J. Mol. Biol. (1990) 215, 183–198

Dynamic Monte Carlo Simulations of Globular Protein Folding

Model Studies of *in Vivo* Assembly of Four Helix Bundles and Four Member β-Barrels

Andrzej Sikorski

Department of Chemistry
University of Warsaw
Pasteura 1 02-093, Warsaw, Poland

and Jeffrey Skolnick†

Department of Molecular Biology
Research Institute of Scripps Clinic
10666 North Torrey Pines Road
La Jolla, CA 92037, U.S.A.

(Received 12 December 1989; accepted 20 April 1990)

As part of an ongoing series of dynamic Monte Carlo simulations of globular protein folding, the nature of the folding pathway, of model four-member β-barrels and four-helix bundles, under highly idealized conditions *in vivo*, has been examined. The ribosome is crudely modeled as an inert hard wall on to which the model protein chain is attached. Three cases are considered in detail. The first corresponds to post-translational assembly in which the fully synthesized chain is tethered to the wall and starts out under strongly denaturing conditions. The system is cooled down, and the chain is allowed to fold. Interestingly, the helical motif prefers to assemble parallel to the wall, whereas the β-barrel, predominantly assembles with its principal axis perpendicular to the wall. In the former case, the dominant intermediate, the helical hairpin, is different from that in free solution, a three-helix bundle. The wall acts to reduce the expanse of configuration space that must be searched and aids in folding. Two situations that might lead to co-translational folding are also simulated. In the first case, to eliminate wall effects, the chain is slowly synthesized in free solution, and in the second case, it is slowly synthesized from the wall. In all cases, the chains are observed to fold post-translationally. While partially folded intermediates are observed during synthesis, they lack the stability to survive until chain synthesis is complete. The implications of these results for the folding *in vivo* of real protein chains is discussed, and a model of multiple domain protein folding is proposed.

1. Introduction

At present the process by which the nascent protein chain folds up into the native conformation *in vivo* is not well understood (Tsou, 1988). Does the protein chain fold up during synthesis, i.e. is folding a co-translational process, or is the entire protein chain synthesized first, with folding occurring after chain synthesis is completed? While early workers envisioned co-translational folding as the likely mechanism (Phillips, 1967; De Coen, 1970), subsequent workers considered this unlikely (Anfinsen & Scheraga, 1975; Baldwin & Creighton, 1980; Creighton, 1984). This was based on the belief that the denatured state is a purely statistical random coil, devoid of secondary structure (Creighton, 1985), and that the presence of the C terminus in a number of proteins is necessary for biological activity (Taniuchi, 1970). However, with the recent recognition that proteins may have fluctuating secondary structure of marginal stability in the denatured state (Bundi *et al.*, 1976, 1978; Shoemaker *et al.*, 1985, 1987; Hruby, 1985; Dyson *et al.*, 1988*a*,*b*) the co-translational hypothesis (with the possibility of constant conformational readjustments as synthesis proceeds) has been revived (Bergman & Kuehl, 1979; Purvis *et al.*, 1987; Tsou,

† Author to whom all correspondence should be addressed.

0022-2836/90/170183-16 $03.00/0

1988). There is the additional and fundamental question of whether the native conformation corresponds to the global free energy minimum structure; and if this is the case, is the global free energy minimum structure in the presence of the ribosome the same as in the absence of the ribosome? This paper adopts the viewpoint that the native conformation is at a global free energy minimum (Anfinsen, 1972), and in the context of a series of dynamic Monte Carlo simulations (Binder, 1984), explores the nature of the folding process in an idealized model of assembly *in vivo*.

There are several important experimental observations of cases *in vitro* and *in vivo* that must be reproduced by any simple model if it is even to be considered as a candidate for providing qualitative insight into the folding process. With respect to the equilibrium properties, for single-domain globular proteins *in vitro*, the conformational transition must be well approximated by a two-state model with intermediates sparsely populated (Creighton, 1985, 1988; Brandts *et al.*, 1975). In the denatured state, there are elements of fluctuating secondary structure of marginal stability (Bundi *et al.*, 1976, 1978; Shoemaker *et al.*, 1985, 1987; Hruby, 1985; Dyson *et al.*, 1988*a*,*b*). The N-terminal sections of a number of proteins, in particular, have significant amounts of secondary structure (Hua *et al.*, 1985). Turning to the nature of the folding kinetics, in a number of cases *in vitro*, the transition state is near the native conformation (Creighton, 1985, 1988), with the rate of unfolding being more sensitive to changes in conditions than is the refolding rate (Tsong & Baldwin, 1978; Brems *et al.*, 1982; Goldenberg & Creighton, 1985). The actual folding process is rather fast, but the time it takes to fold is of the order of seconds or minutes (Garel & Baldwin, 1973; Kuwajima *et al.*, 1985; Creighton, 1988), i.e. the protein spends a substantial fraction of its time in the denatured state before undergoing a relatively rapid conformational transition to the native conformation. (We neglect here complications introduced by the presence of a slow refolding phase (Lin & Brandts, 1987), e.g. due to the presence of proline (Brandts *et al.*, 1975; Lang *et al.*, 1987).)

With respect to assembly *in vivo*, the folding studies of ribonuclease A (Taniuchi, 1970; Andria & Taniuchi, 1978) seem to suggest that the conformation of a protein readjusts subsequent to the completion of synthesis. Other proteins, however, exhibit full biological activity without the entire C-terminal portion (Liao & Khorana, 1984). The formation of some of the disulfide bridges occurs cotranslationally, for example in the case of a cross-link between Cys35 and Cys100 in the nascent immunoglobin light-chain. This observation led Bergman & Kuehl (1979) to suggest that folding *in vivo* starts from the N terminus.

Previously, we have developed a simplified class of models of globular proteins that reproduce all of the qualitative features of globular protein folding under *in vitro*, i.e. free solution, conditions (Kolinski *et al.*, 1986*a*,*b*, 1987*a*,*b*; Skolnick *et al.*, 1988, 1989*a*,*b*; Skolnick & Kolinski, 1990; Sikorski & Skolnick, 1989*a*,*b*, 1990). These features are not *a priori* built into the model, but emerge as a consequence of the physics. In an extensive series of dynamic Monte Carlo (MC†) simulations of a diamond lattice, α-carbon representation of globular proteins, we demonstrated that the folding of all the known four-helix bundle topologies and of a plastocyanin-like, β-protein analog could be reproduced. To obtain the desired structures, it is necessary to have the appropriate pattern of hydrophobic- and hydrophilic-type residues (actually, hydrophobic-type residue pairs experience an attractive potential of mean force, hydrophilic/hydrophobic pairs experience a repulsive potential of mean force, and hydrophilic pairs may be either repulsive or indifferent), a marginal preference for secondary structure in the denatured state and, depending on the particular topology, the presence of residues that, at the very least, are neutral to turn formation (i.e. weakly hydrophilic residues for which the native turn is but one of a number of equally likely conformations).

We subsequently analyzed the mechanism of assembly *in vitro* (Skolnick & Kolinski, 1990; Sikorski & Skolnick, 1990). For β-barrel proteins, folding is seen to initiate at or near a β-strand, followed by the formation of a β-hairpin. The next β-strand then zips up; this is followed by the zipping up of subsequent β-strands until the final folded conformation is reached. In the case of α-helices, because an isolated helix is more stable than an isolated β-strand, initiation not only occurs at a turn; in addition folding also begins from an α-helix onto which the adjacent helix in the hairpin assembles and from a hydrophobic pair of residues located away from the native turn. Then both helices assemble, with possible shifting *via* an inchworm-like mechanism to produce the appropriate registration of the helices. Folding then proceeds by the sequential on-site construction of the remaining helices.

In all cases, in the early stages of folding, the existing assembled secondary structure acts as a scaffolding that aids in the assembly of subsequent secondary structure. However, in the later stages, assembled structure can hinder subsequent assembly; for example, if the unfolded tail is on the wrong side of the assembled protein, the tail must rattle around until it finds the channel into which it must snake up in order to assemble the native structure. An example of this is shown in Figure 1, where a representative pathway of the assembly for a four-helix bundle having the cytochrome *c*′ topology, i.e. a four-helix bundle with tight bends, is presented. Near the native conformation there is a pause associated with finding the native conformation of the unassembled tail, which is mainly of entropic origin. Similar considerations hold for the

† Abbreviations used: MC, Monte Carlo; *t*, *trans*; *g*, *gauche*.

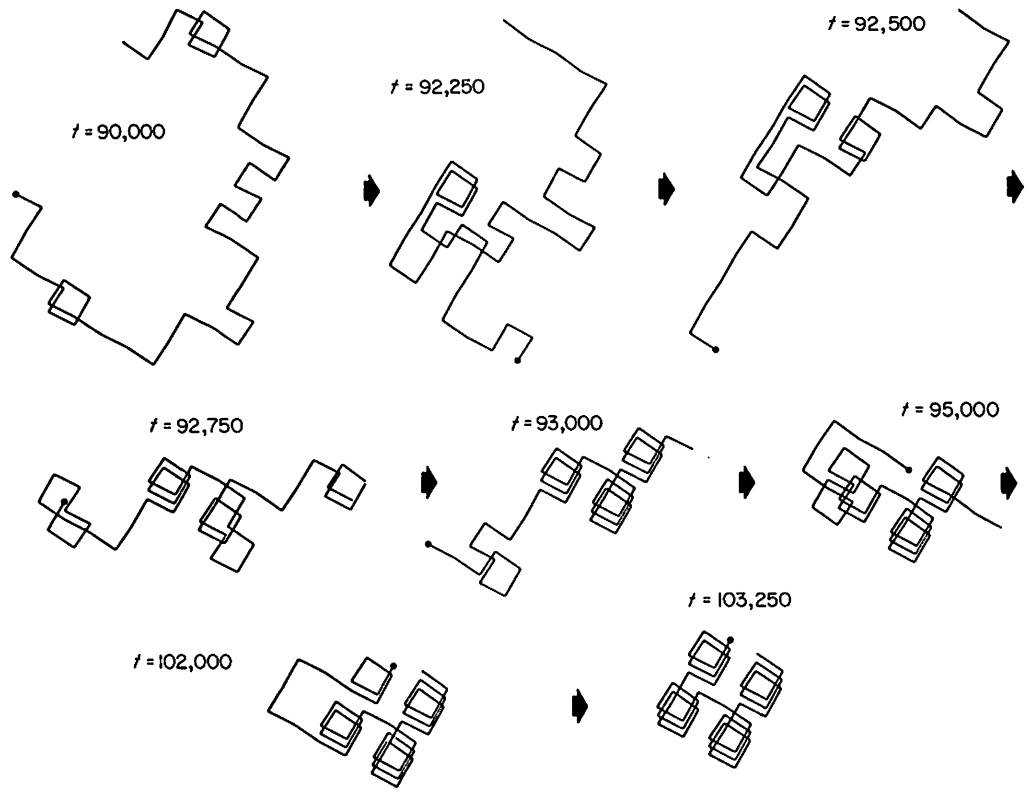

Figure 1. Representative folding trajectory of a 4-helix bundle in solution for a model having a central turn-neutral region flanked by 2 amphipathic tails. The time indicated in the Figure is from the start of the simulation run. The simulation is performed at $T^* = 0{\cdot}667$. See the text for further details.

assembly *in vitro* of the β-protein models. In both cases there are manifold spatial trajectories consistent with a given overall pathway, but the number of choices is reduced the closer one is to the native state. There is not a unique pathway, but a relatively small number of pathways that can be described in a general sense. For example, in the case of the four-helix bundle assembly, folding initiates at one of the three turns. This is followed by the assembly, *via* an on-site mechanism, of the third helix. The unassembled random-coil tail thrashes about before the fourth and final helix assembles.

While we certainly recognize that, in real proteins, the process of assembly *in vivo* is extremely complicated (Tsou, 1988), here we propose to examine the nature of the folding process in the context of a series of *highly* simplified models. In particular, the folding of four-helix bundles and four-member β-barrels will be examined. First, we examine the simplest case, i.e. post-translational assembly, and approximate the ribosome as a hard planar wall. The C terminus of the model protein is attached to the wall, and we examine the effect(s) of the wall on the folding pathway and on the resulting folded structure(s). Among the questions explicitly addressed are the differences, if any, between the folding intermediates in free solution and when a chain end is attached to a wall. Do the resulting folded structures prefer to assemble parallel or perpendicular to the wall, and, if so, is the preference the same for α-helical and β-proteins? Next, the possibility of co-translational assembly is explored; here two physical situations are treated. The first is closest to the post-translational case. The model chain is grown from a hard wall with the rate of synthesis chosen to be slow relative to the rate of folding observed in the post-translational and *in vitro* cases. (Actually, in practice, a whole range of synthesis rates has been simulated.) For this situation, there are two not necessarily separable influences on the folding pathway. The first is the effect of the wall itself on the folding pathway, and the second is the assumption of slow synthesis relative to the rate of folding *in vitro*. To separate these two effects we have considered one additional physical situation, slow synthesis of the model protein in solution. In this case, there is no hard wall, but the C-terminal end is pinned at a point in space and the synthesis of the chain is slow relative to the observed (from the simulation) rate of refolding *in vitro*.

One might expect a number of distinct physical effects to occur in assembly *in vivo*. Unlike the *in vitro* case, one of the ends is no longer free to rattle about but is now pinned. Thus, the two ends are no longer dynamically equivalent. If co-translational folding obtains, as the chain is synthesized, the newly synthesized section of chain will be pulled into the native-like, N-terminal portion until there

is a taut piece of chain between it and the ribosome. For folding to proceed further, one of two things can happen; in one scenario, slack must be generated between the ribosome and the folded portion of the chain, i.e. synthesis of a number of residues occurs (the slow step). These newly synthesized residues must then fold. They could assemble either by an "on-site" construction mechanism or by the preformation of secondary structure that then diffuses to the assembled native-section of the chain (Karplus & Weaver, 1976, 1979; Lee *et al.*, 1987). However, if synthesis is slow relative to folding, it is difficult to see why slack in the newly synthesized chain would be generated. Thus, in an alternative mechanism, little, if any, slack is generated. If this condition obtains, then the assembled N-terminal portion must move back and forth relative to the ribosome as assembly occurs, much as a needle moves back and forth in a sewing machine. Such a folding mechanism would be substantially different from the on-site folding mechanism seen in the solution simulations (Skolnick & Kolinski, 1990; Sikorski & Skolnick, 1990). In the former case, the unfolded pieces of the chain are dragged towards the assembled tertiary structure. Here, the tertiary structure moves towards the denatured piece. A potential problem with co-translational assembly is that it implies that folding intermediates must be long lived, i.e. the assembled tertiary structures never dissolve during the entire course of amino acid synthesis. If so, one would expect to see substantial intermediate populations during *in vitro* assembly as well. The only way out of this quandary would be to assume that the ribosome itself stabilizes the folding intermediates.

The outline of the remainder of the paper is as follows. In the next section, the model of the protein is discussed in detail, including the nature of the allowed interactions and the amino acid sequences employed. We next present an overview of the MC algorithm, followed by the simulation results. Finally, we conclude with a discussion of the range of the validity of the co-translational *versus* post-translational folding hypotheses.

2. Description of the Model

(a) *Interactions*

The model protein-chain is assumed to be confined to a tetrahedral (diamond) lattice. The lattice is constructed from a set of vectors obtained by the permutations of the vectors $(\pm 1, \pm 1, \pm 1)$, such that adjacent bond vectors form a tetrahedral bond angle (109·47°). Hence, an α-carbon representation of the protein-chain is employed. While this is a very crude approximation because of the discretization of space and the highly reduced description of the protein, this approximation is advantageous in that it permits a significant speed-up in the calculations, and, it is hoped, preserves the essence of the physics of folding of real globular proteins. For the considered *in vitro* cases, this has been shown to be true (Skolnick *et al.*, 1989*a*,*b*; Sikorski & Skolnick, 1989*a*,*b*).

The model chain consists of n amino acid residues located on a consecutive sequence of adjacent lattice sites. For many of the *in vivo* simulations described below, it is assumed that one end of the chain is immobile (the last residue only), and that this residue is located in the neighborhood of a hard, impenetrable wall that, very crudely, mimics the presence of a ribosome. We consider this model to be a first approximation of folding condition *in vivo*, in which we merely take into consideration the relatively large size of a ribosome compared with the protein-chain and, thus, neglect any curvature of the wall. Apart from the hard core repulsion of the wall, there are no other attractive or repulsive interactions between the wall and the amino acid residues. These more sophisticated models will be studied in the future. Here, our objective is to examine the influence on folding of the simplest possible case, a hard wall.

The fact that the protein-chain occupies a finite volume is easily implemented on a lattice by forbidding the multiple occupancy of all lattice sites. This definition of short-range repulsions is also computationally very tractable. Every internal bond (connecting 2 residues), is found in one of three conformational states: the planar *trans* (t) and either of the two out-of-plane *gauche* states (g^+ and g^-). For a chain of n residues, the number of such rotational states is $n-3$. As in previous work (Skolnick *et al.*, 1989*a*,*b*; Sikorski & Skolnick, 1989*a*,*b*), a β-strand is represented on the tetrahedral lattice as a sequence of t states, and the right-handed α-helix consists of a sequence of g^- states. In order marginally to stabilize the presence of α and β secondary structure, short-range interactions are included in the model. In the case of β-protein models, it is assumed that the t state corresponds to the lowest energy state (equal to zero) and both *gauche* states have the same energy ($\varepsilon_g > 0$). This parameter forms the basis of a reduced temperature scale $T^* = k_B T/\varepsilon_g$, where k_B is the Boltzmann constant and T is the absolute temperature.

In the case of α-helical proteins, it is possible to introduce a local preference for g^- states in the same way as is described above for β-proteins. We have opted for helical wheel-type interactions (Schiffer & Edmunson, 1967) on the basis that they are more physical, but qualitatively identical results are found if a local g^- preference is used. They mimic hydrogen bonds between peptide groups in the backbone, as well as the possibility of side-chain–side-chain interactions. More specifically, it is assumed that the ith and $(i+4)$th residues interact with an attractive potential of mean force, $\varepsilon_c < 0$, when all of the residues between them are g^- states, i.e. a single helical turn is formed. Each additional g^- state that is added to the helical sequence causes a decrease in energy equal to ε_c. For helical proteins the reduced temperature scale is $T^* = -\varepsilon_c/k_B T$.

For both α and β-proteins, long-range interactions are limited to nearest-neighbor pairs of non-bonded

residues. There are only two kinds of residue in a chain: hydrophobic and hydrophilic (they are introduced so that the interior and exterior parts of the native conformation can be distinguished). Hydrophobic residues interact with other hydrophobic residues with an attractive potential of mean force (Hill, 1956) of $\varepsilon_h < 0$. Non-bonded nearest-neighbor hydrophobic and hydrophilic residue pairs interact with a repulsive potential of mean force $\varepsilon_w > 0$. Hydrophilic pairs of residues interact with a potential of mean force that can be repulsive or zero, with no difference in folding pathways or in the resulting folded structures. Thus, hydrophobic and hydrophilic interactions are independent of the backbone chain configuration and are determined by the type of residue alone. The potentials of mean force characterizing these interactions have spherical symmetry. The names used above as "hydrophobic" and "hydrophilic", actually indicate the relative preference of forming contacts. They do not have to be, in fact, purely hydrophobic or hydrophilic interactions. For example, the attraction arising from the formation of a salt bridge can be included in the hydrophobic type of potential as well.

There is a further kind of long-range interaction introduced for β-proteins (Kolinski *et al.*, 1987*a*). A co-operative potential $\varepsilon_c' < 0$ is introduced to mimic the conformational coupling between non-bonded conformational states; the nature of this interaction is shown schematically in Figure 2. Basically, if two non-bonded t states are nearest neighbors, the conformation is stabilized by a factor $-\varepsilon_c'$; if the non-bonded nearest neighbors' states involve three t states, it is stabilized by a factor $-2\varepsilon_c'$, and if four t states are involved, a stabilization factor of $-4\varepsilon_c'$ is introduced. An equivalent co-operative potential has been introduced for helical proteins with no qualitative change in behavior (Sikorski & Skolnick, 1989*a*).

It should be pointed out that all of the interactions described above are valid for any conformation of a model chain and not merely the native case, i.e. non-native interactions are allowed as well. Thus, no target potential is introduced.

Previously (Skolnick *et al.*, 1989*a*,*b*; Sikorski & Skolnick, 1989*a*,*b*), we found the set of conditions necessary to obtain unique native states of α and β-proteins *via* an all-or-none conformational transition from the denatured state. Here, we assume that the same, or a very similar, primary sequence should lead to the folding *in vivo* of the same topology. In the case of a diamond lattice representation of a four-helix bundle, the residues belonging to the α-helical stretches should display the hydrophobic/hydrophilic pattern in which residues i, $i+1$ and $i+3$ are hydrophobic and residue $i+2$ is hydrophilic. A detailed description of this pattern is presented elsewhere (Sikorski & Skolnick, 1989*a*); we remind the reader that a g^- helix contains four residues per turn. The above pattern allows two of the four faces of each helix to interact, and the other two faces to be exposed to solvent. Of course, the

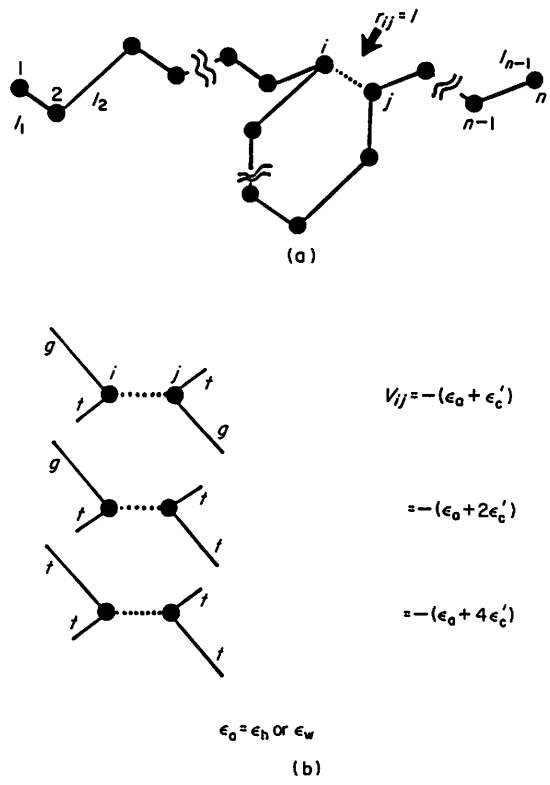

Figure 2. Schema of the allowed long-range interactions in β-barrel proteins. (a) Residues i and j and their relative positions along the chain. (b) The ij pair experience a hydrophobic/hydrophilic interaction embodied in the parameter ε_a that depends on the kinds of residue involved. ε_c' is a co-operativity parameter (see the text).

system has to fold to the helical conformation. We reiterate that interactions between all non-bonded pairs of residues occur for all, and not necessarily native, conformations.

Residues located in short bends between helical stretches have somewhat different statistical properties. In the case of helical bundles with tight bends, the presence of a "turn-neutral" central bend is crucial for obtaining the *unique* native state. The term turn-neutral means that based on short-range interactions alone, no particular configuration of a bend is preferred. The other interactions can be simply turned down ($\varepsilon_c = \varepsilon_h = \varepsilon_w = 0$), or only the local interactions are zero ($\varepsilon_c = 0$) and the bend is hydrophilic ($\varepsilon_h = 0$). The two remaining turns do not have to be specified at all in the primary sequence; rather, they can be part of the amphipathic pattern of the adjacent helical stretches, i.e. $\varepsilon_c < 0$, and based on local interactions alone the helical conformation is preferred.

For four-member, β-barrels, all three bend regions need to be specified as turn-neutral in the primary sequence (Skolnick *et al.*, 1988, 1989*a*). The properties of these turn-neutral bends are the same as described above, but, additionally, here the co-operative interactions (ε_c') are equal to zero. The stronger condition (3 bend regions) required for folding to unique four-member β-barrels can be

explained by the much lower energy of an α-helical hairpin ($-17{\cdot}8\ k_BT$ at the transition temperature), as compared to that of a β-hairpin ($-12{\cdot}3\ k_BT$). Because the entropy loss on folding is comparable in both cases, the free energy of an α-helical hairpin is lower than the β-hairpin. This means that the formation of out-of-register conformations in helical structures is much less favorable than out-of-register conformations in β-sheets for the choice of parameters that we employed.

As in our previous papers (Skolnick *et al.*, 1989*a*,*b*; Sikorski & Skolnick, 1989*a*,*b*), we use the following notation to specify a primary sequence. $\mathbf{H}_i(k)$ represents a part of the chain which should form the ith α-helical stretch in the native state that consists of k residues. $\mathbf{B}_i(k)$ corresponds to the ith putative β-strand that contains k residues. $\mathbf{b}_i^0$ corresponds to a short-bend neutral region consisting of three residues. When $\mathbf{b}_i^0$ is not specified, the hydrophobic/hydrophilic pattern is assumed to be consistent with that of the adjacent α-helices or β-strands.

We have examined three variants of the primary sequence of four-member, α-helical bundle models of folding *in vivo*. The difference between these models lies in the location of the putative bend neutral regions. The first case with primary sequence $\mathbf{H}_1(12)\mathbf{H}_2(10)\mathbf{b}_2^0\mathbf{H}_3(11)\mathbf{H}_4(12)$ has a central turn neutral region. The second case has the primary sequence $\mathbf{H}_1(12)\mathbf{H}_2(12)\mathbf{H}_3(10)\mathbf{b}_3^0\mathbf{H}_4(11)$, with the neutral bend located closest to the C terminus (which is connected to the ribosome), and the third case, with the primary sequence $\mathbf{H}_1(10)\mathbf{b}_1^0\mathbf{H}_2(9)\mathbf{b}_2^0\mathbf{H}_3(9)\mathbf{b}_3^0\mathbf{H}_4(11)$, has three bends that are turn-neutral. The main objectives of the study of these three models are to evaluate the minimum set of conditions required to fold a model chain to the unique native state and to check whether differences in the primary sequence can have an influence on folding pathways and on the resulting globular conformation.

Two different primary sequences were used in the examination of *in vivo* folding models of β-barrel proteins. The first case has the primary sequence $\mathbf{B}_1(9)\mathbf{b}_1^0\mathbf{B}_2(8)\mathbf{b}_2^0\mathbf{B}_3(9)\mathbf{b}_3^0\mathbf{B}_4(11)$, i.e. all three bends are turn-neutral. The second case has the primary sequence $\mathbf{B}_1(9)\mathbf{b}_1\mathbf{B}_2(8)\mathbf{b}_2^0\mathbf{B}_3(9)\mathbf{b}_3^0\mathbf{B}_4(11)$, where bend 1 (closest to the C terminus connected to the ribosome) has its native conformation $(g^+g^-g^+)$ preferred with an energy, $-3\varepsilon_g$ (actually all *gauche* conformations are iso-energetic).

The proportions between the various interaction potentials were fixed at each temperature, having values $\varepsilon_c = \varepsilon_h/2 = -\varepsilon_w/2$ in the models of α-helical proteins. Previously, these particular ratios were found to be very useful because the contributions to the total configurational energy due to the local and long-range interactions are comparable in magnitude. A change in the temperature leads to the proportional changes of all the potentials. In the case of β-proteins, we employ the same set of parameters as in our previous papers, $\varepsilon_g = -2\varepsilon_c' = -4\varepsilon_h = \varepsilon_w$ and proceed as in the α-helical case.

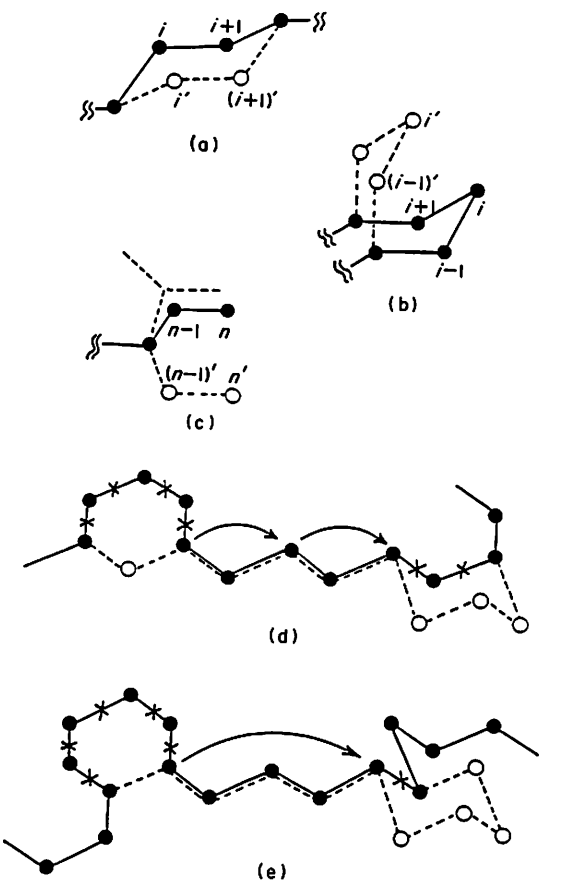

Figure 3. (a) Three-bond kink motion. (b) Four-bond kink motion. (c) Two-bond end motion. (d) Four-bond wave motion. (e) Five-bond wave motion. The filled circles and continuous lines are the initial configuration and the open circles and broken lines are the final configuration.

(b) *Monte Carlo algorithm*

The dynamic MC simulation technique has been applied to solve a stochastic kinetics master equation (Binder, 1984, 1987; Skolnick & Kolinski, 1990). The description of the theoretical foundations of this kind of MC simulation was described by Binder (1987). The most important features of the algorithm are: (1) the possibility of hunting through the entire configurational space with a high efficiency in the random coil state, as well as in the vicinity of the native state; (2) the possibility of rapidly passing through numerous local minima on the free-energy surface.

In particular, to model the internal dynamics, the model chain undergoes a series of local rearrangements (Kremer *et al.*, 1981). In the chain interior there are the following kinds of rearrangements. (1) As shown in Figure 3(a), there is a three-bond flip motion that changes the conformation $g^\pm \rightarrow g^\mp$. This involves the displacement of three bonds located in one-half of a chair conformation of a cyclohexane molecule to the other half of the chair conformation. These serve to diffuse local orientations down the chain. (2) As shown in Figure 3(b),

there are four-bond kink motions that create a new local orientation in the chain by the interchange of the sequence (3). As depicted in Figure 3(c), there are the chain-end modifications that randomly change the orientation of the two bonds on the free end of the chain.

It has been shown that this set of micromodifications spans the entire available configurational space (Iwata & Kurata, 1969) and produces the correct local (Valeur *et al.*, 1975) and long-distance random coil dynamics (Kolinski *et al.*, 1987*b*); with the assumption that hydrodynamic interactions can be neglected. For that part of the simulations where there is no need to maintain the correct time scale (e.g. in the calculation of the equilibrium properties of the system and the sampling of conformational space at temperatures above the folding transition), two additional micromodifications were used (Skolnick *et al.*, 1989*a*). (4) There are four-bond wave motions (see Fig. 3(d)), where four bonds from the inner part of the chain composed of the sequence $g^{\pm}g^{+}$ are interchanged with a randomly chosen, consecutive pair of bonds located elsewhere in the chain. (5) There are also five-bond wave motions (see Fig. 3(e)), where five of the six consecutive bonds that form a cyclohexane-like ring are interchanged with a single, randomly chosen bond, located in another part of the chain. The residues are renumbered after these moves to maintain the hydrophobic/hydrophilic pattern of interactions. These micromodifications are very important in the low temperature regions where there is a significant amount of secondary structure in the chain. The latter two kinds of move can dissolve incorrectly folded species and can shift deformations along the chain. These motions play the role of an internal reptation (Wall & Mandel, 1975) move. Interestingly, apart from the reduction in simulation time required to fold when the moves shown in Figure 3(d) and (e) are included, the folding pathways are qualitatively identical with those when only the moves shown in Figure 3(a) to (c) are included.

A single MC step (the fundamental time unit) consists of an attempt to modify both ends and $P_{3b}(n-3)$ attempts at a three-bond motion and $P_{4b}(n-4)$ attempts at four-bond motions, where the conformations to be modified are chosen randomly. P_{3b} and P_{4b} are the intrinsic probabilities of three and four-bond motion, respectively. It has been shown by Kolinski *et al.* (1987*b*) that the choice of $P_{3b}=0{\cdot}25$ and $P_{4b}=0{\cdot}75$ gives the correct Rouse dynamics of a random coil chain (random coil dynamics without hydrodynamic interactions), i.e. the scaling of the diffusion constant and largest internal relaxation times with chain length are reproduced In addition, the local dynamics as probed by orientational correlation functions are also reproduced (Valeur *et al.*, 1975). Other values work as well, but this choice gives Rouse dynamics (Yamakawa, 1969) for very short chain lengths. The frequency of wave motions, (4) and (5), varied with the temperature, but usually they were applied once for every time unit (if only equilibrium sampling is desired) and not at all when folding pathway information was desired. Interestingly, the mechanism of folding is the same whether or not four- and five-bond wave motions are invoked.

The MC procedure works as follows. The total energy of the system is calculated after every time unit. Having the energy, E_{old}, before the attempted MC cycle and the new energy, E_{new}, after the chain conformation has been modified, then the probability of transition to the new configuration is:

$$P = \min\{1, \exp[(E_{new}-E_{old})/k_BT]\}.$$

This procedure is repeated many times to allow the model to hunt freely through conformational space, and to achieve thermodynamic equilibrium. It has been proven that in the limit of a large number of MC cycles, the averages over the numbers of cycles converge to the ensemble averages. A detailed discussion concerning the ergodicity of the algorithm has been been presented by Binder (1984, 1987). Here, we want only to point out that in order to sample representative conformational states and to be sure that a model chain is not locked in a long-lived metastable state (i.e. in a deep local, free energy minimum), the simulation of the system under a given set of conditions is repeated many times. The system is started from a number of different initial chain configurations, using different strings of random numbers, and heating and cooling sequences are undertaken in the folding–unfolding transition region. The number of MC cycles in a single simulation run varies, but it is of the order of 10^5 to 10^7 MC time steps.

During the simulation runs, cumulative time averages of some parameters describing the equilibrium properties of a model chain are calculated. The mean-square radius of gyration is calculated from:

$$\langle s^2\rangle = \frac{1}{n}\sum_{i=1}^{n}(r_i-r_{cm})^2, \qquad (1)$$

where r_i is the co-ordinate of the ith residue and r_{cm} is the co-ordinate vector of the center of mass. The normalized helix content, θ_h, is defined as:

$$\theta_h = \frac{f(T^*)-f_{coil}}{f_{nat}-f_{coil}}, \qquad (2)$$

where $f(T^*)$, f_{coil} and f_{nat} are the fraction of helical (g^-) states at the temperature T^*, in the denatured state and in the native state. f_{nat} is 0·8667 for four-member helical bundles with $n=48$, with the remainder of the residues involved in turns, and f_{coil} is assumed to 1/3 (in the high temperature limit). θ_h takes on values between zero (the denatured state) and unity (the fully native state). In the case of β-barrels, an analogous quantity, θ_β, the normalized fraction of $\beta(t)$ states, was calculated. The fraction of t states in the denatured conformation is assumed to be 1/3, and in the native state it is 0·79, with the remainder of the residues involved in bends.

In order to elucidate the influence of a hard wall (ribosome) on the properties of a model protein, we

have also examined the components of the mean-square radius of gyration parallel and perpendicular to the wall. Placing the wall on the XY plane, the perpendicular component of the mean-square radius of gyration is:

$$\langle s_z^2 \rangle = \frac{1}{n} \sum_{i=1}^{n} (z_i - z_{cm})^2, \quad (3)$$

where z_i is the z co-ordinate of the ith residue, and z_{cm} is the z co-ordinate of the center of the mass. α-Helices on a diamond lattice can be parallel to, or perpendicular to, the wall located in the XY plane.

Because of lattice constraints, β-strands are oriented along different lattice directions from α-helices. In order to maintain the same influence of the ribosome, the hard wall was located on a diagonal of the MC box. The wall is defined by the equation $(x+z) = L$, where L is an edge of the MC box. This orientation of the wall enables us to have β-strands parallel or perpendicular to the wall, just as in the case of α-helices. Thus, the perpendicular component of the mean-square radius of gyration is calculated according to:

$$\langle s_\perp^2 \rangle = \frac{1}{n} \sum_{i=1}^{n} ((x_i + z_i - L)/\sqrt{2} - z'_{cm})^2, \quad (4)$$

where z'_{cm} is the component of the center of mass perpendicular to the wall co-ordinate and is given by:

$$z'_{cm} = \frac{1}{n\sqrt{2}} \sum_{i=1}^{n} (x_i + z_i - L). \quad (5)$$

The model chain is placed in a part of a MC box whose size is chosen to be the smallest possible box that can contain the entire (even a fully extended) chain. In the case of α-helices, the chain is attached to the middle of the XY plane at $z = 0$, and the MC box corresponds to a rectangular parallelepiped of dimensions $100 \times 100 \times 50$. In the case of β-proteins, the chain is attached at the center of a cubic box of dimensions $100 \times 100 \times 100$.

3. Results

(a) *Equilibrium properties*

The equilibrium properties of both the four-helix bundle and four member β-barrel models are monitored by following the mean-square radius of gyration as a function of time. This parameter is very sensitive to the conformation of the molecule as its components change their values dramatically during the folding transition. The denaturated (athermal, $T^* =$ infinity) system is cooled down step by step, just as in a real renaturation experiment. The final configuration of a simulation run at temperature T_i serves as the initial configuration for the $i+1$th run at a temperature T_{i+1}. Each run is completed after the system reaches equilibrium. We have obtained the equilibrium value of the mean-square radius of gyration, $\langle s^2 \rangle$ *versus* the reduced temperature T^*. As these transition curves are qualitatively identical with those presented by Skolnick *et al.* (1988, 1989*a*) and Sikorski & Skolnick (1989*a*), we summarize the results below. For all models under consideration, the folding transition for the *in vivo* case is slightly shifted toward lower temperatures. This is probably a kinetic effect; in the transition region only relatively few folding and unfolding transitions are obtained in a single simulation run, and the statistics in this region are not very good. The presence of the wall does not appear to exert an important influence on the equilibrium aspects of the folding. If we remove the wall and leave one chain end pinned, exactly the same configurational properties are observed. Unlike the properties of a chain in the transition region (where the sampling statistics are poor), the configurational properties of the denatured state (at high temperature) and the native state (at low temperature) are more adequately sampled. The chain dimensions (as characterized by the mean-square radius of gyration) are basically the same as in the case of a free chain (folding *in vitro*). The native states are exactly the same as for an *in vitro* folded protein model, with the single exception of the end connected to the wall. The configuration of a chain adjacent to this point can be slightly distorted, especially in the case when α-helices or β-strands are pointed in the direction parallel to the wall. The conformational transitions are observed to be very steep, as was the case for folding *in vitro*. The confirmation of the all-or-none character of the folding transition can be found from the analysis of the number of native hydrophobic contact pairs, N_c, *versus* time. In the denatured state, the number of contacts fluctuates near zero because the contacts are rare and short-lived. In the low temperature region, the number of contacts is almost constant (21 and 20 for the pure native state for α- and β-proteins, respectively); slightly distorted native structures can have one contact less, with minor fluctuations around these values. In the transition region, however, the number of contacts are, for a certain time, essentially constant (with fluctuations) and then change very rapidly from zero to the value characteristic of the native state. This behavior is the same as in the *in vitro* folding model and as in real small proteins. It should be pointed out, however, that in spite of a very co-operative and steep transition, some marginally populated intermediate conformations do exist. Their exact nature will be discussed below.

For the cases examined, the denatured state does not correspond exactly to the random-coil state that is entirely devoid of secondary structure. Especially just above the folding transition, there is a considerable amount of fluctuating secondary structures. For example, the fraction $\theta_h = 0{\cdot}271$ at the temperature $T^* = 0{\cdot}714$ for the four-helix bundle having the primary sequence $\mathbf{H}_1(12)\mathbf{H}_2(10)$ $\mathbf{b}_2^0\mathbf{H}_3(11)\mathbf{H}_4(12)$, and the fraction of t states, $\theta_\beta = 0{\cdot}41$ at $T^* = 1{\cdot}21$ for the four-member, β-barrel having the primary sequence $\mathbf{B}_1(9)\mathbf{b}_1^0\mathbf{B}_2(8)$ $\mathbf{b}_2^0\mathbf{B}_3(9)\mathbf{b}_3^0\mathbf{B}_4(11)$. The relatively large amount of

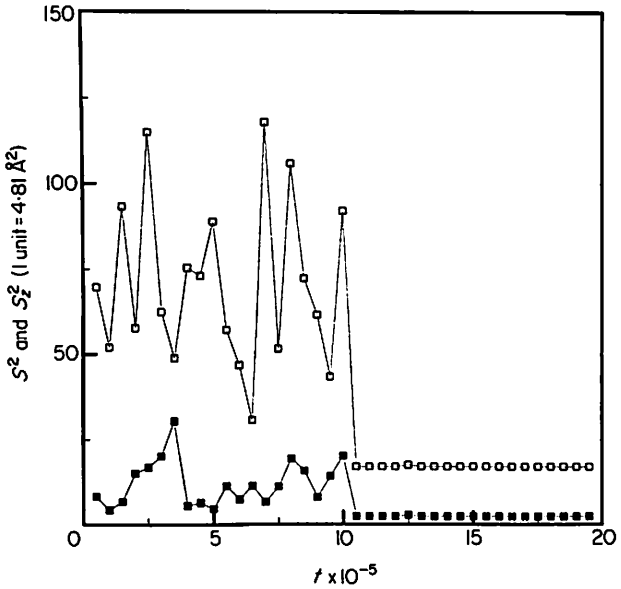

Figure 4. Plot of S^2 and S_z^2 *versus* time (see the text for details), for the assembly of a 4-helix bundle having primary sequence $\mathbf{H}_1(12)\mathbf{H}_2(12)\mathbf{H}_3(10)\mathbf{b}_3^0\mathbf{H}_4(11)$ and that assembles to a native state whose principal axis is parallel to the plane of the wall. (□) S^2; (■) S_z^2.

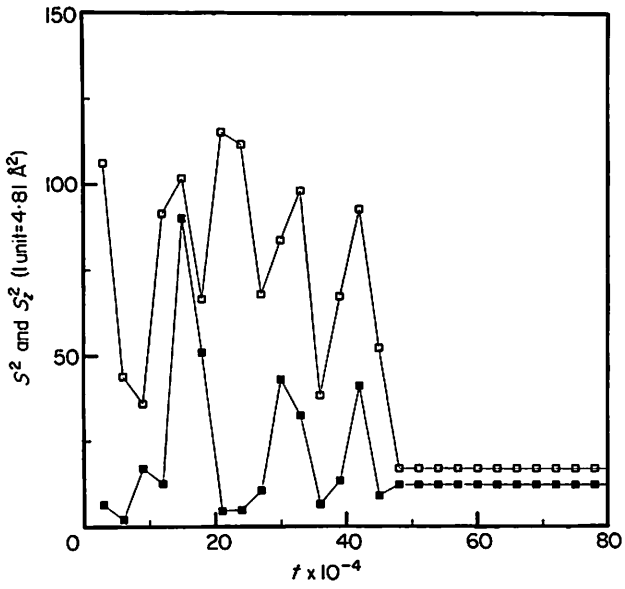

Figure 5. Plot of S^2 and S_z^2 *versus* time (see the text), for the assembly of a 4-helix bundle having primary sequence $\mathbf{H}_1(12)\mathbf{H}_2(12)\mathbf{H}_3(10)\mathbf{b}_3^0\mathbf{H}_4(11)$ and that assembles to a native state whose principal axis is perpendicular to the plane of the wall. (□) S^2; (■) S_z^2.

secondary structure found in the unfolded state is included to bias the case towards co-translational folding. We remind the reader that the secondary structure present in the denatured state is constantly forming and dissolving. All three cases, the *in vivo* case, the *in vitro* case and the pinned chain in solution model, exhibit the same amount of secondary structure. This provides a further confirmation that the wall does not exert a substantial influence on the equilibrium properties of the model protein-chain.

(b) *Dynamic properties*

(i) *Post-translational folding simulations*

In order to determine when folding occurs, changes in the instantaneous value of the square of the radius of gyration, S^2 and the component of S^2 perpendicular to the wall, S_z^2 are plotted *versus* time in Figures 4 and 5, for the assembly of a four-helix bundle having the primary sequence $\mathbf{H}_1(12)\mathbf{H}_2(12)\mathbf{H}_3(10)\mathbf{b}_3^0\mathbf{H}_4(11)$, and where the final folded structure is respectively parallel and perpendicular to the plane of the wall. In the case of assembly parallel to the wall, in the native conformation, S_z^2 is substantially smaller than S^2, whereas, for the case of assembly perpendicular to the wall, these two quantities are comparable in magnitude.

In Figure 6, the instantaneous value of the square of the radius of gyration, S^2, and the component of S^2 perpendicular to the wall, S_z^2, are plotted *versus* time for the case of a β-barrel having the primary sequence $\mathbf{B}_1(9)\mathbf{b}_1^0\mathbf{B}_2(8)\mathbf{b}_2^0\mathbf{B}_3(9)\mathbf{b}_3^0\mathbf{B}_4(11)$, whose principal axis in the native conformation lies perpendicular to the plane of the wall. In Figure 6, there are huge fluctuations of S^2, between 20 and 100, as the model chain randomly thrashes about in configuration space. At the time $t = 53{,}250$, the radius of gyration diminishes substantially and the fluctuations have much smaller amplitude. The same behavior can be noted in Figures 4 and 5, at $t = 10^6$ and $4{\cdot}5 \times 10^5$, respectively, for the cases of assembly parallel and perpendicular to the wall. For both kinds of folding motifs, this suggests that some kind of intermediate states are formed.

As time further progresses, at $t = 53{,}500$ for the β-barrel, and at $t = 11 \times 10^5$ and 5×10^5 for α-helical proteins assembling parallel and perpendicular to the wall, the model chains have radii of gyration corresponding to their respective native states ($\langle S^2 \rangle = 26$ for the β-protein and $\langle S^2 \rangle = 17$ for the four-helix bundle). In other words, they have successfully folded.

For the case of four-helix bundles, the native state assembles predominantly parallel to the wall, while for β-proteins, assembly occurs predominantly perpendicular to the wall. The origin of the orientation dependence of assembly will be discussed below.

During each simulation run, the conformations of the chain are stored. A detailed analysis of a large number of folding trajectories leads to some general conclusions about pathways of folding. In Figures 7 and 8, representative folding pathways are displayed for α- and β-proteins having a primary sequence identical with those in Figures 4 and 6, respectively. These are typical folding pathways observed for all the primary sequences, and thus, we focus on these two cases in the interest of brevity.

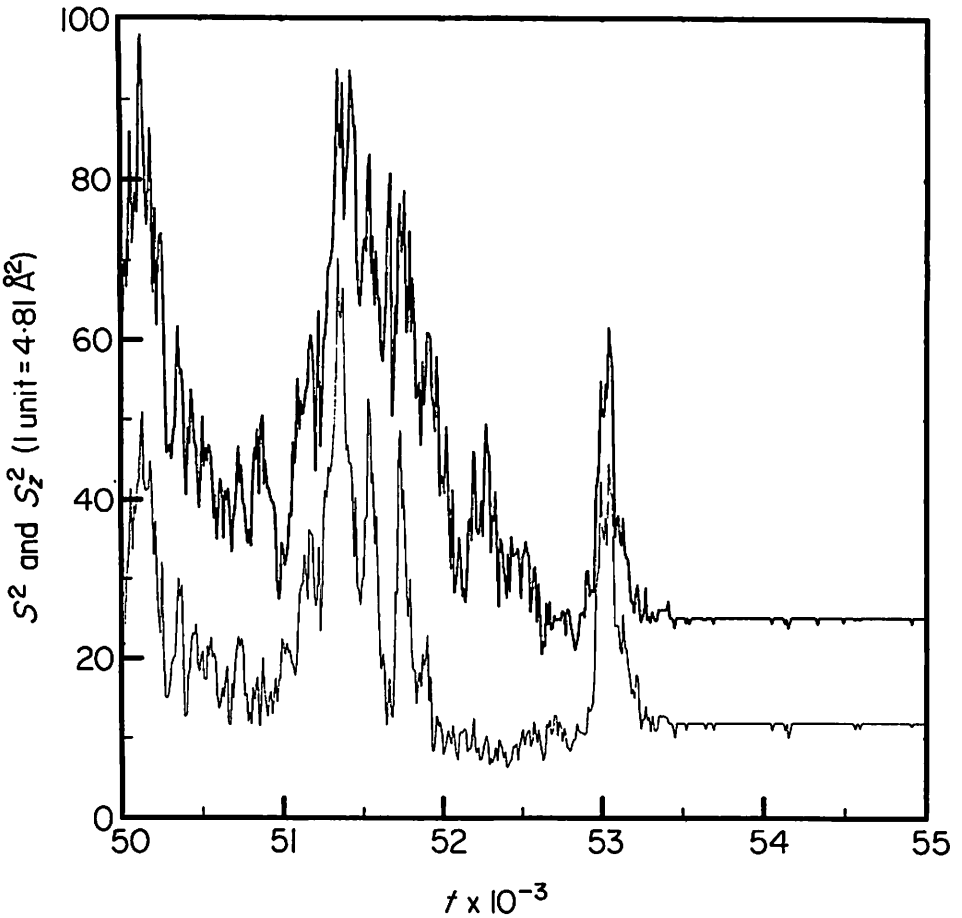

Figure 6. Plot of S^2 and S_z^2 *versus* time for the assembly of a 4-member β-barrel having primary sequence $B_1(9)b_1^0B_2(8)b_2^0B_3(9)b_3^0B_4(11)$ and that assembles to a native state whose principal axis is perpendicular to the plane of the wall. Top line, S^2; bottom line, S_z^2.

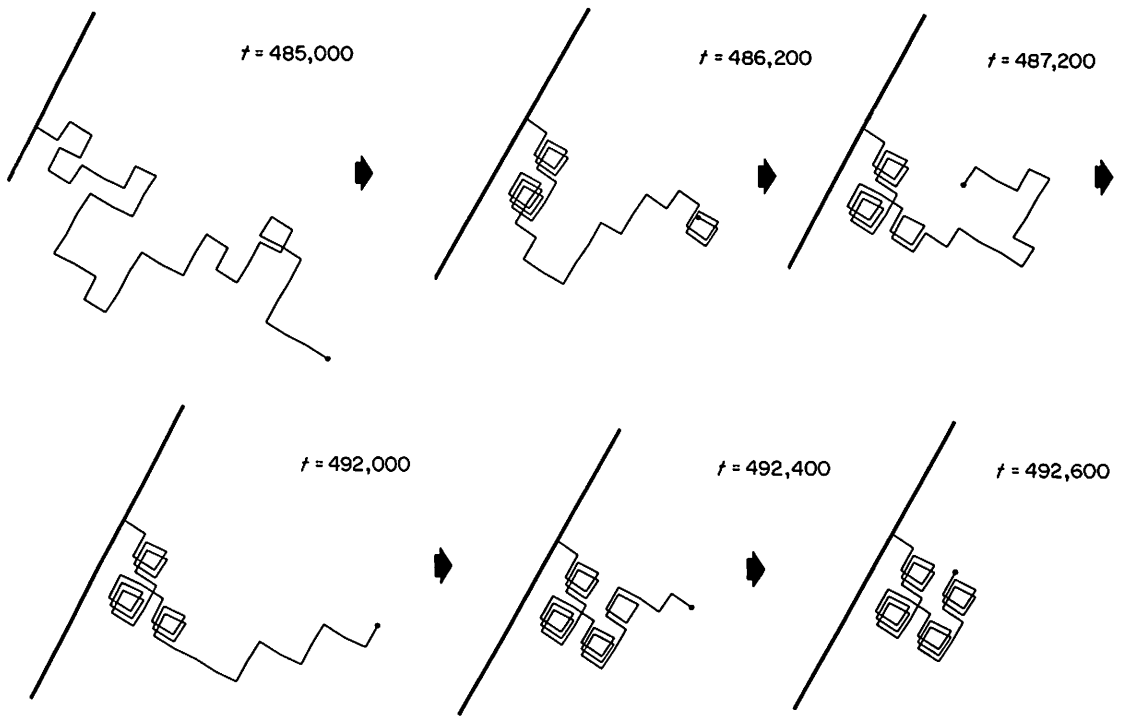

Figure 7. Four-helix bundle that assembles parallel to the plane of the wall. The times indicated in the Figure are from the start of the simulation run.

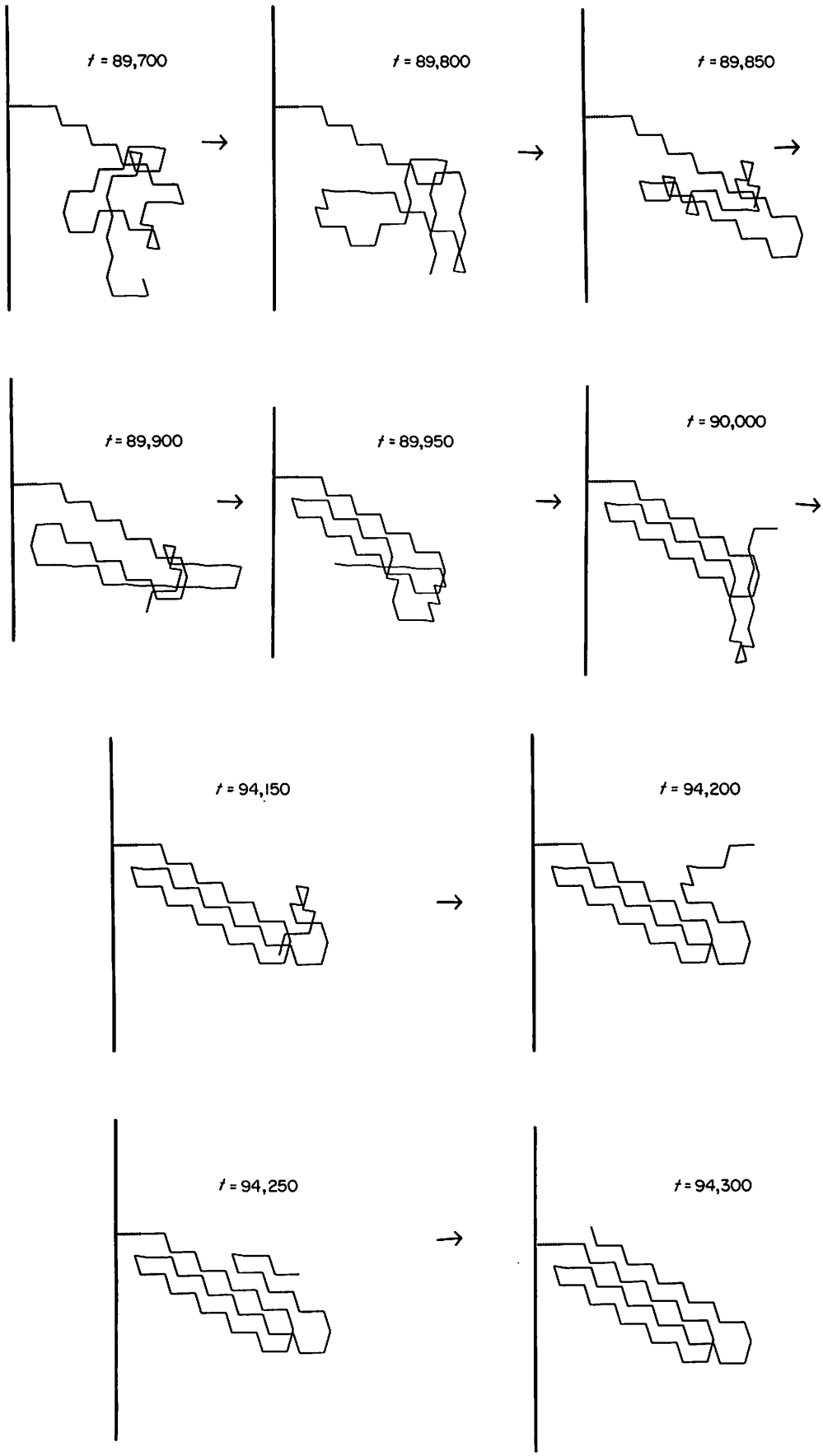

Figure 8. Representative folding trajectory of a 4-member β-barrel that assembles perpendicular to the plane of the wall. The times indicated in the Figure are from the start of the simulation run.

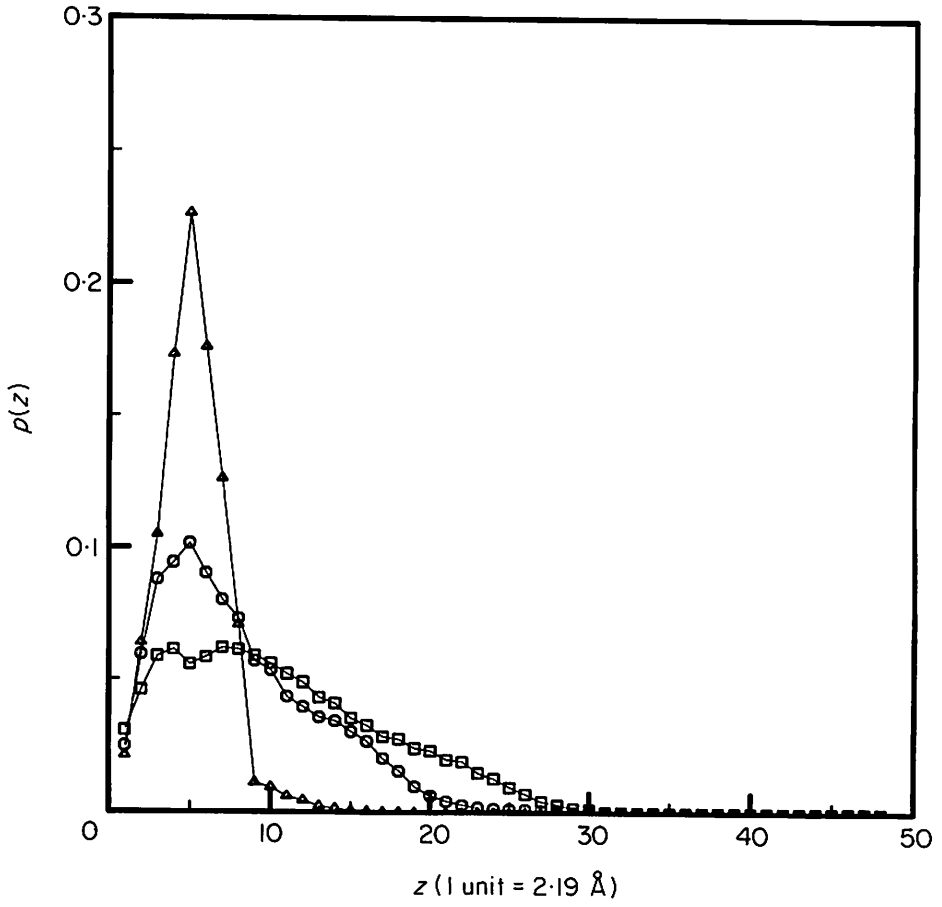

Figure 9. Density distribution of residues, $p(z)$, *versus* distance from the wall (see the text) for the case of a 4-helix bundle at the times indicated for the case of assembly parallel to the plane of the wall. (△) $t = 500{,}000$; (○) $t = 490{,}000$; (□) $t = 80{,}000$.

It should be pointed out that, during most of the time, the chain undergoes conformational changes that are mostly irrelevant to folding, and that the important conformational changes happen over a very short period. The very first structure formed in both α and β cases is a hairpin. The formation of a hairpin occurs between $t = 485{,}000$ and $t = 486{,}200$ for the α-protein (see Fig. 7) and between $t = 89{,}700$ and $t = 89{,}850$ for the β-protein (see Fig. 8). Only the assembly of the hairpin that is closest to the connection point can lead to the formation of the entire native state. The assembly of a hairpin at the free end of a chain cannot be successful because, due to the lattice constraints, folded secondary structure cannot diffuse. The same situation occurs with the formation of a central hairpin. Thus, compared to folding *in vitro*, the number of pathways is limited; in "free solution", any hairpin can lead to a native state, and the average waiting time for the start of successful folding is considerably longer, compared with the *in vitro* model. Moreover, the formation of the hairpin closest to the wall is constrained. For folding *in vitro*, a hairpin starts to assemble beginning from a bend or in its vicinity (α-proteins only), and then the hairpin zips up. (This pathway also occurs when more general moves that permit helix and hairpin motion are included. In either case it will not affect the conclusions about the relative importance of co-translational *versus* post-translational folding.) For folding *in vivo*, the α-helical stretch (or β-strand) closest to the wall has to be formed first, to place the bend region at a certain distance from the wall. This means that there is a huge entropic barrier to folding initiation, and any incorrectly folded structure leads to the tension near the point of connection that has to be released.

Compared to the *in vitro* model of folding of helices, the lifetime of the helical hairpin (closest to the wall) is greater. Because this hairpin is predominantly oriented parallel to the wall (in about 70% of the simulation runs that lead to successful folding), it seems that the presence of the wall stabilizes the hairpin. In the case of β-proteins, the β-strands are oriented in the direction perpendicular to the wall in about 60% of the simulation runs. The difference between the orientations of the native states in α- and β-proteins is caused by the difference in the stabilization energy of a hairpin. The orientation parallel to the wall causes an increase in energy, as there must be a bend that allows the chain to attach to the wall. For α-helices, this is a much smaller fraction of the total energy of hairpin formation than for the β-protein. Given the

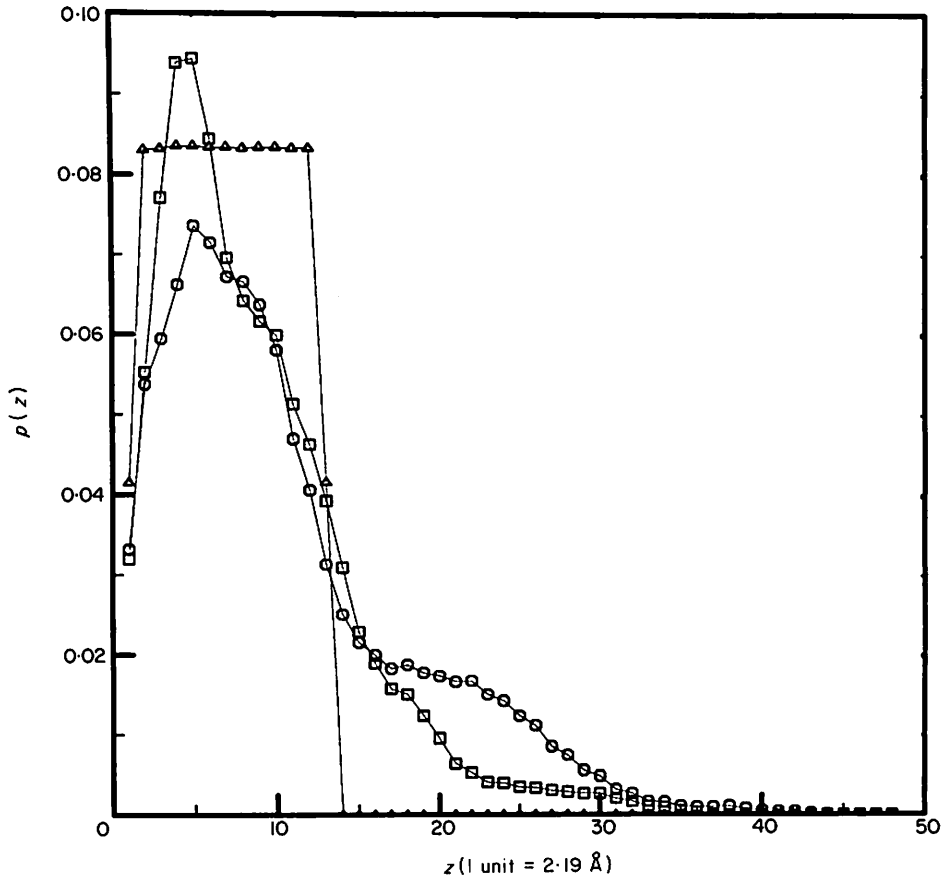

Figure 10. Density distribution of residues, $p(z)$, *versus* distance from the wall (see the text) at the times indicated for the case of a 4-helix bundle assembly perpendicular to the plane of the wall. (△) $t = 810{,}000$; (○) $t = 450{,}000$; (□) $t = 30{,}000$.

smaller stability of β-hairpins as compared to helical hairpins, this is sufficient to favor the perpendicular orientation of the native β-barrel.

The next step along the folding pathway involves the assembly of the adjacent helix (β-strand) onto the existing hairpin. This process is relatively fast, it occurs between $t = 486{,}200$ and $t = 492{,}000$ for the α-protein and between $t = 89{,}850$ and $t = 90{,}000$ for the β-proteins (see Figs 7 and 8, respectively). The mechanism of assembly is the same as in the *in vitro* folding model and involves the zipping up of secondary structure on site, beginning with the bend. In this stage of the folding, there is a much smaller entropic barrier and the energy term prevails. This places the system (3-member helical bundle or 3-member β-barrel) in a local minimum of free energy. Starting from this point, the excluded volume hinders folding by blocking a number of accessible conformations for the last helix (β-strand). Thus, these intermediate states are relatively long-lived with the remaining (unfolded) part of a chain thrashing around. In the last step, the chain end starts to assemble by snaking its way into the native conformation. This step occurs between $t = 492{,}000$ and 492,600 for the helix bundle and between $t = 90{,}000$ and $t = 94{,}300$ for the β-barrel.

The unfolding process is basically the reverse of the folding described above. The helix (β-strand) closest to the free end dissolves. Then the next helix undergoes dissolution, followed by the subsequent helix, and then those closest to the wall dissolve.

Additional insight into the folding process is afforded by Figures 9 and 10, where the probability of finding a residue at a distance z from the wall, $p(z)$ is plotted *versus* z at various times following the initiation of folding for the case of four-helix bundle assembly parallel and perpendicular to the wall. Due to entropic effects at short times, there is a density depletion layer close to the wall. This tends to push residues further into solution when they are in the random coil state. As folding proceeds for the case of parallel assembly (Fig. 9), there is a spike corresponding to the distance of the assembled bundle from the wall. In Figure 10, there is a plateau region at long times corresponding to the location of each of the 11 residues located in the helices that are oriented perpendicular to the wall.

(ii) *Co-translational folding simulations*

We next summarize the results from our co-translational simulations. We begin with the case of synthesis from a hard wall using the identical local

dynamics as in the post-translational case. Synthesis proceeds by the sequential addition of pairs of residues; pairs are added because this allows for chain growth without the assembled portions of the chain having to be moved. This is a feature of the diamond lattice.

We chose a range of pausing times up to 20,000 steps between the subsequent addition of pairs of residues. The time between residue pair addition is about the mean assembly time for the models in free solution. Thus, the synthesis of a chain takes on the order of 4×10^5 time steps. This is comparable in the *in vitro* case to the total simulation time that allows these systems to fold fully. For example, for the four-helix bundle with three neutral bends, the mean time from initiation of assembly to the formation of the fully native conformation is about 3×10^4 time units. In no case did we observe successful folding to the native conformation until synthesis was completed. Then the system folded in a manner identical with the post-translational case; it, in fact, folded post-translationally. Identical behavior was observed in the case of synthesis from a point in solution.

This is not to say that no tertiary structure formation occurred during synthesis. Rather, these intermediates are marginally stable and are observed to form and dissolve many times during the course of chain synthesis. It is not at all surprising to find that the intermediates must be marginally stable, otherwise a two-state model would not adequately describe the conformational transition. Thus, these intermediates fold and dissolve many times during the time the chain is synthesized.

It might legitimately be argued that, since the algorithm employs only three and four-bond kinks and end motions, movement of assembled secondary structure cannot occur, and hence, the algorithm artificially biases towards post-translational assembly. This limitation is particularly severe in the case of α-helices; β-strands can be moved about by a kink defect mechanism (Skolnick & Kolinski, 1990). Thus, we introduced into the algorithm the possibility of single helix and helical hairpin rotations and translations (Sikorski & Skolnick, 1990). Just as in the free solution case, including these moves did not change the mechanism of assembly from post-translational, on-site construction.

It might be argued that these results are due to the particular choice of a lattice, which in turn immediately specifies the nature of the small-scale local moves. However, for the free solution case we have observed identical folding pathways on a 24 nearest neighbor lattice that employs a completely different set of local moves and a more detailed chain representation that includes side-chains (Skolnick & Kolinski, unpublished results). This argues that the folding pathways that are observed are independent of the choice of lattice and may, in fact, be universal, at least for the class of all lattice models.

4. Summary and Discussion

In the context of a highly simplified model of *in vivo* protein folding, we have examined the validity of the co-translational *versus* post-translational folding hypothesis. Modeling the ribosome as an inert hard wall, we find that folding proceeds by a punctuated on-site construction mechanism, just as is found for a free solution. For the case of α-helical proteins, there is a marked preference for assembly with the principal axis of the helices parallel to the plane of the wall. Rather than the three-helix conformation being the long-lived intermediate prior to native state assembly as in the *in vitro* case, the relatively long-lived intermediate is the helical hairpin. Basically, the wall accelerates folding by eliminating wrong intermediates. Since it is possible to place an α-helix parallel to the wall without distorting the ends, this kinetic effect dominates and results in about a 3:1 preference of assembly parallel over perpendicular to the wall. This effect is aided, somewhat, by the presence of a density depletion layer that is of entropic origin near the wall (see Figs 9 and 10). That is, it is more likely that the helices can be pulled from the solution without running into other pieces of chain that could hinder assembly by acting as obstacles to the movement of the random-coiled segment that will later form helices.

Turning to the assembly of four member β-barrels, unlike the helix case, approximately 60% fold perpendicular to the wall. This is due to the fact that if the axis of the planar β-strand is parallel to the wall, the end must be kinked in order to attach to the wall. The β-hairpin, for our choice of parameters, is substantially less stable than a helical hairpin that does not have to distort to attach to the wall (the virtual bonds are perpendicular to the helix axis; a β-hairpin parallel to the wall is an inherently less stable conformation). Whether this will hold for more realistic models remains to be seen.

In spite of substantial effort, we never observed a case of successful co-translational folding. Perhaps this is an artifact of the algorithm; however, we believe this not to be the case for the following reasons. First, while partially assembled nascent folding conformations are observed, they are simply not stable enough to survive during the entire course of synthesis without dissolving. In fact, they form and dissolve many times during the course of synthesis. However, if disulfide bonds could form (they are absent in the present model but will be implemented later), then when the partially folded intermediate did occur, this could be accompanied by disulfide bond formation. Thus, disulfide bond formation during synthesis (Bergman & Kuehl, 1979) does not prove co-translational folding, rather, just as in these simulations, it indicates that partially folded forms are populated, and more to the point, it does not prove that in partially folded conformations without cross-links the conformation would not unfold again. Second, because substantial

intermediate populations are absent *in vitro*, unless there are entities in the cell that stabilize such late folding conformers *in vivo*, one would not expect to see intermediate populations having a lifetime of the order of seconds or longer *in vivo*. That is, just as in the free solution case, one expects to find secondary structures of marginal stability that are in dynamic equilibrium with their unfolded forms. Thus, although these models are highly simplified, they are very suggestive that post-translational folding obtains for single-domain proteins.

In the case of multiple-domain proteins, we could expect the first domain to fold and then, depending upon its intrinsic stability, the first folded domain might persist until synthesis of the second domain is completed. This second domain might fold post-translationally or co-translationally if the first domain acts to stabilize the folding intermediates present in the second domain. These simulations are suggestive, however, that a wall, even if it is inert, might stabilize the first domain. A key point to remember is that, even if the folded conformation is at the global free energy minimum in the presence of the constraint (the tether attaching it to the ribosome), it might not be at the global free energy minimum in the absence of the constraint, and it might be kinetically trapped. Overall, these simulations are supportive of perhaps a punctuated, post-translational folding model. Single domain proteins fold post-translationally, but for a multiple domain protein, the entire molecule need not be synthesized before folding occurs.

This research was supported, in part, by NIH grant GM-37408 from the Division of General Medical Sciences, United States Public Health Service. We are grateful to Professors John Taylor and Larry McLaughlin for originally suggesting the problem.

References

Andria, G. & Taniuchi, H. (1978). *J. Biol. Chem.* **253**, 2262–2270.

Anfinsen, C. B. (1972). *Biochem. J.* **128**, 737–749.

Anfinsen, C. B. & Scheraga, H. A. (1975). *Advan. Protein Chem.* **29**, 205–300.

Baldwin, R. L. & Creighton, T. E. (1980). In *Protein Folding* (Jaenick, R., ed.), pp. 217–260, Elsevier/North Holland, Amsterdam.

Bergman, L. W. & Kuehl, W. M. (1979). *J. Biol. Chem.* **254**, 8869–8876.

Binder, K. (1984). Editor of *Applications of the Monte Carlo Method in Statistical Physics*, pp. 213–269, Springer, Berlin.

Binder, K. (1987). Editor of *Monte Carlo Methods in Statistical Physics*, pp. 1–45, Springer, Berlin.

Brandts, J. F., Halvorson, H. R. & Brennan, M. (1975). *Biochemistry*, **14**, 4953–4963.

Brems, D. N., Cass, R. & Stellwagen, E. (1982). *Biochemistry*, **21**, 1488–1493.

Bundi, A., Andreatta, R. H., Rittel, W. & Wuthrich, K. (1976). *FEBS Letters*, **64**, 126–129.

Bundi, A., Andreatta, R. H. & Wuthrich, K. (1978). *Eur. J. Biochem.* **91**, 201–208.

Creighton, T. E. (1984). *Advan. Biophys.* **18**, 1–20.

Creighton, T. E. (1985). *J. Phys. Chem.* **89**, 2452–2459.

Creighton, T. E. (1988). *Proc. Nat. Acad. Sci., U.S.A.* **85**, 5082–5086.

De Coen, J. L. (1970). *J. Mol. Biol.* **49**, 405–414.

Dyson, H. J., Rance, M., Houghton, R. A., Lerner, R. A. & Wright, P. E. (1988*a*). *J. Mol. Biol.* **201**, 161–200.

Dyson, H. J., Rance, M., Houghten, R. A., Lerner, R. A. & Wright, P. E. (1988*b*). *J. Mol. Biol.* **201**, 201–217.

Garel, J. R. & Baldwin, R. L. (1973). *Proc. Nat. Acad. Sci., U.S.A.* **70**, 3347–3351.

Goldenberg, D. P. & Creighton, T. E. (1985). *Biopolymers*, **24**, 167–182.

Hill, T. L. (1956). *Statistical Mechanics*, pp. 143–191, McGraw-Hill, New York.

Hruby, V. J. (1985). *Pept. Anal. Synth. Biol.* **7**, 1–14.

Hua, Q. X., Qian, Y. Q. & Tsou, C. L. (1985). *Sci. Sin. (Engl. edit.)*, **28B**, 854–862.

Iwata, K. & Kurata, M. (1969). *J. Chem. Phys.* **50**, 4008–4013.

Karplus, M. & Weaver, D. L. (1976). *Nature (London)*, **160**, 404–406.

Karplus, M. & Weaver, D. L. (1979). *Biopolymers*, **18**, 1421–1427.

Kolinski, A., Skolnick, J. & Yaris, R. (1986*a*). *J. Chem. Phys.* **85**, 3585–3597.

Kolinski, A., Skolnick, J. & Yaris, R. (1986*b*). *Proc. Nat. Acad. Sci., U.S.A.* **83**, 7267–7271.

Kolinski, A., Skolnick, J. & Yaris, R. (1987*a*). *Biopolymers*, **26**, 937–962.

Kolinski, A., Skolnick, J. & Yaris, R. (1987*b*). *J. Chem. Phys.* **86**, 1567–1585.

Kremer, K., Baumgartner, A. & Binder, K. (1981). *Phys. A*, **15**, 2879–2892.

Kuwajima, K., Hiraoka, Y., Ikeguchi, M. & Sugai, S. (1985). *Biochemistry*, **24**, 874–881.

Lang, K., Schmid, F. X. & Fischer, G. (1987). *Nature (London)*, **329**, 268–270.

Lee, S., Karplus, M., Bashford, D. & Weaver, D. (1987). *Biopolymers*, **26**, 481–506.

Liao, M. J. & Khorana, H. G. (1984). *J. Biol. Chem.* **259**, 4194–4199.

Lin, L. N. & Brandts, J. F. (1987). *Biochemistry*, **26**, 3537–3543.

Phillips, D. C. (1967). *Proc. Nat. Acad. Sci., U.S.A.* **57**, 484–495.

Purvis, I. J., Bettany, A. J. E., Santiago, T. C., Coggins, J. R., Duncan, K., Eason, R. & Brown, A. J. P. (1987). *J. Mol. Biol.* **193**, 413–417.

Schiffer, M. D. & Edmunson, A. (1967). *Biophys. J.* **1**, 121–135.

Shoemaker, K. R., Kim, P. S., Brems, D. N., Marqusee, S., York, E. J., Chaikin, I. M., Stewart, J. M. & Baldwin, R. L. (1985). *Proc. Nat. Acad. Sci., U.S.A.* **82**, 2349–2353.

Shoemaker, K. R., Kim, P. S., York, E. J., Stewart, J. M. & Baldwin, R. L. (1987). *Nature (London)*, **326**, 563–567.

Sikorski, A. & Skolnick, J. (1989*a*). *Biopolymers*, **28**, 1097–1113.

Sikorski, A. & Skolnick, J. (1989*b*). *Proc. Nat. Acad. Sci., U.S.A.* **86**, 2668–2672.

Sikorski, A. & Skolnick, J. (1990). *J. Mol. Biol.* **212**, 819–836.

Skolnick, J. & Kolinski, A. (1990). *J. Mol. Biol.* **212**, 787–817.

Skolnick, J., Kolinski, A. & Yaris, R. (1988). *Proc. Nat. Acad. Sci., U.S.A.* **85**, 5057–5061.

Skolnick, J., Kolinski, A. & Yaris, R. (1989*a*). *Biopolymers*, **28**, 1059–1095.

Skolnick, J., Kolinski, A. & Yaris, R. (1989*b*). *Proc. Nat. Acad. Sci., U.S.A.* **86**, 1229–1233.
Taniuchi, H. (1970). *J. Biol. Chem.* **245**, 5459–5468.
Tsong, T. Y. & Baldwin, R. L. (1978). *Biopolymers*, **17**, 1669–1678.
Tsou, C. L. (1988). *Biochemistry*, **27**, 1809–1812.
Valeur, B., Jarry, J. P., Geny, F. & Monnerie, L. (1975). *J. Poly. Sci. Poly. Phys.* **13**, 667–674.
Wall, F. T. & Mandel, F. (1975). *J. Chem. Phys.* **63**, 4592–4595.
Yamakawa, H. (1969). In *Modern Theory of Polymer Solutions*, pp. 101–187, Harper and Row, New York.

Edited by T. Richmond