β-Sheet Coil Transitions in a Simple Polypeptide Model

Kanthi Yapa,[1] David L. Weaver,[1] and Martin Karplus[2]
[1]Department of Physics, Tufts University, Medford, Massachusetts 02155, and [2]Department of Chemistry, Harvard University, Cambridge, Massachusetts 02138

ABSTRACT   A simplified model of a polypeptide chain is used to study the dynamics of the β-sheet–coil transition. Each amino acid residue is treated as a single quasiparticle in an effective potential that approximates the potential of mean force in solution. The model is used to study the equilibrium and dynamic aspects of the sheet–coil transition. Systems studied include ones with both strands free to move (two-strand sheet), and ones with either strand fixed in position (multistrand sheet). The equilibrium properties examined include sheet–coil equilibrium constants and their dependence on chain position. Dynamic properties are investigated by a stochastic simulation of the Brownian motion of the chain in its solvent surroundings. Time histories of the dihedral angles and residue–residue cross-strand distances are used to study the behavior of the sheet structure. Auto- and cross-correlation functions are calculated from the time histories with relaxation times of tens to hundreds of picoseconds. Sheet–coil rate constants of tens of ns^{-1} were found for the fixed strand cases.

Key words: β-sheet-coil transition, β-hairpin, Langevin dynamics, equilibrium properties, quasiparticle, effective potential, autocorrelations, cross-correlations, time histories, rate constants

INTRODUCTION

The biological activity of a globular protein is determined by its average three-dimensional conformation (the native structure) and by the internal flexibility of this structure. The native structure can fold spontaneously under appropriate conditions without any information other than that contained in the linear sequence of amino acid residues. Since the purpose of the genetic code is to specify the sequence of amino acids, the prediction of the native structure of a protein from its amino acid sequence is an essential part of molecular biology. Because of the very large number of natural ($\approx 10^9$) and possible artificial (a chain of 100 amino acids has $20^{100} \approx 10^{130}$ different possible sequences) proteins, it will not be possible to determine most of the native structures by physical methods. A rational approach to

protein design and engineering, thus, has as one of its elements the design of a computational model for protein folding. An understanding of protein folding involves two related parts, namely, a determination of how the peptide elements interact dynamically to produce the final fold; and a method for predicting the nature of the final three-dimensional structure. It is the dynamic aspect of the folding problem which is the concern of this paper. Some years ago, we proposed a simplified model for the folding process, which is called the diffusion-collision model.[1] In this model, it is assumed that the formation of the tertiary structure of a protein takes place by the diffusion, collison, and coalescence of fluctuating structural elements (microdomains), such as α-helices and β-strands that are not stable (or often only marginally stable) in the absence of contacts with other elements. The existence of microdomains in the folding polypeptide chain and their importance in the kinetics of folding provides a way for the folding protein to avoid examining the entire set of conformational alternatives. The microdomains move diffusively under the influence of internal and systematic and random external forces, and microdomain–microdomain collisions take place. Collisions sometimes lead to coalescence into microdomain pairs and higher aggregates (higher order microdomains or subdomains). This can occur if both microdomains have at least part of their secondary structure (i.e., that in the contact region), and the collision involves an appropriate orientation. The diffusion–collision portion of the folding process could start with the entire polypeptide chain in an extended random-coil state or a more collapsed state. Overall folding would consist of a series of such steps leading to a conformation with portions of the backbone close to the native conformation. It is possible that this leads to the molten globule state,[2] where some secondary structural elements are loosely associated. The final step of the folding process would then be the formation of the exact tertiary structure, including close packing of the side chains. During this last phase, proline isomerization[3] or other slow events could be rate limiting.

As outlined above, the diffusion–collision model provides a description of the large-scale kinetic events in small globular proteins and in the domains of larger proteins. It also implies that the principal, early dynamic events of the folding process are concerned with the properties of microdomains and their interactions, rather than with the individual amino acids of the polypeptide chain, as in a random search. The diffusion–collision model differs from other descriptions of folding, which are generally qualitative. It is possible to develop a scheme for calculating the folding pathway and the time course of species formed during the folding reaction. If other events (e.g., the determination of the tertiary structure from the molten globule) are rate limiting, the diffusion–collision results would be related only to the kinetics of the search process and not the overall folding rate.

Two methods have been used to calculate the diffusion–collision dynamics involved in the folding of peptides and proteins. In the first method, which is essentially analytic, the dynamics of folding is simulated by a set of diffusion equations that describe the motion of the microdomains in aqueous solution, and by coupled boundary conditions that provide for their collision and possible coalescence. Detailed calculations[4] showed that the diffusion–collision dynamics of a multimicrodomain protein reduces to a network of two-microdomain steps in which the calculable rate constants depend on physical properties of microdomains (size, shape, reactivity, orientation) and the equilibrium state of the system. The folding kinetics are approximated by solving kinetic equations that couple the elementary steps. This method is the "chemical kinetics" approximation applied to the diffusion–collision model. The evaluation of the individual rate constants for the coalescence of two entities (e.g., two elementary microdomains or microdomain complexes) is straightforward. They can be written analytically in terms of the physical parameters of the system. In most calculations, the geometric parameters in the rate constants have been evaluated in a spherical approximation of the type often used to simplify diffusion calculations. In the second method, an alternative to the description of the microdomains as spherical entities, a somewhat more detailed model of the polypeptide chain is introduced. Each amino acid residue is approximated by a single interaction center linked by virtual bonds.[5] The diffusion equation, which is used in the first method and which provides a continuum approximation to the motion of the microdomains, is replaced by the discrete, coupled Langevin equations of motion for the residues with a correspondingly simplified potential energy function and a random force to approximate the effect of solvent. These equations can be solved numerically to treat the full folding problem or to obtain information related to the chemical kinetics approximation. The validity of

some of the approximations of the analytical model has been assessed by this approach. This includes the effect of intervening chain segments on the diffusion of microdomains. Also, the values of some of the parameters used to calculate folding rates (e.g., stabilities of helices and sheets) and the kinetic constants for microdomain–microdomain collisions can be examined. The method has been applied in simulating the helix–coil transition of an α-helical[5] microdomain and the diffusion–collision of two α-helices (microdomains) connected by a coil segment.[6] In the two-helix simulation,[6] the problem was simplified to reduce the cost of the simulations by contraining the helices to remain helical and allow dihedral angle transitions to occur only in the coil segments, i.e., the helix–coil transition of residues in the helical portion was not permitted to occur. Thus, the calculation simulated the diffusion–collision dynamics of two stable microdomains connected by a chain segment. Helix–coil transitions, coupled to collisional events, will be introduced in subsequent simulations on the basis of studies of the rates of the coil-to-secondary-structure transitions.[5]

The diffusion–collision model was proposed to isolate the factors important in protein folding dynamics and to be able to obtain the time scale of the folding process with simple assumptions and without lengthy calculations. Such a model is of particular interest at present because an increasing number of experiments, making use of techniques such as nuclear magnetic resonance[7–10] and circular dichroism,[11–17] are being applied to whole proteins and protein fragments. In previous papers, various aspects of the folding of α-helical proteins were considered. They include the helix–coil transition,[5] the coalescence of two α-helices connected by a coil segment,[6] and the overall folding kinetics of apomyoglobin[18] and the λ-repressor,[19] as well as the initial step in folding of cytochrome c.[20]

A number of other models have been proposed for globular protein folding, e.g., the framework model,[21] which proposes that H-bonded secondary structure is formed early in folding; the jigsaw puzzle model,[22] which suggests that proteins fold by large numbers of different, parallel pathways rather than by a single defined sequence of events; and the sequential model,[23] in which a subunit of the polypeptide chain (e.g., a microdomain) serves as a nucleus for chain propoagation to obtain the native structure. Each of these models is a possible realization of the diffusion–collision mechanism in protein folding (a review of the connection of other models of folding to the diffusion–collision model is in preparation[24]), because the details of folding rates and pathways depend on the properties of the microdomains of the system.

In the present paper, we begin to examine the folding behavior of β-sheets. This complements the earlier work on α-helices. The information from this
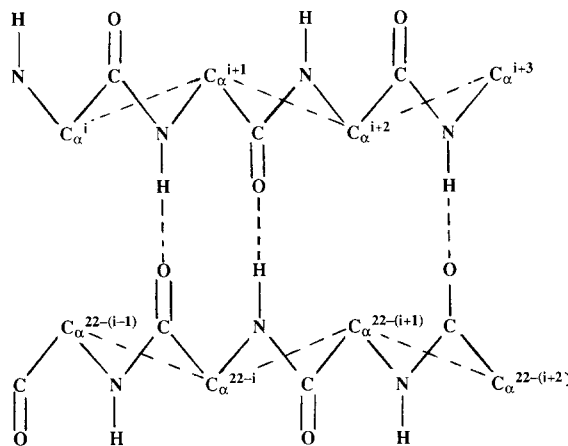
Fig. 1. The geometry of the simplified polypeptide model compared with that of the atomic structure for an antiparallel β-sheet. Virtual bonds between the α carbons are shown with dashed lines. In addition, some of the cross-strand hydrogen bonds are indicated by dashed lines between the amide hydrogen and the carbonyl oxygen.

type of simulation is essential for refining models of protein folding of the type we have outlined above. We describe an equilibrium study and a Langevin dynamics simulation of the behavior of the terminal residues of an existing β-hairpin. The model used is similar to that employed in simulating the helix–coil transitions of an α-helix.[5] Each amino acid is treated as a single quasiparticle in an effective potential that approximates the potential of mean force in solution. Subsequent papers will consider other aspects of folding involving β-sheets. The second section describes the polypetide model used in this work. The equilibrium properties of the β-sheet fragment are considered in the third section, and the dynamics are presented in the fourth section. A concluding discussion is then given.

POLYPEPTIDE MODEL

Polyalanine was chosen as the model for the antiparallel β-sheet because of the simplicity of the side chains, although it is more stable in α-helical form.[25] The polyalanine chain was represented by a simplified model based on the earlier work of Flory[26] and a simulation of helix–coil transitions.[5,27] Each alanine residue was replaced by a soft sphere with a radius of 2.75 Å and center at the C_i^α position of the ith residue in the chain.[28] The radius is that of a sphere whose volume equals that of the alanine side chain. The spheres are linked by virtual bonds and bond angles with harmonic restoring potentials.[26] The geometry of the simplified representation and its connection with the polypeptide chain structure is shown in Figure 1. The two strands of the antiparallel β-sheet were built using the atomic coordinates given by Arnott, Dover, and Elliott[29] for β-poly-L-alanine fibers; the parameters are $\phi = 41.4°$, $\psi = -45.3°$, and $\omega = 1.5°$. The interchain hydrogen-bond length is 2.83 Å. The two strands are connected at one end by a tight turn to

make one polypeptide chain. Only the residues at the open end of the chain were studied in this work. The calculations were performed for a 22-residue chain with 10 residues in each strand and two additional residues for the turn. Even though the bond and bond angle interactions do not play a significant role in our investigation, we used flexible potentials in order to simplify the force calculations. The values of the bond and bond angle force constants k_b and k_θ are such that the bond length b and bond angle θ do not change appreciably from their equilibrium values. The virtual dihedral angle α_i (see Fig. 1) is defined in terms of the residue centers i, $i+1$, $i+2$, and $i+3$ ($\alpha_i = 0$ for the eclipsed conformation), and is associated with an intrinsic torsional potential. In order to fit the two antiparallel strands into the β-hairpin, the dihedral angles in the stationary part of the hairpin have the values $\alpha_7 = 172.73°$, $\alpha_8 = 127.34°$, $\alpha_9 = 6.172°$, $\alpha_{10} = 4.72°$, $\alpha_{11} = 50.86°$, $\alpha_{12} = 151.80°$, and $\alpha_{13} = 159.85°$. Interactions between residues separated by three or more virtual bonds are given the effective central force potentials which represent excluded-volume and net attractive effects present in aqueous surroundings. Additional torsional potential terms are used to stabilize the β-sheet structure. They can be regarded as due to hydrogen bonds between residues opposite to one another in the two strands. The choice of a single interaction center per residue replaces the conventional H-bond pattern of the antiparallel strands by virtual H-bonds between opposite pairs of residues. Two parameters are used to determine whether residues are in a sheet. First, the dihedral angle α is allowed to have a broad angular width of $90°$,[24] $135° \leq \alpha \leq 225°$, because the ϕ, ψ pairs of real strands show large deviations from ideal values. The second parameter is the distance $r_{i,22-(i-1)}$ between residues opposite each other on the two strands. A range of values, 4.0–5.7 Å, was chosen

from the $C^\alpha \cdots C^\alpha$ distances found between adjacent antiparallel strands in existing β-sheet proteins (22 β hairpins were examined in 14 proteins). Any pair of opposite residues is considered to be in the sheet conformation when the corresponding dihedral angels and the distance between them are within the ranges given above.

The total energy E describing the polyalanine chain can be written as the sum of six terms:

$$E = E_b + E_\theta + E_\alpha + E_{ev} + E_{sol} + E_\beta. \quad (1)$$

The bond and bond angle energies are harmonic

$$E_b = \sum_b k_b (b - b_0)^2 \quad (2)$$

where k_b = 40 kcal/mol-Å² and b_0 = 3.8 Å, and

$$E_\theta = \sum_\theta k_\theta (\theta - \theta_0)^2 \quad (3)$$

where k_θ = 40 kcal/mol-rad² and q_0 = 130.0° (2.2698 rad). The dihedral angle contribution to the energy has the form

$$E_\alpha = \sum_\alpha V(\alpha) \quad (4)$$

where $V(\alpha)$ is the Fourier series for the L-Ala dipeptide given by Levitt.[28] The excluded volume term is given by

$$E_{ev} = \sum_{i>j} V_{ev}(r_{ij}) \quad (5)$$

with the summation running over all pairs of residues separated by three or more bonds. Here r_{ij} is the distance between residues i and j and

$$V_{ev} = \begin{cases} \epsilon\left[3\left(\dfrac{r^0}{r}\right)^8 - 4\left(\dfrac{r^0}{r}\right)^6 + 1\right] & (r<r^0) \quad (6) \\ 0 & (r\geq r^0) \end{cases}$$

with ϵ = 0.21 kcal/mol and r_0 = 5.5 Å, the van der Waals parameters for alanine residue pairs. The excluded volume term is purely repulsive to avoid overcounting the attractive interaction.[5] The attractive van der Waals and solvent contribution has the form

$$E_{sol} = \sum_{i>j} \sigma g(r_{ij}) \quad (7)$$

with σ = −1.0 kcal/mol for alanine residues, and the summation runs over all pairs of residues separated by three or more bonds. The function $g(r)$ is a sigmoid function that varies from $g(0)$ = 1 to $g(r)$ = 0 for $r > 9$ Å.[28] The sheet-stabilization energy is

$$E_\beta = \sum_{\alpha_i} A_\alpha f(\alpha_i) \quad (8)$$

where the summation runs over all the virtual dihedral angles in the two strands. The function $f(\alpha)$ is bell-shaped with a maximum value at α = 180°, and

with vanishing first derivaties at α = 135° and α = 225°. Its detailed form is

$$f(\alpha) = \begin{cases} 4.86\,(\alpha - \alpha_-)^2 - 4.13\,(\alpha - \alpha_-)^3 \\ \qquad\qquad\qquad 135° < \alpha < 180° \\ 4.86\,(\alpha - \alpha_+)^2 + 4.13\,(\alpha - \alpha_+)^3 \\ \qquad\qquad\qquad 180° < \alpha < 225° \quad (9) \\ 0 \qquad\qquad\qquad \text{otherwise} \end{cases}$$

where α_- = 135° (2.356 rad) and α_+ = 225° (3.927 rad). The coefficient A_{α_i} for each dihedral angle depends on the neighboring dihedral angles and the distances $r_{i,22-(i-1)}$ according to the set of rules described below.

We consider three cases in this study: both strands free to move and either strand separately fixed in position. The former represents an isolated β-sheet, and the latter represents sheets in which one or the other of the two strands is participating in another sheet structure and is consequently bound in position. We need to consider both of the cases with one strand fixed, because the strands are antiparallel and their residue centers lie on planes which are not parallel, but rather meet at about a 10° angle. This leads to some differences in the potential of mean force for the two strands (see below). The dependence of the sheet stabilization coefficient A_{α_i} for each dihedral angle on other dihedral angles is described below in detail for the three cases considered; i.e., both strands free to move and strand one or two fixed in position. Values were chosen to yield equilibrium constants near unity for the free residues. Calculation of equilibrium constants beyond the residue pair (3,20) was very time consuming in the case where both strands were allowed to move. Therefore in the calculation of equilibrium properties described below, only the three sets of pairs (1,22), (2,21), and (3,20) were computed for this case. A simplified picture of the model is shown in Figure 2.

Case I: Both Strands Free to Move

If residues $i+1$, $i+2$, ..., $22-i$ and the two residues $i-1$ and $22-(i-2)$ are in the sheet conformation (Fig. 3a), then residues i and $22-(i-1)$ may be considered to be in the interior of the sheet structure, and $A_{\alpha_i} = A_{\alpha_{19-(i-1)}}$ = −6.0 kcal/mol. This reflects the difficulty of nucleating a coil in the interior of the sheet. If residues $i+1$, ..., $22-i$ are in the sheet conformation, but $i-1$ and/or $22-(i-2)$ are not (Fig. 3a), then residues i and $22-(i-1)$ are in the sheet–coil interface, and $A_{\alpha_i} = A_{\alpha_{19-(i-1)}}$ = −1.43 kcal/mol. This value is chosen to yield an equilibrium constant near unity. If a residue is neither in the sheet interior nor in the sheet–coil interface, then A_{α_i} = 0.

Case II: Strand Two Fixed in Position

Consider Figure 3b with strand two fixed in position (i.e., $\alpha_{14} = \cdots = \alpha_{19}$ = 180°). If residues $i+1$,
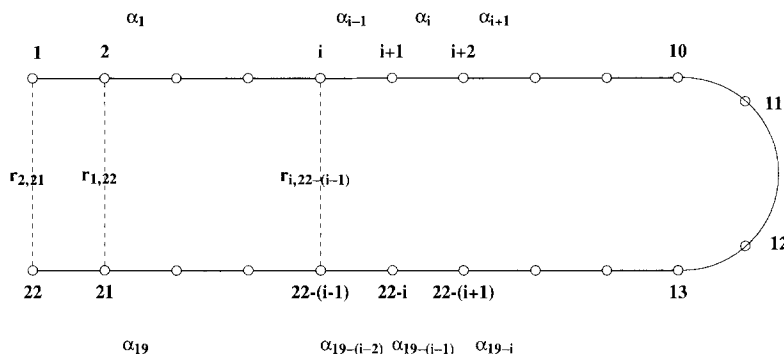
Fig. 2. A schematic picture of the model polypeptide chain. The dihedral angles α_i, defined by the residue centers i $i+1$, $i+2$, and $i+3$ are indicated and some of the cross-strand hydrogen bonds between residues i and $22-(i-1)$ are shown as dashed lines.

... , 9 and $i-1$ are in the sheet conformation, then $A_{\alpha_i} = -3.0$ kcal/mole. If residues $i+1, \ldots, 9$ are in the sheet conformation, but residue $i-1$ is not, then residue i is in the sheet–coil interface and $A_{\alpha_i} = -0.53$ kcal/mole. The value -0.53 is chosen to yield an equilibrium constant s_i or $s_{22-(i-1)}$ near unity for the case of strand one fixed. If residue i is neither in the sheet interior nor in the sheet-coil interface, then $A_{\alpha_i} = 0$.

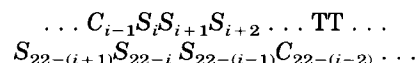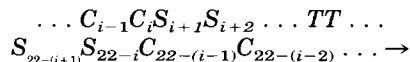## Case III: Strand One Fixed in Position

Consider Figure 3c with strand one fixed in position (i.e., $\alpha_1 = \alpha_2 \cdots = \alpha_6 = 180°$). If residues $22-i$, ... , 14 and $22-(i-2)$ are in the sheet conformation, then $A_{\alpha_i} = -3.0$ kcal/mol. If residues $22-i, \ldots, 14$ are in the sheet conformation, but residue $22-(i-2)$ is not, then residue $22-i$ is in the sheet–coil interface and $A_{\alpha_i} = -0.53$ kcal/mol. The value -0.53 is chosen to yield an equilibrium constant s_i or $s_{22-(i-1)}$ near unity for the case of strand one fixed. If residue i is neither in the sheet interior nor in the sheet–coil interface, then $A_{\alpha_i} = 0$.

EQUILIBRIUM PROPERTIES

The methodology employed here is similar to that used by McCammon et al.[5] in their study of the helix–coil transitions at the terminus of a helix in a small polyvaline peptide. The equilibrium constant and the mean energy for residues at the interface of the sheet–coil transition have been computed for the three cases.

Case I: Both Strands Free to Move

Assume that the residues $i+1, i+2, \ldots, 22-i$ are in the sheet structure and residues $1, 2, \ldots, i-1$ and $22-(i-2), \ldots, 22$ are in a random coil sequence. The equilibrium constant $s_{i,22-(i-1)}$ for the transformation of the two residues i and $22-(i-1)$ from the coil state to the sheet state corresponds to the reaction

$$\ldots C_{i-1}C_iS_{i+1}S_{i+2} \ldots TT \ldots$$
$$S_{22-(i+1)}S_{22-i}C_{22-(i-1)}C_{22-(i-2)} \ldots \rightarrow$$

$$\ldots C_{i-1}S_iS_{i+1}S_{i+2} \ldots TT \ldots$$
$$S_{22-(i+1)}S_{22-i} S_{22-(i-1)}C_{22-(i-2)} \cdots$$

where C, S, and T represent coil, sheet, and turn residues, respectively. The equilibrium constant for the reaction, which corresponds to the addition of two residues to an existing sheet (see, for example, ref. 5) is

$$s_{i,22-(i-1)} = \frac{[S_i] [S_{22-(i-1)}]}{[C_i] [C_{22-(i-1)}]}. \tag{10}$$

To calculate $s_{22-(i-1)}$, the residues $i+1, \ldots, 22-i$ were fixed in the sheet configuration (i.e., the dihedral angles $\alpha_{i+1}, \ldots, \alpha_6$ and $\alpha_{14}, \ldots, \alpha_{19-i}$ were fixed at 180°), so that they form a β hairpin. The dihedral angle α_i was varied at 10° intervals over the angles 0°, 10°, 20°, ... ,350°. For each value of α_i, the angles of the random coil residues $\alpha_1, \ldots, \alpha_{i-1}$ and $\alpha_{19-(i-1)}, \ldots, \alpha_{19}$ were varied at 30° intervals over the angles 0°, 30°, 60°, ... ,330° and the Boltzmann factors corresponding to all other varied angles were added to give a relative probability

$$P'(\alpha_i) = \sum e^{\frac{-E(\alpha_i, \alpha_1, \ldots, \alpha_{i-1}, \alpha_{19-(i-1)}, \ldots, \alpha_{19})}{RT}} \tag{11}$$

where the summation runs over angles $\alpha_1, \ldots, \alpha_{i-1}$ and $\alpha_{19-(i-1)}, \ldots, \alpha_{19}$. $P'(\alpha_i)$ is proportional to the probability of finding residue i in the state defined by dihedral angle α_i while residues $i+1, \ldots, 22-i$ are in the sheet conformation and residues $1, \ldots, i-1$ and $22-(i-1), \ldots, 22$ sample all coil configurations. The bond lengths and bond angles were fixed at the equilibrium values, so the energy $E(\alpha_i, \alpha_1, \ldots, \alpha_{i-1}, \alpha_{19-(i-1)}, \ldots \alpha_{19})$ includes only the torsional functions and the nonbonding interactions which change upon variation of these dihedral angles. In accordance with the rules given above, the coefficient A_{α_i} in the sheet stabilization term has the value $A_{\alpha_i} = -1.43$ kcal/mol and $A_{\alpha_i} = \cdots = A_{\alpha_{i-1}} = A_{\alpha_{19-(i-1)}} = \cdots A_{\alpha_{19}} = 0$. As stated above, the magnitude -1.43 kcal/mol was chosen to lead to equilib-
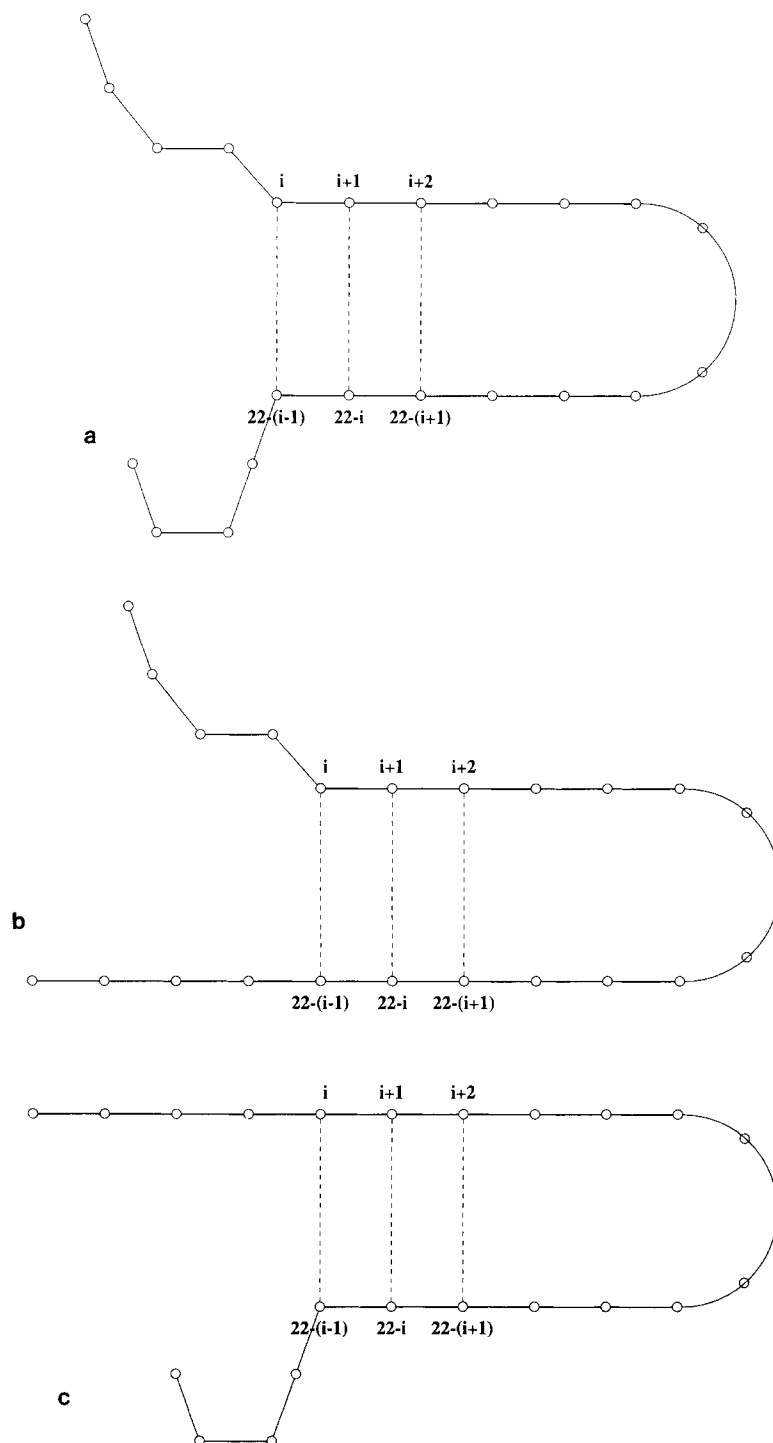
Fig. 3. **(a)** Schematic picture of the model peptide chain with both strands free to move. Residues $i+1, \ldots, 22-i$ are in the sheet interior, residues i and $22-(i-1)$ are at the sheet–coil interface, and residues $1, \ldots, i-1$ and $22-(i-2), \ldots, 22$ are in the coil state. **(b)** Strand two fixed. Residues $i+1, \ldots, 22$ are fixed in the initial sheet state, residue i is at the sheet–coil interface, and residues $1, \ldots, i-1$ are in the coil state. **(c)** Strand one fixed. Residues $1, \ldots, 22-i$ are fixed in the initial sheet state, residue $22-(i-1)$ is at the sheet–coil interface, and residues $22, \ldots, 22-(i-2)$ are in the coil state.

rium constants near unity. The absolute probability $P(\alpha_i)$ is given by

$$P(\alpha_i) = \frac{P'(\alpha_i)}{\sum_{\alpha_i} P'(\alpha_i)} \quad (12)$$

where the summation runs over all values of A_{α_i}. Similarly $P'(\alpha_{19-(i-1)})$, which is associated with residue $22-(i-1)$, is calculated by varying $\alpha_{19-(i-1)}$ over angles 0°, 10°, 20°, . . . ,350° while varying angles $\alpha_1, \ldots, \alpha_i$ and $\alpha_{19-(i-2)}, \ldots, \alpha_{19}$ over angles 0°, 30°, . . . , 330° and summing the corresponding Boltzmann factors. In this case, $A_{\alpha_{19-(i-1)}} = -1.43$ kcal/mol and $A_{\alpha_1} = \cdots A_{\alpha_i} = A_{\alpha_{19-(i-2)}} = \cdots = A_{\alpha_{19}} = 0$ and

$$P'(\alpha_{19-(i-1)}) = \sum e^{\frac{-E(\alpha_{19-(i-1)}, \alpha_1, \ldots, \alpha_i, \alpha_{19-(i-2)}, \ldots, \alpha_{19})}{RT}}$$
$$(13)$$

where the summation runs over angles $\alpha_1, \ldots, \alpha_i$ and $\alpha_{19-(i-2)}, \ldots, \alpha_{19}$. In this case the absolute probability $P(\alpha_{19-9i-1})$ is

$$P(\alpha_{19-(i-1)}) = \frac{P'(\alpha_{19-(i-1)})}{\sum_{\alpha_{19-(i-1)}} P'(\alpha_{19-(i-1)})} \quad (14)$$

where the summation is over all the values of $\alpha_{19-(i-1)}$. The equilibrium constant $s_i,22-(i-1)$ is the probability of finding residues i and $22-(i-1)$ in the sheet state over the probablility of finding residue i and/or residue $22-(i-1)$ in the coil state. Thus,

$$s_{i,22-(i-1)} = \frac{\sum P(\alpha_i) P(\alpha_{19-(i-1)})}{1 - \sum P(\alpha_i) P(\alpha_{19-(i-1)})}. \quad (15)$$

The summation runs over the values of α_i and $\alpha_{19-(i-1)}$) which fall within the range 135°–225° with the requirement that for each set of α_i and $\alpha_{19-(i-1)}$, the distance $r_{i,22-(i-1)}$ is within the range 4.0–5.7 Å.
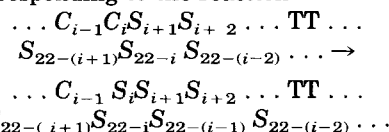
The mean energy and its elements were calculated for residues at the sheet–coil interface. The pth element of the mean energy is denoted by $<E_p(\alpha_i)>$ where E_p can be E or any component of E as defined in Eqs. (1)–(9). For a particular value of α_i, $<E_p(\alpha_i)>$ is obtained by fixing residues $i+1, \ldots, 22-i$ in the sheet and sampling residues $1, \ldots, i-1$ and $22-(i-1), \ldots, 22$ in all coil configurations and then averaging to get

$$<E_p(\alpha_i)> =$$

$$\frac{\sum E_p(\alpha_i, \alpha_1, \ldots, \alpha_{i-1}, \alpha_{19-i-1}, \ldots, \alpha_{19}) e^{\frac{-E(\alpha_i, \alpha_1, \ldots, \alpha_{i-1}, \alpha_{19-i-1}, \ldots, \alpha_{19})}{RT}}}{\sum e^{\frac{-E(\alpha_i, \alpha_1, \ldots, \alpha_{i-1}, \alpha_{19-(i-1)}, \ldots, \alpha_{19})}{RT}}}$$
$$(16)$$

with the summations over angles $\alpha_1, \ldots, \alpha_{i-1}$, $\alpha_{19-9i-1}, \ldots, \alpha_{19}$.

Case II: Strand Two Fixed in Position

The dihedral angles $\alpha_{14}, \ldots, \alpha_{19}$ corresponding to residues 17, . . . , 22 are held fixed at 180° and residues $i+1, \ldots, 16$ are in the sheet conformation. Assume that residues $1, \ldots, i-1$ are in a coil conformation (Fig. 3b). The equilibrium constant for the transformation of residue i from the coil to a sheet state corresponding to the reaction

$$\ldots C_{i-1} C_i S_{i+1} S_{i+2} \ldots \text{TT} \ldots$$
$$S_{22-(i+1)} S_{22-i} S_{22-(i-2)} \ldots \rightarrow$$
$$\ldots C_{i-1} S_i S_{i+1} S_{i+2} \ldots \text{TT} \ldots$$
$$S_{22-(i+1)} S_{22-i} S_{22-(i-1)} S_{22-(i-2)} \ldots$$

is given by

$$s_i = \frac{[S_i]}{[C_i]}. \quad (17)$$

The relative probability of finding residue i in the state α_i while residues $1, \ldots, i-1$ sample all coil configurations and all other residues are held in the sheet conformation is

$$P'(\alpha_i) = \sum e^{\frac{-E(\alpha_i, \alpha_1, \ldots, \alpha_{i-1})}{RT}} \quad (18)$$

where the summation is over angles $\alpha_1, \ldots, \alpha_{i-1}$. The ceofficient A_{α_i} in the sheet stabilization energy term is -0.53 kcal/mol (chosen to yield equilibrium constants near unity) and $A_{\alpha_i}, \ldots, A_{\alpha_{i-1}} = 0$. the absolute probability $P(\alpha_i)$ is

$$P(\alpha_i) = \frac{P'(\alpha_i)}{\sum_{\alpha_i} P'(\alpha_i)} \quad (19)$$

The equilibrium constant s_i for residue i at the sheet–coil interface is the probability of finding residue i in the sheet state over the probability of finding it in the coil state,

$$s_i = \frac{\sum P(\alpha_i)}{1 - \sum P(\alpha_i)}. \quad (20)$$

Here the summations run over angles 170°, 180°, 190°, and 200° (with the dihedral angle $\alpha_{19-(i-1)}$ fixed at 180°, the distance $r_{i,22-(i-1)}$ falls in the range 4.0–5.7 Å only for these angles). The mean energy $<E_p(\alpha_i)>$ is given by

$$<E_p(\alpha_i)> =$$

$$\frac{\sum E_p(\alpha_i, \alpha_1, \ldots, \alpha_{i-1}) e^{\frac{-E(\alpha_i, \alpha_1, \ldots, \alpha_{i-1})}{RT}}}{\sum e^{\frac{-E(\alpha_i, \alpha_1, \ldots, \alpha_{i-1})}{RT}}} \quad (21)$$

with the summations over angles $\alpha_1, \ldots, \alpha_{i-1}$ with the dihedral angles varied over the range 0°, 30°, . . . , 330°.

Case III: Strand One Fixed in Position

Following the procedure corresponding to that used for Case II, the equilibrium constants for strand one fixed in position can be calculated.

Using the methods discussed above, the equilibrium constants were calculated for residue pairs (1, 22), (2, 21), and (3, 20) with both strands free to move; for residues 1, 2, 3, 4, and 5 with strand two fixed in position; and for residues 18, 19, 20, 21, and 22 with strand one fixed in position. The resulting equilibrium constants are listed in Table I. The s values are sensitive to the value chosen for the coefficient A_α at the sheet–coil interface. The values $A\alpha_i = -1.43$ kcal/mol (with both strands free to move) and $A_{\alpha_i} = -0.53$ kcal/mol (with either strand fixed in position) were chosen to yield equilibrium constants near unity. The equilibrium constants for both strands free to move and for strand two fixed in position are similar (Tables Ia and b), whereas the equilibrium constants for strand one fixed in position are somewhat different reflecting the fact that the two strands, being antiparallel, are distinguishable.

The equilibrium constants show an approxiately linear dependence on the coefficient A_{α_i}. For example, one finds $s = 1.2$ for residue 3 for $A_{\alpha_i} = -0.70$ kcal/mol, and $s = 0.66$ for residue 3 for $A_{\alpha_i} = -0.25$ kcal/mol, with strand two fixed in position. A similar variation is found in the other two cases.

It is seen from Table I that the equilibrium constants have an alternating behavior. This is a result of the alternating behavior of the mean energy of residues at the sheet–coil interface (discussed in detail below). Values for the equilibrium constants in Table Ib (strand two fixed) and Table Ic (strand one fixed) are not identical because the strands are antiparallel. The different values for residues 1 and 22 (in Table Ib and c) reflect the fact that the two residues have different mean energy patterns (see Fig. 4), even though $A_{\alpha_i} = -0.53$ kcal/mol in both cases.

The mean energies for residue 1 and residue 22 at the sheet–coil interface (with both strands free to move) calculated with Eq. (16) are shown in Figure 4. While the results for residues 3, 5, 7 and 17, 19, 21 are similar to those shown for residue 1, residues 2, 4, 6 and 16, 18, 20 have mean energies similar to those for residue 22, i.e., the neighboring residues (adjacent ones or opposite residues in two strands) are in different environments. The mean energy curves for the residues with either strand fixed in position are similar (the minimum at 180° is about −0.53 kcal/mol) to the ones shown in Figure 4. The mean energy for residue 1 in the angular range 220° < α < 330° is dominated by repulsive excluded volume contributions, while in the range of angles 30° < α < 130° the intrinsic torsional potentials are important. For residue 22, the dominant potentials are reversed in the same angular ranges. When residue 1 is in the dihedral angular range 220° < α < 330°,

TABLE I. Sheet–Coil Equilibrium Constants

a. Both Strands Free to Move	
Residue pair	s
1,22	1.05
2,21	0.82
3,20	0.90
b. Strand Two Fixed in Position	
Residue	s
1	1.08
2	0.89
3	1.00
4	0.88
5	1.01
c. Strand One Fixed in Position	
Residue	s
22	0.92
21	0.82
20	0.70
19	0.92
18	0.70

residue 22 is within van der Waals radial distance of residue 1 and has a large excluded volume barrier. Similarly, residue 1 is within van der Waals radial distance of residue 22 when its dihedral angle is in the range 30° < α < 130°.

DYNAMIC PROPERTIES

The dynamic properties of the β-hairpin model have been studied by carrying out a stochastic dynamics simulation of the Brownian motion of the system in aqueous solvent. The motion of a residue is largely determined by the time variation of its nonbonded interactions with neighboring residues. These interactions produce forces that act to speed up or slow down the motion in a given direction. In the simulations, the hydrodynamic interactions among residues are not included; the diffusion tensor is diagonal and constant. The form of the Langevin equation which describes the motion of residues in the diffusive limit is

$$\frac{d\mathbf{r_i}}{dt} = -\frac{D}{k_bT}\frac{\partial V}{\partial \mathbf{r_i}} + \frac{D}{k_bT}\mathbf{R_i} \tag{22}$$

where D is the diffusion constant with the value 8.9138×10^{10} Å^2/s, corresponding to alanine residues with Stokes radius 2.75 Å in water at 300 K; the viscosity equals 0.8513 cp; $\mathbf{r_i}$ is the position vector of the ith residue and $\mathbf{R_i}$ represents the random force exerted on the ith residue by the surrounding water molecules. The Langevin equations were integrated using the algorithm of Ermak and McCammon.[30] This yields the position of the ith residue at time $t + \Delta t$ as

$$\mathbf{r_i}(t + \Delta t) = \mathbf{r_i}(t) - \frac{D}{k_b T}\frac{\partial}{\partial \mathbf{r_i}} V[\mathbf{r_i}(t)]\Delta t + \chi_i \quad (23)$$

where χ_i is the displacement due to the random force. The quantities on the right-hand side of Eq. (23) are evaluated at the beginning of the time step. The time step Δt was chosen to be small enough (0.03 ps) so that the systematic forces on the residues do not change appreciably during a step.

The simulations were performed for the polypeptide chain for the three cases considered in the previous section; the two cases with one strand fixed in position and from one to five free residues in the other strand and the case where residues in both strands are free to move. The time histories of the dihedral angles α_i and distances $r_{i,22-(i-1)}$ were used to study the behavior of the sheet structure. In the simulations, the residues in both strands were initially in the sheet conformation.

Strand 1: 2 Residues

A trajectory of 90 ns was performed for the first two residues in strand one while all other residues were held in their initial positions [i.e., $\alpha_3 = \cdots = \alpha_6 = \alpha_{14} = \cdots = \alpha_{19} = 180°$, $\alpha_7 = \cdots = \alpha_{13}$ in the hairpin configuration (see above)]. This simulation is labeled $S_1(1,2)$. The total energy for the two residues as a function of their dihedral angles α_1 and α_2, respectively, is shown in Figure 5. The excluded volume peak is cut off at 10 kcal/mol. The minimum M1 in Figure 5b corresponds to the conformation with all residues in the sheet, and the minimum M2 corresponds to the relatively stable conformation in which residue 1 is in the coil state and residue 2 is in the sheet conformation. This is the potential which residues 1 and 2 follow when they move (see Fig. 4a). The time histories of parameters α_1, α_2 and $r_{1,22}$, $r_{2,21}$ are plotted in Figure 6a–l at 120 ps intervals for the entire simulation. The high potential barrier caused by the excluded volume term in the potential energy function (see Fig. 5b) in the range $30° < \alpha_2 < 160°$ keeps α_2 above 160° for all times, while almost all angular space is available to α_1. Figure 6 shows that residue 1 fluctuates mainly between 0° and 200° and the distance varies mainly between the minimum of 4.0 Å and 13 Å. It is apparent from Figures 6 and 7, that the motion of residue 2 is bound to a narrow region in both linear and angular space because it is at the sheet–coil inteface. It is limited to smaller angles by the excluded volume term in the potential energy function, and to higher angles by the intrinsic torsional energy (Fig. 5). The angular contributions to the potential energy of residue 2 are similar to those shown in Figure 4b. When residue 1 unravels, residue 2 enters the sheet–coil interface and is most likely to follow residue 1, and break its hydrogen bond with residue 21. To clarify this behavior, a part of the simulation in

the time interval 9–11 ns is plotted in Figure 7a and b at 1.2 ps intervals. Sequential motion is apparent here, especially in the range 9–10 ns. Both residues 1 and 2 have their dihedral angles mainly in the sheet region (horizontal dashed lines in Fig. 7b). However, their similar distance traces are primarily outside the sheet region (horizontal dashed lines in Fig. 7a). It is also clear that residue 2 moves back to the sheet conformation more often, while residue 1 spends more time in coil conformations. This is shown particularly well in the time interval 10–11 ns. Residue 2 stays in the sheet conformation from 10 to 10.6 ns and then enters the coil region, whereas residue 1 is the coil region throughout most of this time interval.

The probability distributions of the simulation parameters α_1, α_2, $r_{1,22}$, and $r_{2,21}$ are shown in Figure 8 for the complete 90 ns simulation. Figure 8 confirms that residue 2 is less mobile than residue 1; i.e., α_2 is constrained by the excluded volume potential energy barrier (see Fig. 5) to remain mainly above 180°, whereas α_1 can take on most values in the angular space. The major peaks in Figure 8c and d near 200° correspond to the global minimum M1 in Figure 5b. The probabilities of separately finding α_1, α_2, $r_{1,22}$ and $r_{2,21}$ in the β-sheet region (namely, 4.0 Å $\leq r \leq$ 5.7 Å, 135° $\leq \alpha \leq$ 225°) are, respectively, $P_{\alpha_1} = 0.58$, $P_{\alpha_2} = 0.87$, $P_{r_{1,22}} = 0.25$, and $P_{r_{2,21}} = 0.55$. The probability of finding residue one in the sheet state is $P_1^{sheet} = 0.22$ and the probability of finding both residues in the sheet conformation (that is, 4.0 Å $\leq r_{1,22} \leq$ 5.7 Å, 135° $\leq \alpha_1$ and $\alpha_2 \leq$ 225°) is 0.19. The small peaks around 60° for α_1 and 9 Å for $r_{1,22}$ correspond to the conformation in which residue 1 is in the coil state and residue 2 is still in the sheet conformation. This is a relatively stable conformation with potential energy corresponding to the second minimum M2 in Figure 5.

We can further quantify the dynamic behavior by calculating the autocorrelation function $C_{AA}(\tau)$ that shows how the value of some dynamic parameter A at time t is correlated with the value of A at time $t + \tau$:

$$C_{AA}(\tau) = \langle A(t)A(t+\tau)\rangle \quad (24)$$

where the average is taken, in principle, over all times t. For a trajectory (time-series) calculated at N discrete time steps t_i, Eq. (24) is approximated by

$$C_{AA}(t_k) = \frac{\frac{1}{N}\sum_{i=1}^{N-k} A(t_i)A(t_i + t_k)}{\frac{1}{N}\sum_{i=1}^{N}[A(t_i)^2]} \quad (25)$$

the denominator normalizes the autocorrelation function to unity at $\tau = 0$. The cross-correlation
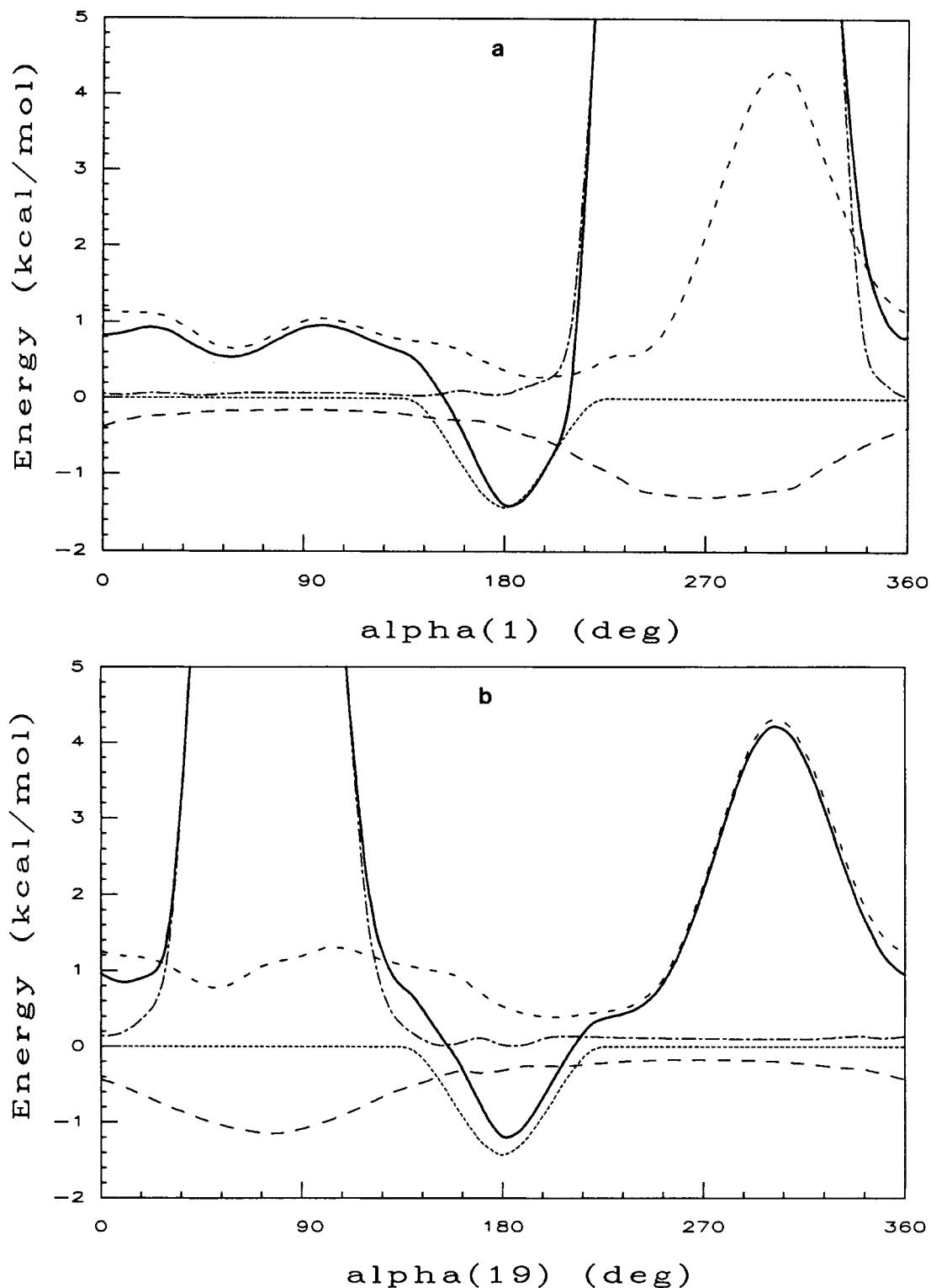
Fig. 4. (a) Mean energy [calculated from Eq. (16)] and its components for residue 1 at the sheet–coil interface. The notation is E (solid line), E_α (dotted line), E_{ev} (dashed–dot–dashed line), E_{sol} (dash–dot–dot–dot–dashed line), and E_a (dashed line). (b) Mean energy and its components for residue 22 at the sheet–coil interface. The notation is E (solid line), E_α (dotted line), E_{ev} (dash–dot–dashed line), E_{sol} (dash–dot–dot–dot–dashed line), and E_β (dashed line).

a

b
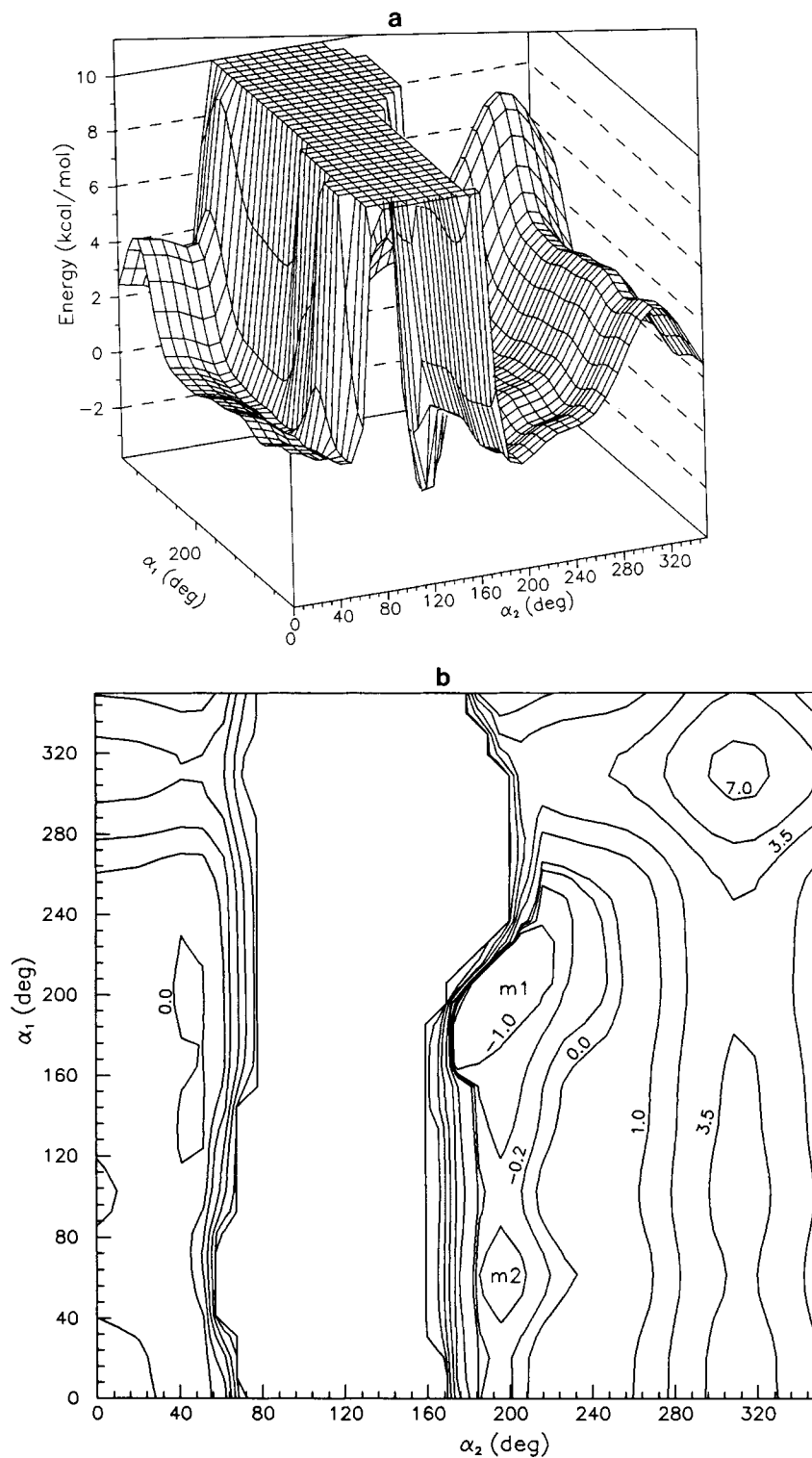
Fig. 5. **(a)** Total energy [calculated from Eq. (1)] versus α_1 and α_2 for the unraveling of residues 1 and 2 in strand one with all other residues fixed in the initial sheet position. The global minimum at (α_1 = 180°, α_2 = 180°) is −3.53 kcal/mol. E is cut off at 10 kcal/mol. **(b)** Contour plot of total energy E as a function of α_1 and α_2. M1 is the global minimum at −3.53 kcal/mol and M2 is a local minimum at −0.53 kcal/mol.

6a

6b

6c

6d

6e

6f

6g

6h
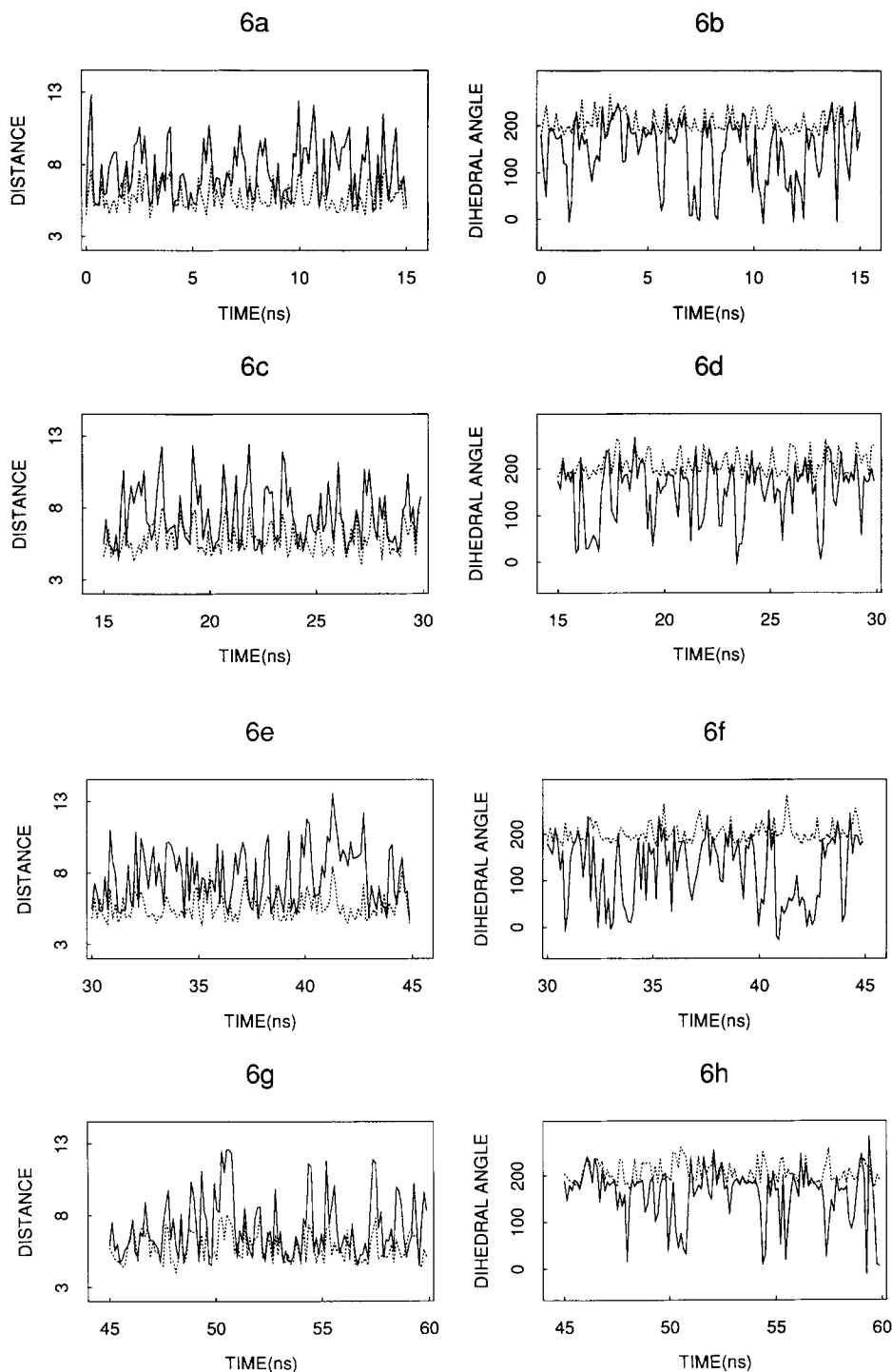
Fig. 6a–h.   Legend appears on page 249.

function C_{AB} (τ) of two dynamic parameters A and B is defined by the generalization of Eqs. (24) and (25) to

$$C_{AB} (\tau) = <A\ (t)\ B\ (T + \tau) > \qquad (26)$$

$$C_{AB}\ (t_k) = \frac{1}{N} \sum_{i=1}^{N-k} \frac{A\ (t_i)\ B\ (t_i + t_k)}{\sqrt{\frac{1}{N} \sum_{i=1}^{N} [A\ (t_i)^2]\ \frac{1}{N} \sum_{i=1}^{N} [B\ (t_i)^2]}} \qquad (27)$$
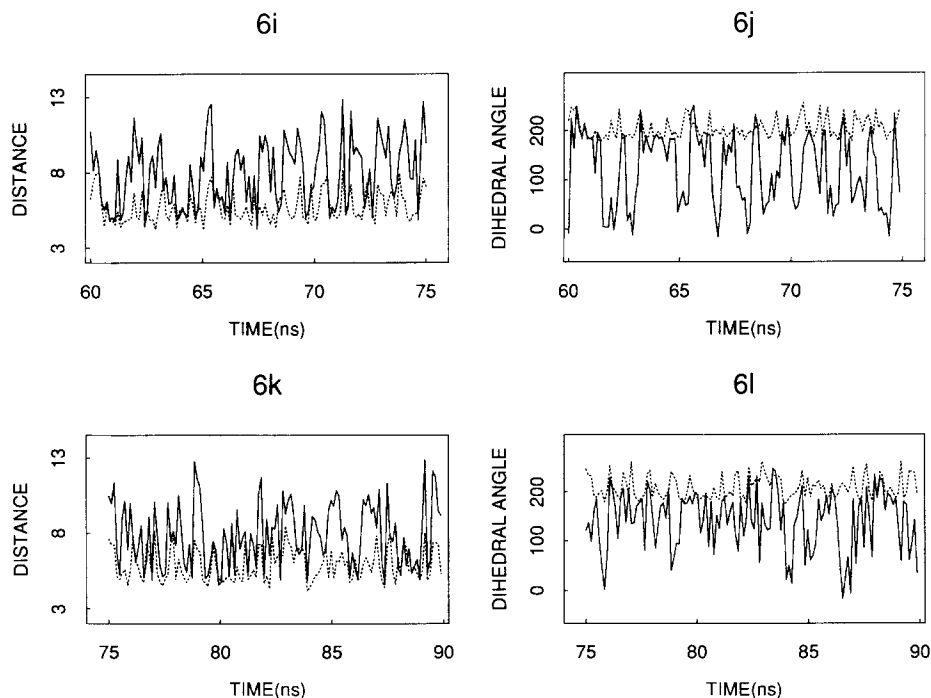
Fig. 6. Ninety nanosecond time histories (in 15-ns increments) of the distances and dihedral angles for simulation S_1 (1,2). In the distance versus time trajectories (a,c,e,g,i,k), the solid line is $r_{1,22}$ and the dotted line is $r_{2,21}$. In the dihedral angle plots, the solid line is α_1 and the dotted line is α_2. The time points are 120 ps apart.

Note that C_{AB} (τ) is not necessarily equal to one at τ = 0. In the discussion below, the dynamic parameters used will be taken to be the fluctuations relative to the mean values. The auto- and cross-correlation functions for the stimulation parameters $r_{1,22}$, $r_{2,21}$, α_1, and α_2 are shown in Figure 9a–d. The distance autocorrelation functions (Fig. 9a) have mean relaxation times t^{relax}, defined by

$$t^{relax} = \int_o^\infty C\,(\tau)\,d\tau \qquad (28)$$

of 105 ps ($r_{1,22}$) and 46 ps ($r_{2,21}$). The dihedral angle autocorrelation functions (Fig. 9c) have similar relaxation times of 135 ps (α_1) and 47 ps (α_2). This is not surprising since α_1, $r_{1,22}$ and α_2, $r_{2,21}$ are coupled. That the correlation functions associated with residue 2 relax more rapidly is interesting; it is due to the presence of smaller more rapid fluctuations (see Fig. 6). The distance cross-correlation functions are shown in Figure 9b. There is a strong initial correlation between the two distances of 0.46. The cross-correlation function is not symmetrical and $<r_{1,22}$ (t) $r_{2,21}$ (t+τ)> decays more rapidly to zero than $<r_{2,21}$ (t) $r_{1,22}$ (t+τ)>. The α_1, α_2 cross-correlation functions are shown in Figure 9d. In this case the initial correlation is weak and fluctuates near zero. In helix–coil simulations of simplified residue systems,[5,27] the helical (and coil) state of a residue is determined by its dihedral angle alone. In this system, more than 10% of the trajectories show strong

dihedral angle correlations in that the unwinding (to the coil state) of a residue is simultaneous with or closely followed by the unwinding of the next residue in the helix.

Strand 2: 2 Residues

A similar procedure was followed to study the unraveling of the last two residues in strand two, residues 21 and 22, while the rest of the system remains fixed in the sheet conformation (i.e., α_1 = ⋯ = α_6 = α_{14} = ⋯ α_{17} = 180°, α_7 = ⋯ = α_{13} in the hairpin configuration). This simultation was also carried out for a period of 90 ns. The variable parameters for this simulation are the dihedral angles α_{18}, α_{19} and the distances $r_{1,22}$, $r_{2,21}$. This simulation is labeled S_2 (21,22). Unlike residue 2 in the simulation discussed above, residue 21, at the sheet–coil interface, is not bound to a narrow angular range. The potential energy function for the unraveling, which is shown in Figure 10, exhibits a smooth surface for α_{18} except for the angular range 200° < α_{18} < 300°. This is due to a lower torsional barrier for smaller angles. The potential of mean force for residue 21 is similar to the potential function shown in Figure 4a. One can clearly see from the time histories of this simulation (Fig. 11) that residue 21 stays in coil conformations for longer periods before drifting back to the sheet state. Moreover, sequential unraveling is more visible in this
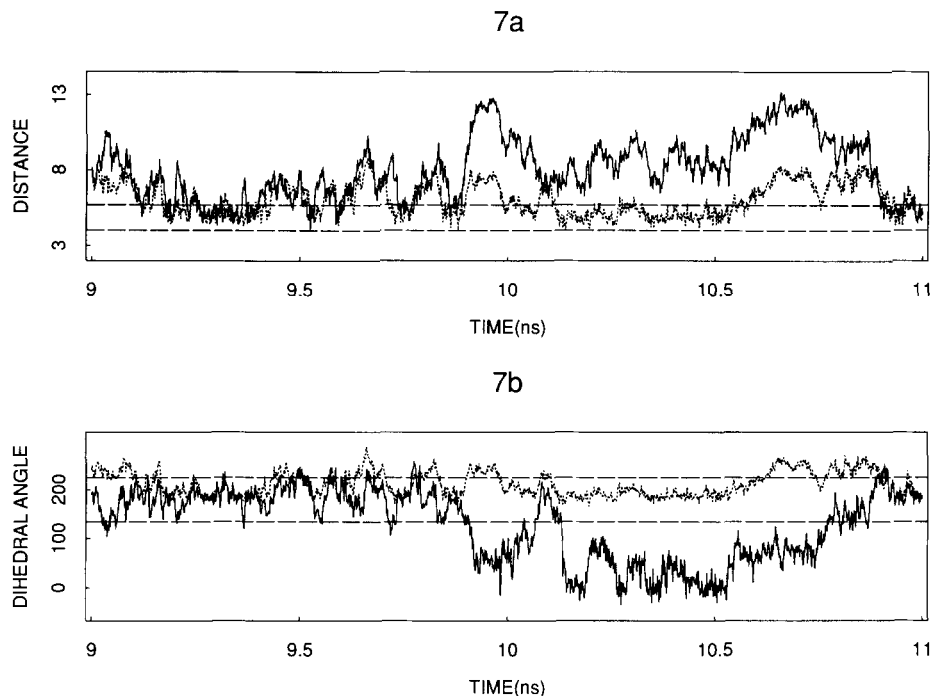
7a

7b

Fig. 7.   (a) A blow-up of a portion of Figure 6a for the time range 9–11 ns. (b) A blowup of a portion of Figure 6b for the time range 9–11 ns.

simulation than in the strand two fixed case. Often, when residue 22 moves away from the sheet, residue 21, which is now at the interface readily enters the coil state and the system moves toward the minimum M3 in Figure 10, where both α_{18} and α_{19} are around 50°. This is apparent in Figure 12 where the portion of the simulation between 0.3 and 0.9 ns is shown with points separated by 1.2 ps. The probability distributions for the distances $r_{1,22}$, $r_{2,21}$ and dihedral angles α_{19}, α_{18} are shown in Figure 13. That for α_{19} and $r_{1,22}$ is similar to the strand one results. However, the secondary peaks in the probability distributions of $r_{2,21}$ and α_{18} shown in Figure 13 correspond to simultaneous stable coil conformations of residues 21 and 22. The peak around 50° for α_{19} is partly due to these two-residue coil conformations, and partly due to configurations where residue 22 is in the coil state while residue 21 is still in the sheet conformation (corresponding to minimum M2 in Fig. 10). The probability of finding residue 22 in the sheet state is $P_{22}^{sheet} = 0.144$ and the joint probability of finding both residues in the sheet is 0.138.

The time correlation functions for the fluctuations of the angles and distances are shown in Figure 14. In Figure 14a, the autocorrelation function for the fluctuations in the distances $r_{1,22}$ and $r_{2,21}$ corresponds to a single decaying expontial with very small oscillations. The relaxation times calculated

according to Eq. (28) are $t_{r_{1,22}}^{relax} = 207$ ps and $t_{r_{2,21}}^{relax} = 255$ These times are longer than those found for the equivalent simulation with strand two fixed. Also, in contrast to S_1 (1,2), the two relaxation times are very similar. The cross-correlation functions of the two distances are shown in Figure 14b. The two correlation functions $<r_{1,22}\,(t)\,r_{2,21}\,(t + \tau)>$ and $<r_{2,21}\,(t)\,r_{1,22}\,(t + \tau)>$ are almost identical in shape and magnitude. The relaxation times are $t_{r_{1,22,2,21}}^{relax} = t_{r_{2,21,1,22}}^{relax} = 288$ ps to decay to $1/e$ of the initial value of 0.76. These distance cross-correlations are substantially higher, and persist longer than the corresponding correlations with strand two fixed [S_1 (1,2)] discussed above.

The autocorrelation functions for the dihedral angles α_{19} and α_{18} are shown in Figure 14c. These functions have an essentially exponential character with relaxation times [computed with Eq. (28)] of $t_{\alpha_{19}}^{relax} = 178$ ps and $t_{\alpha_{18}}^{relax} = 337$ ps. The autocorrelation of α_{19} is similar to that of α_1 in S_1 (1,2) (see Fig. 9c). However, the autocorrelation of α_{18} is much more persistent than that of the comparable dihedral angle, α_2, of S_1 (1,2). As is the case for S_1 (1,2), the dihedral cross correlation functions (Fig. 14d) are small (0.16 at $t = 0$) with minor oscillatory behavior around zerio. Thus, in both of the simulations [S_1 (1,2) and S_2 (21,22)], the main cross-correlation occurs between the distance functions $r_{1,22}$ and $r_{2,21}$.
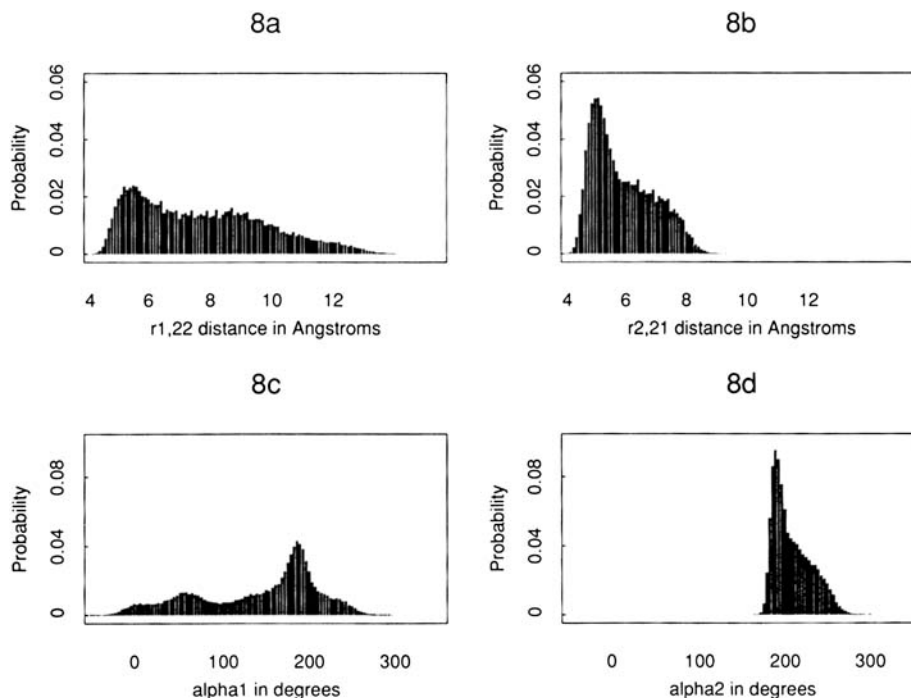
Fig. 8. Probability distributions of the distances and dihedral angles for the 90-ns simulation S_1 (1,2).

Strand 1: 4 Residues

To investigate the mechanism of unfolding of the sheet further, the unraveling of the first four residues in strand one was studied, while the rest of the system remained fixed in the sheet conformation [simulation S_1 (1,2,3,4)]. The variable parameters of the simulation are the dihedral angles α_1, α_2, α_3, α_4 and the distances $r_{1,22}$, $r_{2,21}$, $r_{3,20}$, $r_{4,19}$. A portion of the time history of this simulation is shown in Figure 15. In this time period, residues 1 and 2 had cross-strand distances mainly out of the sheet range, and distance $r_{2,21}$ tends to follow distance $r_{1,22}$ (see below for a quantitative description of this tendency in terms of the cross-correlation function for this pair of distances). Residues 3 and 4 have distances mainly within the sheet range. From Figure 15b, we see that all the dihedrals are mainly within the sheet range, so residues 3 and 4 were in the sheet state and residues 1 and 2 were mainly in the coil state in the time period. The probability distributions of the distances and dihedral angles for S_1 (1,2,3,4) are shown in Figure 16a–h. Figure 16c and d shows that the $r_{3,20}$ and $r_{4,19}$ distributions for residues 3 and 4 are almost identical to Figure 8a and b for the $r_{1,22}$ and $r_{2,21}$ distributions for residues 1 and 2 in S_1 (1,2). This result is reasonable, because residue 4 in S_1 (1,2,3,4)] and residue 2 in S_1(1,2) are at the sheet–coil interface and residue 3 [S_1(1,2,3,4)] and residue 1 [S_1 (1,2)] are in a comparable position

with respect to the sheet–coil interface. For the same reason, Figure 16g and h for the α_3 and α_4 distributions in S_1 (1,2,3,4) and Figure 8c and d for the α_1 and α_2 distributions in S_1 (1,2) are almost identical. Residues 3 and 4 in S_1 (1,2,3,4) and residues 1 and 2 in S_1 (1,2) experience almost the same potential of mean force (see Figs. 4 and 5).

In S_1 (1,2,3,4), residues 1 and 2 are further from the sheet–coil interface and, consequently have more freedom of movement. This is apparent in Figure 16a and b, which presents the probability distributions for $r_{1,22}$ and $r_{2,21}$ (note the lengthened distance scale in this figure relative to those for $r_{3,20}$ and $r_{4,19}$). Residue 1 can be almost anywhere (from the sheet region to more than 20 Å from residue 22), and residue 2, although more localized, can still be found more than 15 Å from residue 21. The dihedral angle distributions of α_1 and α_2 shown in Figure 16e and f resemble one another and also the α_3 distribution in Figure 16g, although the latter distribution has sharper peaks near 50° and 180°. The angular potentials of mean force experienced by residues 1 and 3 in S_1 (1,2,3,4) are somewhat similar and also similar to that experienced by residue one in S_1 (1,2); residues 2 and 4 correspond to residue 2 in S_1 (1,2).

Figure 17a–g shows the auto- and cross-distance correlation functions for this simulation. From Figure 17a, $r_{1,22}$, $r_{2,21}$, and $r_{3,20}$ have relatively long
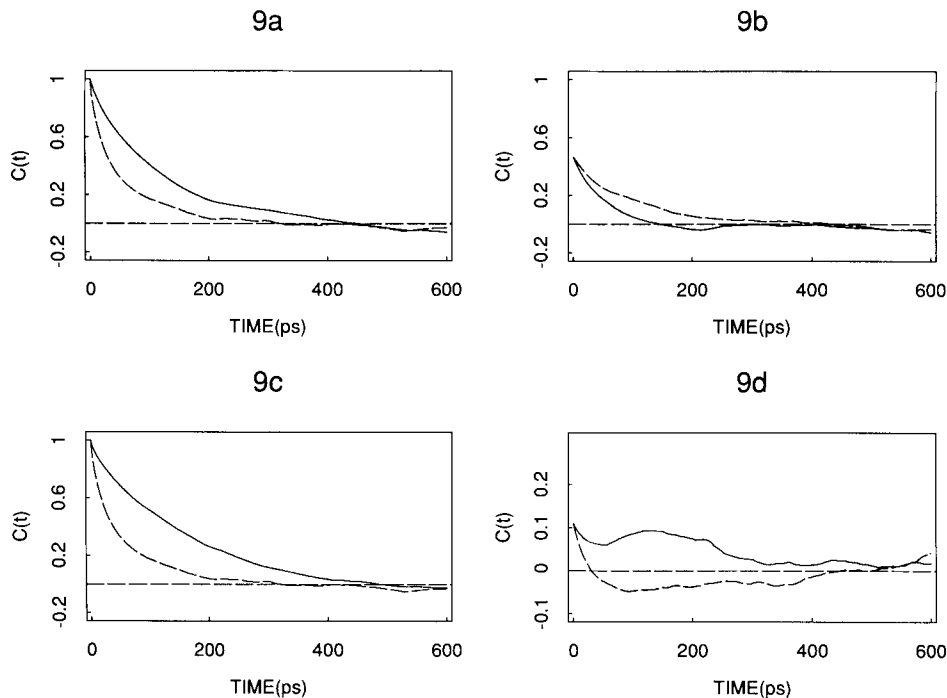
9a

9b

9c

9d

Fig. 9.   Auto- and cross-correlation functions [calculated using Eqs. (25) and (27)] for the simulation S_1 (1,2). (a) Autocorrelation functions versus time for $r_{1,22}$ (solid line) and $r_{2,21}$ (dashed line). (b) Cross-correlation functions versus time for $<r_{1,22}$ (t) $r_{2,21}$ ($t+\tau$)> (solid line) and $<r_{1,21}$ (t) $r_{1,22}$ ($t+\tau$)> (dashed line). (c) Autocorrelation functions versus time for α_1 (solid line) and α_2 (dashed line). (d) Cross-correlation functions versus time for $<\alpha_1$ (t) α_2 ($t+\tau$)> (solid line) and $<\alpha_2$ (t) α_1 ($t+\tau$)> (dashed line).

autocorrelation relaxation times (values are, respectively, 502, 471, and 447 ns) while $r_{4,19}$ has a shorter relaxation time (133 ns). The cross-correlation of $r_{1,22}$ and $r_{2,21}$ (Fig. 17b) is very strong, starting at 0.90 at zero time and persisting in a manner similar to the autocorrelations of the individual distances. The cross-correlation of $r_{1,22}$ and $r_{3,20}$ (Fig. 17c) starts at 0.79 at zero time and decays like the autocorrelation functions. The distance cross-correlation for residues 1 and 4, namely, that between $r_{1,22}$ and $r_{4,19}$ (Fig. 17d), is the weakest of the distance cross-correlations in this simulation. It begins at 0.27 for zero time and decays more quickly than the other cross-correlations of the $r_{1,22}$ distance. The cross-correlation of the distances $r_{2,21}$ and $r_{3,20}$ (Fig. 17e) begins at 0.86 for zero time and relaxes qualitatively and quantitatively in a manner similar to the cross-correlation of $r_{1,22}$ and $r_{2,21}$. The cross-correlation of $r_{2,21}$ and $r_{4,19}$ (Fig. 17f) starts at 0.40 at zero time and relaxes to zero like the $r_{1,22}:r_{4,19}$ cross-correlation. In Figure 17g, the cross-correlation of $r_{3,20}$ and $r_{4,19}$ begins at 0.44 at zero time and relaxes to zero like the other auto- and cross-correlations involving residue 4. All the distance correlations are symmetrical.

It is clear from Figure 17a–g that the nearest-neighbor residue distances are very strongly correlated and that some second neighbors also have

strong correlations. If residue 1 moves from the sheet conformation, by having $r_{1,22} > 5.7$ Å, then residues 2 and 3 and to a lesser extent residue 4 will immediately follow and leave the sheet distance range. The weaker distance cross-correlation between residues 3 and 4 (Fig. 17g) is similar (with a longer relaxation time) to that between residues 1 and 2 (Fig. 9b) in simulation S_1 (1,2). In the latter case, however, the symmetry is not as pronounced, with the cross-correlation with $r_{1,22}$ decaying more rapidly. The auto- and cross-distance correlations illustrated in Figure 17a–g show that the residue at the sheet–coil interface is less likely to follow the neighboring residues, and less likely to follow its own prior behavior than the coil residues away from the interface, which all behave similarly and are strongly self- and cross-correlated.

The correlation functions for the four dihedral angles α_1, α_2, α_3, and α_4 in S_1 (1,2,3,4) are shown in Figure 18a–g. With the exception of α_3, the dihedral angle correlation functions relax in a similar way with relaxation times of, respectively, 70, 97, and 135 ns. C_{α_3, α_3} (t) has a rapid exponential decay like the other dihedrals, but subsequently relaxes in an almost linear manner. Using Eq. (28), its relaxation time is 269 ns. All of the dihedral cross-correlation functions are small for all times, tending to oscillate around zero with small amplitudes. Thus,
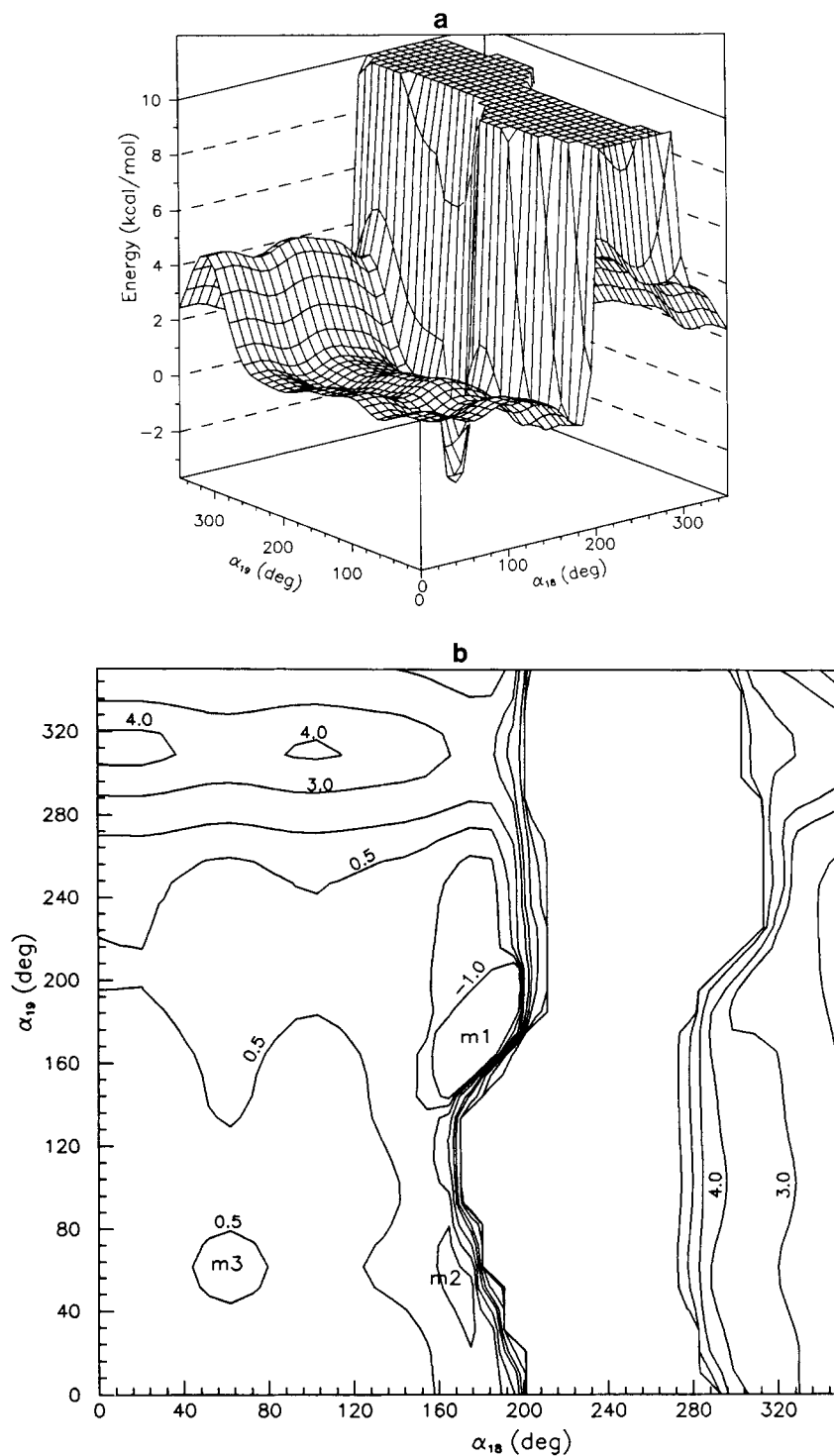
Fig. 10.   (a) Total energy [calculated from Eq. (1)] versus α_{19} and α_{18} for the unraveling of residues 22 and 21 in strand two with all other residues fixed in the initial sheet position. The global minimum at ($\alpha_1 = 180°$, $\alpha_2 = 180°$) is −3.66 kcal/mol. E is cut off at 10 kcal/mol. (b) Contour plot of total energy E as a function of α_{19} and α_{18}. M1 is the global minimum at −3.66 kcal/mol and M2 (−0.53 kcal/mol) and M3 (1.00 kcal/mol) are local minima.

254 K. YAPA ET AL.
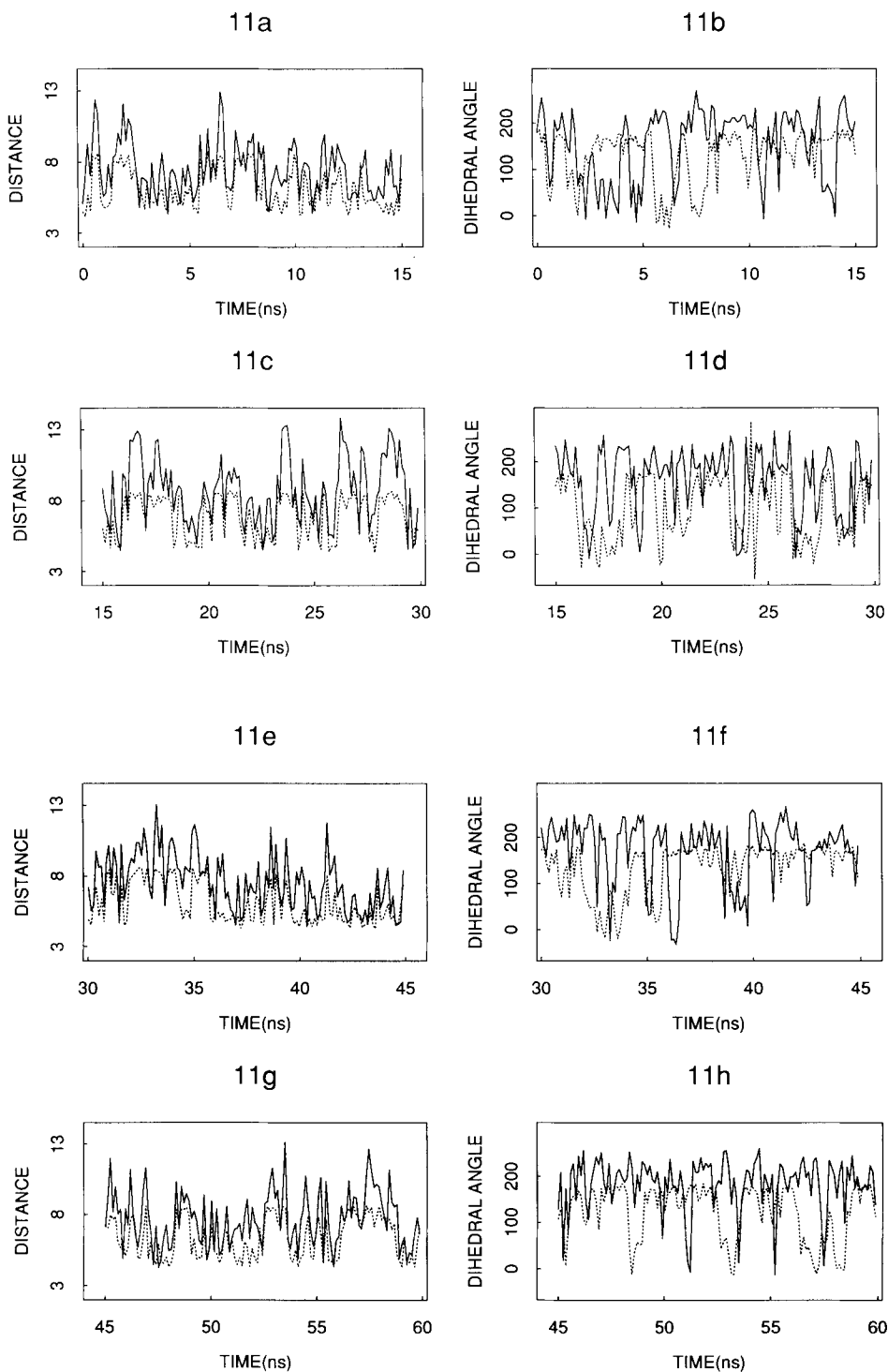
11a 11b 11c 11d 11e 11f 11g 11h

Fig. 11a–h. Legend appears on page 255.

the strong distance correlations which cause the residue positions to follow one another are essentially absent for the dihedral angles.

These correlations indicate that the completely folded sheet unfolds in a concerted way by unzippering from residue 1 to residue 4. Refolding involves the distances falling into the sheet range in a concerted manner (Fig. 17) with the dihedrals passing into the sheet range relatively independently from one another (Fig. 18).
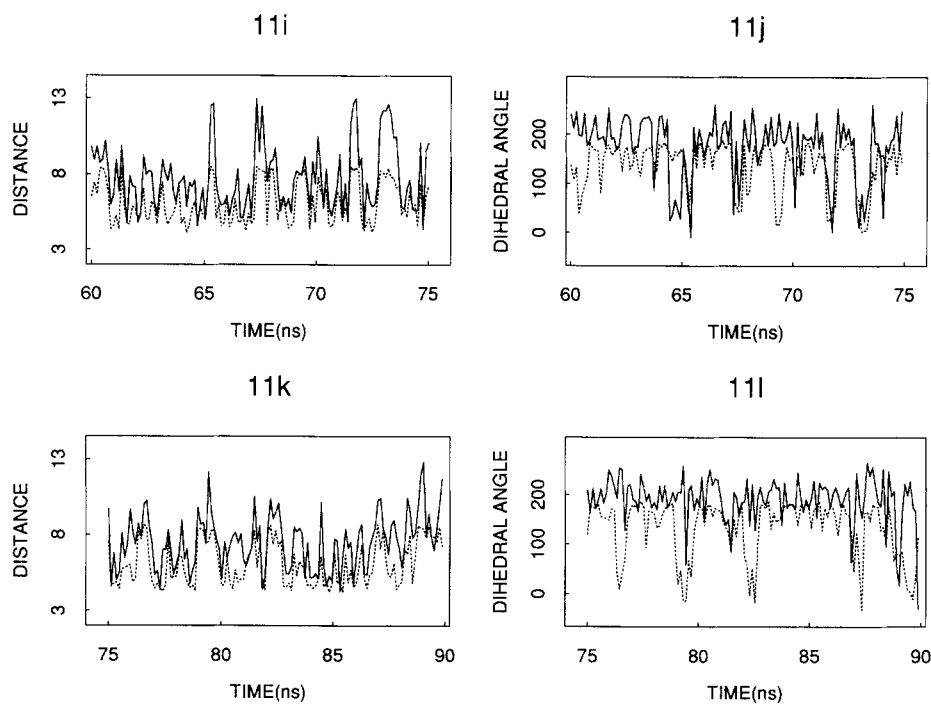
11i

11j

DISTANCE

13

8

3

60   65   70   75
TIME(ns)

DIHEDRAL ANGLE

200

100

0

60   65   70   75
TIME(ns)

11k

11l

DISTANCE

13

8

3

75   80   85   90
TIME(ns)

DIHEDRAL ANGLE

200

100

0

75   80   85   90
TIME(ns)

Fig. 11.   Ninety nanosecond time histories in 15-ns increments of the distances and dihedral angles for simulation S_2 (22,21). In the distance versus time trajectories (**a,c,e,g,i,k**), the solid line is $r_{1,22}$ and the dotted line is $r_{2,21}$. In the dihedral angle plots, the solid line is α_{19} and the dotted line is α_{18}. The time points are 120 ps apart.

12a

DISTANCE

13

8

3

0.3   0.5   0.7   0.9
TIME(ns)

12b

DIHEDRAL ANGLE

200

100

0

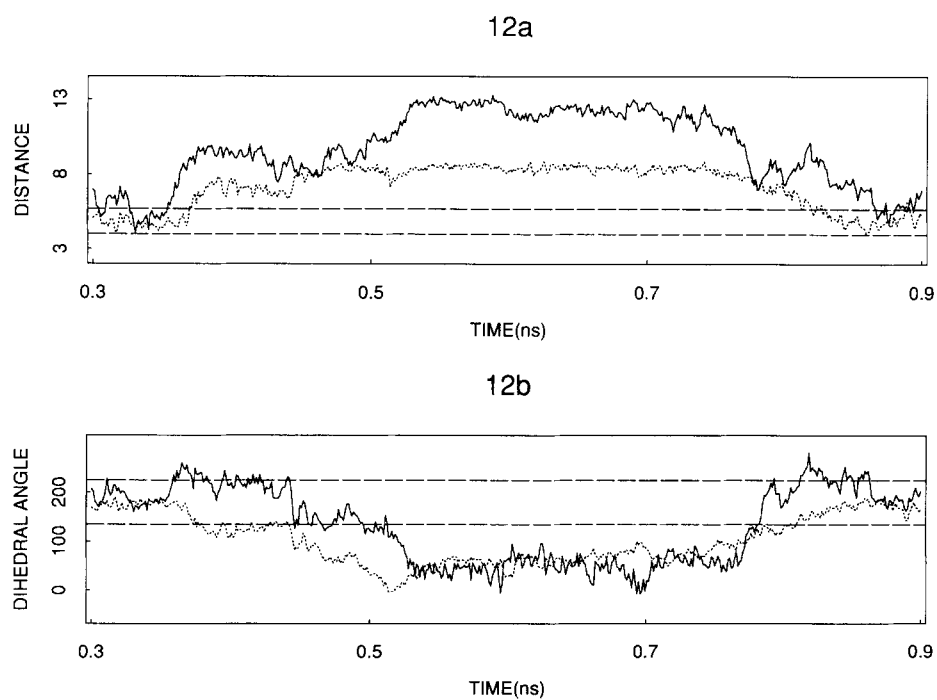0.3   0.5   0.7   0.9
TIME(ns)

Fig. 12.   (a) A blow-up of a portion of Figure 11a for the time range 0.3–0.9 ns. (b) A blowup of a portion of Figure 11b for the time range of 0.3–0.9 ns.
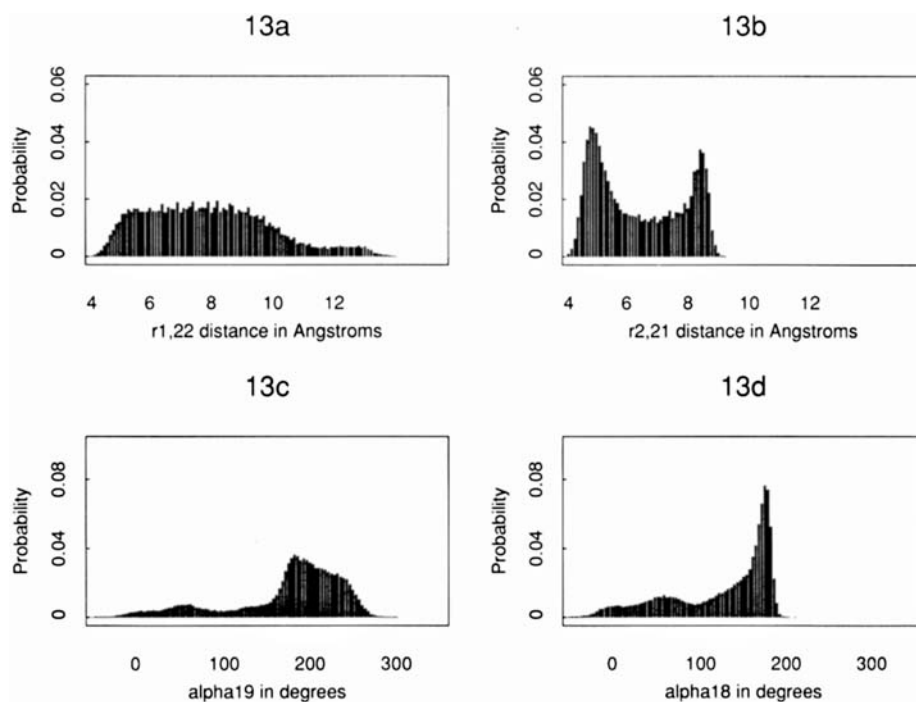
Fig. 13. Probability distributions of the distances and dihedral angles for the 90-ns simulation S_2 (22,21).
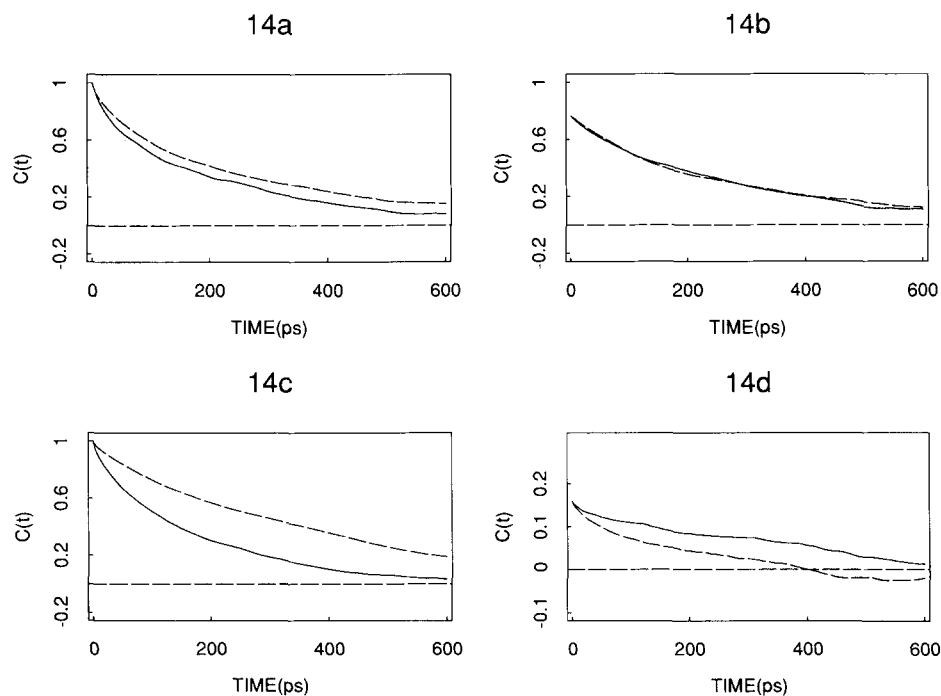

Fig. 14. Auto- and cross-correlation functions [calculated using Eqs. (25) and (27)] for the simulation S_2 (22,21). (a) Autocorrelation functions versus time for $<r_{1,22} (t) \, r_{2,21} (t+\tau)>$ (solid line) and $<r_{2,21} (t) \, r_{1,22} (t+\tau)>$ (dashed line). (b) Cross-correlation functions versus time for > (dashed line) $<r_{1,22} (t) \, r_{2,21} (t+\tau)>$ (solid line) and $< r_{2,21} (t) \, r_{1,22} (t+\tau)$ (c) Autocorrelation functions versus time for α_{19} (solid line) and α_{18} (dashed line). (d) Cross-correlation functions versus time for $<\alpha_{19} (t) \, \alpha_{18} (t+\tau)>$ (solid line) and $<\alpha_{18} (t) \, \alpha_{19} (t+\tau)>$ (dashed line).
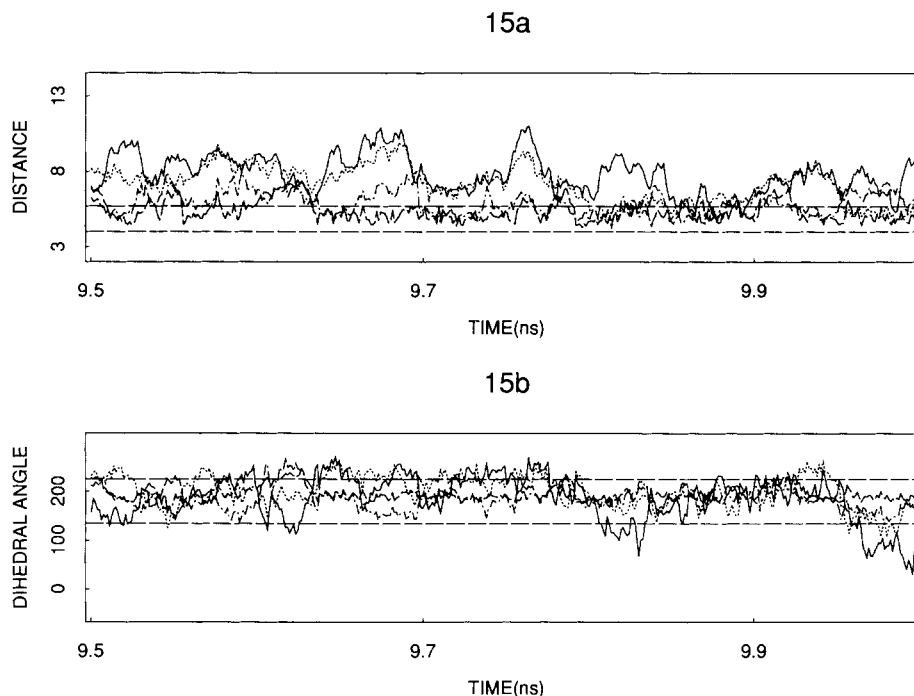
15a

15b

Fig. 15. A 0.5 ns portion of the 90-ns simulation S_1 (1,2,3,4). (a) Distance versus time trajectories for $r_{1,22}$ (solid line), $r_{2,22}$ (dotted line), $r_{3,20}$ (dash–dot–dashed line), and $r_{4,19}$ (dashed line). (b) Dihedral angle versus time for α_1 (solid line), α_2 (dotted line), α_3 (dash–dot–dashed line), and α_4 (dashed line).

Transition Rate Constants

Ten 90-ns simulations were carried out to obtain rate constants for a residue at the sheet—coil interface. These simulations studied the dynamics of residues 1 through 5 in strand one (with strand two fixed) and residues 22 through 18 in strand two (with strand one fixed). The simulations are labeled (see Tables II and III) $S_1(1)$, $S_1(1,2)$, $S_1(1,2,3)$, $S_1(1,2,3,4)$, and $S_1(1,2,3,4,5)$ for fixed strand two and $S_2(22)$, $S_2(22,21)$, $S_2(22,21,20)$, $S_2(22,21,20,19)$, and $S_2(22,21,20,19,18)$ for fixed strand one. For each simulation, the trajectories of all the residues listed in parentheses were followed for the full 90 ns with the other residues constrained in their initial sheet positions. The values of the variable cross-strand distances and dihedral angles were obtained at 1.2 ps intervals for the full 90 ns. This corresponds to 75,000 time steps for each simulation. The sheet–coil rate constants for the unconstrained residue furthest from the free end (strand one for S_1 simulations or strand two for S_2 simulations) were found from the trajectories of the corresponding simulation. For example, from $S_2(22,21,20)$ we obtained the rate constants for residue 20. To estimate the rate constants, the trajectory data were sorted, and the time points at which the residues nearest the free end (residues 22 and 21 in the example above) were in the sheet state were eliminated. The remaining data were further sorted so that the residue

at the sheet–coil interface at a given time point (residue 20 in the example above) was in the sheet or coil state. The chronologically ordered time steps of the resulting two files were analyzed for gaps of greater than 1.2 ps to produce a vector of times spent in the coil states (folding times, τ_f) or in the sheet states (unfolding times τ_u), and arithmetic mean values determined. Weighting the folding and unfolding times with the distribution of time values (see below) leads to smaller mean values, so the rate constants reported must be considered upper limits. Typical sequences [from $S_1(1,2)$] of times spent folding and unfolding are shown in Figure 19a and b. The distributions of τ_f and τ_u values for the same simulation are shown in Figure 19c and d. One sees from Figure 19a and b that the length of a particular time (τ_f or τ_u) does not depend on the lengths of the intervals that precede it, and from Figure 19c and d that the distributions of folding and unfolding times are exponential. This indicates that the residue (residue 2 in this example) switches back and forth between sheet and coil states in a stochastic (Poisson) manner. The rate constants data are summarized in Table II for the simulations with strand two fixed and in Table III for the simulations with strand one fixed.

Kinetic equilibrium coefficients s' were computed from the ratio k_f/k_u in Tables II and III. These values may be compared with the true equilibrium con-
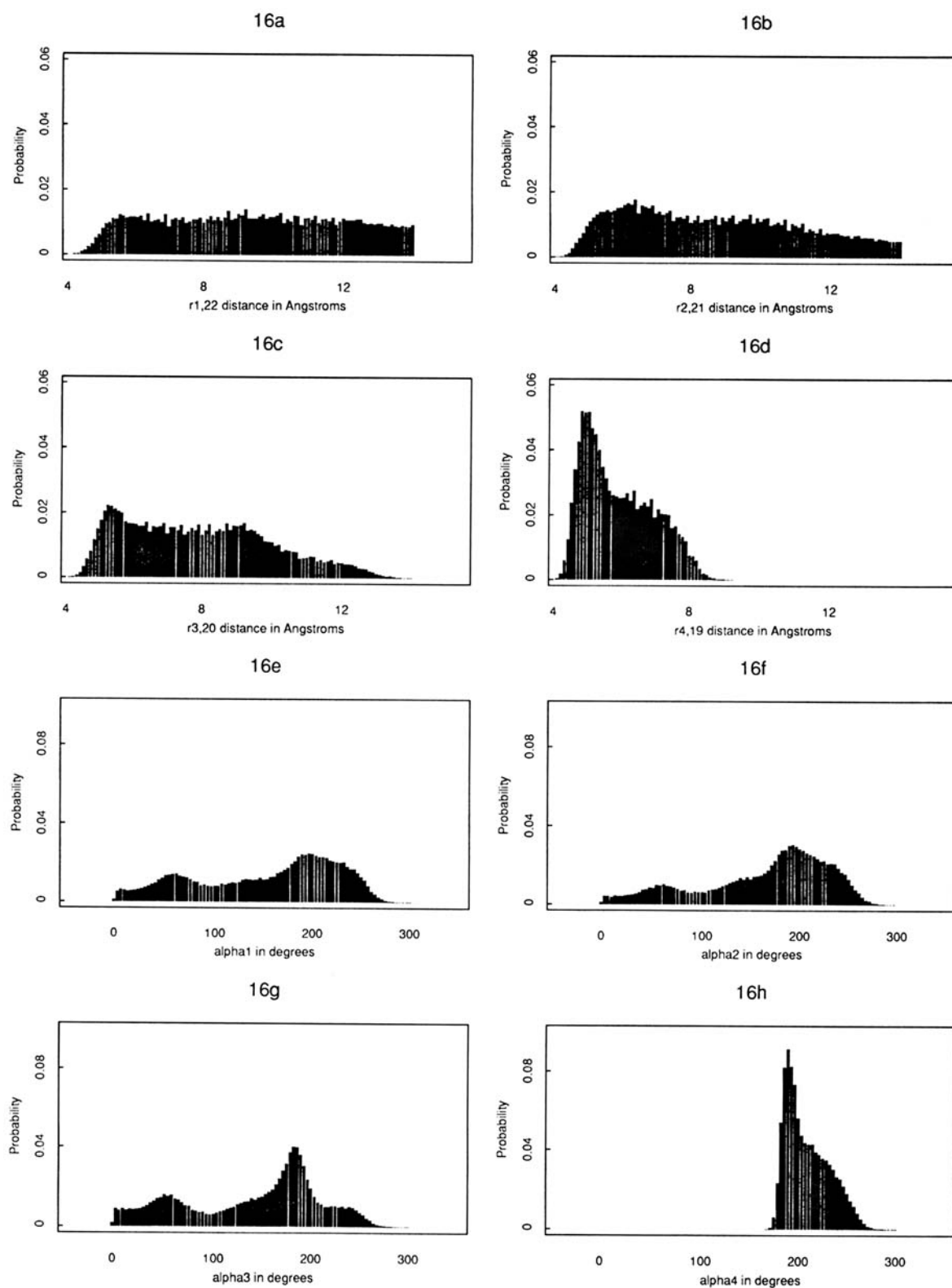
K. YAPA ET AL.

Fig. 16.   Probability distributions of the distances and dihedral angles for the 90-ns simulation S_1 (1,2,3,4).

17a

C(t)

TIME(ps)

17b

C(t)

TIME(ps)

17c

C(t)

TIME(ps)

17d

C(t)

TIME(ps)

17e

C(t)

TIME(ps)

17f

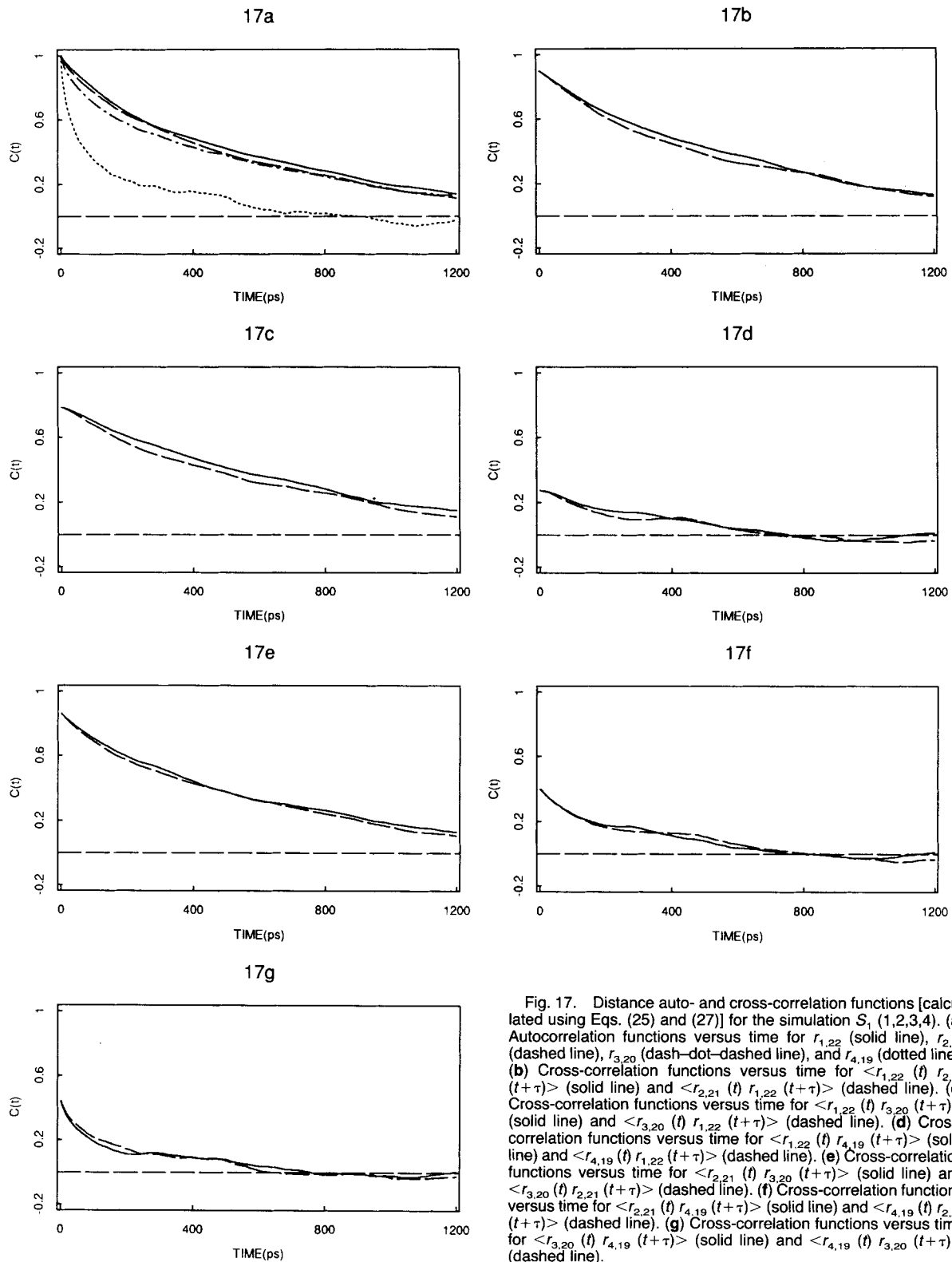C(t)

TIME(ps)

17g

C(t)

TIME(ps)

Fig. 17. Distance auto- and cross-correlation functions [calculated using Eqs. (25) and (27)] for the simulation S_1 (1,2,3,4). (a) Autocorrelation functions versus time for $r_{1,22}$ (solid line), $r_{2,21}$ (dashed line), $r_{3,20}$ (dash–dot–dashed line), and $r_{4,19}$ (dotted line). (b) Cross-correlation functions versus time for $<r_{1,22}$ (t) $r_{2,21}$ $(t+\tau)>$ (solid line) and $<r_{2,21}$ (t) $r_{1,22}$ $(t+\tau)>$ (dashed line). (c) Cross-correlation functions versus time for $<r_{1,22}$ (t) $r_{3,20}$ $(t+\tau)>$ (solid line) and $<r_{3,20}$ (t) $r_{1,22}$ $(t+\tau)>$ (dashed line). (d) Cross-correlation functions versus time for $<r_{1,22}$ (t) $r_{4,19}$ $(t+\tau)>$ (solid line) and $<r_{4,19}$ (t) $r_{1,22}$ $(t+\tau)>$ (dashed line). (e) Cross-correlation functions versus time for $<r_{2,21}$ (t) $r_{3,20}$ $(t+\tau)>$ (solid line) and $<r_{3,20}$ (t) $r_{2,21}$ $(t+\tau)>$ (dashed line). (f) Cross-correlation functions versus time for $<r_{2,21}$ (t) $r_{4,19}$ $(t+\tau)>$ (solid line) and $<r_{4,19}$ (t) $r_{2,21}$ $(t+\tau)>$ (dashed line). (g) Cross-correlation functions versus time for $<r_{3,20}$ (t) $r_{4,19}$ $(t+\tau)>$ (solid line) and $<r_{4,19}$ (t) $r_{3,20}$ $(t+\tau)>$ (dashed line).

K. YAPA ET AL.

18a

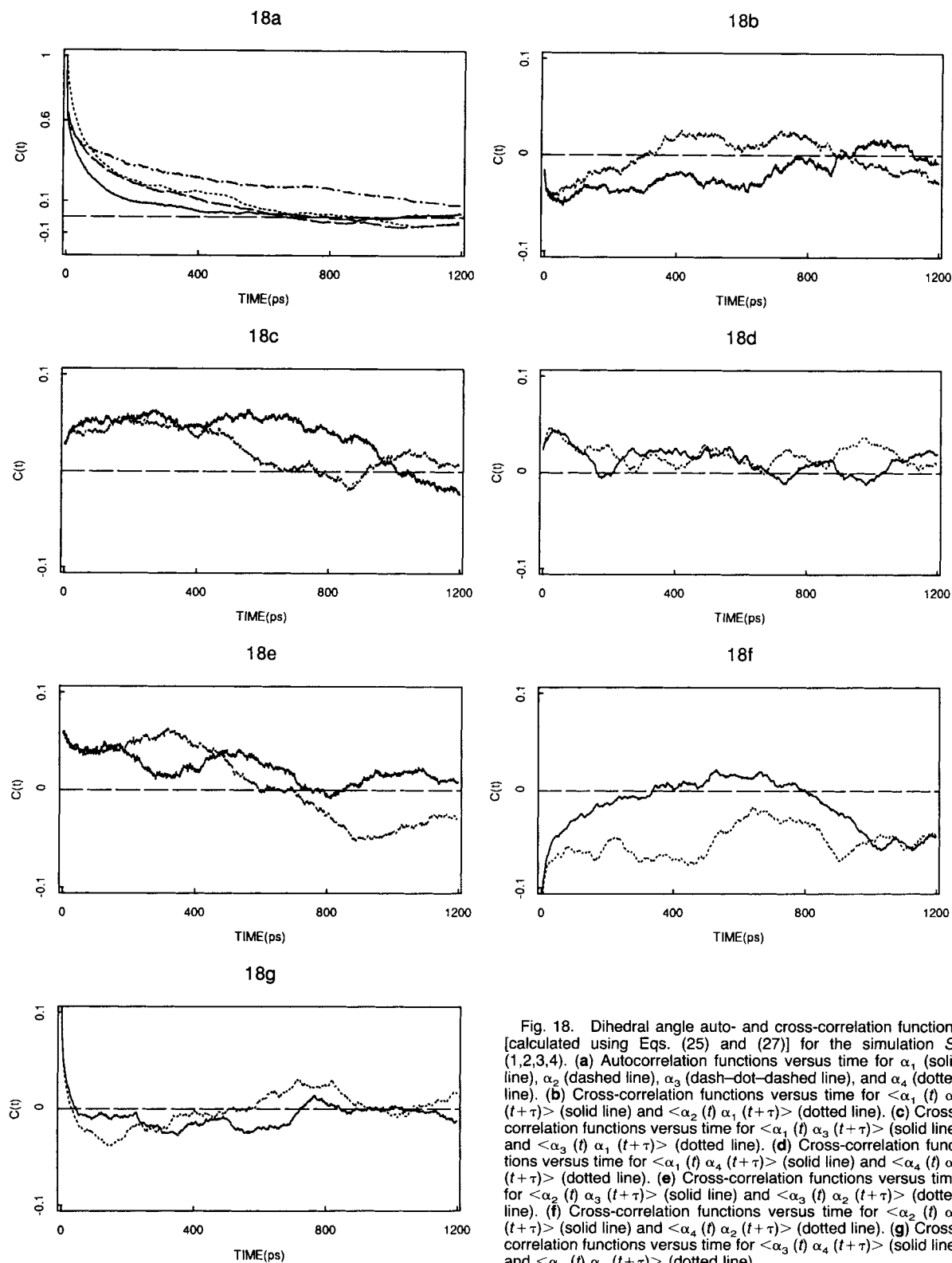

18b

18c

18d

18e

18f

18g

Fig. 18. Dihedral angle auto- and cross-correlation functions [calculated using Eqs. (25) and (27)] for the simulation S_1 (1,2,3,4). (a) Autocorrelation functions versus time for α_1 (solid line), α_2 (dashed line), α_3 (dash–dot–dashed line), and α_4 (dotted line). (b) Cross-correlation functions versus time for $<\alpha_1$ (t) α_2 $(t+\tau)>$ (solid line) and $<\alpha_2$ (t) α_1 $(t+\tau)>$ (dotted line). (c) Cross-correlation functions versus time for $<\alpha_1$ (t) α_3 $(t+\tau)>$ (solid line) and $<\alpha_3$ (t) α_1 $(t+\tau)>$ (dotted line). (d) Cross-correlation functions versus time for $<\alpha_1$ (t) α_4 $(t+\tau)>$ (solid line) and $<\alpha_4$ (t) α_1 $(t+\tau)>$ (dotted line). (e) Cross-correlation functions versus time for $<\alpha_2$ (t) α_3 $(t+\tau)>$ (solid line) and $<\alpha_3$ (t) α_2 $(t+\tau)>$ (dotted line). (f) Cross-correlation functions versus time for $<\alpha_2$ (t) α_4 $(t+\tau)>$ (solid line) and $<\alpha_4$ (t) α_2 $(t+\tau)>$ (dotted line). (g) Cross-correlation functions versus time for $<\alpha_3$ (t) α_4 $(t+\tau)>$ (solid line) and $<\alpha_4$ (t) α_3 $(t+\tau)>$ (dotted line).

TABLE II. Rate Constants: Strand Two Fixed in Position

Run	Res	τ_f(ns)	τ_u(ns)	k_f(ns^{-1})	k_u(ns^{-1})	s
S_1 (1)	1	0.0139	0.0237	71.7	42.2	1.70
S_1 (1,2)	2	0.0167	0.0143	59.9	69.7	0.86
S_1 (1,2,3)	3	0.0144	0.0287	69.4	34.8	1.99
S_1 (1,2,3,4)	4	0.0173	0.0163	57.8	62.2	0.94
S_1 (1,2,3,4,5)	5	0.0140	0.0317	71.4	31.6	2.26

TABLE III. Rate Constants: Strand One Fixed in Position

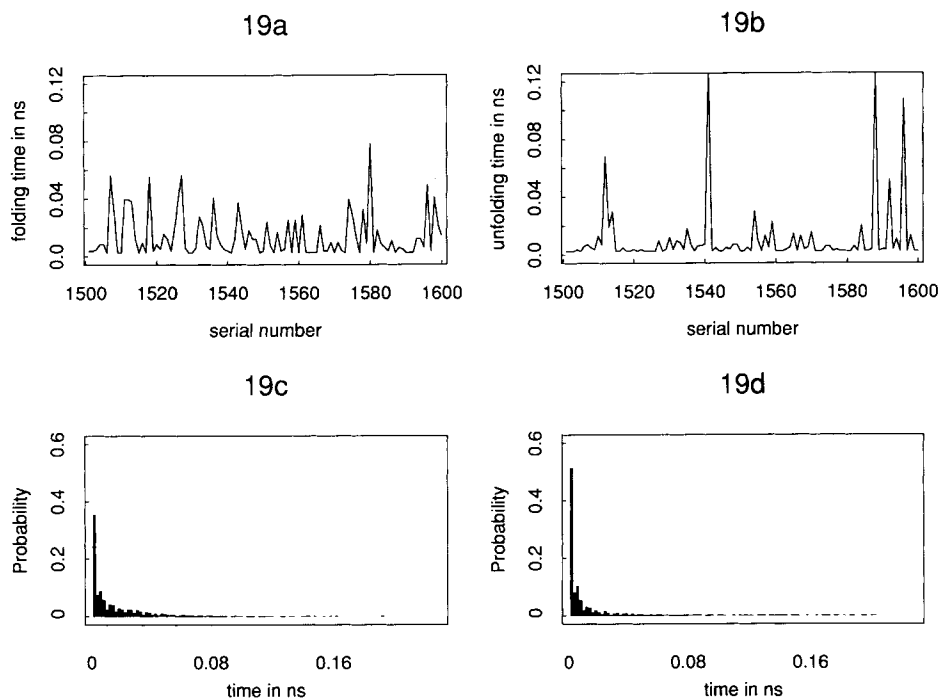| Run	Res	τ_f(ns)	τ_u(ns)	k_f(ns^{-1})	k_u(ns^{-1})	s
S_2 (22)	22	0.0103	0.0155	97.1	64.7	1.50
S_2 (22,21)	21	0.0227	0.0266	44.1	37.6	1.17
S_2 (22,21,20)	20	0.0166	0.0186	60.2	53.8	1.12
S_2 (22,21,20,19)	19	0.0254	0.0253	39.4	39.5	1.00
S_2 (22,21,20,19,18)	18	0.0178	0.0161	56.2	62.1	0.90

Fig. 19.   (a) Sequence of folding times (τ_f) from the S_1 (1,2) simulation. (b) Sequence of unfolding times (τ_u) from the S_1 (1,2) simulation. (c) Probability distribution of folding times from the S_1 (1,2) simulation. (d) Probability distribution of unfolding times from the S_1 (1,2) simulation.

stants s in Table I. Since the 90-ns simulations are not long enough to sample the entire configuration space used to calculate the equilibrium constants s, we would not expect equality of the s' and s values. However, the differences are a factor of two or less in most of the cases studied. This indicates that the kinetic equilibrium coefficients s' and the rate constants k_f and k_u are reasonably well established by the simulations.

From the simulation data described above, it was observed that at some time points, inner residues were found in the coil state, while the outer residues had cross-strand distances and dihedral angles in the sheet range. This behavior may be related to the occurrence of bulges in β-sheets in some proteins (termed β bulges).

Bootstrap analyses[31] were performed on some of the time vectors of τ_f and τ_u values to estimate the statistical error in the rate constants reported in Tables II and III. The results indicate that the rate constants have statistical errors of a few percent.

DISCUSSION

Simplified Model

A simplified model for a polypeptide chain with one quasiparticle per amino acid has been used to study the equilibrium and dynamic properties of an anti-parallel two-stranded β-sheet. In the model, the backbone torsion angles φ and ψ are replaced by the virtual dihedral angles α; α_i is approximately related[32] to $\phi_{i+1} + \psi_i$. Use of α is appropriate, because the path of the adjacent C^α atoms depends mainly on $\phi_{i+1} + \psi_i$, while the rotation of the peptide group between C^α atoms depends mainly on $\phi_{i+1} - \psi_i$, the latter having a relatively small effect on the chain's overall direction. Hence, by replacing these two degrees of freedom with α, one retains most of the conformational flexibility of the chain. With this simplification, the energy function E used to represent the chain corresponds to the potential of mean force which is averaged over amide plane orientations.

A limitation of the model is that the detailed hydrogen bonding is not treated. In a real antiparallel sheet, alternate pairs of residues form two hydrogen bonds between the β strands of the β-hairpin considered here. With a single interaction center per residue, only an effective stabilizing interaction corresponding to hydrogen bond formation is used. Further, every residue in the ideal model of the two strands has the same value for the dihedral angle α regardless of hydrogen bonds.

Although there are no experimental values for comparison, parameters for the model were chosen so that the sheet–coil equilibrium constants are close to one for all cases studied. This makes the sheet–coil equilibria approximately equal to the helix–coil equilibria.[5] The s parameter, defined by Eq. (10) for a pair of residues in the case with both strands free, is approximately given by the average of the parameters of the corresponding residues in the cases with either strand fixed in position (Table Ib and c). There are interesting periodicities in the s parameters along the chain which can be rationalized in terms of the alternating mean energy of neighboring residues (adjacent in a strand or opposite on the other strand) as discussed in the text. The β-sheet stabilization term in the energy function contributes up to -2.86 kcal/mol upon addition of a pair of residues to the sheet structure; this is the average effective energy of formation of a hydrogen bond in the model. The coil state is stabilized by near-neighbor nonbonded interactions (E_{ev}), and to a lesser extent by hydrophobic interactions.

Analysis of the mean energy of residue 1 and residue 22 as a function of the dihedral angle α shows significant differences (see Fig. 4). The intrinsic torsional barrier of ~3 kcal/mol dominates in the angular range 250°–350° for residue 22, while the excluded volume term contributes to the very high barrier for residue 1. The differences in the energy

curves are mainly due to interactions with the other strand. There are no essential changes in energy when one strand is fixed in position and the overall mean energy curves are similar to the ones in Figure 4 with a reduced potential well at 180°.

Dynamic Properties

The simulations show that the residues at the ends of the strands unravel quite readily, with calculated rates in the range 30–70/ns. When they leave the sheet structure, the end residues often move far away from the opposite residues (e.g., the $r_{1,22}$ probability distributions of Figs. 8a, 13a, and 16a have substantial probabilities at >7.0 Å). As expected, the residues at the sheet–coil interface show significantly less mobility than the coil residues. The probability distributions provide an essential tool for understanding the behavior of these unraveling residues. One can see from them that the unraveling residues often pass through the sheet conformation. Together with the time histories and effective energy plots, these distributions offer a comprehensive picture of the dynamics. From Figures 8 [$S_1(1,2)$], 13 [$S_2(22,21)$], and 16 [$S_1(1,2,3,4)$], one sees that only the residue at the sheet–coil interface maintains its cross-strand distance mostly in the sheet range with the rest of the residues having substantial positional freedom.

The potential energies shown in Figures 5 and 10 are useful in understanding the time histories of unraveling residues. For example, the differences in the potential surfaces for α_1 and α_{19} indicate that the fluctuations of these residues will be different even though both residues are at the interface in the simulations and have other similarities (e.g., Figs. 8c and 13c). Since the potential surface is fairly flat over much of the conformational space (limited by excluded volume interactions) for the last two residues in strand two (with strand one fixed in position), they diffuse relatively freely in most of the space.

A soft repulsive potential allows the simplified residues to approach closer than their combined van der Waals radius of 3.5 Å, to model the ability of residue side chains to move aside in response to close approach. This response is lost in a spherical residue approximation. In known protein structures, the C^α in β-sheets can be as close as 4.0 Å. In our model a residue–residue distance of 4.0 Å has a repulsive energy contribution of 2.6 kcal/mol [the excluded volume contribution of Eqs. (5) and (6)].

As expected, the relaxation times observed for the auto- and cross-correlation functions in the $S_2(22,21)$, $S_1(1,2)$, $S_2(22,21,20,19)$, and $S_1(1,2,3,4)$ are much longer than those observed for individual atoms in proteins[32] and for torsional fluctuations of tyrosine rings in proteins.[3] Typical relaxation times in these systems are fractions of picoseconds to picoseconds, compared to tens to hundreds of picoseconds observed in the sheet–coil simulations. This is
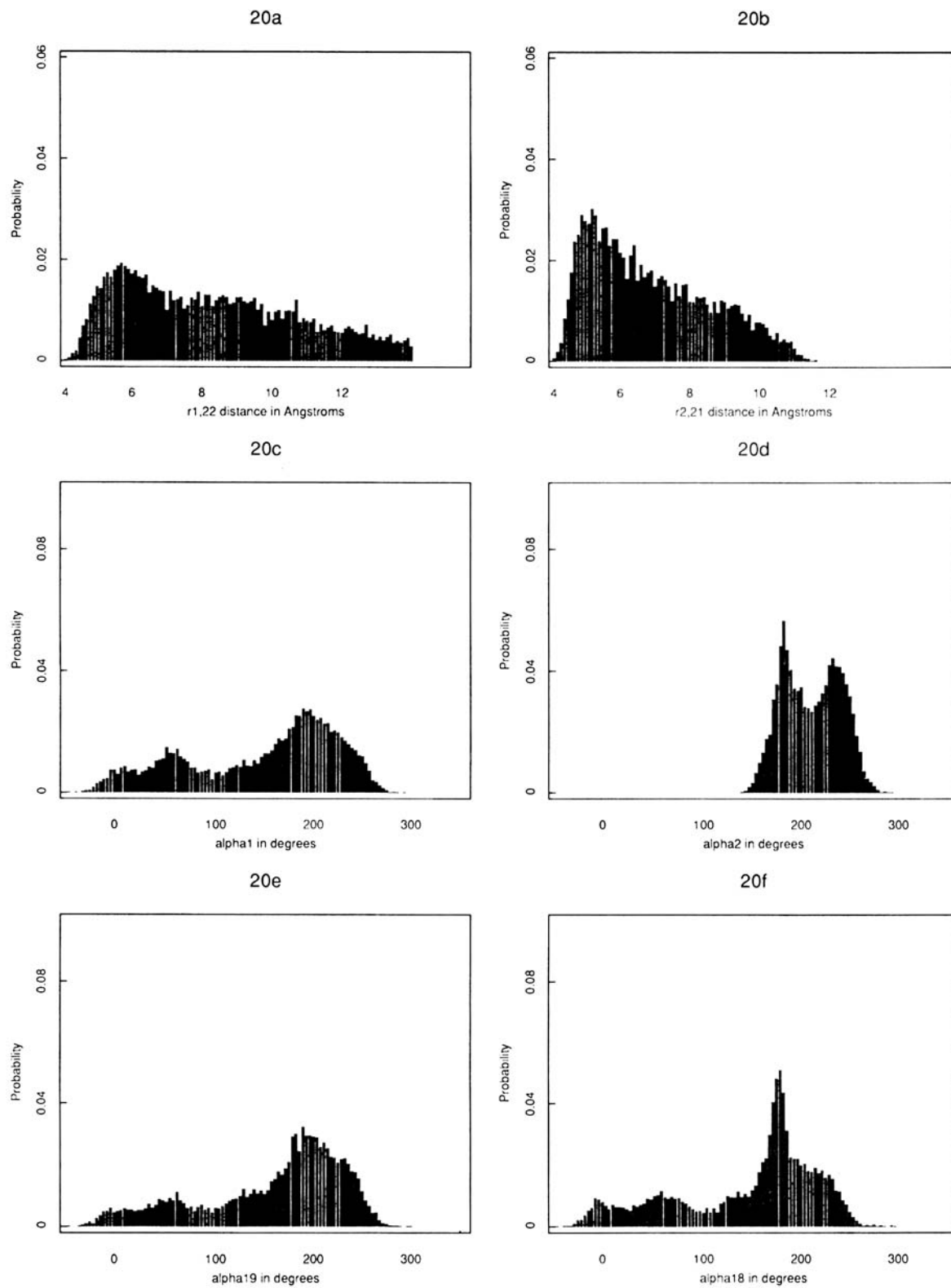
Fig. 20. Probability distributions of the distances and dihedral angle for the 15-ns simulation S_{12} (1,2,22,21).

true because larger scale motions (including dihedral angle transitions) of larger quasiparticles are being considered.

The sheet–coil transition rates of inner residues in the strands are approximately the same as the values of the end residues in all cases (see Tables II and III). This behavior differs from helix–coil transition rates,[5] where the rate constants for internal residues rapidly drop to a limiting value much smaller than that of the terminal residue. Also, the values obtained here are an order of magnitude larger than the helix–coil transition rates of the end residues in the helix. The roughly equal rate constants for all residues in the sheet–coil simulations may be due to the fact that unlike helices, a sheet conformation is less dependent on adjacent residues. Though the values obtained here may depend on the parameters chosen, they suggest that the nucleation parameter for an antiparallel β-sheet is close to one.

Simulations were also performed for cases where both strands were free to move. This corresponds to two-strand antiparallel sheets in proteins in which the strands do not interact with adjacent strands. Four independent simulations were performed for the unraveling of residue pairs (1,22), (1–22,2–21), (1–22,2–21,3–20), and (1–22,2–21,3–20,4–19), each for a time of 15 ns. The time histories of dihedral angles and distances show behavior similar to the cases with one strand fixed. For example, the dynamics of the first two and last two residues of the sheet is related to the combined picture of the two simulations with either strand fixed (see Fig. 20 for probability distributions). This is expected because the mean energies of every opposite pair is similar to that of every adjacent pair in each strand. The cross-strand distances (Fig. 20a and b) are flatter and extend to larger distances than the fixed strand cases (Figs. 8 and 13), because both strands are moving at the same time and residues can move further away. Further analyses of the dynamics of the simulations in the cases with both strands free, will be given in a separate publication.

The calculation of the sheet–coil rate constants in this paper and the corresponding calculation of helix–coil rate constants[5] shows that it is possible to use simulations to determine some of the parameters important for models of protein folding. For the diffusion–collision model,[1] detailed calculations[4] have shown that two-microdomain processes can be used to compute the dynamics of a multimicrodomain protein, using rate constants depending on measurable physical properties of the microdomains and the equilibrium state of the system (see Introduction). For the β-hairpin, the parameters calculated in this work provide the equilibrium probability of finding a microdomain in the folded state [given by $k_f/(k_f + k_u)$] and the characteristic time for fluctuations between the folded and unfolded states [given by $1/(k_f + k_u)$]. Of course, amino acid residues

are more complicated than the simplified representation used here, but the present results should be of use for overall folding estimates which concern themselves primarily with the backbone conformation and nonspecific residue interactions.

Related work has been done by Skolnick and co-workers[34] who have carried out a number of equilibrium and dynamic Monte carlo simulations of a quasiparticle model of the type used in the α-helix and β-sheet studies described here. However, rather than doing Langevin dynamics in the full conformational space, they study the behavior of a chain of beads with the quasiparticles confined to a tetrahedral, diamond lattice. An algorithm involving the master equation approach to stochastic dynamics is used to change the conformation of the chain. In their study of β-sheet proteins, Skolnick and Kolinski[35] found that with parameters that distinguish turn and β-sheet regions, folding typically starts at or near a β-turn, and then extends to the rest of the sheet. They argue that this is not in accord with a diffusion–collision mechanism. However, since assembled secondary structure is immobile in their model, a diffusion–collision mechanism is prohibited. To determine the realtive contributions of the diffusion–collision and "on site" assembly, it is necessary to do calculations with a model that combines the two with appropriate relative rates, i.e., one where secondary structural elements can diffuse while undergoing coil to secondary structure transitions. The present Langevin simulations and their extensions provide some of the parameters required for such a treatment of secondary and tertiary structure formation.

ACKNOWLEDGMENTS

REFERENCES

1. Karplus, M., Weaver, D.L. Protein-folding dynamics. Nature (London) 260:404–406, 1976.
2. Dolgikh, D.A., et al., α-Lactalbumin: Compact state with fluctuating teritary structure?. FEBS Lett. 136:311–315, 1981.
3. Brandts, J.F., Halvorson, H.R. Brennan, M. Consideration of the possibility that the slow step in protein denaturation reactions is due to cis-trans isomerism of proline residues. Biochemstiry. 14:4953–4983, 1975.
4. Weaver, D.L. Alternative pathways in diffusion-collision controlled protein folding. Biopolymers 23:675–694, 1984.
5. McCammon, J.A., Northrup, S.H., Karplus, M., Levy, R.M. Helix-coil transitions in a simple polypeptide model. Biopolymers 19:2033–2045, 1980.
6. Lee, S., Karplus, M., Bashford, D., Weaver, D.L. Brownian

dynamics simulation of protein folding: A study of the diffusion-collision model. Biopolymers 26:481–506, 1987.

7. Wright, P.E., Dyson, H.J., Lerner, R.A. Conformation of peptide fragments of proteins in aqueous solution: Implications for initiation of protein folding. Biochemistry 27: 7167–7175, 1988.

8. Roder H, Elove, G.A., Englander, S.W. Structural characterization of folding intermediates in cytochrome c by H-exchange labelling and proton NMR. Nature (London) 335: 700–704, 1988.

9. Udganokar, J.B., Baldwin, R.L. NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A. Nature (London) 335:694–699, 1988.

10. Oas, T.G., Kim, P.S., A peptide model of a protein folding intermediate. Nature (London) 336:42–48, 1988.

11. Labhardt, A.M. Kinetic circular dichroism shows that the S-peptide α-helix of ribonuclease S unfolds fast and refolds slowly. Proc. Nat. Acad. Sci. U.S.A. 81:7674–7678, 1984.

12. Gilmanshin, R.I., Ptitsyn, O.B. An early intermediate of refolding α-lactalbumin forms within 20 ms. FEBS Lett. 223:327–329, 1987.

13. Kuwajima, K., Hiraoka, Y., Ikeguchi, M., Sugai, S. Comparisons of the transient folding intermediates in lysozyme and α Lactalbumin. Biochemistry 24:874–881, 1985.

14. McCoy, L.F., Rowe, E.S., Wong, K.P. Multiparametric kinetic study on the unfolding and refolding of bovine carbonic anhydrase B. Biochemistry 19:4738–4743, 1980.

15. Dolgikh, D.A., Kolomiets, A.P., Bolotina, I.A., Ptitsyn, O.B. 'Molten-globule' state accumulates in carbonic anhydrase folding, FEBS Lett. 165:88–92, 1984.

16. Semisotnov, G.V., et al. Sequential mechanism of refolding of carbonic anhydrase B. FEBS Lett. 224:9–13, 1987.

17. Kuwajima, K., Yamaya, H., Miwa, S., Sugai, S., Nagamura, T. Rapid formation of secondary structure framework in protein folding studied by stopped-flow circular dichroism. FEBS Lett, 221:115–118, 1987.

18. Bashford, D., Cohen, F.E., Karplus, M., Kuntz, I.D., Weaver, D.L. Diffusion-collision model for the folding kinetics of myoglobin. Proteins 4:211–227, 1988.

19. Bashford, D., Karplus, M., Weaver, D.L. Diffusion-collision model for the folding kinetics of the λ-repressor operator-binding domain. J. Biomol. Str. Dyn. 1:1243–1255, 1984.

20. Bashford, D., Karplus, M., Weaver, D.L. The diffusion-collision model of protein folding. *Protein Folding, Deciphering the Second the Second Half of the Genetic Code:* AAAS Publication No. 89–18S, 1990.

21. Kim, P.S., Baldwin, R.L. Specific intermediates in the folding of small proteins and mechanism of folding. A. Rev. Biochem. 51:459–489, 1982.

22. Harrison, S.C., Durbin, R. Is there a simple pathway for the folding of a polypeptide chain. Proc. Nat. Acad. Sci. U.S.A. 82:4028–4030, 1985.

23. Tsong, T.Y., Baldwin, R.L., McPhie, P. A sequential model of nucleation-dependent protein folding: Kinetic studies of ribonuclease A. J. Mol. Biol. 63:453–475, 1972.

24. Karplus, M., Weaver, D.L. Protein folding dyanmics: A current perspective of the diffusion-collision model. In preparation, 1991.

25. Arnott, S., Dover, S.D., Elliott, A. Structure of β-poly- L-alanine: refined atomic coordinates for an anti-parallel beta-pleated sheet. J. Mol. Biol. 30:201–208, 1967.

26. Flory, P.J. "Statistical Mechanics of Chain Molecules." New York: Wiley, 1969.

27. Pear, M.R., Northrup, S.H., McCammon, J.A., Karplus, M., Levy, R.M. Correlated helix-coil transitions in polypeptides. Biopolymers 20:629–632, 1981.

28. Levitt, M. A simplified representation of protein conformations for rapid simulation of protein folding. J. Mol. Biol. 104:59–107, 1976.

29. Arnott, S., Dover, S.D., Elliott, A.J. Refinement of bond angles of an α-helix. J. Mol. Biol. 30:209–212, 1967.

30. Ermak, D.L., McCammon, J.A. Brownian dynamics with hydrodynamic interactions. J. Chem. Phys. 69:1352–1360, 1978.

31. Efrom, B. The jackknife, the bootstrap and other resampling plans. SIAM Monograph No. 38, 1982.

32. Swaminathan, S., Ichiye, T. van Gunsteren, W., Karplus, M. Time dependence of atomic fluctuations in proteins: Analysis of local and collective motions in bovine pancreatic trypsin inhibitor. Biochemistry 21:5230–5241, 1982.

33. McCammon, J.A., Wolynes, P.G., Karplus, M. Picosecond dynamics of tyrosine side chains in proteins. Biochemistry 18:927–942, 1979.

34. Skolnick, J., Kolinski, A., Sikorski, A. Dynamic monte carlo simulations of globular protein folding, structure and dynamics. Comments Mol. Cell. Biophys. 6:223–247, 1990.

35. Skolnick, J., Kolinski, A.J. Dynamic monte carlo simulations of globular protein folding/unfolding pathways I. Six-member, Greek key β-barrel proteins. J. Mol. Biol. 212: 787–817, 1990.