J. Mol. Biol. (1991) 221, 499–531

Dynamic Monte Carlo Simulations of a New Lattice Model of Globular Protein Folding, Structure and Dynamics

Jeffrey Skolnick†

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

and Andrzej Kolinski

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

(Received 1 June 1990; accepted 29 April 1991)

A long-standing problem of molecular biology is the prediction of globular protein tertiary structure from the primary sequence. In the context of a new, 24-nearest-neighbor lattice model of proteins that includes both α and β-carbon atoms, the requirements for folding to a unique four-member β-barrel, four-helix bundles and a model α/β-bundle have been explored. A number of distinct situations are examined, but the common requirements for the formation of a unique native conformation are tertiary interactions plus the presence of relatively small (but not irrelevant) intrinsic turn preferences that select out the native conformer from a manifold of compact states. When side-chains are explicitly included, there are many conformations having the same or a slightly greater number of side-chain contacts as in the native conformation, and it is the local intrinsic turn preferences that produce the conformational selectivity on collapse. The local preference for helix or β-sheet secondary structure may be at odds with the secondary structure ultimately found in the native conformation. The requisite intrinsic turn populations are about 0·3% for β-proteins, 2% for mixed α/β-proteins and 6% for helix bundles. In addition, an idealized model of an allosteric conformational transition has been examined. Folding occurs predominantly by a sequential on-site assembly mechanism with folding initiating either at a turn or from an isolated helix or β-strand (where appropriate). For helical and β-protein models, similar folding pathways were obtained in diamond lattice simulations, using an entirely different set of local Monte Carlo moves. This argues strongly that the results are universal; that is, they are independent of lattice, protein model or the particular realization of Monte Carlo dynamics. Overall, these simulations demonstrate that the folding of all known protein motifs can be achieved in the context of a single class of lattice models that includes realistic backbone structures and idealized side-chains.

Keywords: protein folding; folding pathways; secondary structures; bends and turns; assembly mechanism

1. Introduction

The development of algorithms capable of predicting the tertiary structure of a globular protein, given the primary sequence of amino acid residues that it comprises, is a long-sought objective of theoretical biophysical chemistry (Anfinsen, 1973; Creighton, 1985, 1988). However, because of the very large number of degrees of freedom involved in the folding of a globular protein, the straightforward application of standard techniques, such as molecular dynamics or Brownian dynamics, is impossible (Skolnick & Kolinski, 1989). These techniques, at best, sample time scales of the order of tens of nanoseconds, and the time required to fold

† Author to whom all correspondence should be addressed.

0022-2836/91/180499-33 $03.00/0

a protein *in vitro* is typically of the order of seconds or minutes (Garel & Baldwin, 1973; Creighton, 1988). Thus, alternative methods that render the problem tractable are required (Ueda *et al.*, 1978; Taketomi *et al.*, 1988; Chan & Dill, 1989; Skolnick & Kolinski, 1989). All involve, in one way or another, the parsing of the conformational space that must be searched. Very much in this spirit, we have developed a series of diamond lattice models of β- and α-proteins that successfully reproduce the essential qualitative features of both the thermodynamics and kinetics of protein folding (Kolinski *et al.*, 1986, 1987*a*,*b*; Sikorski & Skolnick, 1989*a*,*b*, 1990; Skolnick & Kolinski, 1989, 1990*a*; Skolnick *et al.*, 1988, 1989*a*,*b*). However, a number of the structural features of real globular proteins cannot be recovered on a diamond lattice. Among the more salient deficiencies are the inability to pack α-helices and β-sheets in a parallel orientation, the requirement that helices contain four residues per turn, the inability of β-sheets to exhibit a twist and the low packing density characteristic of a diamond lattice. Here, we introduce a new class of lattice models that does not suffer from these limitations and, by way of illustration, examine the folding of four-member β-barrels, four-helix bundles and mixed α/β-proteins.

Unlike in the diamond lattice model, where each α-carbon is represented by a single lattice site, it is desirable that the α-carbon should occupy a finite size; that is, it should have a surface. This permits the role of packing to be examined directly. By creating a surface of interaction, as opposed to a point, the expanse of configuration phase space that must be searched is effectively reduced. One might expect that two surfaces can roll over each other and thereby aid in protein folding initiation. However, the lattice representation must remain computationally tractable, so that a reasonable number of independent folding and unfolding runs can be performed, thereby demonstrating that the results are reproducible. Furthermore, this also permits partial exploration of the range of parameters that will fold a given protein motif.

As in the diamond lattice simulations (Kolinski *et al.*, 1986, 1987*a*,*b*; Sikorski & Skolnick, 1989*a*,*b*, 1990; Skolnick & Kolinski, 1989, 1990*a*; Skolnick *et al.*, 1988, 1989*a*,*b*), we adhere to the following criteria. First, the native conformation cannot be specified by a target potential in the algorithm; that is, interactions between any, and not necessarily a native, pair of residues can occur when the residues are sufficiently close. To avoid building in a particular motif, interactions between pairs of side-chains must be independent of the backbone conformation. With respect to the thermodynamics of the conformational transition, for small single domain proteins, the transition that emerges and which is not *a priori* built into the simulation must be reasonably well approximated by a two-state model (Brandts *et al.*, 1975; Creighton, 1988). If kinetic information is required, the simulation must be in at least qualitative accord with experiments that indicate that the transition state is near the native state (Creighton, 1985, 1988).

Perhaps the most stringent test of this class of models is the ability to fold mixed α/β-proteins. In the case of a diamond lattice, this is simply not possible. Indeed, the acquisition of the ability to pack α-helices and β-sheets in parallel provided the initial impetus for the development of a more sophisticated lattice description. The lattice representation we employ consists of the cyclic permutation of vectors of the type $(\pm 2, \pm 1, 0)$ that join contiguous α-carbon atoms. These lattice vectors are like a "knight's walk" in chess, except that, here, the walk occurs in three dimensions. We shall refer to this model as a "210 lattice', for short.

We summarize, here, some of the more salient properties of this 24-nearest-neighbor lattice. If the range of distances observed in real proteins between second nearest-neighbor α-carbon atoms is allowed, each interior α-carbon has 18 accessible conformations. This lattice has a rather high degree of conformational freedom, and the question is "can it reproduce the local geometry of helices and sheets"? Otherwise, it might reasonably be argued, why not use a lower co-ordination number lattice. As shown below (see Fig. 2, for example), the 210 lattice does a remarkably good job of reproducing the local protein geometry. Elsewhere, we show that it reproduces the known global folds of real proteins at the level of 1·5 to 2 Å (1 Å = 0·1 nm) root-mean-square (M. Milik & J. Skolnick, unpublished results).

Given a lattice description, one must possess algorithms capable of folding to the desired native conformation from the unfolded state. As in the diamond lattice case, we employ dynamic Monte Carlo techniques (Binder, 1984, 1987). This method entails the solution of a stochastic kinetics master equation that describes the time evolution of the system (Binder, 1987). In the limit of a very long series of steps, each of which involves conformational modifications of the chain, the system will sample an equilibrium distribution of states. Whether or not the folding trajectories are physical depends on the kinds of move employed, and care must be taken, if pathway information is also desired, to make these elemental moves as physical as possible to avoid significant time scale distortions. We chose local conformational transitions as well as some longer wavelength moves that permit secondary structure translations and rotations. Among the questions addressed below are whether the folding rules for a given motif are the same on the 210 lattice and the diamond lattice. It is also important to determine whether the folding and unfolding pathways are similar. If both of the folding rules and pathways are similar, then given the vast differences between the two kinds of lattice and the particular kinds of elemental motion employed, this would strongly argue for the lattice independence of the results. These are among the points explored below.

The outline of the remainder of the paper is as follows. In Overview of the Model, we present the

model, including the details of the backbone and side-chain representations, and the nature of the interactions that are employed. In Approximate Analytic Theory of Folding, we present an approximate analytic theory that allows for a reasonably good prediction of the thermal transition curve, given the folded conformation. This is useful both for assessing the validity of the all-or-none approximation as well as identifying the transition region. In Results, we present results for the folding of β-barrels, four-helix bundles, and a mixed motif of α/β proteins, and in Summary and Conclusions, we summarize the conclusions obtained from these simulations.

2. Overview of the Model

(a) *Chain geometry*

The protein consists of an α-carbon atom plus side-chain (β-carbon) representation of each amino acid. As shown in Figure 1(a), the entire system is embedded in an underlying cubic lattice, constructed from vectors of the type $(\pm 1, 0, 0)$; that is, the distance between any two adjacent cubic lattice points is unity (1·7 Å) in model units. Adjacent α-carbon atom positions are connected by vectors of the type $(\pm 2, \pm 1, 0)$. Each α-carbon atom occupies a central lattice site plus the six adjacent cubic lattice sites. Thus, as is evident in Figure 1(b), a hard core envelope of the chain having a finite backbone thickness is created. That this is of reasonable size is verified by the fact that fits to the Brookhaven Protein Data Bank crystal structures differ by 0·1 Å between fits with just the α-carbon atoms and with the full seven-point hard core envelope. The backbone conformation of the ith α-carbon is specified by the square of the distance between α-carbons $i-1$ and $i+1$, r_θ^2, with θ the bond angle between three consecutive α-carbon atoms. There is a one-to-one mapping of distances and bond angles; we use r_θ^2 for convenience. In model units, where the distance between consecutive α-carbon atoms equals $5^{1/2}$, (3·785 Å), physically reasonable values of r_θ^2 are 6, 8, 10, 12, 14, 16 and 18, as r_θ^2 spans the range from 7 to about 19 in real proteins. A planar β-strand can be constructed from a sequence of 16 states; a β-strand with a twist can be built from a sequence of 18 states. An α-helix having four residues per turn can be constructed from a sequence of 12 states, and a helix having 3·5 residues per turn can be constructed from a sequence of 6 and 14 states, to name just one possibility.

In Figure 2, we display the superimposed structures of the 12-state helix and the 16-state β-strand on real elements of (a) an α-helix and (b) β-sheet. These lattices clearly do a very good job of matching the local structural elements found in the secondary structure of real proteins.

From the central α-carbon atom vertex, the side-chain is formed from four lattice points. Three are face-centered cubic-type lattice vectors; that is, vectors of the type $(\pm 1, \pm 1, 0)$. The fourth vector is a diamond lattice vector, $(\pm 1, \pm 1, \pm 1)$; this vector is the center of the hydrophobic or hydrophilic interaction. These four vectors comprise a volume slightly larger than a methyl group. A schematic representation of a side-chain is shown in Figure 1(c). The orientation of the side-chain depends on the backbone conformation, i.e. r_θ^2. At least two of the three face-centered cubic vectors comprising the side-chain are in the L-conformation, and the diamond lattice type vector is always in the L-conformation. For the calculations described below, either the residues are glycine, in which case there is no side-chain, or the residues have a side-chain of uniform size. Generalization to varying

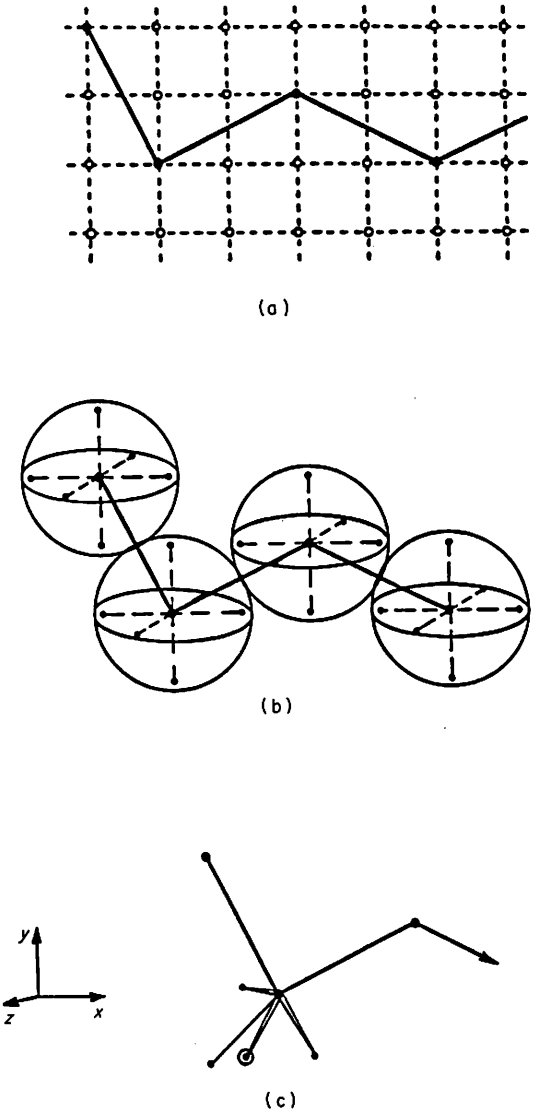

Figure 1. (a) Schematic representation of the underlying cubic lattice and 210-type vectors connecting consecutive α-carbon atoms of the protein backbone. (b) Schematic representation of main-chain backbone showing the hard core envelope. (c) Schematic representation of protein backbone plus a side-chain.

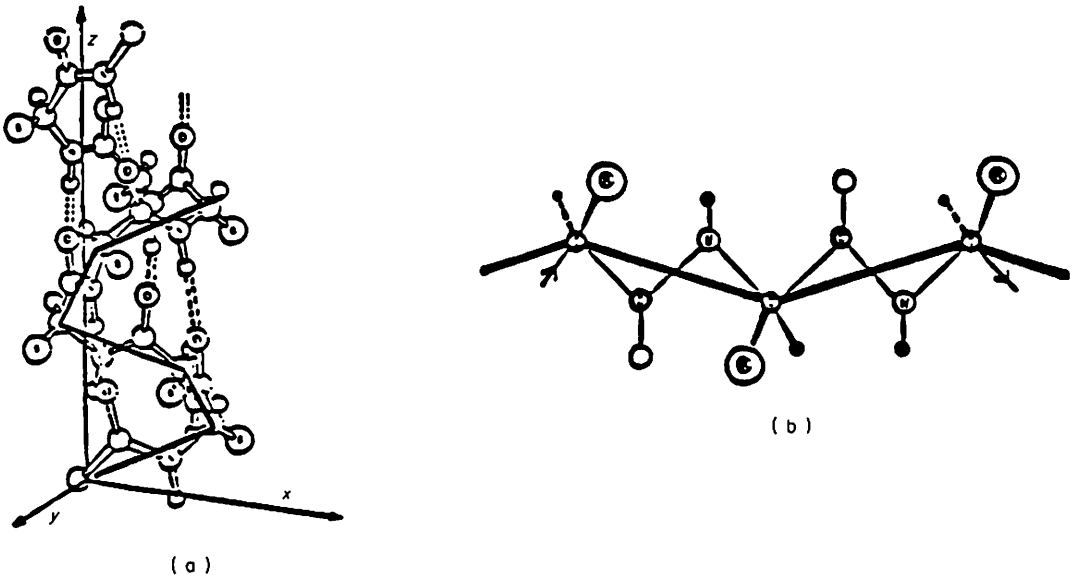

Figure 2. (a) A state-12-type helix, superimposed onto a real α-helix. (b) A state-16-type β-strand superimposed onto a real planar β-strand.

side-chain size is straightforward and will be undertaken in future work.

(b) *Interactions*

For a chain composed of n residues in an α-carbon representation, the conformation is described by $n-2$ bond angles, θ, and $n-3$ torsional angles, ϕ. In order that the first and last residues have a defined conformation, we append to each end of the chain a virtual residue. These virtual residues, which are devoid of side-chains, are completely inert, except that they occupy space. Thus, with the addition of the two fictitious end residues, there are n bond angles and $n-1$ torsional angles that specify the backbone conformation of the chain. For convenience, the chain is numbered from 1 to N.

The intrinsic conformational preference in this class of models is represented by the individual residue preference for the various bond angle states. These are conveniently described in the lattice model by r_θ^2. Since for every residue i, there are seven distinct values of r_θ^2, corresponding to 18 distinct local conformational states, the local energetic preference for each of the r_θ^2, $\varepsilon_\theta(r_{\theta i}^2)$ must be specified. In these early model calculations, to keep the number of adjustable parameters to a minimum, apart from a few (helix and/or sheet and/or turn) conformations where $\varepsilon_\theta(r_{\theta i}^2) = 0$, all of the other conformations are made isoenergetic and assigned an $\varepsilon_\theta > 0$.

We next turn to the torsional potential associated with the ith residue that involves residues $i-1$ to $i+2$. Thus, a dihedral angle potential must be specified for residues 2 (real residue 1) to $N-2$ (real residue, $n-1$). There are 324 rotational states that exist for each internal bond; these must all be assigned a relative energy ε_ϕ. In principle, the statistical weights of all such rotational states must be specified. This will be done in subsequent application to real proteins. Here the majority of the conformations will be taken to be isoenergetic, sometimes with a small bias towards a small subset of these conformations that are native, and at other times towards a manifold of conformers, some of which are decidely non-native. For example, helices can be favored in native β-sheet regions. Thus, the short and intermediate range preferences are accounted for by $\varepsilon_\theta(r_\theta^2)$ and ε_ϕ, respectively.

These models also have a finite backbone thickness. In addition to the hard core repulsion there is a soft core repulsion between non-bonded α-carbon atom backbone centers that extends up to a distance of $5^{1/2}$. If r_{kl} is the distance between the kth and lth such centers, then the soft core repulsive energy between the pair, E_{rep} is of the form:

$$E_{rep} = \begin{bmatrix} \infty; & r_{kl}^2 = 0, 1 \\ 3\varepsilon_{rep}; & r_{kl}^2 = 3 \\ \varepsilon_{rep}; & r_{kl}^2 = 5 \\ 0; & \text{otherwise} \end{bmatrix}. \quad (1)$$

ε_{rep} will typically take on the value of 6 in the calculations that follow.

We now specify the tertiary interactions. To mimic the effect of hydrogen bonding and dipolar type interactions, we have introduced a co-operative interaction that, at present, allows for secondary structure stabilization when any part of the α-carbon, hard core envelope of the mth residue is at a distance of 3 units from the α-carbon center of the kth residue. Then,

$$E_{ckl} = \varepsilon_c\{\delta_{|b_k|:|b_m|} + \delta_{|b_{k+1}|:|b_m|} + \delta_{|b_k|:|b_{m+1}|} + \delta_{|b_{k+1}|:|b_{m+1}|}\} \quad (2)$$

δ_{ij} is the Kronecker delta. Here, ε_c is applied uniformly to all residue pairs independent of their conformation.

Interactions between any pair of side-chains are allowed if the interacting side-chain sites lie at a distance of $\sqrt{2}$ from each other; there are 12 such sites on the lattice. While a full hydrophobicity scale (Miyazawa & Jernigan, 1985) can be (and in fact already has been (Skolnick & Kolinski, 1990*b*)) implemented in the initial stages of the calculations described here, we opt for a simplified hydrophobicity index. Side-chains can be hydrophobic, hydrophilic, or inert. Pairs of hydrophobic side-chains interact with an attractive potential of mean force (Hill, 1956), hydrophobic/hydrophilic pairs interact with a repulsive potential of mean force, and hydrophilic pairs may be weakly attractive or repulsive with no change in qualitative behavior.

Glycine residues lack side-chains and are assigned a hydrophobic index, $h(i)$, of zero. Hydrophobes are assigned a negative hydrophobicity index, $(h(i) < 0)$, and hydrophilic residues are assigned a positive hydrophobicity index $(h(i) > 0)$. For all side-chains that are greater than two residues apart down the chain, the side-chain–side-chain interaction matrix element between the ith and jth pair, $am(i, j)$, is of the form:

$$am(i, j) = -h(i) \cdot h(j) \cdot \varepsilon. \qquad (3)$$

Here $\varepsilon = \varepsilon_{\text{phobe-phobe}}$ (>0), if $h(i)$ and $h(j)$ are both negative (that is, both are hydrophobic), $\varepsilon = \varepsilon_{\text{phobe-phil}}$ (>0) if one residue is hydrophobic and the other is hydrophilic, and $\varepsilon = -\varepsilon_{\text{phil-phil}}$ with $\varepsilon_{\text{phil-phil}} > 0$ if both $h(i)$ and $h(j)$ are positive; that is, both side-chains are hydrophilic.

At this juncture, it is important to emphasize that both native and non-native interactions are allowed between any non-bonded pair of residues that are spatially close enough to interact. The algorithms described below have no long-range target potential that drives the system towards the native conformation. Indeed, in a number of cases considered below, based on local interactions, the native conformation is one of a large number of isoenergetic states that may include both native and non-native preferences. For example, in the case of β-barrels, we also consider the case where helical states are locally favored, but where the primary sequence has an alternating odd/even hydrophobic/hydrophilic pattern consistent with β-barrel formation. It is the juxtaposition of short, medium and long-range interactions that produce the final, stable, folded conformation.

In the simulation itself, all of the energetic parameters, ε_θ, ε_ϕ, ε_{rep}, $\varepsilon_{\text{phobe-phobe}}$, $\varepsilon_{\text{phil-phob}}$ and $\varepsilon_{\text{phil-phil}}$ are uniformly scaled by a reduced temperature factor, T^*. Thus, we make the simplifying approximation that the relative ratios of the various interactions are temperature independent. Strictly speaking, this is incorrect. This is not an inherent limitation of the model; once these parameters become known, then realistic values can be employed. Moreover, we make no attempt to divide up these potentials of mean force into their enthalpic and entropic contributions. Rather, we fix their relative ratio and examine the consequences on their ability to fold to native conformations. We should point out that, even if we take the free energy of the native state as being entirely of enthalpic origin (an incorrect assumption) at the transition midpoint where the free energy of the folded and unfolded states are equal, the native conformation of a four-helix bundle is stabilized by about $2{\cdot}2k_BT$ per residue. This value is comparable in magnitude to that in metmyoglobin of $1{\cdot}6k_BT$ (Privalov & Kechinashvili, 1974). In these models, about one-third to one-half of the native state free energy is due to inherent secondary structural preferences. We also mention that the present realization of the model probably overestimates the configurational entropy of the unfolded state, as a myriad of non-physical conformations (e.g. left-handed helices) are not penalized as severely as they would be in nature. This will have the effect of requiring larger values of the energy to give the same native population as in the real system. However, the choice of the parameters is not arbitrary. For secondary structural elements in the unfolded state, populations of helices and/or β-sheets cannot exceed a small percentage. Turn populations cannot exceed the values observed in turn peptides, where for β-proteins the native population is of the order of 1 % and in helical proteins it can be substantially larger (J. P. Waltho, H. S. Cheol & P. E. Wright, unpublished results; H. J. Dyson, J. R. Sayre, R. A. Lerner & P. E. Wright, unpublished results).

Consistent with experiment, in the denatured state of these model proteins there are hydrophobic clusters of marginal stability (Bundi *et al.*, 1976, 1978), nascent helices for helical proteins (Shoemaker *et al.*, 1985, 1987; Dyson *et al.*, 1988*a*,*b*) and reverse turns (Dyson *et al.*, 1988*a*). The dimensions of the unfolded state are substantially smaller (about 2/3 the volume) than a statistical random coil exhibiting excluded volume. These systems have a mean-square radius of gyration about three times that of the folded state; thus, they may be somewhat more expanded than thermally denatured real proteins (Privalov *et al.*, 1986). We have erred on the side of more expanded states. Presumably if folding can occur from a more expanded space, then it could occur when the configurations of the denatured state are more severely restricted.

The following convention is employed to specify the sequence. $B_i(k)$ is the ith stretch in the primary sequence composed of k residues. The k residues have identical ε_θ and ε_ϕ; these residues have a marginal short and intermediate range preference for β-states. We emphasize, here, that this is only a marginal preference and that, in the transition region without tertiary interactions, the probability of being in the fully native conformation is small. On a per residue basis, as in real proteins, the native population in the denatured state is of the order of several per cent.

Another possibility occurs when a sequence of k residues is locally indifferent as to whether they

are in an α-helix or in a β-sheet. Let $AB_i(k)$ denote the ith stretch in the amino acid sequence containing k residues such that $\varepsilon_\theta(12) = \varepsilon_\theta(16)$ for all k residues. Finally, there is the possibility that a sequence of k residues locally prefers α-helical states, such that $\varepsilon_\theta(12) = 0$ and $\varepsilon_\theta(16) > 0$. This is denoted by the short-hand notation, $A_i(k)$.

Putative bend regions are denoted by $b_i(j)$ and consist of j residues located at the interface between putative β-stretches i and $i+1$. Again, we emphasize that, unless otherwise indicated, this is just a marginal turn preference: for example, for the β-proteins, typical turn populations in the transition region in the denatured state are 0·3 to 1·0%. This is substantially larger than the purely random value of 18^{-j}; still, turns are negligibly populated in the unfolded state. On the other hand, solution nuclear magnetic resonance studies indicate that this population is non-random as well (H. J. Dyson, J. R. Sayre, R. A. Lerner & P. E. Wright, unpublished results; J. P. Waltho, H. S. Cheol & P. E. Wright, unpublished results).

(c) *Dynamic Monte Carlo algorithm*

The dynamics of the chain is simulated by a (pseudo) random sequence of the following kinds of conformational rearrangements. In all cases, the bead on which the move is to be performed is chosen at random and cannot violate excluded volume or bond constraints.

(1) Single bead flips or "spike" motions (Downey & Kovac, 1987), which are shown in Figure 3.

(2) Two bead end flips, in which the two end bonds are transformed to a new set of vectors.

Elsewhere, we have demonstrated that moves (1) and (2) satisfy the correct dynamics for the athermal random coil state in the absence of hydrodynamic interactions (Baumgartner, 1984), that is, they obey Rouse-like dynamics with excluded volume (Kolinski *et al.*, 1991). Moreover, they even give the correct character of local conformational relaxation as probed by orientational correlation functions.

While Monte Carlo dynamics can be implemented with only the purely local moves of (1) and (2), these moves suffer from the disadvantage that they cannot move about assembled pieces of secondary structure, in particular α-helices. Thus, we include the possibility of rotations and translations of the chain.

(3) Chain rotations, such as are shown schematically in Figure 4(a), are performed on randomly chosen segments.

(4) Finally, there are internal wave-like motions, such as are shown in Figure 4(b). These moves serve to propagate defects down a subchain by deleting a defect at one end of the subchain and creating the defect at the other end of the subchain.

After each of the elemental steps (1) to (4), the energy of the new conformation, E_{new}, is calculated and compared to the energy of the old conformation, E_{old}, and a standard asymmetric Metropolis sampling scheme is employed.

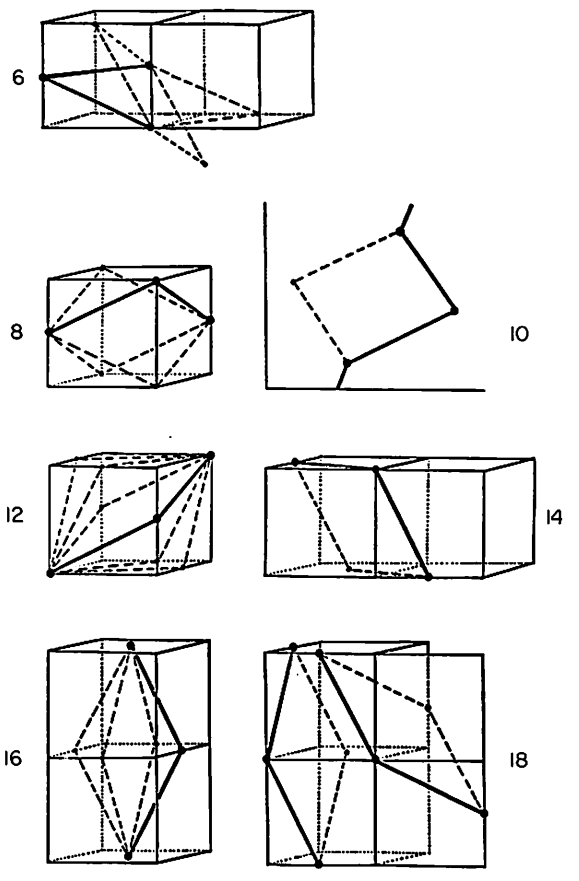

Figure 3. Local elemental single bead flips associated with the various bond-angle states corresponding to $r_\theta^2 =$ 6, 8, 10, 12, 14, 16 and 18, respectively. The continuous lines indicate the initial orientation of the pair of bonds, and the broken lines are the possible outcomes.

A single Monte Carlo dynamics time step consists of N attempts at (1), each of the chain ends are subjected to (2), and one attempt at (3) and (4). Thus, (3) and (4) are somewhat attenuated to avoid significant time scale distortion. However, they are in reality underpenalized. This is done to determine whether preformation of secondary structure dominates as the assembly mechanism. If it turns out that, even when these kinds of move are artificially enhanced, the mechanism does not contribute appreciably to assembly (and this is indeed the case), one can conclude that when they are even more heavily penalized they will certainly not contribute. Only when they are important does one have to worry about their correct weighting (see Sikorski & Skolnick, 1990). Their inclusion does not have any important effect on the dynamics in the unfolded state.

Typically, in the denatured state about 33%, 60%, 7% and 0·1% of attempted moves (1) to (4),

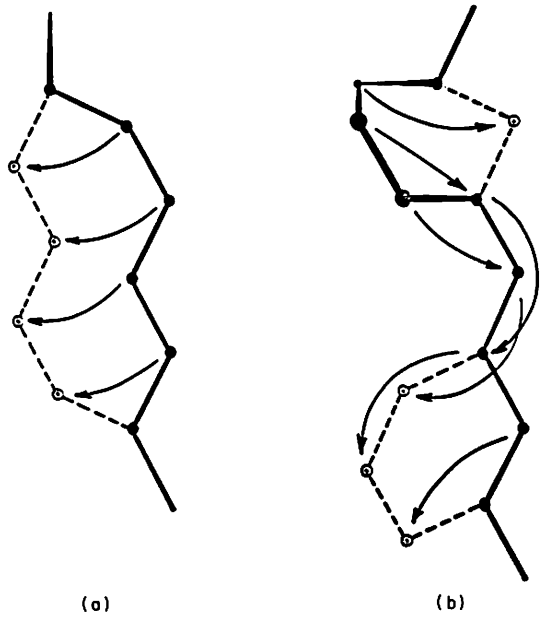

Figure 4. Schematic representation of internal rotations (a) and wavelike motions (b). The continuous line shows the initial chain conformation and the broken line the final chain conformation. The arrows show the direction of the local defect propagation.

respectively, succeed. The system is started out in a randomly generated, high T^* state. It is then cooled down, equilibrated, cooled further, etc., until collapse to a folded conformation occurs. For each run in the transition region, at least $1{\cdot}25 \times 10^6$ Monte Carlo time steps are sampled.

The set of elemental moves employed here satisfies a stochastic kinetics master equation describing the time evolution of the system (Binder, 1987). By construction, in the limit of a large number of steps, because the algorithm satisfies detailed balance, an equilibrium distribution of states will be generated. Whether or not the description of the dynamics reflects the actual folding process is another matter. Elsewhere, it has been demonstrated that these moves produce correct random coil dynamics (Kolinski *et al.*, 1991), and the hope is that they can describe the overall folding process, as well. There is a long history of this class of models in polymer physics that shows that the dynamics are independent of the particular choice of local moves, provided that such moves span configurational space (Baumgartner, 1984). A detailed discussion about questions of ergodicity and the validity of this kind of approach may be found elsewhere (Binder, 1987; Kolinski *et al.*, 1987*a*; Skolnick & Kolinski, 1990*c*). More recent work, using off-lattice Brownian dynamics, indicates that the mechanism of helix hairpin assembly is identical with the 210 lattice simulations, thereby providing further support for the universality of the pathways (A. Rey & J. Skolnick, unpublished results).

3. Approximate Analytic Theory of Folding

In the following, we present an approximate analytic theory for the conformational transition of this class of models. The use of this approximate analytic theory allows for an estimation of the approximate thermal transition region before the simulation is undertaken. Thus, substantial computer time is saved by avoiding the well characterized (and uninteresting) high temperature denatured state as well as the possibility of drastically quenching the system by choosing simulation temperatures that are too low. As shown below, the theory in general works remarkably well and is one indication that the conformational transitions simulated below are well approximated by a two-state model. However, the explicit nature of the conformational transition can be (and is) determined directly from the folding trajectories.

By restricting the protein to a lattice, one merely has a rotational isomeric state model (Flory, 1969) of the protein. Such models have been extremely useful in describing the equilibrium conformational properties of polymers (Flory, 1969), and there is every reason to expect that this description should work for proteins as well. While more sophisticated models can be developed, here we make the simplest of assumptions that, as shown below, provide a semiquantitative description of the conformational transition of these model proteins. First, the conformational transition is treated in the context of a two-state model. Thus, the free energies of the denatured, A_D, and native, A_N, states are required. A_D is calculated by neglecting all long-range tertiary interactions in the denatured state; however, pentane-like effects are included. While there are certainly a few hydrophobic contacts in the unfolded state, since they are of the order of several k_BT in magnitude, their inclusion has a minor effect on the estimated transition temperature. In fact, the model can readily accommodate contacts between residues that are less than five residues apart; since we neglect contacts between first and second neighbors down the chain, their inclusion has a minor effect on results. For the calculation of A_N, we neglect the small local fluctuations about the native state and approximate A_N by the energy of the native state, E_N. A similar set of approximations, when applied to diamond lattice models, worked very well, and provided semiquantitive fits to the simulation results (Skolnick & Kolinski, 1990*a*; Sikorski & Skolnick, 1990). The treatment developed below can be generalized to calculate the free energy along a folding co-ordinate (Skolnick & Kolinski, 1990*a*; Sikorski & Skolnick, 1990).

In the context of a two-state model for folding, the fraction of molecules in the native state, f_N, is given by:

$$f_N = \frac{\exp\{-(E_N - A_D)\}}{[1+\exp\{-(E_N - A_D)\}]}. \qquad (4)$$

An explicit expression for A_D may be found in Appendix A, equation (17).

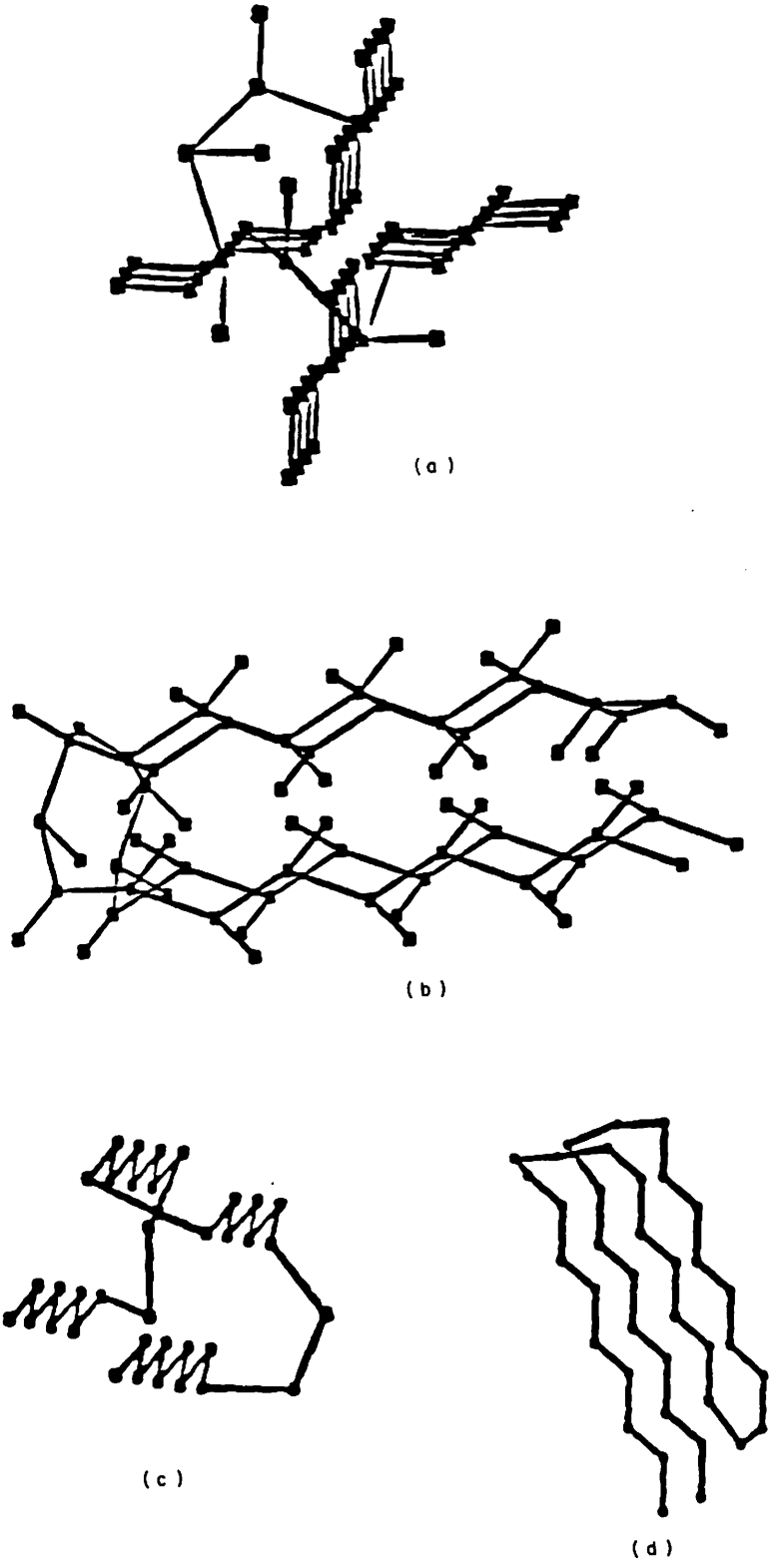

Figure 5. Native conformation of the 4-member β-barrel including the central α-carbon atoms and interacting side-chain sites in (a) a side and (b) a top view. The α-carbon backbone is shown in (c) a side and (d) a top view.

Similarly, the mean-square radius of gyration $\langle S^2 \rangle$ defined as:

$$\langle S^2 \rangle = \frac{\sum_{i=1}^{N} (r_i - r_{cm})^2}{N}, \quad (5)$$

with $|r_i - r_{cm}|$ the distance of the ith bead from the center of mass, r_{cm}, may be obtained from:

$$\langle S^2 \rangle = f_N \langle S_N^2 \rangle + (1 - f_N) \langle S_D^2 \rangle, \quad (6)$$

where $\langle S_N^2 \rangle$ and $\langle S_D^2 \rangle$ are the mean-square radii of gyration in the native and denatured state, respect-

Table 1
Folding results for β-barrel proteins

Sequence†	No. of folding attempts	No. of successful folds	Intrinsic turn probability‡
$\varepsilon_\alpha > \varepsilon_\beta$; 1; 1·75	5	5	0·0046
$\varepsilon_\alpha > \varepsilon_\beta$; 1; 1·5	6	4	0·0021
$\varepsilon_\alpha = \varepsilon_\beta$; 1; 1·5	6	6	0·0025
$\varepsilon_\alpha = \varepsilon_\beta$; 0·5; 1·5	7	5	0·0093
$\varepsilon_\alpha = \varepsilon_\beta$; 0·25; 1·5	7	6§	0·0023
$\varepsilon_\alpha = \varepsilon_\beta$; 0; 1·5	11	5	0·063
$\varepsilon_\alpha = \varepsilon_\beta$; 0; 1·75	10	10	0·14
$\varepsilon_\alpha < \varepsilon_\beta$; 0; 1·6; 0·05	11	11	0·036
$\varepsilon_\alpha = \varepsilon_\beta$; 1; 0	14	0	5×10^{-5}

† See the text for a more detailed discussion of the primary sequence.
‡ Calculated from the analytic theory near the calculated transition midpoint; eqn (7).
§ Five out of the 6, were fully native; 1 had a non-native central turn.

ively, both of which are taken from the simulation. Observe that $\langle S_D^2 \rangle$ is typically 75% of that of a random coil polymer with excluded volume, but no attractive interactions. These quantities are taken from the simulation.

4. Results

(a) *Four-member β-barrels*

In the following, we examine the requirements for folding to the unique four-member β-barrel state. Side (a) and top (b) views of the native conformation, including both α-carbon atoms and the interacting side-chain sites are shown in Figure 5. The top view (b) shows the hydrophobic core with the side-chains nicely interdigitating, just as in the case of real proteins (Richardson, 1981); Figure 5(c) and (d) show the corresponding conformation with the α-carbon atoms alone. In the fully native conformation, there are 20 pairs of contacts between neighboring side-chains.

In all the cases considered here, the pattern of hydrophobic and hydrophilic residues is the same. The model chain consists of $N = 37$ residues. In each β-strand, all of the even (odd) residues are hydrophobic (hydrophilic). Strand 1 consists of residues 1 to 8. The turn residues 9 to 11 are all hydrophilic. Strand 2 runs from residues 12 to 18. The turn residues 19 and 20 are, respectively, hydrophilic and hydrophobic. Strand 3 runs from residues 21 through 26. Turn residues 27 to 29 are hydrophilic. Strand 4 runs from 30 to 37.

(i) *Equilibrium folding simulations*

We begin with the examination of model proteins having the primary sequence $B_1(7)b_1(4)B_2(6)b_2(3)B_3(5)b_3(4)B_4(8)$. In the first case considered, for B_i, the following parameters are used: $\varepsilon_\theta(16) = 0$, and all other $\varepsilon_\theta = 0{\cdot}25/T^*$. For 16 states, $\varepsilon_\phi = -0{\cdot}6/T^*$ and is zero for all other states. For the turns, b_i, $\varepsilon_\theta = 0$ for the native conformation, and $\varepsilon_\theta = 0{\cdot}25/T$ for all other conformations. Similarly, $\varepsilon_\phi = -0{\cdot}6(1{\cdot}75)/T^* = -1{\cdot}05/T^*$, $\varepsilon_{\text{phil-phil}} = 0{\cdot}25/T^*$, $\varepsilon_{\text{phil-phob}} = 1/T^*$ and $\varepsilon_{\text{phob-phob}} = -0{\cdot}75/T^*$. The co-operativity parameter $\varepsilon_c = -0{\cdot}15/T^*$. Thus, in the native conformation, the total short range free energy $E_\theta = 0$, the total torsional energy $E_{\text{tor}} = -25{\cdot}8/T^*$, the total side-chain interaction free energy arising from hydrophobic interactions $E_{\text{side}} = -14{\cdot}25/T^*$ and the co-operative interaction free energy $E_c = -11{\cdot}25/T^*$. Thus, the total energy of the native state $E_N = -51{\cdot}3/T^*$, and secondary structural preferences contribute about 50% of the total energy of stabilization of the native state. A summary of the conformational properties of this, as well as all the other types of primary sequences examined, is presented in Table 1. This primary sequence will be designated by a short-hand notation $\varepsilon_\alpha > \varepsilon_\beta$; 1, 1·75. This notation indicates that based on bond angle preferences, β-conformations are locally preferred for the B_i portions of the primary sequence, and that the torsional angle preference, ε_ϕ, for native-like conformations in the B_i region is locally favored by a ratio of 1 : 1·75 over that in the turn regions.

A major point that must be addressed is the intrinsic preference for the native conformation in the absence of long range interactions. Is the native conformation so favored without tertiary interactions that the system has no choice but to find it? If so, then nothing really is learned by performing the simulation. The system has to be "cooled" to very low temperatures to favor the native state. (Actually in this case, native-like fragments become kinetically locked and the native conformation is not recovered.) To address this point, the approximate analytic theory is quite helpful. Using equation (5), we find that the transition midpoint, including tertiary interactions, is predicted to be near $T^* = 0{\cdot}576$. At this temperature, employing equation (17) (see Appendix), $A_D = -88{\cdot}44$. E_N without tertiary interactions equals $-44{\cdot}79$. The fraction of molecules in the native conformation that would be present if all tertiary interactions are turned off (that is, the equilibrium

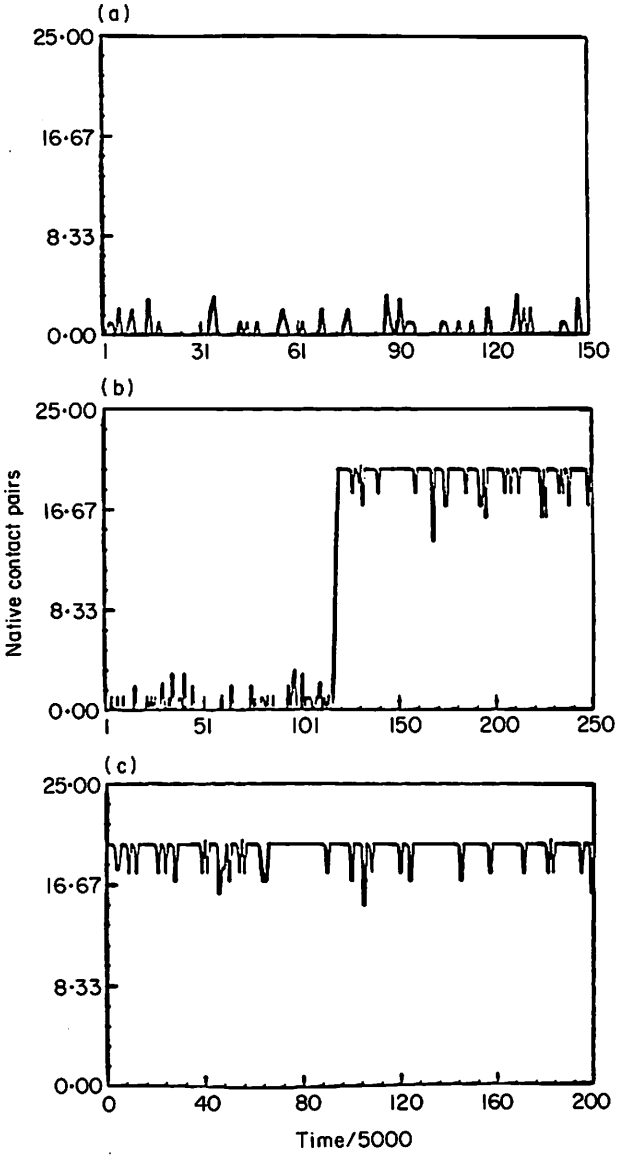

Figure 6. Plot of the number of native contact pairs, N_c, *versus* time under (a) denaturing conditions at $T^* = 0{\cdot}6$, (b) in the transition region at $T^* = 0{\cdot}58$, and (c) under strongly native conditions at $T^* = 0{\cdot}545$.

population based on short and medium range interactions embodied in E_θ and E_{tor} alone) is given by:

$$f_N^0 = \exp(-E_{tor})/\exp(-A_D). \quad (7)$$

For the particular case at hand, using equation (7), we have $f_N^0 = 1{\cdot}11 \times 10^{-19}$, and the partition function of the denatured state is $2{\cdot}46 \times 10^{38}$. These numbers are typical for the cases described below. Thus, there is a negligible preference for the native state in the absence of long range interactions, and furthermore, by recognizing that these systems undergo a highly co-operative transition, it is safe to conclude that finding the native conformation is by no means guaranteed by the above choice of short and medium range interaction parameters. Rather, this system will have to thrash about until it finds the native state, a situation quite close to the real protein case. Similarly, let us calculate the intrinsic probability of a native-like turn, and for definiteness, we focus on the first turn. If this turn is excised from the chain, at the transition midpoint, the native turn population is calculated to be about about 1%.

Starting from randomly generated high temperature denatured states, in a series of five cooling runs, the native conformation has always been obtained for the system having the primary sequence described above; i.e. $\varepsilon_\alpha > \varepsilon_\beta$; 1; 1·75. Folding to the native state occurs directly without the formation of out-of-register intermediates. Typically, for a given set of parameters, about five to ten renaturation runs are performed. This is generally adequate to establish that folding to the same native conformation has occurred, but the scatter in the transition region, at this level of sampling, is quite large.

The next question that must be addressed is the nature of the conformational transition. In Figure 6(a) to (c), the average number of native contact pairs between side-chains, N_c, *versus* time is plotted for a chain under denaturing conditions at $T^* = 0{\cdot}6$, in the thermal transition region at $T^* = 0{\cdot}58$ and under strongly renaturing conditions at $T^* = 0{\cdot}545$. The times indicated in the Figure are in units of 5000 Monte Carlo steps. Under denaturing conditions, N_c fluctuates around zero, characteristic of a relatively short, unfolded chain. In the transition region, the system starts out unfolded, and then around $t/5000 = 118$, it undergoes a rapid transition in about 6500 Monte Carlo time steps to the fully native molecule. For the remainder of this run, it stays in the native state. All other conformational properties also undergo sharp changes in value as is characteristic of an all-or-none transition. This is not to say that there are no intermediates between the denatured state and the fully folded molecule, rather that such intermediates are marginally populated. On further cooling to $T^* = 0{\cdot}545$, the system is fully native, with minor oscillations in N_c arising from the fluctuations of the ends.

While the population of native-like turns is rather small (about 1%) in the denatured state, it is definitely non-random (the random probability is about 10^{-5}), and it is important to examine what happens as the propensity for native-like turns is diminished. Thus, we undertook a series of simulations that had a primary sequence identical with that of the first case, but where the torsional potential for b_i, $\varepsilon_\phi = -0{\cdot}6(1{\cdot}5)/T^*$; that is, a primary sequence of the type $\varepsilon_\alpha > \varepsilon_\beta$; 1; 1·5, is assumed. Not surprisingly, decreasing the turn propensity for native-like states has decreased the stability of the native conformation and decreased the transition temperature. Out of a total of six conformational transitions, four folded directly to the native conformation, and two produced out-of-register states involving an external strand that then dissolved and refolded. This analysis points out that in the absence of some intrinsic preferences for secondary structure, a

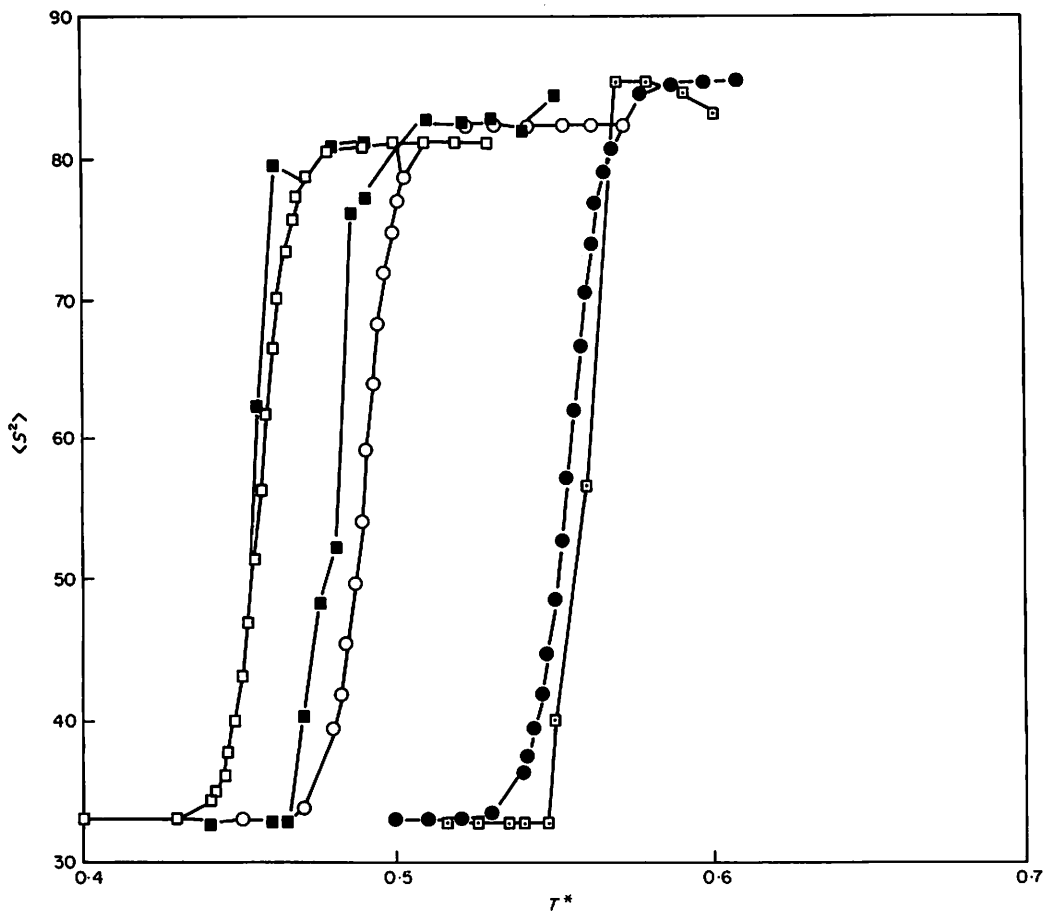

Figure 7. The curves going from right to left denoted by the filled squares, and filled circles present $\langle S^2 \rangle$ *versus* T^* obtained from the simulation for models having primary sequences $\varepsilon_\alpha = \varepsilon_\beta$; 1; 1·5; $\varepsilon_\alpha = \varepsilon_\beta$; 0·5; 1·5; and $\varepsilon_\alpha = \varepsilon_\beta$; 0·25; 1·5; respectively. The open squares, open circles and open dotted squares are the curves generated from the approximate analytic theory using eqns (4) and (6), for the corresponding primary sequences.

manifold of in and out-of-register conformations can be generated, and it is the marginal intrinsic turn propensities that are acting to select out one conformer from the others as the unique folded form.

The local preferences for β-states are further relaxed in primary sequences $AB_1(7)b_1(4)AB_2(6)b_2(3)AB_3(5)b_3(4)AB_4(8)$. For the AB_i stretches the following parameters are used: $\varepsilon_\theta(16) = \varepsilon_\theta(12) = 0$, and all other $\varepsilon_\theta = 0{\cdot}25/T^*$. Thus, based on short-range preferences alone, α and β conformers are isoenergetic. Three cases of non-zero ε_ϕ have been examined in which $\varepsilon_\phi(16, 16, 37) = -0{\cdot}6\gamma/T^*$, with $\gamma = 1$, 0·5 and 0·25 for cases 1 to 3, respectively. All of the other conformational states are isoenergetic. For the turns in b_i, $\varepsilon_\theta = 0$ for the native conformation and $\varepsilon_\theta = 0{\cdot}25/T$ for all other conformations. Similarly, $\varepsilon_\phi = -0{\cdot}6(1{\cdot}5)/T^*$, $\varepsilon_{\text{phil-phil}} = 0{\cdot}25/T^*$, $\varepsilon_{\text{phil-phob}} = 1/T^*$ and $\varepsilon_{\text{probe-probe}} = -0{\cdot}75/T^*$. The co-operativity parameter $\varepsilon_c = -0{\cdot}15/T^*$. The primary sequences in this series of simulations are identified by the short-hand notation $\varepsilon_\alpha = \varepsilon_\beta$; γ; 1·5. The γ and 1·5 refer to the scale factors of ε_ϕ in the AB_i and the turn portions of the sequence, respectively.

In Figure 7, in the curves denoted by the filled squares and circles, $\langle S^2 \rangle$ *versus* T^* obtained from the simulation for models having primary sequences $\varepsilon_\alpha = \varepsilon_\beta$; 1; 1·5, $\varepsilon_\alpha = \varepsilon_\beta$; 0·5; 1·5 and $\varepsilon_\alpha = \varepsilon_\beta$; 0·25; 1·5, respectively are plotted. The open squares, circles and dotted squares are the curves generated from the approximate analytic theory using equations (4) and (6). Quite good agreement between the analytic theory and the simulation is evident. As expected, decreasing the local conformational preference destabilizes the native conformation, with $E_{\text{tor}} = -24/T^*$, $-17{\cdot}4/T^*$ and $-14{\cdot}1/T^*$ for the above three primary sequences, and thus, the transition shifts to lower temperature. In the $\gamma = 0{\cdot}25$ case, intermediate range preferences for native-like conformations contribute about 36% of the total native state energy.

For the primary sequence $\varepsilon_\alpha = \varepsilon_\beta$; 1; 1·5, in each of six renaturation runs the in-register native conformation has been obtained in an all-or-none transition starting from a randomly generated denatured state. Interestingly, for the sequence $\varepsilon_\alpha > \varepsilon_\beta$; 1; 1·5, out-of-register conformations were observed prior to refolding to the native conformation. Note that the energy of the native conformation is identical in

these two cases; they will differ only in the energy of helical states in non-native conformations. The probable explanation for the avoidance of the out-of-register states is a kinetic effect. Because helical and β-conformations are isoenergetic, more than one conformation is locally preferred, and this makes it possible for the system to correct mistakes without getting trapped in a deep local free energy minimum. These two cases have an essentially identical thermal transition region. The analytic theory indicates that the native conformation of the $\varepsilon_\alpha > \varepsilon_\beta$; 1·0; 1·5 sequence is slightly more stable (this is because the free energy of the denatured state is slightly smaller, since α and β states are unequally weighted).

For the primary sequence $\varepsilon_\alpha = \varepsilon_\beta$; 1; 1·5, the population of turn 1 in the denatured state at the transition temperature is estimated from the analytic theory to be 0·25%. Once again, it is not random, but most certainly it is not enforced. In fact, it is in the range observed experimentally (H. J. Dyson, J. R. Sayre, J. R. Lerner & P. E. Wright, unpublished results), in the β-protein plastocyanin. If the local turn preference is decreased, out-of-register states become important, an expected result. We also considered the primary sequences having $\varepsilon_\alpha = \varepsilon_\beta$; 0·5; 1·5 and $\varepsilon_\alpha = \varepsilon_\beta$; 0·25; 1·5. The results are summarized in Table 1.

The next case considered has the identical primary sequence as in the previous cases, except that the torsional potential in the putative β-strands has been turned off. We examined molecules of the type $\varepsilon_\alpha = \varepsilon_\beta$; 0; 1·75. In a series of ten simulation runs, folding to the native conformation was always observed. Note that the local preferences as determined by ε_ϕ comprise 32% of the energy of the native conformation. Plots of $\langle S^2 \rangle$ *versus* T^* for the $\varepsilon_\alpha = \varepsilon_\beta$; 0; 1·5 and $\varepsilon_\alpha = \varepsilon_\beta$; 0, 1·75 sequences are qualitatively identical with those shown in Figure 7, and, since they provide no new insights, they are not displayed. As expected, the system having a larger intrinsic turn propensity denatures at a higher T^* value. The $\varepsilon_\alpha = \varepsilon_\beta$; 0; 1·75 sequence was observed to fold in the region from $T^* = 0{\cdot}40$ to 0·44. The intrinsic population of turn 1, calculated on the basis of equation (7), varies from about 35% to 10% over this temperature range. Thus, if the native turn propensity is sufficiently augmented, it appears that tertiary interactions plus intrinsic turn propensities are sufficient to yield the unique native state. This is verified by the fact that $\varepsilon_\alpha = \varepsilon_\beta$; 0; 1·5 sequences folded to in-register native state in six of 11 simulations. Note, however, that if the short-range interactions favoring β-strand formation are decreased, turn formation at a non-native location becomes more likely, and thus the intrinsic turn propensity must be appreciably augmented (see Table 1) to ensure the recovery of a unique conformational state.

Next, $A_1(7)b_1(4)A_2(6)b_2(3)A_3(5)b_3(4)A_4(8)$ type sequences were considered. The β-strands in the denatured-state locally favor α-helix, but the amino acid pattern still consists of alternating hydrophobic and hydrophilic residues. For the particular case examined, in all the A_i, $\varepsilon_\theta(12) = 0$, $\varepsilon_\theta(16) = 0{\cdot}05/T^*$, and all other $\varepsilon_\theta = 0{\cdot}25/T^*$. Furthermore, $\varepsilon_\phi = 0$ for all the residues in A_i. This particular series of simulations deals with molecules having the sequence $\varepsilon_\alpha < \varepsilon_\beta$; 0, 1·6; 0·05. In all 11 simulation runs, the four-member, β-barrel native conformation has been uniquely obtained. The transition is of an all-or-none character, and at the estimated transition midpoint of $T^* = 0{\cdot}37$, the native population of turn 1 in the denatured state is 0·037. These systems appear to be locally kinetically frustrated, in that the local preference is for α-helix, whereas the global free energy minimum conformation is the β-strand. At temperatures estimated from analytic theory to be under strong renaturating conditions, the simulation times prior to obtaining the folded conformation typically are two to six million time steps, as opposed to about a 0·5 million time steps for the other cases considered. These systems clearly spend substantial time trapped in relatively deep, local minima. As the local preference for helical conformations is increased in the putative β-strand-forming regions, while the unique four-member β-barrel is sometimes obtained, at other times the system thrashes about for over 50 million time steps without finding a unique folded form.

The importance of turns in producing the native conformation having been demonstrated, it is also of interest to examine what happens if β-states are preferred everywhere in the sequence. Thus, a series of simulations on a primary sequence $AB_1(37)$ was undertaken. In this case, the pattern of hydrophobes and hydrophilic residues is the same as in the above case, but *every* residue locally favors β-states by a factor $\varepsilon_\phi = -0{\cdot}6/T^*$, with $\varepsilon_\alpha = \varepsilon_\beta$. The short-hand notation for this sequence is $\varepsilon_\alpha = \varepsilon_\beta$; 1; 0. In a series of 14 independent renaturation turns, a manifold of β-barrel states was generated. The majority (11) of the low temperature collapsed states were native-like four-member barrels, with varying degrees of registration. Two involved a four-member β-barrel where the two outside β-hairpins are orthogonal to each other, and one state involved a three-member β-barrel, each of whose strands are approximately equal in length. Thus, intermediate range interactions can code for the kind of folded secondary structure that the native conformation can adopt (β in this case), and tertiary interactions can serve to dictate, to a great extent, the realization of the motif (4-member β-barrel); however, they are insufficient to produce a unique native state. Once again, when a marginal, local turn propensity is introduced, this serves to select-out the desired conformation, and thus, as shown above, one then obtains folding to a unique native-state.

(b) *Folding/unfolding pathways*

Following the suggestion of Wright *et al.* (1988), it is convenient to divide the folding process into early and late folding events. In all cases, folding initiates

at or near one of the β-turns. Out of 22 successful folding events, 18 initiated at the turn between β-strands 2 and 3, followed by rapid assembly of the β-hairpin. Three initiated at or near the β-turn between strands 1 and 2 or strands 3 and 4. One initiated at the turn between the β-hairpin at 1 to 2; the external strand then dissolved, followed by growth of the central hairpin. The preference for initiation at the turn between strands 2 and 3 originates from two complementary effects. From Figure 5 (see view (c) in particular), it is apparent that in the 2, 3 hairpin, the two adjacent β-strands are anti-parallel. The addition of each successive residue in the β-conformation is energetically favored by the co-operative interaction (E_c) between the β-strands, as well as by the fact that successive residues are next to each other. Thus, the backbone α-carbon atoms can nestle, thereby providing a kinetic enhancement of the folding process. As is shown in Figure 5(d), the distance between the residues in strands 1 and 2 alternates, with the hydrophobic residues being closer to one another (at a distance of $\sqrt{10}$) and the hydrophilic residues being further from each other (at a distance of $\sqrt{26}$). In the 1, 2 hairpin, after a pair of hydrophobic side-chains pack, the addition of a hydrophilic residue in the β-conformation costs free energy. The hydrophilic residue neither experiences tertiary interactions nor is there the nestling of consecutive α-carbon atoms that serves to accelerate the assembly process. Hence, while the external strands do serve as initiation sites, it is less likely. Finally, the central hairpin can be stabilized by the subsequent growth of either of the two external strands, whereas an external strand can be stabilized only by the growth of one strand to form the central hairpin.

Since the qualitative mechanism of assembly is the same for all the above cases, we pick a representative example having the primary sequence, $\varepsilon_\alpha > \varepsilon_\beta$; 1; 1·75. In Figure 8, a representative folding trajectory (identical with that in Fig. 6(b)) is shown. It is important to recognize that the actual assembly process is extremely fast. The system spends 1·6% of its time undergoing the conformational transition and the remainder of its time either in the unfolded or in the folded state. The conformations at different times are shown at different orientations that aid in the visualization of the folding process. At $t = 585{,}800$, folding is seen to initiate from the central turn between the β-hairpin composed of strands 2 and 3. The secondary and tertiary structures found in the native-state form almost simultaneously. As is evident from the pictures, the existing tertiary structure acts as scaffolding onto which further secondary structure is added. Moreover, folding is not unidirectional; β-strands dissolve as well as form during the course of assembly. Compare the $t = 585{,}900$ snapshot with the $t = 586{,}000$ snapshot where a slight dissolution of the β-hairpin is evident. By $t = 586{,}300$, the first β-hairpin is almost fully assembled. However, by $t = 586{,}550$, the majority of one of the two strands in the β-hairpin dissolves and then reforms at $t = 586{,}600$. Then, there is a pause as the random coil tail thrashes about until the next native-like turn forms. By $t = 587{,}700$, three of the four β-strands are essentially in place. Thus, assembly to the three-member β-barrel intermediate takes 1900 time steps from the beginning of folding. Now, the excluded volume of the chain hinders assembly. Most of the configurations of the denatured tail are non-productive, and the remaining tail thrashes about until $t = 591{,}800$, when the tail has worked its way into a position(s) that will permit native state assembly. After this, the assembly is very rapid, and by $t = 592{,}250$ the fully folded molecule is observed. Thus, the three-member β-barrel is the long lived intermediate, living for 4550 time steps or 71% of the total elapsed time from the start of folding.

The mechanism of assembly is best described as punctuated, on-site construction. A corresponding plot of the instantaneous value of the square radius of gyration *versus* time shows the rapid fluctuations characteristic of the random coil state for $t < 585{,}000$. This is then followed by the formation of a collapsed, but non-random, conformation that corresponds to the three-member β-barrel intermediate. Finally, the native conformation is seen. After assembly, as indicated by the plot of N_c *versus* time on Figure 6(b), there are minor fluctuations in the value of N_c about the native number of 20 contacts. Examination of the trajectory reveals that this involves local denaturation of the chain ends.

In all cases, unfolding is the reverse of folding. One of the external strands denatures, and the three-member β-barrel intermediate forms. Either the external strand denatures or the internal strand closest to the denatured tail unfolds. Finally, the β-hairpin is left, which then unfolds with a corresponding expansion in the chain dimensions.

Knowledge of the folding pathway also rationalizes the appearance of the out-of-register fourth external strand, when the intrinsic propensity to form a native-like turn is too small. In all cases, the native, three-member β-barrel intermediate forms. However, now there is a sticky hydrophobic interior, and if the chain happens to form a turn at a non-native location it may become kinetically trapped, thereby producing the out-of-register four-member β-barrel. Depending upon the temperature, the out-of-register strand may then dissolve and reform to give the in-register native conformation. Increasing the turn propensity for native-like turns serves to eliminate these undesirable out-of-register states. We hasten to point out that the intrinsic preferences required to avoid out-of-register states are quite small, typically requiring less than a 1% turn population in the denatured state.

(c) *Folding of four-helix bundles*

We now turn to the folding of left-handed, four-helix bundles with tight bends. The desired native conformation with α-carbon atoms and interacting side-chain sites is displayed in Figure 9, which

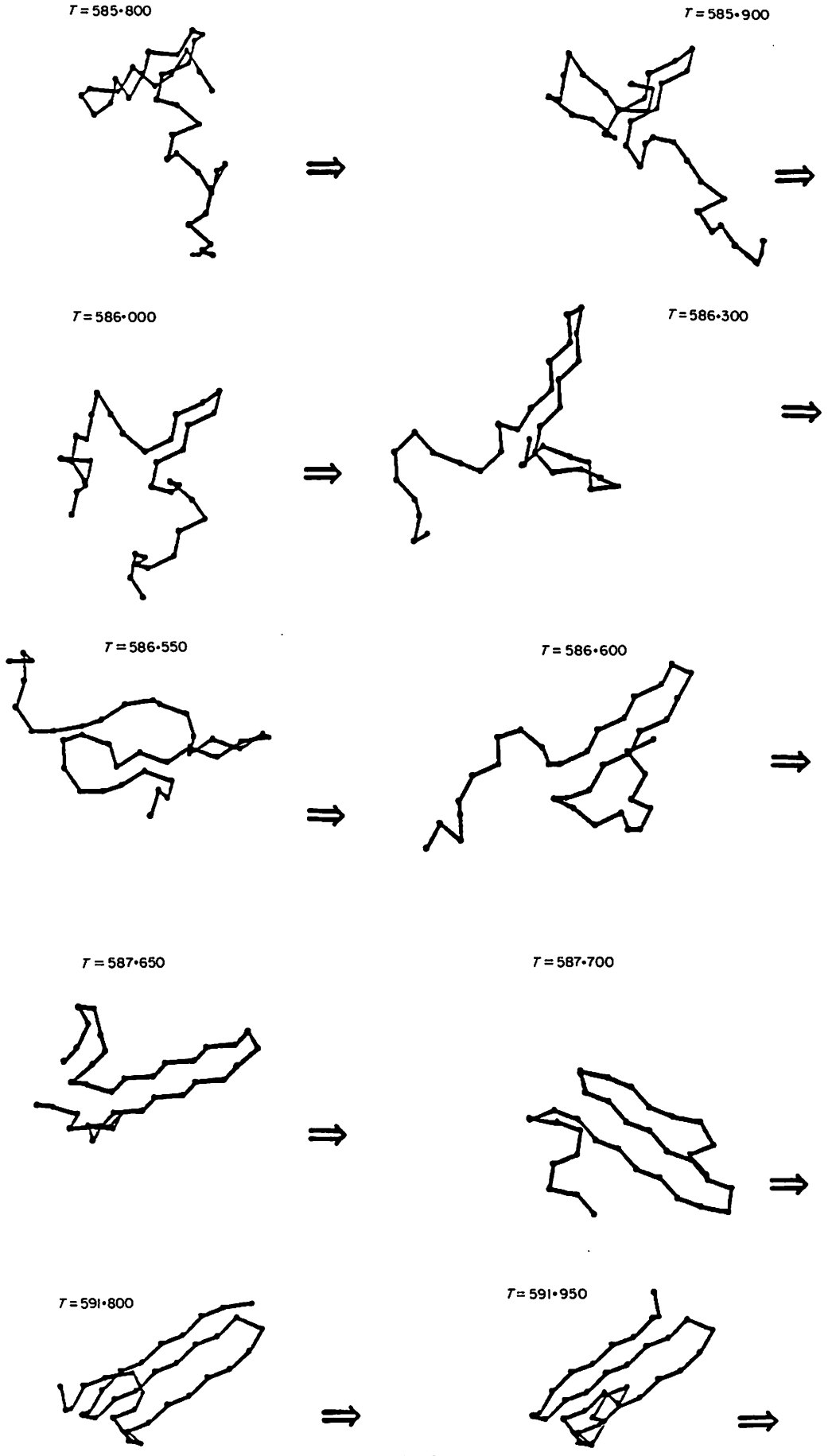
T = 585·800
T = 585·900
T = 586·000
T = 586·300
T = 586·550
T = 586·600
T = 587·650
T = 587·700
T = 591·800
T = 591·950

Fig. 8.

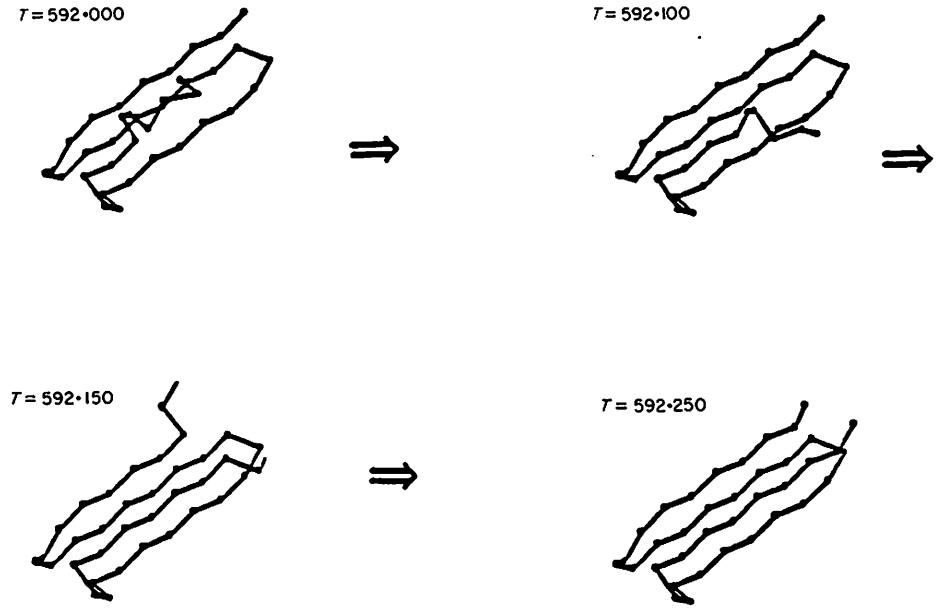

Figure 8. Representative folding trajectory of model protein having primary sequence for $\varepsilon_\alpha > \varepsilon_\beta$; 1·0; 1·75. The times indicated are from the start of the simulation run.

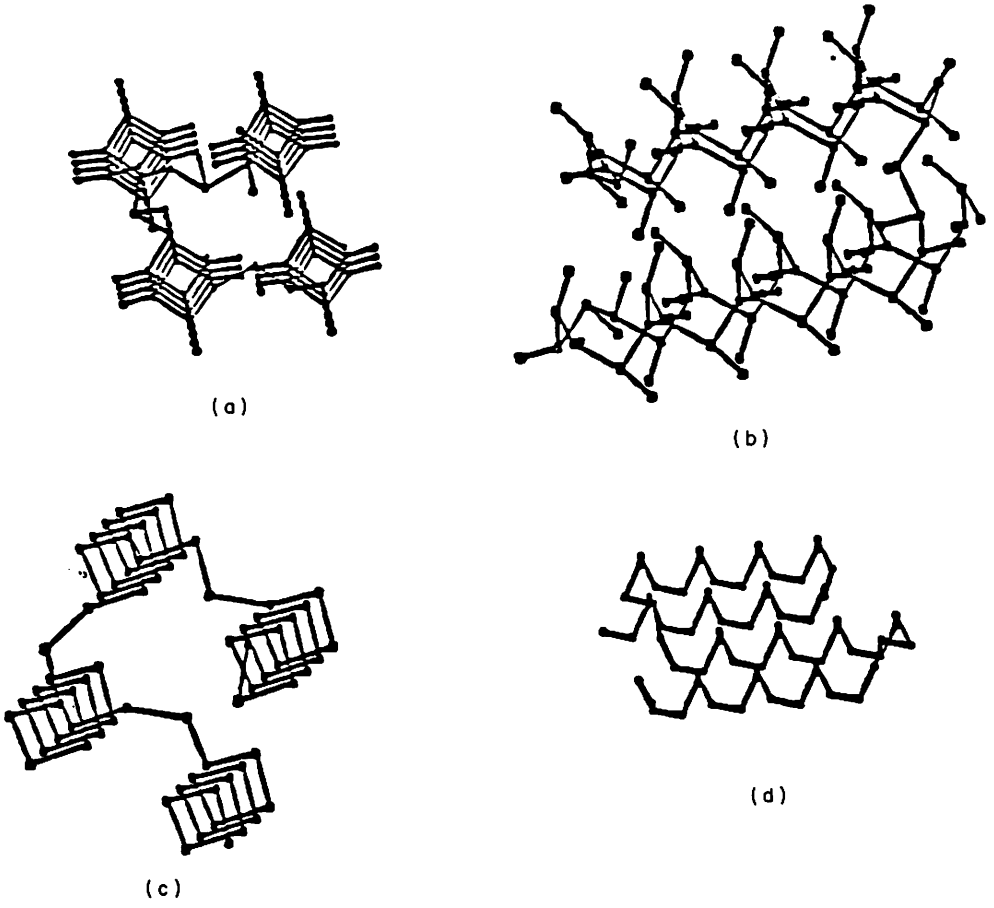

Figure 9. The native conformation of a left-handed, 4-helix bundle, whose α-carbon atoms and interacting side-chain sites are displayed in (a) a top and (b) a side view. Also displayed are the corresponding top (c) and side (d) views of the α-carbon backbone.

Table 2
Folding results for helical models

Sequence†	No. of folding runs	No. of successful folds‡	Intrinsic turn probability§
$\varepsilon_\alpha < \varepsilon_\beta$; 0·15; 1·5	21	18	0·067
$\varepsilon_\alpha < \varepsilon_\beta$; 0·15; 1·75	8	8	0·11
$\varepsilon_\alpha < \varepsilon_\beta$; 0·1; 2; 1·5	7	7	0·277‡
$\varepsilon_\alpha < \varepsilon_\beta$; 0·1; 1·5	9	7	0·86
$\varepsilon_\alpha = \varepsilon_\beta$; 0·15; 2	11	11	0·18
$\varepsilon_\alpha = \varepsilon_\beta$; 0·15; 1·5	6	6	0·053

† See the text for a more detailed discussion of the primary sequence.
‡ Central native turn population; the outer turn population is 0·078.
§ Calculated from the analytic theory near the transition midpoint; see eqn (7).

shows a top (a) and a side (b) view. Also displayed are the corresponding top (c) and side (d) views of the α-carbon backbone. Observe that, unlike the tetrahedral lattice case (Sikorski & Skolnick, 1989*a*), the four residue per turn, four-helix bundle can be oriented with the helices located at the points of a diamond rather than at the corners of a square. This particular arrangement of the helices produces a conformation that is closer to real four-helix bundles (DeGrado *et al.*, 1989). The alternative square geometry can also be accommodated on this lattice. The exclusion of this alternative conformation that produces non-native three-helix bundles that are in deep local free energy minima will necessitate a slightly higher turn propensity than was required for the folding of four-helix bundles on a tetrahedral lattice (Sikorski & Skolnick, 1989*a*). This point is amplified below.

In the idealized geometry displayed in Figure 9, all four helices are parallel, whereas in real four-helix bundles, the front and back pair of helices are tilted with respect to each other (Richardson, 1981; Chothia *et al.*, 1977; Richmond & Richards, 1978; Chou *et al.*, 1988). However, here, just as in real helix bundles, the side-chains nestle to produce a dense hydrophobic core. The model chain consists of $N = 69$ residues, with the end residues being virtual residues. The amphipathic sequence of residues, described when arranged in the four-helix bundle, produces a native conformation having 13 pairs of contacts between interacting hydrophobic side-chain sites.

(i) *Equilibrium results*

A concise compilation of the folding results described below is found in Table 2. We begin with an examination of a primary sequence of the type $A_1(13)b_1(4)A_2(13)b_2(4)A_3(15)b_3(4)A_4(16)$, in which for the A_i, $\varepsilon_\theta(12) = 0$ and all $\varepsilon_\theta = -0{\cdot}6/T^*$. For the non-native turn conformers in b_i, $\varepsilon_\theta = 0{\cdot}25/T^*$, except that to help to avoid the presence of non-native three-helix bundle intermediates (see Fig. 10), $\varepsilon_\theta(18) = 1/T^*$ for residues 16, 33 and 52. $\varepsilon_\phi = -0{\cdot}6(2)/T^*$ for the second and third residues in native-like conformations in the turn, and $\varepsilon_\phi = -0{\cdot}6(\gamma)/T$ for the first and fourth residues in native-like conformations in the turn. $\gamma = 1{\cdot}5$ and 1·75 in the two cases considered. Finally, for all residues, $\varepsilon_c = -0{\cdot}15/T^*$. The two primary sequences are designated by the short-hand notation $\varepsilon_\alpha < \varepsilon_\beta$; 0·15; 1·5 and $\varepsilon_\alpha < \varepsilon_\beta$; 0·15; 1·75 for $\gamma = 1{\cdot}5$ and 1·75, respectively. This particular class of primary sequences is chosen to mimic the case where all of the local preferences are consistent with the native conformation. For the $\gamma = 1{\cdot}5$ case in the transition region at $T^* = 0{\cdot}43$, the intrinsic probability of a turn being in a native conformation is about 6·7% on the basis of equation (7). This turn population is well within the range of physically reasonable values (J. P. Waltho, H. S. Cheol & P. E. Wright, unpublished results; H. J. Dyson, J. R. Sayre, R. A. Lerner & P. E. Wright, unpublished results).

As in the β-barrel case, the approximate analytic theory obtained by using equation (17) (see Appendix) provides a reasonable approximation to $\langle S^2 \rangle$ *versus* T^* with a difference in transition midpoint of about 5% for the $\gamma = 1{\cdot}5$ case between the simulation and the theory, and for the $\gamma = 1{\cdot}75$ case, this difference is about 3%. We have also examined the uniqueness of the conformational transition when $\varepsilon_\phi = -0{\cdot}6(1{\cdot}5)/T^*$ for all the turn residues and where $\varepsilon_\theta(18) = 0{\cdot}25/T^*$ for residues 16, 33 and 52. The native four-helix bundle was recovered in eight out of 11 independent renaturation runs. The other three runs produced the three-helix bundle of the type shown in Figure 10(a).

For the $\gamma = 1{\cdot}5$ case, a total of 21 independent renaturation runs were performed in which the native four-helix bundle was obtained 18 times. For the other three cases, a mixed three-helix bundle (in which 2 of the helices are in the square orientation, and the other pair is in the diamond orientation, with the remainder of the molecule unfolded) was obtained. This conformation is in a deep local free energy minimum. A representative of these non-native, collapsed states is shown in Figure 10(a), which has 14 contact pairs. Just as in the β-barrel case, there are a manifold of side-chain conformations having a similar number of favorable tertiary contacts consistent with a given element of secondary structure, and the short-range interactions serve to choose one out of the manifold of

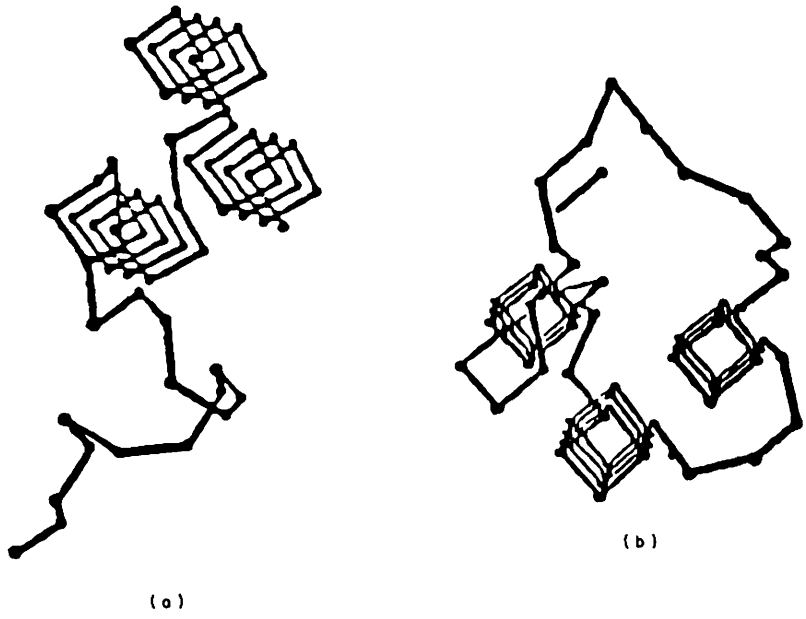

Figure 10. Representative 3-helix bundles obtained from folding when the native turn propensity is too low. (a) Outer hairpin in square geometry; (b) central hairpin in square geometry.

such conformations as native. In one of the three times that this three-helix bundle appeared, it subsequently unfolded, and, for the other two times, it persisted throughout the course of the simulation run.

To eliminate this off-pathway intermediate, we augmented slightly the turn propensity by setting $\gamma = 1{\cdot}75$. In eight simulation runs, the desired native conformation was uniquely obtained. No evidence of the undesirable three-helix intermediate was seen. A plot of the number of native contact pairs *versus* time reveals an all-or-none transition.

Then, primary sequences $AB_1(13)b_1(4)AB_2(13)b_2(4)AB_3(15)b_3(4)AB_4(16)$ that are locally indifferent to helix and sheet were examined. In the AB_i, $\varepsilon_\theta(12) = \varepsilon_\theta(16) = 0$ and all other $\varepsilon_\theta = 0{\cdot}25/T^*$, and in all the native-like conformers, $\varepsilon_\phi = -0{\cdot}6/T^*$. For the non-native turn conformers in b_i, $\varepsilon_\theta = 0{\cdot}25/T^*$, except that $\varepsilon_\theta(18) = 1/T^*$ for residues 16, 33 and 52. $\varepsilon_\phi = -0{\cdot}6(\gamma)/T^*$ for all the native-like conformers of the turn residues. In the first case examined, for all the turn residues $\gamma = 2$, and in the second case considered, for each of the putative turn residues $\gamma = 1{\cdot}5$, for all but the third turn residue where $\gamma = 2$. Finally, for all residues, $\varepsilon_c = -0{\cdot}15/T^*$. These two primary sequences will be described by the short-hand notation $\varepsilon_\alpha = \varepsilon_\beta$; 0·15; 2·0 and $\varepsilon_\alpha = \varepsilon_\beta$; 0·15; 1·5, respectively.

The most fully characterized case corresponds to $\varepsilon_\alpha = \varepsilon_\beta$; 0·15; 2·0, where in all of 11 independent simulations the fully native conformation is uniquely obtained. For the $\varepsilon_\alpha = \varepsilon_\beta$; 0·15; 1·5 case, in a series of six independent renaturation runs, the native conformation was uniquely obtained as well. Both cases exhibit an all-or-none transition, and there is no evidence whatsoever of the formation of the non-native three-helix bundle intermediate. Note that for the $\varepsilon_\alpha < \varepsilon_\beta$; 0·15; 1·5 case where helical states are locally favored, the non-native, three-helix bundle was formed in three out of 21 times. These results are consistent with the folding studies of four-member β-barrels, where a similar situation obtained. Thus, it appears once again that the presence of locally degenerate conformations allows the model protein to undo folding mistakes, before it gets trapped in deep local minima.

We now examine the situation where the helical regions in the primary sequence locally favor β-strand formation, but where the amino acid pattern is consistent with four-helix bundle formation. Thus, we considered primary sequences of the type $B(13)b_1(4)B_2(13)b_2(4)B_3(15)b_3(4)B_4(16)$ in which for the B_i, $\varepsilon_\theta(16) = 0$, $\varepsilon_\theta(12) = 0{\cdot}35/T^*$ and all other $\varepsilon_\theta = 0{\cdot}25/T^*$. All of the native-like $\varepsilon_\phi = -0{\cdot}6/T^*$. For the non-native turn conformers in b_i, $\varepsilon_\theta = 0{\cdot}25/T^*$, except that $\varepsilon_\theta(18) = 1/T^*$ for residues 16, 33 and 52 and $\varepsilon_\theta(16) = 0$ for residues 14, 31, and 50. $\varepsilon_\phi = -0{\cdot}6(2)/T^*$ for the native-like conformers in second and third residues in the putative native turns, and $\varepsilon_\phi = -0{\cdot}6(1{\cdot}5)/T$ for native-like conformers in the first and fourth residues of the putative native turns. For this particular choice of parameters, in the native conformation, $E_\theta = 19{\cdot}25/T^*$, $E_{tor} = -45/T^*$, $E_{side} = -9{\cdot}75/T^*$ and $E_c = -13{\cdot}95/T^*$. The total energy of the native conformation is $-49{\cdot}45/T^*$. The energy of a single residue in the helical state is $0{\cdot}35/T^*$, that of a pair of residues equals $0{\cdot}1/T^*$ and that of three residues is $-0{\cdot}15/T^*$. Thus, there is a significant local energetic cost to initiate helices, and it is not until a helical turn is complete that a helical state becomes energetically favorable. At $T^* = 0{\cdot}33$, the probability of an isolated helical state is 0·04, that of a pair of helical states is 0·017 and for a consecutive

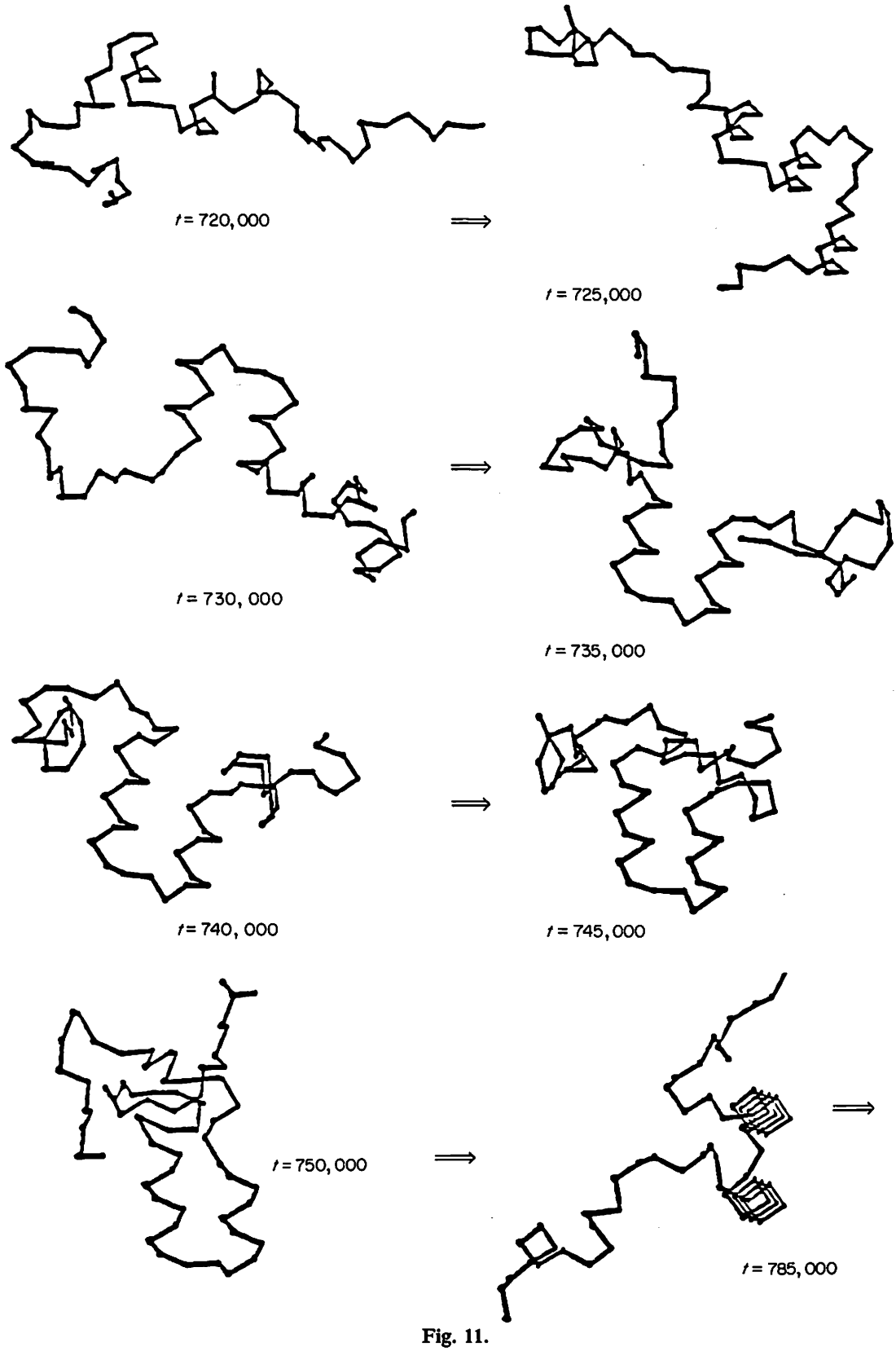

Fig. 11.

series of three-helical residues, it is 0·004. This primary sequence is designated by the short-hand notation, $\varepsilon_\alpha > \varepsilon_\beta$; 0·15; 1·5.

If a slight thermalization is employed, the four-helix bundle was obtained in each of 11 independent renaturation runs. The thermalization works as follows. To help the system surmount the local free energy barriers, for 1% of the time, on average, the

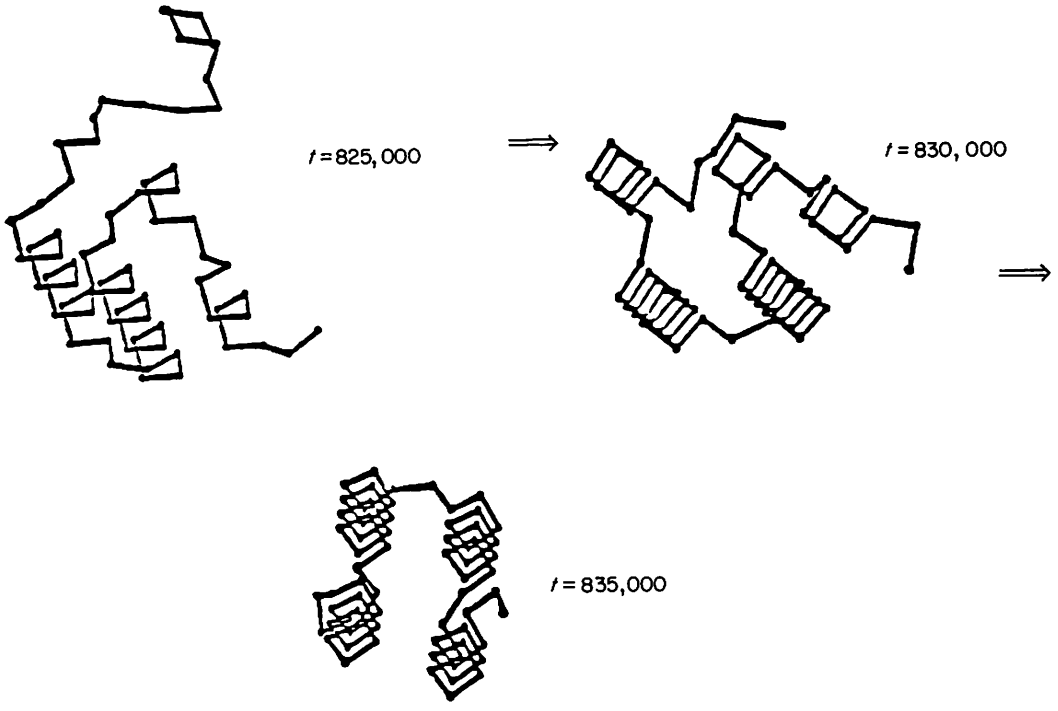

Figure 11. Representative folding trajectory of primary sequence type $\varepsilon_\alpha < \varepsilon_\beta$; 0·15; 1·75, where a native turn serves as the folding initiation site and lives to the end of the conformational transition. Different orientations are chosen at various times to highlight various features. The times indicated in the Fig. are from the start of the simulation.

temperature is raised by 3%. If the thermalization is removed from the algorithm, then the native four-helix bundle was obtained in four out of 6 independent renaturation runs. These non-native conformers occurred at quite low temperatures, and their presence partially reflects quenching effects.

To eliminate these unwanted intermediates, primary sequences of the type $\varepsilon_\alpha > \varepsilon_\beta$; 0·15; 1·75 were examined where $\varepsilon_\theta(12) = 0{\cdot}3/T^*$ in the putative helical regions. Thus, single residues strongly favor β-conformations, pairs of helical states are isoenergetic with β-conformations and triplets favor helical states by $0{\cdot}3/T^*$ The native conformation was uniquely obtained in seven out of seven attempts without thermalization. However, in many cases, prior to obtaining the native state, non-native three and four-helix bundle conformers were observed. These sequently dissolved. As would be expected, these systems are kinetically frustrated. On the basis of local conformational preferences, they desire to be in a completely different region of configuration space from that found in the native state.

We now summarize below the results of a series of simulations where turns are not at all favored. In all ten simulation runs, the desired native conformation was never obtained. Common misfolded structures, obtained three times, were three-helix bundle intermediates such as are shown in Figures 10(a) and (b). In some cases, a three-helix bundle with two helices of equal length and a third helix of approximately twice the length was obtained. Helical hairpins, perhaps with a dangling third helix, were obtained four times. In two other cases, unfolded conformations containing non-interacting helices were obtained. In one case, a four-helix bundle was observed. It consisted of a central short hairpin flanked by one longer and one shorter helix, the former being in the correct helix bundle geometry. Once again, these observations demonstrate that, while tertiary interactions alone can induce helix bundle formation, the resulting tertiary structures are non-unique. The turn locations are variable as would be expected from the fact that it costs the same free energy to make a turn independent of turn location, and the various lengths of helices in the three-helix bundle do not differ by a sufficient amount in energy to insure unique turn locations. Finally, we considered the case where the central turn is preferred. The turn preferred region served to localize the central turn, but the three-helix bundle intermediate was obtained in each of three independent runs.

A final series of simulation runs were attempted in which $\varepsilon_\phi = -0{\cdot}6/T^*$ for α-states everywhere in the primary sequence, but where $\varepsilon_\phi = -0{\cdot}6(2)/T^*$ for the turn conformations in the putative native turn regions as well. Now, for four out of four independent renaturation runs, the native conformation has been uniquely obtained *via* an all-or-none transition. We then reduced the propensity somewhat so that the 8, 14, 18 states in the turn have $\varepsilon_\phi = -0{\cdot}6(2)/T^*$ and the flanking 12 states have $\varepsilon_\phi = -0{\cdot}6(1{\cdot}5)/T^*$. In all four cases, the native conformation was again uniquely recovered. These simulations demonstrate that turn conformations

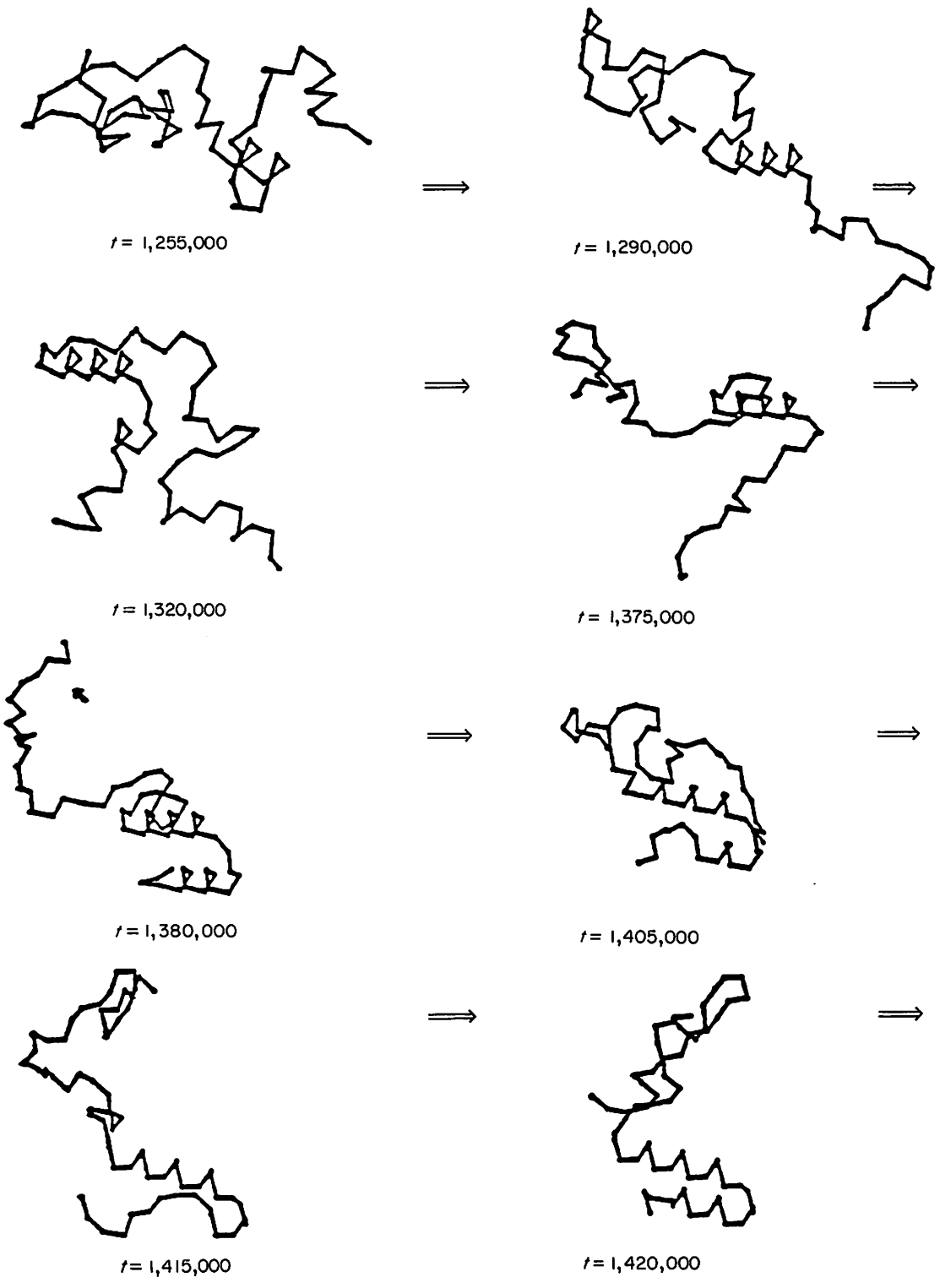

Fig. 12.

need not be the only states that are locally favored, but when they are sufficiently populated the native conformation can be found.

(ii) *Folding pathways*

Two kinds of early folding event were seen. In the first and simplest type, folding initiates at a native turn, and the native helices are constructed more or less simultaneously from the turn. Prior to initiation at the turn, the secondary structure that will be present in the native state has not yet appeared. Figure 11 shows such a folding transition for an $\varepsilon_\alpha < \varepsilon_\beta$; 0·15; 1·75 sequence, at $T^* = 0{\cdot}45$. Different orientations are chosen, at different times, to highlight different features that are important. The times indicated in the Figure are from the start of

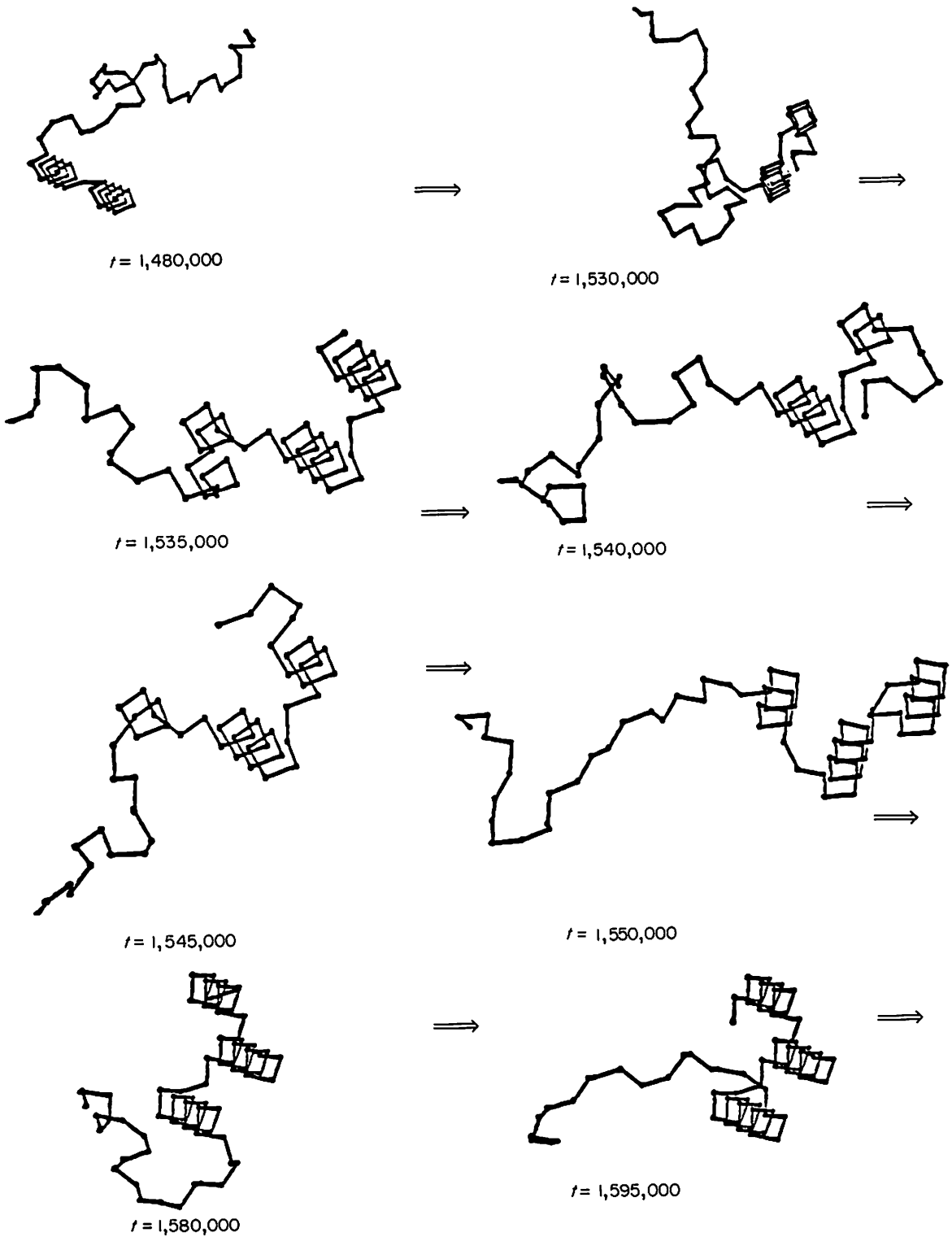

Fig. 12.

the run. Folding is seen to initiate directly from the central turn at $t = 720{,}000$. As is evident in the series of snapshots preceding $t = 745{,}000$, both interior helices form *via* side-by-side in-place construction. That is, an on-site mechanism obtains. The two tails then thrash about until $t = 825{,}000$, when turn three finally snaps into place. By $t = 830{,}000$, helix 4 is partially constructed, and the native turn is finally assembled. Interestingly, the first helix is also partially non-native, having adopted, near its end, the square bundle geometry. In the following 5000 time steps, this helix dissolves by a kink defect mechanism and is replaced by the native helix so that by $t = 835{,}000$, the four-helix bundle having 12 of 13 native contacts is first observed. Subsequently, the system undergoes the

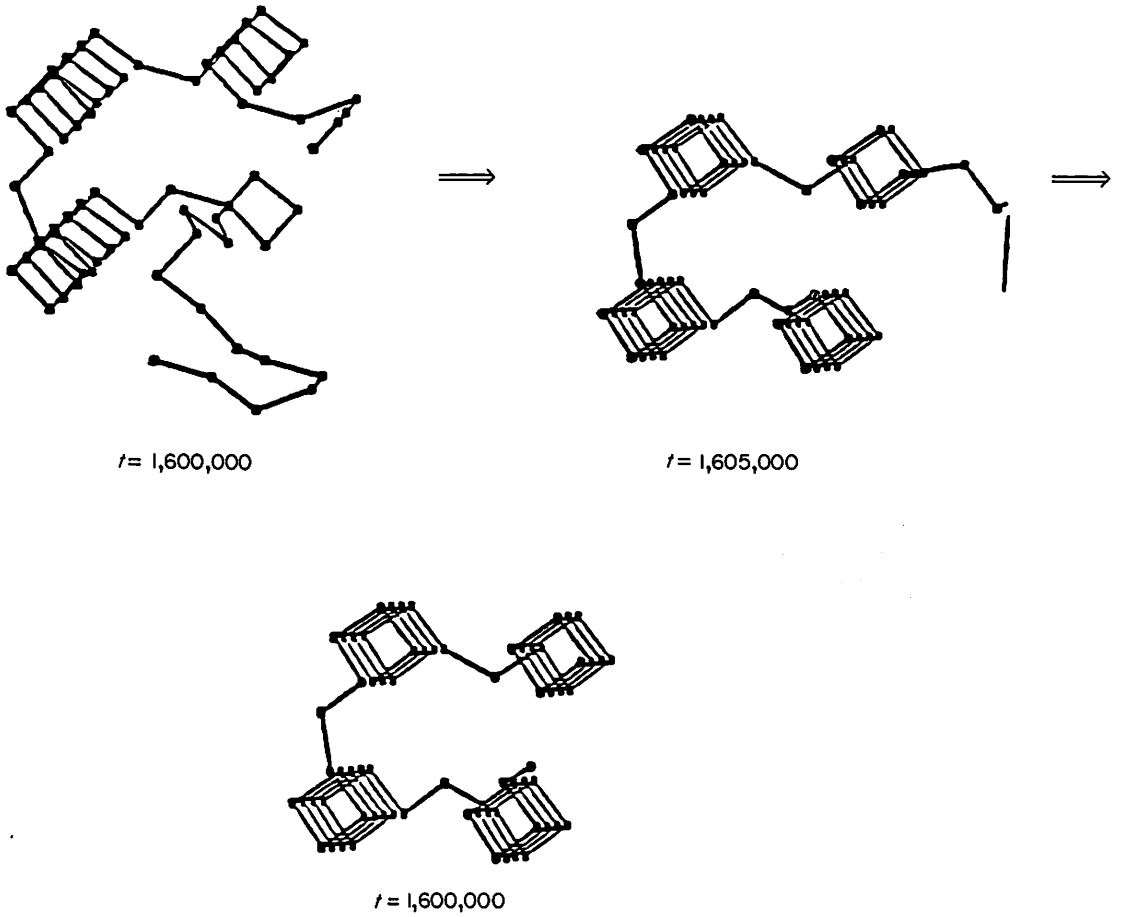

Figure 12. Representative folding trajectory of primary sequence type $\varepsilon_\alpha < \varepsilon_\beta$; 0·1; 2·0; 1·5. Different orientations are chosen at various times to highlight various features. The times indicated in the Figure are from the start of the simulation. See the text for further details.

minor thrashing of the tails. The overall folding process takes 115,000 time steps, with the longest-lived intermediate being the central helical hairpin.

The next kind of folding initiation event involves the formation of an isolated helix after which the native turn forms. This is then followed by formation of the helical hairpin. Sometimes the same hairpin that initially forms corresponds to the long-lived intermediate. At other times, the hairpin that initially forms does not survive until the end of folding. Rather, one of the helices dissolves, and the remaining helix is associated with a different hairpin, which then successfully folds. In the interest of brevity, only an example of the former is presented. We examine the folding of a molecule with primary sequence $\varepsilon_\alpha < \varepsilon_\beta$; 0·1; 2·0; 1·5 at $T^* =$ 0·415, in Figure 12. At $t = 1{,}255{,}000$, a portion of helix 2 first appears. An isolated three-turn helix persists up to $t = 1{,}375{,}000$, when native turn 1 is first seen. By $t = 1{,}380{,}000$, exterior helix 1 has partially formed. While the native turn persists, helix 1 is seen to dissolve (see the $t = 1{,}405{,}000$ and $t = 1{,}415{,}000$ snapshots) and then reform by $t = 1{,}420{,}000$. This conformation, with fluctuations, survives until $t = 1{,}530{,}000$. At $t = 1{,}535{,}000$, helix 3 has started to form from the turn. This third helix entirely dissolves by $t = 1{,}540{,}000$. By $t = 1{,}545{,}000$, the first turn of helix 3 is evident. This helix is then constructed on-site, a process that is essentially finished by $t = 1{,}580{,}000$. At $t = 1{,}600{,}000$, the first turn of helix 4 is evident, and by $t = 1{,}620{,}000$, the native molecule is observed. From the appearance of the helical hairpin (that survives until folding finishes) until the appearance of the three-helix bundle has taken approximately 125,000 time steps. The three-helix bundle then lives an additional 75,000 time steps before the fully native molecule is first obtained. Thus, the lifetime of the hairpin and that of the three-helix bundle are comparable. Note, however, that isolated helix 2 lived 120,000 time steps before folding initiation (involving native contacts) occurred. In other words, the lifetime of the helix is certainly comparable to the overall folding process, and yet, diffusion–collision–adhesion involving preformed helices did not occur. Moreover, isolated small fragments are seen to move about; they are not made artificially immobile by the algorithm. Qualitatively similar folding is seen in our off-lattice Brownian dynamics trajectories (Rey & Skolnick, 1991); thus, these pathways are not an artifact of the lattice nor of the dynamic Monte Carlo algorithm.

Interestingly, the mean folding times, given that successful folding will occur, are about a factor of 10

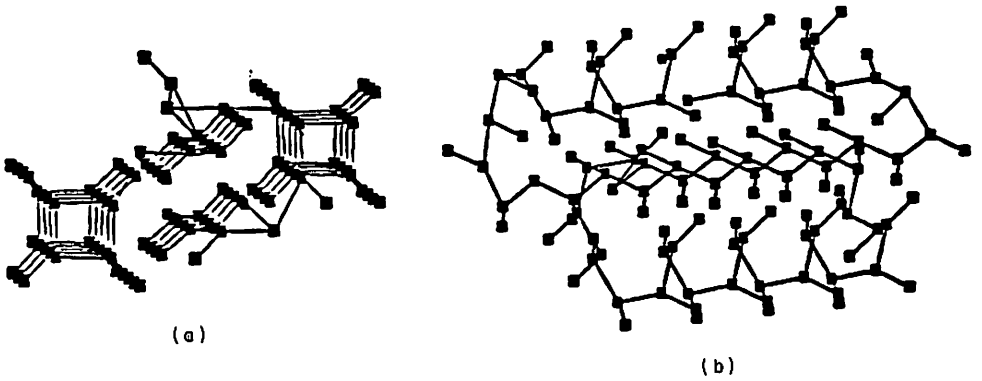

Figure 13. Native conformation of a β/α/β/α-4-member barrel including both α-carbon atoms and interacting side-chain sites is shown in top (a) and side (b) views.

to 20 longer on the 210 lattice than on the diamond lattice. While some slowdown in folding would certainly be expected given the greater number of accessible internal states per residue (3 *versus* 18), the fact that the times are this close indicates that more general considerations based on local entropic reduction, by initiating folding from turns or from a single helix are more important. Folding does not take 6^n times longer (n being the number of residues), and the system does not search through all of the configuration space. These effects would be expected to be far less sensitive to the difference in the total number of degrees of freedom between the two lattices. Moreover, it rationalizes the insensitivity of the folding pathways to lattice details (or even the use of a lattice), protein realization and the choice of local Monte Carlo moves.

One of the striking conclusions obtained from the folding trajectories is that overall initiation from the central helix turn is most likely. Considering those cases where all the turns are isoenergetic, out of a total of 38 successful folding events, in 24, or 61%, turn 2 is the first native turn that appears; 11 out of 38 had turn 1 as the first native turn that appears, and 3 out of 38 had turn 3 as the first native turn that appears. The origin of the preference for the central helix hairpin as the most favored initiation site is due to the fact that we are interested in the subpopulation of hairpins that fold *successfully*. Based on the relative numbers of interior *versus* exterior turns, each should occur 1/3 of the time. However, if the central hairpin appears, there are two ways that it can be further stabilized, either by the construction of outer helix 1 or outer helix 4. If an external hairpin serves as the initiation site, then there is only one helix that can attach to form the three-helix bundle intermediate. This implies that the lifetime of the central hairpin should be shorter than that of the exterior hairpin. Where a sufficient number of runs has been made to give somewhat meaningful statistics, this appears to be the case. For example, for $\varepsilon_\alpha < \varepsilon_\beta$; 0·15; 1·5, the mean lifetimes of the central hairpin and the exterior hairpins are 98,000 and 141,000, respectively.

There is an apparent preference for exterior helix initiation at the N-terminal side of the helix bundle; this occurred 11 times, as compared to three times for the C-terminal side. This result may partially reflect poor statistics. However, there is another effect that may contribute. For our choice of parameters, it is easier to initiate an N-terminal helix from a given turn, than a C-terminal helix. Thus, formation of helix 1 of turn 1 is more probable, and helix 3 of turn 3 is also more probable. However, if helix 1 (or a portion of it) is present, then initiation of helix 2 from the turn is more likely; that is, turn 1 will be the first turn that appears. Consider, now, helix 3. If it continues to grow, it can serve to initiate either a central turn or an exterior turn. Thus, exterior turn 3 is relatively less likely to appear as the first native turn than is turn 1. We point out, however, that another choice of parameters could just as easily have reversed the situation and made the helices on the C-terminal side of the turn easier to initiate. Nevertheless, differences in intrinsic propensities of helix initiation appear to be exerting a number of influences on the folding pathways.

(d) *Mixed α/β bundles*

We examine the nature of the conformational transition in four-member, β/α/β/α-barrels composed of $n = 62$ residues. Figure 13 displays the native conformation including both α-carbon atoms and interacting side-chain sites in (a) top and (b) side views. The β and helical regions have a pattern of hydrophilic and hydrophobic residues consistent with the conformation present in the native state.

(i) *Equilibrium results*

The folding transition has been examined for those primary sequences having local conformational preferences in consonance with the native conformation, indifferent to the global fold, and exhibiting local conformational preferences for the wrong (non-native) kind of secondary structure. In the last case, for example, extended chain (helix) is

favored in regions of the sequence that adopt a helical (β-sheet) conformation in the native state. First, folding was attempted in sequences whose local conformational preferences are in consonance with the native state. That is, we examined sequences of the type $B_1(9)b_1(3)A_1(17)b_2(4)B_2(8)b_2(3)A_2(17)$. Here, $\varepsilon_\theta = 0{\cdot}25/T^*$ for non-native bond angle conformations, and $\varepsilon_\theta = 0$ for conformations locally consistent with the native state. With the exception of the turn residues, $\varepsilon_\phi = -0{\cdot}6/T^*$ for local conformers consistent with the global fold. $\varepsilon_{\text{phil-phob}} = 1/T^*$, $\varepsilon_{\text{phil-phil}} = 0{\cdot}25/T^*$, and $\varepsilon_{\text{phob-phob}} = -0{\cdot}75/T^*$. Two particular situations were explored in detail. In the first, all of the $\varepsilon_\phi = -2(0{\cdot}6)/T^*$ for all of the putative turn forming conformations. This primary sequence is denoted by the short-hand notation, $\varepsilon_\alpha \neq \varepsilon_\beta$; 0·15; 2. Here $\varepsilon_c = -0{\cdot}15/T^*$. In the second case, the turn propensities for the central turn are the same as above, but all the exterior turns have their native conformations locally favored by a factor $\varepsilon_\phi = -1{\cdot}5(0{\cdot}6)/T$. Moreover, $\varepsilon_c = -0{\cdot}2/T^*$. This sequence will be denoted by the short-hand notation $\varepsilon_\alpha \neq \varepsilon_\beta$; 0·2; 1·5. For all the situations discussed below, residues 28 to 30 have $\varepsilon_\phi = -2(0{\cdot}6)/T^*$. If ε_ϕ of these residues is decreased substantially below this value, long lived intermediates involving the C-terminal helix, which interacts with β-strand 2, but not β-strand 1, are observed. In most runs, the native conformation is eventually recovered, but in order to eliminate these undesirable intermediates, the central turn propensity must be augmented. However, as shown below, in no case does the central turn propensity for native-like conformers exceed a few per cent in the denatured state.

In 6(3) out of 6(3) independent renaturation runs, the unique native conformation for the primary sequence $\varepsilon_\alpha \neq \varepsilon_\beta$; 0·15; 2($\varepsilon_\alpha \neq \varepsilon_\beta$; 0·2; 1·5) was obtained. For the former (latter) case, the estimated transition midpoint $T^*_{1/2} = 0{\cdot}47$ (0·492), whereas, the analytic theory (eqn (7)) provides a value of 0·488(0·50). Thus, reasonably good agreement between theory and simulation is evident.

The population of native turn conformers in the denatured state, calculated from the analytic theory assuming $\varepsilon_\phi = -1{\cdot}5(0{\cdot}6)/T^*$ for all of the turns, does not exceed 2%. (The corresponding quantity assuming $\varepsilon_\phi = -2(0{\cdot}6)/T^*$ for all of the turns does not exceed 7% in the transition region, although such an enhanced preference is not required for unique folding.) This is midway between the requisite turn propensities of a few tenths of 1% for a four-member β-barrel and about 5% in four-helix bundles. In all cases discussed below, an all-or-none transition is observed.

Next, sequences were examined where on an individual residue basis α-helical and β-states are isoenergetic. The first case considered has an identical set of turn propensities as in $\varepsilon_\alpha \neq \varepsilon_\beta$; 0·15; 2, but in all the putative helical and sheet regions $\varepsilon_\theta(12) = \varepsilon_\theta(16) = 0$. This sequence is denoted by the short-hand notation $\varepsilon_\alpha = \varepsilon_\beta$; 0·15; 2. For this particular situation, the native conformation was obtained in all five independent renaturation attempts.

Then, the intrinsic turn propensity of the native turn regions was reduced to $\varepsilon_\phi = -1{\cdot}5(0{\cdot}6)/T^*$. This sequence is denoted by the short-hand notation, $\varepsilon_\alpha = \varepsilon_\beta$; 0·15; 1·5. In all eight independent renaturation runs, the native conformation was obtained uniquely.

Finally, we considered primary sequences where β-conformations are favored in the putative helical portions of the primary sequence, and helical conformations are favored in the putative β-forming regions of the primary sequence. In all cases, $\varepsilon_c = -0{\cdot}15/T^*$. A number of different turn propensities were considered, where $\varepsilon_\phi = -\gamma(0{\cdot}6)/T^*$, with $\gamma =$ 1·5 and 1·75. For these primary sequences, the conformation that obtains in the native state is disfavored by a factor of $\varepsilon_\theta = \delta/T^*$. These primary sequences will be denoted by the short-hand notation, $\varepsilon_\beta <> \varepsilon_\alpha$; 0·15; γ; δ. Among the cases considered were primary sequences of the type $\varepsilon_\beta <> \varepsilon_\alpha$; 0·15; 1·5; 0·34 and $\varepsilon_\beta <> \varepsilon_\alpha$; 0·15; 1·75; 0·34. For both cases, the native conformation has been recovered in all five renaturation runs. At the transition midpoint for the latter sequences, based on intrinsic conformational preferences, an individual residue *disfavors* the native conformation by $0{\cdot}94k_BT$, a pair of residues *disfavors* the native conformation by $0{\cdot}22k_BT$ and a consecutive triplet of residues in native conformations, without tertiary interactions, *favors* the native state by $0{\cdot}77k_BT$. Once again, these calculations show that the native conformation can be found by the present algorithm if the helical and β-strand forming regions locally disfavor the conformation found in the native state. On the other hand, the putative turn-forming regions must have an intrinsic preference for native-like turns that give rise to populations in the unfolded state of a few per cent; otherwise, out-of-register and other types of misfolded conformers result. In other words, the portions of the primary sequence that are ultimately found in the native turns are more sensitive to the intrinsic preferences than are those regions that are found in helix or sheet. This may rationalize the observation that in many cases groups of five consecutive residues can be found in more than one kind of secondary structure (Kabsch & Sander, 1984; Zielenkiewicz & Rabczenko, 1988). Furthermore, this conclusion indicates that local interactions alone need not dictate the global fold. Rather, the identification of turn-forming regions along with medium and long-range interactions are sufficient to produce the unique native state in these model, mixed motif proteins and perhaps in real proteins, as well.

(ii) *Folding pathways*

The nature of the folding pathways in $\beta/\alpha/\beta/\alpha$-bundles is explored below. With regard to the early folding events, one of the crucial questions is the type(s) of secondary structures that form first

Table 3
Summary of folding initiation sites for α/β-models

Sequence	$T^*_{1/2}$	Turn 1	Turn 2	Helix 1	β-Sheet 2
$\varepsilon_\alpha \neq \varepsilon_\beta$; 0·15; 2	0·47	0	3	1	0
$\varepsilon_\alpha = \varepsilon_\beta$; 0·15; 1·5	0·44	1	5	0	2
$\varepsilon_\alpha = \varepsilon_\beta$; 0·15; 2	0·46	0	2	0	1
$\varepsilon_\alpha <> \varepsilon_\beta$; 1·5; 0·3	0·365	1	4	0	0
$\varepsilon_\alpha <> \varepsilon_\beta$; 1·5; 0·34	0·365	0	4	0	1
$\varepsilon_\alpha <> \varepsilon_\beta$; 1·75; 0·31	0·362	0	4	0	0
$\varepsilon_\alpha <> \varepsilon_\beta$; 1·75; 0·34	0·355	0	4	0	0

and persist to the end of folding (Wright *et al.*, 1988). Does folding initiate predominantly from a native-like turn, or perhaps a helix or an extended conformation forms first, which is followed by the turn formation? Are these results sensitive to the class of primary sequences involved, i.e. those whose local secondary structural preferences are in consonance with the global fold, as compared to those where local and global preferences are incompatible? What is the general mechanism of assembly? Does punctuated, on-site construction obtain or perhaps is preformed construction dominant (Karplus & Weaver, 1976, 1979; Lee *et al.*, 1987)? Are the longest lived intermediates closer to the folded than the unfolded state (Creighton, 1985, 1988)? For these simple models, answers to these questions are presented below.

Table 3 presents a compilation of the first identifiable element of structure found in the native state that is present initially and persists to the end of folding. For sequences whose local and global secondary structure preferences are in consonance, i.e. $\varepsilon_\alpha \neq \varepsilon_\beta$ type sequences, while turn 2 is the dominant folding initiation site, both helix 1 and β-sheet 2 have appeared prior to the formation of a native turn. While only a relatively small number of runs has been performed, it appears that, as the intrinsic preference for native sheet and helix diminishes, initiation at turn 2 becomes even more dominant. This makes intuitive sense. The putative turn regions retain their intrinsic preferences, while the β-strand and helix regions are locally disfavored. This is not to say that initiation from either a helix or a β-sheet cannot occur; indeed, we have seen examples of such behavior. However, out of a total of 33 independent attempts, 26 involve initiation at turn 2. Turn 1 rarely appears as the folding initiation site. For an $\varepsilon_\phi = 1{\cdot}5$ type sequence in the transition region, we find that turn 2 has the highest probability if $T^* < 0{\cdot}465$, and all the sequences of this type have a transition midpoint below this temperature. Thus, at first glance, it is not surprising that turn 2 serves in these cases as the primary initiation site. However, intrinsic preferences alone cannot account for this disparity. Turn 3 never serves as the primary initiation site, and turn 1 only rarely serves as the initiation site. Even more striking is the fact that for $\varepsilon_\phi = 2$ type sequences, turn 2 is always the least populated.

Why then is the central turn the dominant initiation site? First of all, it must be recognized that only the subpopulation of states that lead to a successful fold is being examined. All things being equal, it is best to initiate folding from the central α/β-element. This is because there are two ways to stabilize this intermediate, either by the formation of an external helix or an external sheet. Furthermore, and most importantly, due to the particular choice of parameters employed, the most probable initiation site is at the C-terminal end of helix 1. This preference is mostly responsible for the overwhelming preference of turn 2 as the primary initiation site.

A representative example of a folding trajectory that illustrates growth from the central turn is shown in Figure 14. The times indicated in the Figure are from the start of the run. At $t = 375{,}000$, folding initiates at the central turn. Observe the presence of a β-strand at right angles to the helix. As confirmed by the snapshots at $t = 380{,}000$ to $t = 395{,}000$, both the helix and β-strand form *via* a side-by-side, on-site assembly process. By $t = 400{,}000$, native turn 1 and the external β-strand 1 have assembled. The remaining unassembled portions of the chain thrash about until $t = 445{,}000$, when turn 3 is evident for the first time. It takes another 40,000 steps for the entire native conformation to appear. The total elapsed time from folding initiation to the first appearance of all of the native contacts is 105,000 time steps and to complete assembly is 110,000 time steps.

In Figure 15, a representative trajectory illustrating the appearance of a β-sheet as the first identifiable element of secondary structure that persists until the end of folding is shown. The folding trajectory was obtained from the primary sequence $\varepsilon_\alpha = \varepsilon_\beta$; 0·15; 1·5 at $T^* = 0{\cdot}47$. Prior to the formation of turn 2 at $t = 645{,}000$, β-sheet 2 has appeared. By $t = 650{,}000$, helix 1 initiates from the central turn and completes assembly by $t = 655{,}000$. By $t = 665{,}000$, native turn 1 appears. This is followed by assembly of external β-sheet 1, a process that is finished by $t = 670{,}000$. The remaining random coil tail will continue to thrash about until $t = 755{,}000$, when turn 3 first appears. The external helix then rapidly assembles on-site, so that by $t = 765{,}000$ the fully native molecule is observed for the first time. The total elapsed folding time is 120,000 time steps, with the β/α/β-intermediate living for 95,000 time steps prior to the formation of the native conformation.

Figure 16 presents a folding trajectory in which

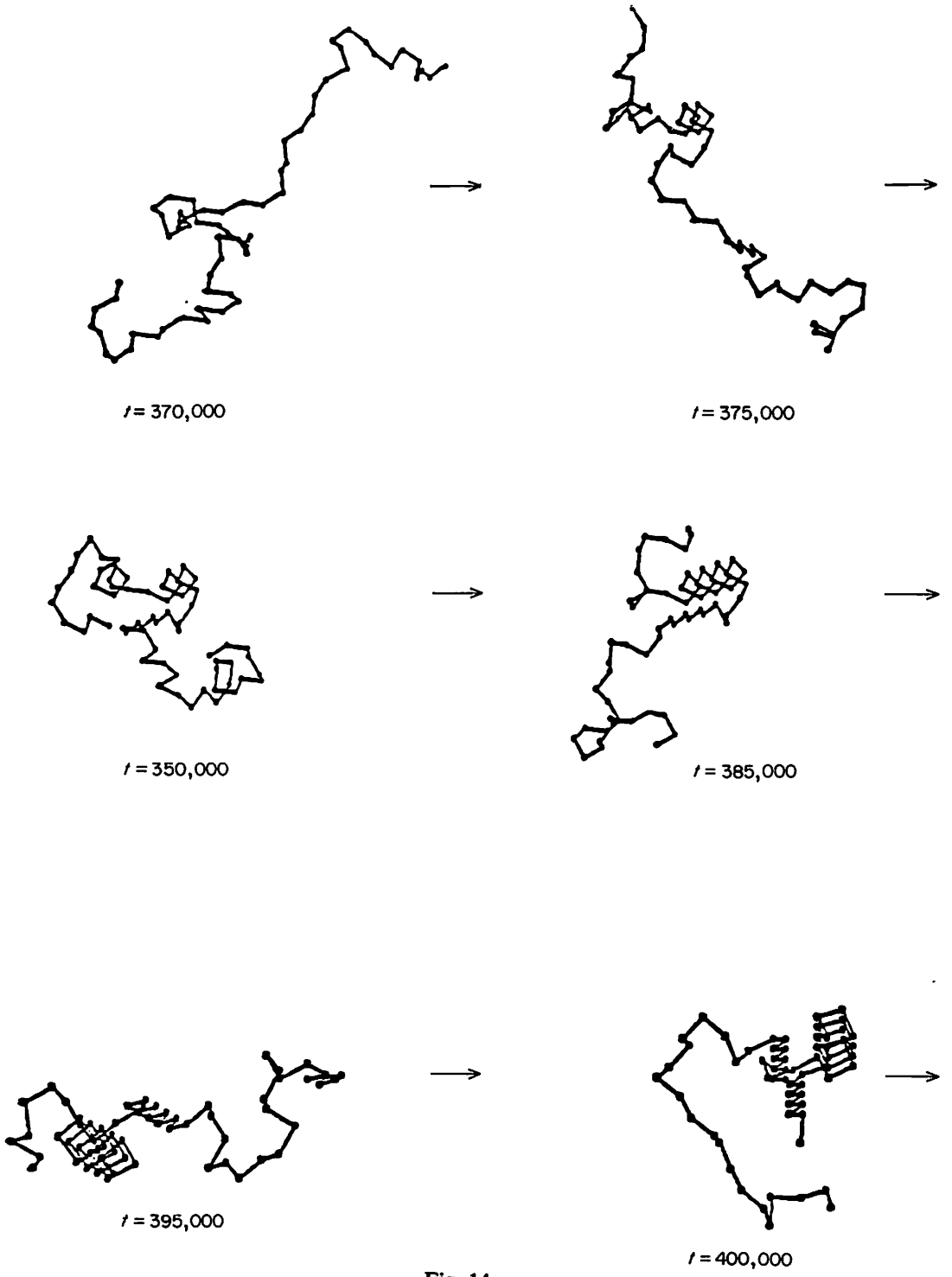

Fig. 14.

helix 1 forms early in the folding process and corresponds to $\varepsilon_\alpha = \varepsilon_\beta$; 0·15; 2 at $T^* = 0{\cdot}48$. At $t = 385{,}000$, folding is seen to initiate from an almost-formed helix plus turn 2. Within 5000 time steps, not only has the first contact between helix 1 and β-strand 2 formed, but the entire helix 1 has zipped up into place. In this case, the construction of the second β-sheet lags behind that of the helix. This sheet in fact will not finish its assembly until about $t = 400{,}000$. In the meantime, turn 1 appears, dissolves and reforms, so that by $t = 400{,}000$, exterior turn 1 (not shown) has formed and will persist until assembly is complete. By $t = 405{,}000$, β-strand 1 begins to zip up from the turn. This process is completed for the first time by $t = 410{,}000$. The external strand 1 will continue to dissolve and reform throughout the remainder of the assembly process. The remaining random coil tails will continue to thrash about until $t = 595{,}000$, when the folding of the exterior helix initiates. By

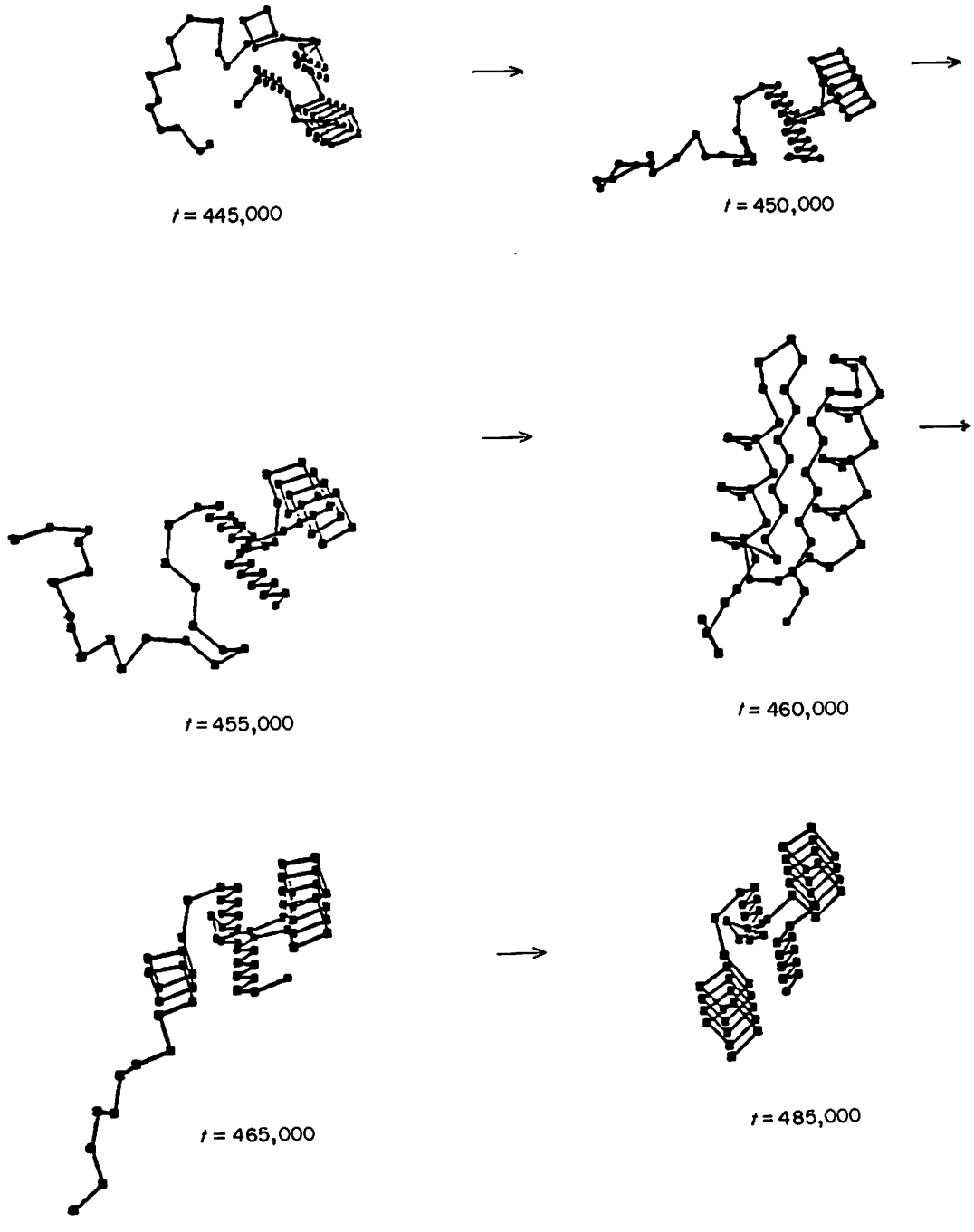

Figure 14. A representative folding trajectory of an $\varepsilon_\alpha = \varepsilon_\beta$; 0·15; 1·5 type sequence illustrating on-site growth from a central turn. The times indicated are from the start of the simulation run at $T^* = 0{\cdot}4375$. Different perspectives are shown at various times.

$t = 610{,}000$ folding is essentially complete, with all 16 native side-chain contacts in place with the fully native molecule first observed at $t = 615{,}000$ (not shown). In all of these cases, the $\beta/\alpha/\beta$-intermediate lies closer to the folded than to the unfolded state, as appears to be required by real experiments (Creighton, 1985, 1988).

(iii) *Summary of a model of an allosteric transition*

During the course of identifying sufficient conditions to obtain the native conformation uniquely, we occasionally observed conformations where helix 1 oscillated between two locations. This occurred under conditions where the turn propensity for the native conformations of the turns was rather weak. This result suggests that an allosteric transition (Monod *et al.*, 1965) involving helix 1 could be controlled simply by modulating the turn propensity associated with the two conformations of the helix. This allosteric conformational transition is somewhat analogous to the helix-shear interface movements identified as the conformational change

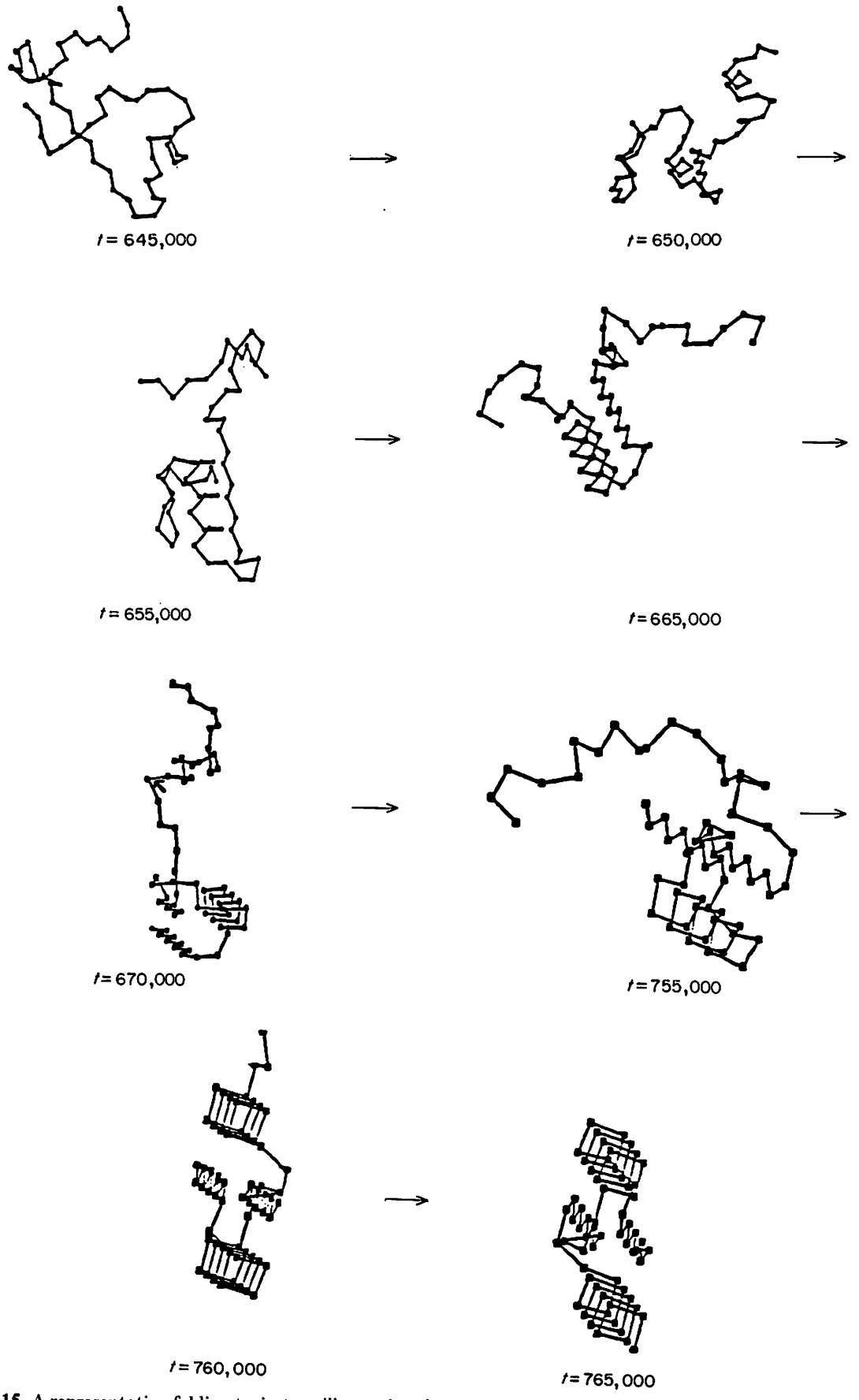

Figure 15. A representative folding trajectory illustrating the appearance of β-sheet 2 as the 1st identifiable element of secondary structure that persists to the end of folding. The folding trajectory was obtained from primary sequence $\varepsilon_\alpha = \varepsilon_\beta$; 0·15; 1·5 at $T^* = 0{\cdot}47$.

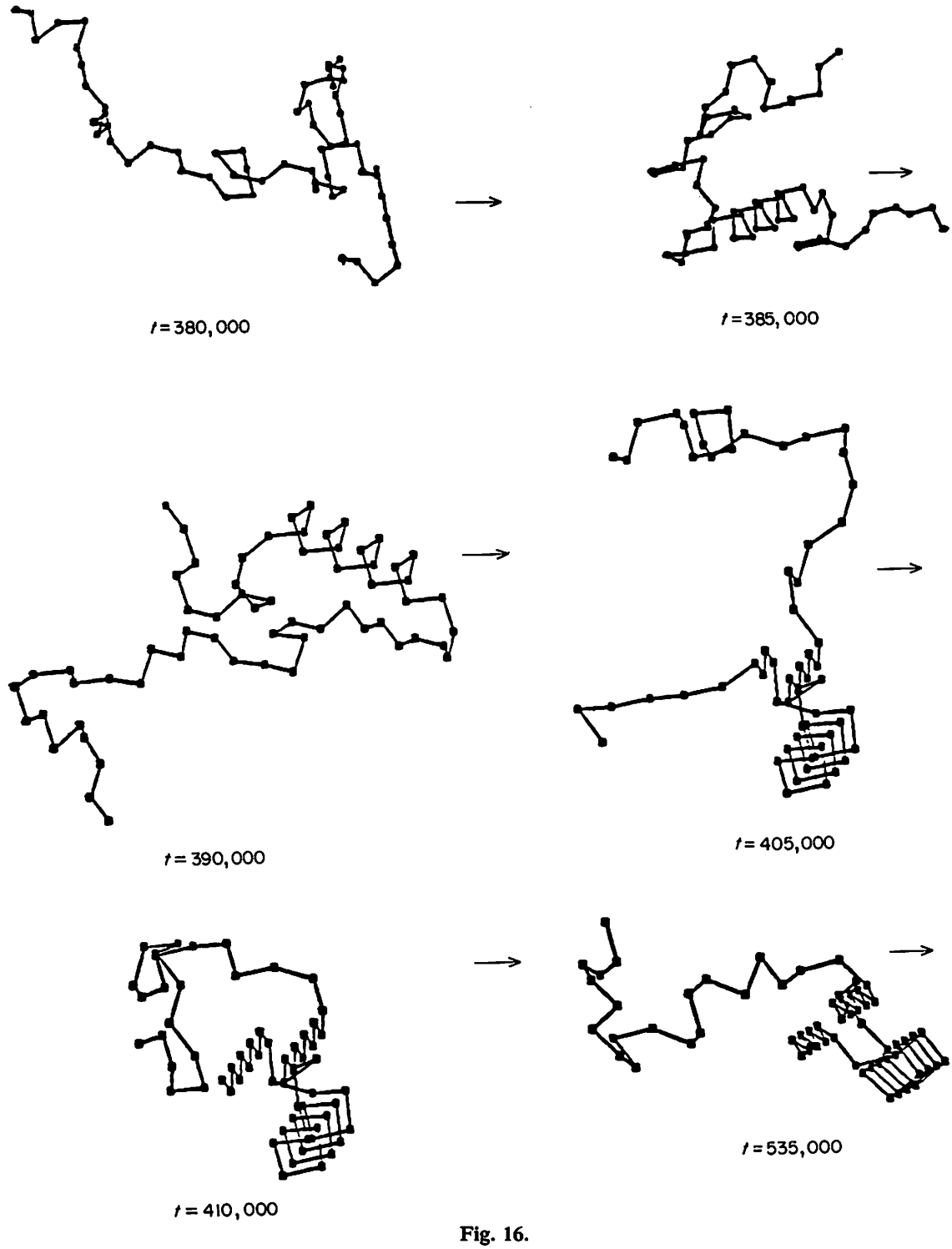

Fig. 16.

in citrate synthase (Lesk & Chothia, 1984) and in insulin (Chothia *et al.*, 1983).

Quite quickly after the turn conditions are changed, the turn conformation that was originally present (and previously favored), quite quickly dissolves and is only occasionally repopulated. In all cases, the system alternates between the two forms of the folded conformation. These results suggest that if the conformational preferences of a turn in real molecules could somehow be modified by an external change in conditions, then it might be possible literally to switch enzymatic activity on and off as desired. One possible way to realize this is by engineering into the native turns natural or synthetic amino acids that undergo appropriate isomerization on the application of light or as a result of a change in pH.

Examination of the trajectories reveals that the system undergoes the allosteric transition involving helix translation not by dissolving the entire helix

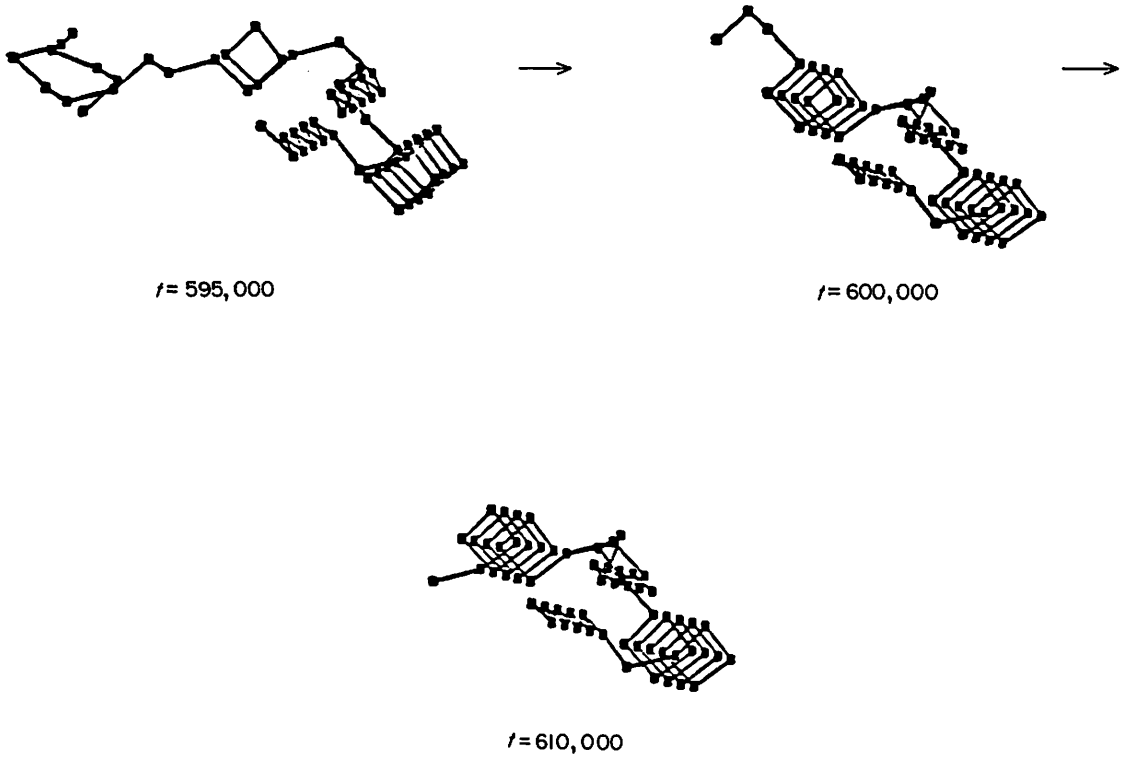

Figure 16. A representative trajectory folding illustrating the appearance of helix 1 early in the folding pathway and persisting to the end of folding. The folding trajectory was obtained from primary sequence $\varepsilon_\alpha = \varepsilon_\beta$; 0·15; 2 at $T^* = 0{\cdot}48$.

and reforming it, but rather by the local denaturation and renaturation of the helix, with the concomitant defect propagation that effectively translates the helix from between the two allosteric positions. The defect propagation mechanism costs relatively little free energy compared to complete helix denaturation, which would destroy both secondary and tertiary structure. Moreover, the application of a switch in turn propensities aids in this process by dissolving the least stable element of the helix, the end, and moving it into the (now) lower free energy orientation of the turn. Thus, these simulations are suggestive of a mechanism of helix or sheet rearrangement in allosteric transitions that involves local defect propagation.

5. Summary and Conclusions

Here, a new lattice model of globular proteins has been applied to the folding of model four-member β-barrels, four-helix bundles and mixed-motif α/β-proteins. For all major motifs, we find that folding to a unique native conformation can occur even if those residues found in the putative β-strands and α-helical regions locally favor the wrong element of secondary structure. For example, in the case of a model $\beta/\alpha/\beta/\alpha$-bundle, locally the native β-regions can rather strongly favor α-helix, and the native α-helical regions can strongly prefer extended conformations. While the folding is certainly slowed down relative to the situation where local and global conformational preferences are in consonance, it is found that a marginal propensity for native-like turns plus tertiary interactions are sufficient to produce unique native conformations. Removal of the marginal, but certainly not irrelevant, native-like turn propensities (which give rise in the unfolded state to turn populations of a few per cent or less) produces mixed motif bundles that are non-unique. We remind the reader that the tertiary interactions between side-chains are independent of backbone conformation. Thus, these systems fold to native conformations that are in the lowest free energy state due to the juxtaposition of hydrophobic/hydrophilic pattern information and short-range preferences that select the native conformation from the manifold of collapsed, native-like states.

From this series of model studies, it appears that the nature of the folding pathways are qualitatively the same for α, β and mixed α/β-motifs. Folding typically initiates at or near a turn present in the native state. In some cases, portions of one of the β-strands or isolated helices form first. The overall mechanism of folding is the same (on-site construction) with the particular pathway(s) chosen dependent on the topology of the native state. For the α and β-proteins, we find that the pathways are qualitatively identical with those seen on the diamond lattice (Skolnick & Kolinski, 1990*a*; Sikorski & Skolnick, 1990) and in the off-lattice folding of helical hairpins (A. Rey & J. Skolnick, unpublished results). This provides further support

that the many features of the pathways depend on topology and not details. (Of course, if a particular turn is favored in one instance and not another, it may serve as the folding initiation site under the favorable circumstances.)

As indicated above, turns play an important role in the initiation of folding. β-Proteins appear to require the least intrinsic propensity to form native-like turns (about 0·3% turn populations in the unfolded state) in order to yield unique folding to the native conformation. Mixed-motif α/β-proteins require slightly larger turn propensities (about 2%), and helical proteins require the largest turn propensities (about 5 to 8%). The relative turn propensities of helical and β-proteins are consistent with a number of experimental observations of Dyson and co-workers (H. J. Dyson, J. R. Sayre, R. A. Lerner & P. A. Wright, unpublished results; J. P. Waltho, H. S. Cheol & P. E. Wright, unpublished results). For all motifs, we have observed that if locally the residues are indifferent to helix or sheet, the turn propensity can be smaller than if helical or sheet regions are preferred. This is due to the fact that, when two conformations are degenerate, if an incorrect turn has occurred, the system can correct the mistake more easily than if the sheet or helix is preferred. While it is also true that prior to initiation the system spends less time in the native-like states, this does not pose a major problem. Fluctuations into and out of sheet and helix are rapid compared to successful initiation events. Once folding begins and the stabilization of the conformation exceeds several k_BT, it is important that channels exist that permit dissolution of undesirable intermediates before they are trapped into very deep local free energy minima. This can also occur if the system locally disfavors the native conformation; however, in the early stages of folding, this means that the initiating site may not have sufficient time to find a native-like piece that is stabilized by tertiary interactions before it dissolves. Thus, successful folding is a compromise between local preferences and tertiary stabilization.

In the series of simulations on $\beta/\alpha/\beta/\alpha$-bundles, turns were also found to play an important role in an allosteric conformational transition. If conditions are changed so that one turn conformation is no longer favored and another is favored, then the central helix can literally hop back and forth between two alternative conformations. This hopping can be switched on and off by the suitable application of a modest turn potential. These results suggest that some allosteric transitions, especially those involving helix shear interface movements (Lesk & Chothia, 1984; Chothia *et al.*, 1983) may be due to multiple conformations of turns. While we realize that the case considered here is extremely idealized, nevertheless these simulations suggest that it may be possible to engineer enzymes that switch on and off by a change in external conditions. For example, light could cause a conformational transition in a turn region that then induces migration of an α-helix or β-sheet from an inactive to an active conformation. Thus, these models may prove useful in guiding the genetic engineering of real proteins.

Overall, the present series of simulations on the 210 lattice model of globular protein folding are encouraging. The folding of all the major motifs has now been achieved, and the crucial test of being able to fold mixed-motif α/β-proteins has been passed. Furthermore, side-chains have been included, and, rather than complicating the situation, they aid in the folding by disfavoring mirror image conformers that are isoenergetic in an α-carbon representation. Therefore, while these models of a globular protein are still highly simplified, the results, thus far, provide the impetus for additional refinements, and work is now in progress on more sophisticated models.

This research was supported in part by NIH grant no. GM-37408 from the Division of General Medical Sciences, United States Public Health Service. Stimulating discussions with Drs H. Jane Dyson, Mariusz Milik, Fred Richards, Antonio Rey, Lee Walters and Peter Wright are gratefully acknowledged.

Appendix A

Calculation of the Denatured State Free Energy

In this Appendix, an approximate analytical expression for the free energy of the unfolded state of a model protein confined to the 210 lattice is calculated.

(a) *General considerations*

We begin with a discussion of the geometrical properties of the 210 lattice. Let us associate each of the 24 possible vectors connecting a given 210 lattice site with a number, 1 through 24:

1 = (2, 1, 0)	**13** = (0, −1, −2)	
2 = (2, 0, 1)	**14** = (0, −2, −1)	
3 = (2, −1, 0)	**15** = (0, 1, −2)	
4 = (2, 0, −1)	**16** = (0, −2, 1)	
5 = (1, 2, 0)	**17** = (−1, 2, 0)	
6 = (1, 0, 2)	**18** = (−1, 0, 2)	
7 = (1, −2, 0)	**19** = (−1, −2, 0)	(8)
8 = (1, 0, −2)	**20** = (−1, 0, −2)	
9 = (0, 1, 2)	**21** = (−2, 1, 0)	
10 = (0, 2, 1)	**22** = (−2, 0, 1)	
11 = (0, −1, 2)	**23** = (−2, −1, 0)	
12 = (0, 2, −1)	**24** = (−2, 0, −1).	

Thus, to specify the conformation of the chain, given the location of the first bead, one needs only to give a sequence of $N-1$ numbers, ranging from 1 to 24. For the first bond vector, the choice of which vector to use is completely arbitrary. For convenience, let us choose vector **1**. The second vector must satisfy the constraint that $6 \leq r_0^2 \leq 18$. In general, for a given vector **i**, there are 18 allowed vectors; the allowed vectors vary depending on the particular vector that precedes them.

This is suggestive that we define a pseudo inner

product by analogy to ortho-normal basis sets as follows:

$$\langle \mathbf{i}, \mathbf{j} \rangle = 1 \tag{9}$$

if the two vectors **i** and **j** are allowed and:

$$\langle \mathbf{i}, \mathbf{j} \rangle = 0, \tag{10}$$

if the two two vectors **i** and **j** are not allowed.

(b) *Calculation of the partition function*

The chain is divided into statistcal weight matrices associated with pairs of bonds. That is, the partition function is calculated as:

$$\mathbf{Z}_D = \mathbf{J}^*_{576} \prod_{i=2;\,\text{even}\,i}^{l_u} U_i^{\phi} \mathbf{J}_{576}. \tag{11}$$

Here, $\mathbf{J}^*_{576}$ is a row vector of dimensionality 576 whose first term is unity and the remainder of the terms are zero. $\mathbf{J}_{576}$ is a column vector of dimensionality 576, all of whose elements are unity. $l_u = N$ if N is even, and $l_u = N-1$ if N is odd. U_i^{ϕ} is a 576 by 576 matrix. For convenience in setting up U_i^{ϕ}, the torsional angles are labeled from 3 to $N-1$, rather than from 2 to $N-2$ as in the text. For $i = 2$, one merely has to account for the bond angle associated with the second residue. Choosing the first bond as vector **1**, the only non-zero elements of U_2^{ϕ} are

$$U_2^{\phi}(1, \mathbf{j}) = \langle 1, \mathbf{j} \rangle \exp(-\varepsilon_{\theta,2}(1, \mathbf{j})/k_B T). \tag{12}$$

We next consider the case where $2 < i < l_u$. Let the bond vectors associated with residues $i-3$, $i-2$, $i-1$ and i be labeled **j**, **k**, **l**, **m**, respectively. The jth bond vector connects residue $i-3$ with $i-2$. The rows of U_i^{ϕ} (row, column) are obtained from **j** and **k** by

$$\text{row} = (\mathbf{j}-1)24 + \mathbf{k}, \tag{13a}$$

$$\text{col} = (\mathbf{l}-1)24 + \mathbf{m}. \tag{13b}$$

Before presenting the elements of the statistical weight matrix $U_i^{\phi}(j, k, l, m)$, it proves convenient to define the statistical weight matrix associated with the torsional potential $U_{\phi}(j, k, l, i)$ due to a particular sequence of three bonds j, k, l where k goes from vertex $i-1$ to i. Let us consider the distance, $r_{i-2,i+1}$, between residues $i-2$ to $i+1$. If the square of this distance is less than 3, then due to the hard core steric repulsion,

$$U_{\phi}(j, k, l, i) = 0, \tag{14a}$$

If $r^2_{i-2,i+1} = 3$, then:

$$U_{\phi}(j, k, l, i) = \langle j, k \rangle \langle k, l \rangle \times \exp[-(\varepsilon_{\phi}(j, k, l) + 3\varepsilon_{rep})/k_B T]. \tag{14b}$$

If $r^2_{i-2,i+1} = 5$, then:

$$U_{\phi}(j, k, l, i) = \langle j, k \rangle \langle k, l \rangle \times \exp[-(\varepsilon_{\phi}(j, k, l) + \varepsilon_{rep})/k_B T]. \tag{14c}$$

For all other $r^2_{i-2,i+1}$,

$$U_{\phi}(j, k, l, i) = \langle j, k \rangle \langle k, l \rangle \times \exp[-(\varepsilon_{\phi}(j, k, l))/k_B T]. \tag{14d}$$

Thus, local short-range repulsions are also accounted for.

For $2 < i < l_u$, if l_u is even, and for $2 < i \leq l_u$ if l_u is odd, then substituting in equation (11)

$$U_i^{\phi}(j, k, l, m) = \langle j, k \rangle \langle k, l \rangle \langle l, m \rangle \times \exp(-(\varepsilon_{\theta,i-i}(k, l) + \varepsilon_{\theta,i}(l, m)/k_B T) \times U_{\phi}(j, k, l, i-1) U_{\phi}(k, l, m, i). \tag{15}$$

Finally, if $i = l_u$, and l_u is even then, since vertex i is at the end of the chain, we need only account for the last bond angle and torsional angle associated with vertex $N-1$. To make this last matrix comformable with the previous matrices, we must append an arbitrary extra bond at the chain end.

$$U^{\phi}_{l_u=N}(j, k, l, N) = \langle j, k \rangle \langle k, l \rangle \langle l, 1 \rangle \times \exp(-(\varepsilon_{\theta N-1}(k, l)/k_B T) \times U_{\phi}(j, k, l, N-1). \tag{16}$$

The free energy of the denatured state is obtained from

$$A_D = -k_B T \ln (\mathbf{Z}_D). \tag{17}$$

References

Anfinsen, C. B. (1973). Principles that govern the folding of protein chains. *Science*, **181**, 223–230.

Baumgartner, A. (1984). Simulation of polymer motion. *Annu. Rev. Phys. Chem.* **35**, 419–435.

Binder, K. (1984). Editor of *Application of the Monte Carlo Method in Statistical Physics*, chapt. 5, Springer-Verlag, Berlin.

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

Brandts, J. F., Halvorson, H. R. & Brennan, M. (1975). Consideration of the possibility that the slow step in protein denaturation reaction is due to *cis–trans* isomerism of proline residue. *Biochemistry*, **14**, 4953–4963.

Bundi, A., Andreatta, R. H., Rittel, W. & Wütrich, K. (1976). Conformational studies of the synthetic fragment 1-34 of human parathyroid hormone by NMR techniques. *FEBS Letters*, **64**, 126–129.

Bundi, A., Andreatta, R. H. & Wütrich, K. (1978). Characterisation of a local structure in the synthetic parathyroid hormone fragment 1-34 by ^{1}H nuclear resonance techniques. *Eur. J. Biochem.* **91**, 201–208.

Chan, H. S. & Dill, R. A. (1989). Compact polymers. *Macromolecules*, **22**, 4559–4573.

Chothia, C., Levitt, M. & Richardson, D. (1977). Structure of proteins: Packing of α-helices and pleated sheets. *Proc. Nat. Acad. Sci., U.S.A.* **74**, 4130–4134.

Chothia, C., Lesk, A. M., Dodson, G. G. & Hodgkin, D. C. (1983). Structure of proteins: packing of α helices and pleated sheets. *Nature (London)*, **302**, 500–505.

Chou, K. C., Maggiora, G. M., Nemethy, G. & Scheraga, H. A. (1988). Energetics of the structure of the four-α-helix bundle in proteins. *Proc. Nat. Acad. Sci., U.S.A.* **85**, 4295–4299.

Creighton, T. E. (1985). The problem of how and why proteins adopt folded conformations. *J. Phys. Chem.* **89**, 2452–2459.

Creighton, T. E. (1988). Toward a better understanding of protein folding pathways. *Proc. Nat. Acad. Sci., U.S.A.* **85**, 5082–5086.

Degrado, W. F., Wasserman, Z. B. & Lear, J. D. (1989).

Protein design, a minimalist approach. *Science*, **243**, 622–628.

Downey, J. P. & Kovac, J. (1987). Effects of coordination number on the dynamics of lattice models for polymer chains. *Macromolecules*, **20**, 1357–1362.

Dyson, H. J., Rance, M., Houghten, R. A., Lerner, R. A. & Wright, P. E. (1988*a*). Folding of immunogenic fragments of proteins in water solution. I. Sequence requirements for the formation of a reverse turn. *J. Mol. Biol.* **201**, 161–200.

Dyson, H. J., Rance, M., Houghten, R. A., Lerner, R. A. & Wright, P. E. (1988*b*). Folding of immunogenic peptide fragments of proteins in water solution. II. The nascent helix. *J. Mol. Biol.* **201**, 201–217.

Flory, P. (1969). *Statistical Mechanics of Chain Molecules* chapt. 7, Interscience, New York.

Garel, J. R. & Baldwin, R. L. (1973). Both the fast and slow refolding reactions of ribonuclease A yield native enzyme. *Proc. Nat. Acad. Sci., U.S.A.* **70**, 3347–3351.

Hill, T. L. (1956). *Statistical Mechanics*, chapt. 6, McGraw-Hill, New York.

Kabsch, W. & Sander, C. (1984). On the use of sequence homologies to predict protein structure: identical pentapeptides can have completely different conformations. *Proc. Nat. Acad. Sci., U.S.A.* **81**, 1075–1079.

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

Karplus, M. & Weaver, D. L. (1979). Diffusion-collision model for protein folding. *Biopolymers*, **18**, 1421–1427.

Kolinski, A., Skolnick, J. & Yaris, R. (1986). Monte Carlo simulations on an equilibrium globular protein folding model. *Proc. Nat. Acad. Sci., U.S.A.* **83**, 7267–7271.

Kolinski, A., Skolnick, J. & Yaris, R. (1987*a*). Monte Carlo studies on equilibrium globular protein folding. I. Homopolymeric lattice models of β-barrel proteins. *Biopolymers*, **26**, 937–962.

Kolinski, A., Skolnick, J. & Yaris, R. (1987*b*). Does reptation describe the dynamics of entangled polymer systems? A model simulation. *J. Chem. Phys.* **86**, 1567–1585.

Kolinski, A., Milik, M. & Skolnick, J. (1991). Static and dynamic properties of a new lattice model of polypeptide chains. *J. Chem. Phys.* **94**, 3978–3985.

Lee, S., Karplus, M., Bashford, D. & Weaver, D. (1987). Brownian dynamics simulations of protein folding: a study of the diffusion collision model. *Biopolymers*, **26**, 481–506.

Lesk, A. M. & Chothia, C. (1984). Mechanisms of domain closure in proteins. *J. Mol. Biol.* **174**, 175–191.

Miyazawa, S. & Jernigan, R. (1985). Estimation of effective inter-residue contact energies from protein crystal structures: quasi chemical approximation. *Macromolecules*, **18**, 534–552.

Monod, J., Wyman, J. & Changeaux, J. P. (1965). On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* **12**, 88–118.

Privalov, P. L. & Khechinashvili, N. N. (1974). A thermodynamic approach to the problem of stabilization of globular protein structure: a calorimetric study. *J. Mol. Biol.* **86**, 665–684.

Privalov, P. L., Griko, Y. V., Venyaminov, S. Y. & Kutyshenko, V. P. (1986). Cold denaturation of myoglobin. *J. Mol. Biol.* **190**, 487–498.

Rey, A. & Skolnick, J. (1991). Comparison of Brownian dynamics and lattice dynamics folding pathways of α-helical hairpins. *Chem. Phys.* in the press.

Richardson, J. S. (1981). The anatomy and taxonomy of protein structure. *Advan. Protein Chem.* **34**, 167–339.

Richmond, T. J. & Richards, F. M. (1978). Packing of α-helices: geometrical constraints and contact areas. *J. Mol. Biol.* **119**, 537–555.

Shoemaker, K. R., Kim, P. S., Brems, D. N., Marqusee, S., York, E. J., Chaikin, I. M., Stewart, J. M. & Baldwin, R. L. (1985). Nature of the charged-group effect on the stability of the C-peptide helix. *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). Tests of the helix dipole model for stabilization of α-helices. *Nature (London)*, **326**, 563–567.

Sikorski, A. & Skolnick, J. (1989*a*). Monte Carlo studies on equilibrium globular protein folding. III. The four helix bundle. *Biopolymers*, **28**, 1097–1113.

Sikorski, A. & Skolnick, J. (1989*b*). Monte Carlo simulation of equilibrium globular protein folding. α-helical bundles with long loops. *Proc. Nat. Acad. Sci., U.S.A.* **86**, 2668–2672.

Sikorski, A. & Skolnick, J. (1990). Dynamic Monte Carlo simulations of globular protein folding/unfolding pathways. II. α-Helical motifs. *J. Mol. Biol.* **212**, 819–836.

Skolnick, J. & Kolinski, A. (1989). Computer simulations of globular protein folding and tertiary structure. *Annu. Rev. Phys. Chem.* **40**, 207–235.

Skolnick, J. & Kolinski, A. (1990*a*). Dynamic Monte Carlo simulations of globular protein folding/unfolding pathways. I. Six member, Greek key β-barrels. *J. Mol. Biol.* **212**, 787–817.

Skolnick, J. & Kolinski, A. (1990*b*). Simulations of the folding of a globular protein. *Science*, **250**, 1121–1125.

Skolnick, J. & Kolinski, A. (1990*c*). Dynamics of dense polymer systems: computer simulations and analytic theories. *Advan. Chem. Phys.* **78**, 223–278.

Skolnick, J., Kolinski, A. & Yaris, R. (1988). Monte Carlo simulations of the folding of β-barrel globular proteins. *Proc. Nat. Acad. Sci., U.S.A.* **85**, 5057–5061.

Skolnick, J., Kolinski, A. & Yaris, R. (1989*a*). Monte Carlo studies on equilibrium globular protein folding. II. β-Barrel globular protein models. *Biopolymers*, **28**, 1059–1095.

Skolnick, J., Kolinski, A. & Yaris, R. (1989*b*). Dynamic Monte Carlo study of a six stranded Greek key globular protein. *Proc. Nat. Acad. Sci., U.S.A.* **86**, 1229–1233.

Taketomi, H., Kano, F. & Gō, N. (1988). The effect of amino acid substitution on a protein-folding-unfolding transition studied by computer simulation. *Biopolymers*, **27**, 527–559.

Ueda, Y., Taketomi, H. & Gō, N. (1978). Studies on protein folding, unfolding and fluctuations by computer simulation. A. Three dimensional lattice model of lysozyme. *Biopolymers*, **17**, 1531–1548.

Wright, P. E., Dyson, H. J. & Lerner, R. H. (1988). Conformation of peptide fragments of proteins in a aqueous solution. Implications for initiation of protein folding. *Biochemistry*, **27**, 7167–7175.

Zielenkiewicz, P. & Rabczenko, A. (1988). The formation of protein secondary structure. Its connection with amino acid sequence. *Biophys. Chem.* **29**, 139–142.

Edited by P. Wright