Ox_1y_1z_1$. $\mathbf{T}_1$ and $\mathbf{T}_2$ are the transformation matrices given by

$$\mathbf{T}_1 = \begin{bmatrix} \cos\theta_1 & \sin\theta_1 & 0 \\ \sin\theta_1 & -\cos\theta_1 & 0 \\ 0 & 0 & 1 \end{bmatrix};$$

$$\mathbf{T}_2 = \begin{bmatrix} \cos\theta_2 & \sin\theta_2 & 0 \\ \cos\phi_2 \sin\theta_2 & -\cos\phi_2 \cos\theta_2 & \sin\phi_2 \\ \sin\phi_2 \sin\theta_2 & -\sin\phi_2 \cos\theta_2 & -\cos\phi_2 \end{bmatrix}. \quad \text{(A4)}$$

In calculations, it proves convenient to express the square magnitude r_{ij}^2 as

$$r_{ij}^2 = R^2 + 2\mathbf{R} \cdot \mathbf{U}_{ij} + \mathbf{U}_{ij} \cdot \mathbf{U}_{ij} \quad \text{(A5)}$$

with the vectors $\mathbf{R}$ and $\mathbf{U}_{ij}$ given by

$$\mathbf{R} = R \begin{bmatrix} \cos\theta_1 \\ \sin\theta_1 \\ 0 \end{bmatrix} = R\mathbf{u}_R, \quad \text{(A6)}$$

$$\mathbf{U}_{ij} = \mathbf{T}_1 \mathbf{T}_2 \mathbf{b}_j - \mathbf{a}_i = \mathbf{b}_j' - \mathbf{a}_i. \quad \text{(A7)}$$

The four scalar equations of equilibrium given by Eqs. (A1) are written explicitly for the case $p=6$ and $q=12$ as

$$R^3 + \alpha_{k1}R^2 + \alpha_{k2}R + \alpha_{k3} = 0, \quad 1 \leq k \leq 4, \quad \text{(A8)}$$

where

$$\alpha_{11} \equiv 3 \sum_{i,j} g_{ij} \mathbf{U}_{ij} \cdot \mathbf{u}_R \bigg/ \sum_{i,j} g_{ij}, \quad \text{(A9)}$$

$$\alpha_{12} = 3 \sum_{i,j} g_{ij} \left[\mathbf{U}_{ij} \cdot \mathbf{U}_{ij} - \mathbf{U}_{ij} \cdot \mathbf{u}_R + 2(\mathbf{U}_{ij} \cdot \mathbf{u}_R)^2 \right.$$

$$\alpha_{13} = \sum_{i,j} g_{ij} \mathbf{U}_{ij} \cdot \mathbf{u}_R (\mathbf{U}_{ij} \cdot \mathbf{U}_{ij} - \mathbf{U}_{ij} \cdot \mathbf{u}_R) \Bigg/ \sum_{i,j} g_{ij}, \quad \text{(A11)}$$

$$\left. + \frac{kT}{2\langle (R-\bar{R})^2 \rangle_0} (R-\bar{R}) \right] \Bigg/ \sum_{i,j} g_{ij}, \quad \text{(A10)}$$

$$\alpha_{21} = \frac{\Sigma_{i,j} g_{ij} [(\mathbf{U}_{ij} \cdot \boldsymbol{\epsilon} \mathbf{b}_j' + 2\mathbf{U}_{ij} \cdot \mathbf{u}_R (\mathbf{u}_R \cdot \boldsymbol{\epsilon} \mathbf{b}_j' + \mathbf{U}_{ij} \cdot \boldsymbol{\epsilon} \mathbf{u}_R)]}{\Sigma_{i,j} g_{ij} (\mathbf{u}_R \cdot \boldsymbol{\epsilon} \mathbf{b}_j' + \mathbf{U}_{ij} \cdot \boldsymbol{\epsilon} \mathbf{u}_R)}, \quad \text{(A12)}$$

$$\alpha_{22} = \frac{\Sigma_{i,j} g_{ij} [(\mathbf{U}_{ij} \cdot \mathbf{U}_{ij} - r_{eq}^2)(\mathbf{u}_R \cdot \boldsymbol{\epsilon} \mathbf{b}_j' + \mathbf{U}_{ij} \cdot \boldsymbol{\epsilon} \mathbf{u}_R) + 2(\mathbf{U}_{ij} \cdot \mathbf{u}_R)(\mathbf{U}_{ij} \cdot \boldsymbol{\epsilon} \mathbf{b}_j')]}{\Sigma_{i,j} g_{ij} (\mathbf{u}_R \cdot \boldsymbol{\epsilon} \mathbf{b}_j' + \mathbf{U}_{ij} \cdot \boldsymbol{\epsilon} \mathbf{u}_R)}, \quad \text{(A13)}$$

$$\alpha_{23} = \sum_{i,j} g_{ij} [(\mathbf{U}_{ij} \cdot \mathbf{U}_{ij} - r_{eq}^2) \mathbf{U}_{ij} \cdot \boldsymbol{\epsilon} \mathbf{b}_j'] \Bigg/ \sum_{i,j} g_{ij} (\mathbf{u}_R \cdot \boldsymbol{\epsilon} \mathbf{b}_j' + \mathbf{U}_{ij} \cdot \boldsymbol{\epsilon} \mathbf{u}_R), \quad \text{(A14)}$$

$$\alpha_{31} = \sum_{i,j} g_{ij} \left[(\mathbf{U}_{ij} \cdot \mathbf{T}_1 \mathbf{T}_2 \boldsymbol{\epsilon} \mathbf{b}_j + 2(\mathbf{U}_{ij} \cdot \mathbf{u}_R) \mathbf{T}_1 \mathbf{T}_2 \boldsymbol{\epsilon} \mathbf{b}_j \cdot \mathbf{u}_R] \right] \Bigg/ \sum_{i,j} g_{ij} (\mathbf{T}_1 \mathbf{T}_2 \boldsymbol{\epsilon} \mathbf{b}_j \cdot \mathbf{u}_R), \quad \text{(A15)}$$

$$\alpha_{32} = \frac{\Sigma_{i,j} g_{ij} [(\mathbf{U}_{ij} \cdot \mathbf{U}_{ij} - r_{eq}^2) \mathbf{T}_1 \mathbf{T}_2 \boldsymbol{\epsilon} \mathbf{b}_j \cdot \mathbf{u}_R + 2(\mathbf{U}_{ij} \cdot \mathbf{u}_R) \mathbf{U}_{ij} \cdot \mathbf{T}_1 \mathbf{T}_2 \boldsymbol{\epsilon} \mathbf{b}_j]}{\Sigma_{i,j} g_{ij} (\mathbf{T}_1 \mathbf{T}_2 \boldsymbol{\epsilon} \mathbf{b}_j \cdot \mathbf{u}_R)}, \quad \text{(A16)}$$

$$\alpha_{33} = \sum_{i,j} g_{ij} [(\mathbf{U}_{ij} \cdot \mathbf{U}_{ij} - r_{eq}^2) \mathbf{U}_{ij} \cdot \mathbf{T}_1 \mathbf{T}_2 \boldsymbol{\epsilon} \mathbf{b}_j] \Bigg/ \sum_{i,j} g_{ij} (\mathbf{T}_1 \mathbf{T}_2 \boldsymbol{\epsilon} \mathbf{b}_j \cdot \mathbf{u}_R), \quad \text{(A17)}$$

$$\alpha_{41} = \frac{\Sigma_{i,j} g_{ij} [\mathbf{U}_{ij} \cdot \mathbf{T}_1 \mathbf{T}_2 \Gamma \mathbf{b}_j + 2(\mathbf{U}_{ij} \cdot \mathbf{u}_R) \mathbf{T}_1 \mathbf{T}_2 \Gamma \mathbf{b}_j \cdot \mathbf{u}_R]}{\Sigma_{i,j} g_{ij} (\mathbf{T}_1 \mathbf{T}_2 \Gamma \mathbf{b}_j \cdot \mathbf{u}_R)}, \quad \text{(A18)}$$

$$\alpha_{42} = \frac{\Sigma_{i,j} g_{ij} [(\mathbf{U}_{ij} \cdot \mathbf{U}_{ij} - r_{eq}^2) \mathbf{T}_2 \Gamma \mathbf{b}_j \cdot \mathbf{u}_R + 2(\mathbf{U}_{ij} \cdot \mathbf{u}_R) \mathbf{U}_{ij} \cdot \mathbf{T}_1 \mathbf{T}_2 \Gamma \mathbf{b}_j \mathbf{T}_1]}{\Sigma_{i,j} g_{ij} (\mathbf{T}_1 \mathbf{T}_2 \Gamma \mathbf{b}_j \cdot \mathbf{u}_R)}, \quad \text{(A19)}$$

$$\alpha_{43} = \sum_{i,j} g_{ij} [(\mathbf{U}_{ij} \cdot \mathbf{U}_{ij} - r_{eq}^2) \mathbf{U}_{ij} \cdot \mathbf{T}_1 \mathbf{T}_2 \Gamma \mathbf{b}_j] \Bigg/ \sum_{i,j} g_{ij} (\mathbf{T}_1 \mathbf{T}_2 \Gamma \mathbf{b}_j \cdot \mathbf{u}_R), \quad \text{(A20)}$$

with g_{ij}, $\boldsymbol{\epsilon}$, and Γ expressed as

$$g_{ij} = (A_{ij}/r_{ij}^{14})(r_{ij}^4 + r_{eq}^2 r_{ij}^2 + r_{eq}^4), \quad \text{(A21)}$$

$$\boldsymbol{\epsilon} = \begin{bmatrix} 0 & -1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix};$$

$$\Gamma = \begin{bmatrix} 0 & 0 & \sin\theta_2 \\ 0 & 0 & -\cos\theta_2 \\ -\sin\theta_2 & \cos\theta_2 & 0 \end{bmatrix}. \quad \text{(A22)}$$

The coefficients of Eqs. (A8) are proportional to g_{ij} which appears in the derivative $\partial \Delta A / \partial r_{ij}^2 = g_{ij}(r_{ij}^2 - r_{eq}^2)$ $\cdot g_{ij}$ is positive for all values of R, as may be seen from Eq. (A21); whereas the term $(r_{ij}^2 - r_{eq}^2)$ may be either positive or negative, and therefore is the important term in determining the roots of Eq. (A8). Equation (A8) also contains a third type of term corresponding to the gradients of r_{ij}^2 with respect to the four unknowns. These terms are also not the important terms inasmuch as they do not change sign. Thus Eq. (A8) contains all the expressions explicit in R that are

dominant in obtaining the roots. This form is especially suitable for iterative solution using either a simplex or a Newton–Raphson algorithm for nonlinear systems of equations. The solution was checked each time whether it was a local minimum or not by evaluating the energy values in its neighborhood.

For the case of two rods not connected by a loop, the problem simplifies significantly, because the loci of the energy minima is characterized by $\theta_1 = \theta_2 = 90°$, provided that the interaction sites are homogeneously distributed on the rods. Also, we have $\phi = \alpha$, where α is the angle between the two rods. The two unknowns of the problem are R and α which are determined from the simultaneous solution of the first and the fourth of Eqs. (A8).

For two disconnected rigid bodies in two dimensions, the three unknowns of the problem may be taken as the x and y components of $\mathbf{R}$ [Fig. 1(a)], and the orientation θ of the M_2 with respect to M_1. In this case, $\phi = 0$, $R_1 = R \cos\theta_1$, $R_2 = R \sin\theta_1$, $\theta_2 = \theta$, and the three unknowns in the problem can be solved from the first three of Eqs. (A8).

APPENDIX B

Details about the geometry and energy characteristics of the model helices illustrated in Fig. 7 are presented here. These aim at modeling α helices in proteins. Each residue is represented by two sites, B and S located on the backbone (α carbon) and on the sidechain (β carbon or sidechain interaction center), respectively. In the case of an idealized pair of α helices, i.e., disconnected α helices composed of sites exhibiting the geometry and energy characteristics of an average residue in globular proteins, the lengths l of the virtual bonds $B-B$ and the angle Θ between backbone virtual bonds are taken as 3.81 Å and 88°, which conform with the geometry of α helices. Backbone bond dihedral angles are assigned the fixed value of $-128°$ with respect to the planar *trans* conformation of three successive bonds. This angle asserts a right-handed rise of 5.36 Å per helix turn (or 1.49 Å per residue), in conformity with right-handed α helices. In the case of more realistic models, such as that of ROP monomer, the exact bond lengths, bond angles and dihedrals of the known structure are taken, together with residue-specific $S-S$ and $S-B$ potentials.

For the idealized pair of helices, the loci of the interaction sites on the backbone and on the side groups are generated by using the conventional bond-based coordinate systems of polymer statistics.[28] Accordingly, the x_i axis of the local system appended to the ith bond is taken along the bond vector $\mathbf{l}_i$ connecting the α carbons $i-1$ and i, the y_i axis makes an acute angle with the axis x_{i-1} of the preceding bond-based frame, and lies in the plane spanned by x_{i-1} and x_i; the z_i axis completes a right-handed coordinate system. In these local bond-based frames, the backbone bond vectors are expressed as $[3.81\ 0\ 0]^T$, where the superscript T denotes the transpose. The side group bond vector is taken as $[-0.511\ 0.87\ 1.15]^T$, assuming bond lengths of 1.53 Å, and by adopting bond angles typical of C^α–C^β bonds. The configurational energies of the pair of helices are evaluated on the basis of the homogeneous interaction potentials[14] displayed in Fig. 8 as

$$E\{\Phi\} = \sum_{i=1}^{m} \sum_{j=1}^{n} \left[E(S_i - S_j) + E(S_i - B_j) + E(B_i - S_j) \right. $$
$$\left. + E(B_i - B_j) \right]. \tag{B1}$$

Here, $E(S_i-S_j)$ is the interaction potential between side groups S_i and S_j belonging to the respective helices M_1 and M_2, $E(S_i-B_j)$ and $E(B_i-S_j)$ refer to the backbone-sidechain interaction energies of the two helices, and $E(B_i-B_j)$ is the backbone-backbone interaction energy. The summations in Eq. (B1) are performed over all interaction sites on the backbone and side groups of the two helices.

The configuration $\{\Phi\}$ of the pair of helices is characterized by a suitable set of variables, say $\{R,\theta_1,\theta_2,\phi\}$ as above [Fig. 1(c)] or the set $\{R,\vartheta,\alpha,\beta\}$ described in Fig. 7. In the latter case, the axes z_1 and z_2 are chosen along the principal axes of the respective helices 1 and 2. The corresponding unit directional vectors are denoted as δ_{z1} and δ_{z2}. α is the angle between the principal axes of the two helices. ϑ is the angle between R and δ_{z1}, and β is the angle between the x_1 axis and the projection of the principal axis of M_2, on the Ox_1y_1 plane. Two additional angles χ and φ are required to characterize the configuration $\{\Phi\}$ in the case of real α helices. χ refers to the rotation of M_2 about its own axis, and ψ is the azimuthal angle describing the orientation of $\mathbf{R}$ with respect to $Ox_1y_1z_1$. The angles ϑ, α, β, and χ are readily determined from

$$\vartheta = \cos^{-1}(\delta_{z1} \cdot \mathbf{R})/R,$$
$$\alpha = \cos^{-1}(\delta_{z1} \cdot \delta_{z2}),$$
$$\beta = \cos^{-1}\{\delta_{z2} \cdot (\delta_{z2} - \delta_{z1}\cos\alpha)\}/|\delta_{z2} - \delta_{z1}\cos\alpha|, \tag{B2}$$
$$\chi = \cos^{-1}\{\mathbf{T}(-\alpha)\delta_{x2} \cdot \delta_{x1}\},$$

where $\mathbf{T}(-\alpha)$ is the transformation matrix operating on the system $O_2x_2y_2z_2$ so as to superimpose the axes z_1 and z_2.

In the case of the pair of helices connected by a loop, the dihedral angles φ_i of the virtual bonds $1 \leq i \leq s$ located in the loop, and those of the last virtual bond of M_1 and first virtual bond of M_2, φ_0, and φ_{s+1}, are assumed to be variable, while all other degrees of freedom are fixed. As a convenient representation of the coordinates of the sites, we append a bond-based frame, $O\xi\eta\zeta$ to the last virtual bond of M_1. All site coordinates in M_1 are fixed with respect to that frame. Those in M_2 are expressed by operating the product $\mathbf{T}(\varphi_0)\mathbf{T}(\varphi_1)\cdots\mathbf{T}(\varphi_{s+1})$ on their local position vectors $\mathbf{b}_j$. Here, $\mathbf{T}(\varphi_j)$ is the transformation matrix operating between the above described bond-based frames $j+1$ and j; the local position vectors $\mathbf{b}_j$ refer to the bond-based frame appended to the first bond of helix 2. Therefore, the counterpart of Eq. (A3) in the present model reads

$$\mathbf{r}_{ij} = \mathbf{R} + \left[\prod_{k=0}^{s+1} \mathbf{T}(\varphi_k) \right] \mathbf{b}_j - \mathbf{a}_j, \tag{B3}$$

where $\mathbf{a}_i$ is the position vector of the ith site in M_1 expressed in the frame $O\xi\eta\zeta$. The bonds in the loop region are subject to pairwise interdependent torsional potentials of the form

$$E_A(\varphi_i,\varphi_{i+1}) = E_A(\varphi_i) + E_A(\varphi_{i+1}) + \Delta E_A(\varphi_i,\varphi_{i+1}) \tag{B4}$$

which vary with the type of the amino acid A between the consecutive bonds i and $i+1$. Here, $E_A(\varphi_i)$ refers to the torsional potential of the ith virtual bond when the latter is rotated by a dihedral angle φ_i, and is succeeded by a residue of type A. Likewise, $E_A(\varphi_{i+1})$ refers to the torsional potential of a bond preceded by a residue of type A. $\Delta E_A(\varphi_i,\varphi_{i+1})$ is the additional coupling energy between the two consecutive dihedral angles flanking the residue A. The reader is referred to previous work for a detailed presentation of the energies $E_A(\varphi_i)$ and $\Delta E_A(\varphi_i,\varphi_{i+1})$ for the 20 types of naturally occurring amino acids.[16] The overall energy of the system consisting of the pair of helices connected by a loop is found from Eq. (2).

The torsional angles $\{\varphi_0,\varphi_1,\ldots,\varphi_{s+1}\}$ fully define a configuration $\{\Phi\}$ in the loop region composed of s inextensible bonds with constant bond angles. The number of variables for determining the minimum energy configuration in-

creases with the size s of the loop. For $s>4$, the solution is preferably found in terms of the variables $\{R,\vartheta,\psi,\alpha,\beta,\chi\}$ rather than $\{\varphi_0,\varphi_2,\ldots,\varphi_{s+1}\}$. For large values of s a limiting case obtains where the loop chain behavior approaches that of a freely jointed chain and its constraining action on the configurations of the two helices diminishes. In this case, one should expect minimum energy conformations which would be only slightly perturbed compared to those obtained for two disconnected helices.

[1] D. B. Adolf and M. D. Ediger, Macromolecules **24**, 5834 (1991).

[2] I. Bahar, B. Erman, and L. Monnerie, Adv. Polymer Sci. **116**, 145 (1994).

[3] J. R. Gunn, A. Monge, and R. A. Friesner, J. Phys. Chem. **98**, 702 (1994).

[4] A. Monge, R. A. Friesner, and B. Honig, Proc. Natl. Acad. Sci. **91**, 5027 (1994).

[5] A. Monge, E. J. Lathrop, J. R. Gunn, P. S. Shenkin, and R. A. Friesner, J. Mol. Biol. **247**, 995 (1995).

[6] R. Srinivasan and G. D. Rose, Proteins **22**, 81 (1995).

[7] B. Park and M. Levitt, J. Mol. Biol. **258**, 367 (1996).

[8] Z. Li and H. A. Scheraga, Proc. Natl. Acad. Sci. USA **84**, 6611 (1987).

[9] R. A. Abagyan and V. N. Maiorov, J. Biomol. Struct. Dyn. **6**, 1045 (1989).

[10] D. W. Banner, M. Kokkinidis, and D. Tsernoglou, J. Mol. Biol. **196**, 657 (1988).

[11] W. H. Press, S. A. Teukolski, W. T. Vetterling, and B. P. Flannery, *Numerical Recipes in Fortran* (Cambridge University Press, London, 1992).

[12] G. Cesarini and D. W. Banner, Trends Biochem. Sci. **10**, 303 (1985).

[13] I. Bahar and R. L. Jernigan, Folding & Design **1**, 357 (1996).

[14] I. Bahar and R. L. Jernigan, J. Mol. Biol. **266**, 195 (1997).

[15] D. Eisenberg, R. M. Weiss, and T. C. Terwilliger, Nature (London) **299**, 371 (1982).

[16] I. Bahar, M. Kaplan, and R. L. Jernigan, Proteins (in press).

[17] R. Abagyan and P. Argos, J. Mol. Biol. **225**, 519 (1992).

[18] D. Cvijovic and J. Klinowski, Science **267**, 664 (1995).

[19] A. G. Murzin and A. V. Finkelstein, J. Mol. Biol. **204**, 749 (1988).

[20] J. W. Bryson, S. F. Betz, H. S. Lu, D. J. Suich, H. X. Zhou, K. T. O'Neil, and W. F. DeGrado, Science **267**, 935 (1995).

[21] A. Kolinski and J. Skolnick, Proteins **18**, 353 (1994).

[22] R. L. Jernigan and I. Bahar, Current Opin. Struct. Biol. **6**, 195 (1996).

[23] A. Sarai, R. L. Jernigan, and J. Mazur, Biophys. J. **71**, 1507 (1996).

[24] J. D. Bryngelson and P. G. Wolynes, Proc. Natl. Acad. Sci. USA **84**, 7524 (1987).

[25] P. G. Wolynes, J. N. Onuchic, and D. Thirumalai, Science **267**, 1619 (1995).

[26] A. Sali, E. Shakhnovich, and M. Karplus, Nature (London) **369**, 248 (1994).

[27] A. Sali, E. Shakhnovich, and M. Karplus, J. Mol. Biol. **235**, 1614 (1994).

[28] P. J. Flory, *Statistical Mechanics of Chain Molecules* (Interscience, New York, 1969) (also reprinted by Hanser Publishers, Oxford University Press, Oxford, 1988).