



## Researchers Perform Monte Carlo Studies in Globular Protein Folding Research

### Introduction

The ability to predict tertiary structure of a globular protein, given the protein primary sequence, is an important unsolved problem of molecular biophysics.<sup>1</sup> The solution to the protein folding problem is not only of intrinsic interest, but could also be of practical importance by allowing for the rational design of novel proteins by genetic engineering, and for smaller proteins by solid state synthesis techniques.<sup>1</sup> Unfortunately, the elucidation of the factors responsible for folding globular proteins has remained elusive, and at present, there are a large number of unanswered questions. The relative importance of the various kinds of secondary

vs. tertiary interactions in determining the folding pathway and conformation of a native state is controversial.<sup>2</sup> Similarly, the level of detail required for tertiary structure prediction is unknown. Turning to the actual folding process, even less is known. What is the nature of early folding intermediates? Is there a multiplicity of folding pathways,<sup>3</sup> or just one?<sup>1</sup> With regard to the late folding events, why is it that in most cases studied so far, the transition state appears to be close to the native conformation? Why is the rate of folding less sensitive to the change of conditions than is the rate of unfolding?<sup>4</sup> Can one predict the effects of site-directed mutagenesis on the folding and unfolding kinetics? It is precisely the nature of the globular protein folding process which the research described below has attempted to address, with the ultimate goal being protein tertiary structure prediction.

In the following, the development of a series of idealized models designed to elucidate the general features of the equilibrium and dynamic aspects of globular protein folding is described.<sup>5-13</sup> Our objective is to develop schematic models in which the minimum set of interactions required to fold a real protein topology is identified, and complications are only introduced when a simpler realization fails to produce an essential feature of the physics. Most important to the validity of this approach is the requirement that the native state cannot be specified in advance. Interactions between any pair of spatially close residues (and not necessarily those in the native state) must be permitted. Starting from a high temperature denatured state, (that is generated at random) the model must reproducibly fold to and unfold from a unique native

state whose topology is found in real proteins. For small proteins, the thermodynamics of the conformational transition that emerges from the simulation (*and which is not built in*) must be consistent with a two state model.<sup>1</sup> In short, the model should possess all the general features of the conformational transition of real globular proteins.

### Overview of the Model

The model consists of a consecutive sequence of  $n$  beads confined to a tetrahedral lattice. A lattice representation is employed because the Monte Carlo (MC) moves can be performed using integer arithmetic. This provides a substantial speed-up in the calculations. Each bead represents an entire amino acid residue. Thus, an  $\alpha$ -carbon representation of a globular protein is used in which the virtual bond angles have been slightly distorted to fit onto a diamond lattice. A  $\beta$ -sheet is represented as a sequence of coplanar trans ( $t$ ) states, and a right-handed  $\alpha$ -helix is represented as a sequence of out-of-plane, gauche minus ( $g^-$ ) states. There is one additional rotational state, the other out of plane, gauche plus ( $g^+$ ) state.

To account for tertiary interactions, the model primary sequence includes hydrophobic and hydrophilic residues. Non-bonded, nearest neighbor hydrophilic-hydrophobic pairs of residues interact with an attractive potential of mean force, hydrophilic-hydrophobic pairs interact with a repulsive potential of mean force, and hydrophilic-hydrophobic pairs can be indifferent, marginally attractive or repulsive,<sup>14</sup> with no change in qualitative behavior.<sup>10</sup>

*Continued on page 15*

## Index

- 1 Skolnick, Kolinski and Sikorski employ models to explore dynamic features of globular protein folding
- 2 **Musings:** Graphical User Interfaces, Part II: Motif
- 5 **New Briefs**
- 6 **Previews**
- 7 Bartlett talks about "Chemistry with FPS Computers"
- 11 **Meetings Roundup:** Bentley and Salahub report on CAMD aspects of PACIFICHEM
- 12 **New Products**
- 21 **Meetings**
- 24 **Professional Opportunities**
- 27 **Subscription Form**

---

## Globular Protein Research

Continued from page 1

Turns are modelled as weakly hydrophilic residues. On a diamond lattice, a turn consists of three residues in a  $g\pm g\pm g\pm$  conformation. Two kinds of turns have been considered, thus far. In the "turn neutral" case, based on short-range interactions, the  $t$  and  $g$  states are isoenergetic; thus, there are 27 equally likely states. Native turns are favored only by tertiary interactions. In the "strong turn" case,  $g+$  and  $g-$  states are isoenergetic, and based on short-range interactions, a native turn is but one of eight equally likely conformations. Finally, we emphasize that turns in particular, and the native conformation in general, have to be found and are not enforced by any *a priori* restrictions in the algorithm.

## Folding of $\beta$ -Barrel Globular Protein Models

While early work focused on the folding of  $\beta$ -hairpins and four member  $\beta$ -barrels,<sup>5,8</sup> our first attempt to reproduce a native-like  $\beta$ -protein topology<sup>15</sup> focused on an analog of plastocyanin.<sup>9</sup> We examined the conformational transition from the denatured state 1, to the unique six-member, Greek-key structural motif,<sup>15</sup> side and top views are shown in 2a, and 2b of Figure 1 (page 16). This model has the same basic topology as the real plastocyanin, except that, in the interest of computational economy, the end- $\beta$ -strands<sup>15</sup> not involved in the Greek-key, have been clipped-off (more recent work shows that including these extra strands provides no new qualitative insights). The native structure contains six anti-parallel  $\beta$ -strands joined by four tight bends and one long loop. The numbering se-

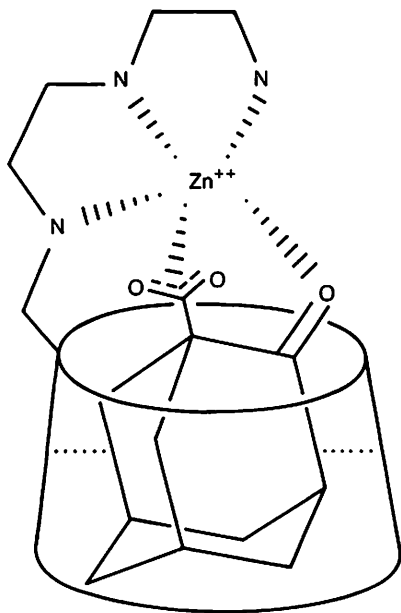
quence of the strands is indicated in 2b of Figure 1, page 16.

In the first case considered, regions that might form a  $\beta$ -stretch have an alternating odd/even hydrophilic-hydrophobic pattern of residues in the primary sequence; the putative tight turn residues are turn neutral, and for the long loop *sans* long-range interactions, the native conformation is one of 16 equally weighted, lowest energy, conformations. The equilibrium conformational transition that emerges is well approximated by a two-state model. However, there is a marginally populated four-member  $\beta$ -barrel intermediate formed from  $\beta$ -strands 2-5, perhaps with  $\beta$ -strand 1 attached. Thus for the first time ever, a complicated native-like topology has been uniquely obtained on folding from the denatured state.

Continued on page 16

---

# Already?



expansion, the Model 500EA grows in incremental additions that match your applications needs.

FPS Computing has been the solutions choice of chemical and biochemical researchers for over ten years. With convenient networking options and a complete UNIX environment, the Model 500EA supports the major codes: Amber, MOPAC, SPARTAN, ACES, Discover, Gaussian, and more. For rapid visualization of the results, FPS Computing offers graphical interfaces to the most popular graphics devices, including the FPS Computing Model 350 and Model 400 Visualization Supercomputers.

In chemistry, FPS Computing is the right choice—for today and tomorrow.

**FPS**  
COMPUTING

## Globular Protein Research

Continued from page 15

The thermal stability can be enhanced by increasing the preference of the residues in the putative tight turn regions for gauche states ( $g+$  and  $g-$  are isoenergetic). Interestingly, we found that the uniqueness of the conformational transition (*i.e.*, folding to one particular native state) can be destroyed by making the loop, based on short-range interactions, totally indifferent to the native state conformation. While, at very low temperature, the unique six-member  $\beta$ -barrel is still observed, in the transition region not only is the

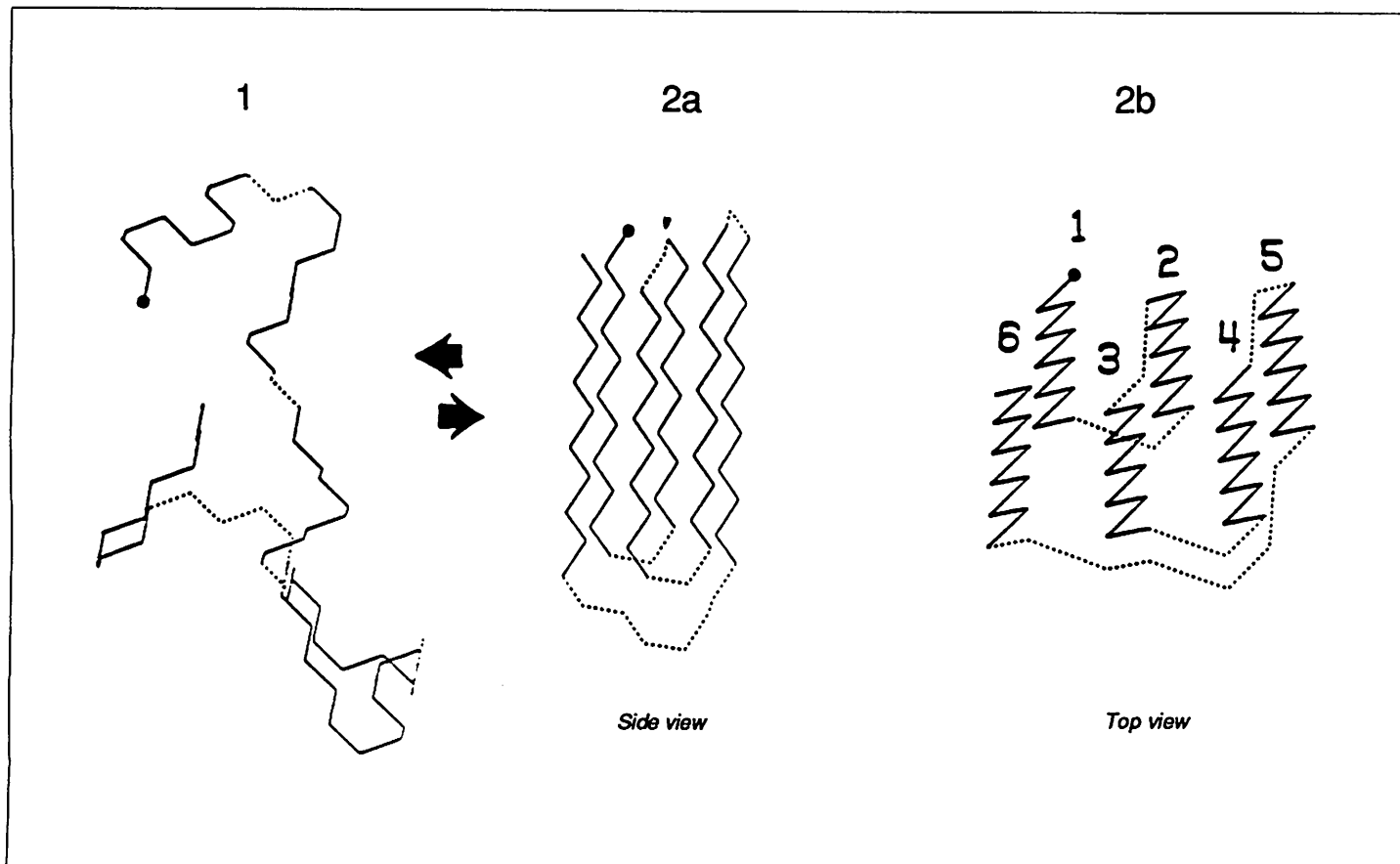
four-member  $\beta$ -barrel intermediate substantially populated, but so are out-of-register conformations of strand 6. Owing to loop entropy, the loop literally tugs at  $\beta$ -strand 6, pulling it into out-of-register conformations. Thus, while their secondary structures are not as regular as  $\alpha$ -helices or  $\beta$ -sheets, loops may play an important role in determining the conformational stability of proteins.

While the previous work employed a series of MC moves that could distort the time scale because both local and longer wavelength moves are employed, recently, a series of dynamic MC simulations<sup>12</sup> was performed in which only local moves that reproduce the correct local and global dynamics for the denatured state, in the absence of hydrodynamic interactions, are allowed.<sup>16</sup> It was hoped that such a simulation could provide some insight into the protein folding pathways. A representative folding trajectory for the

Greek-key is shown in Figure 2. Folding initiates at the turn between strands 2-3, followed by the fairly rapid zipping-up of a  $\beta$ -hairpin. Of course, the nascent structure might dissolve; there is constant competition between structure dissolution and formation during the entire assembly process. All of the  $\beta$ -turns between strands 2-3, 3-4 and 4-5, occur as folding initiation sites, but the turn between strands 2-3 is somewhat more preferred because, owing to the excluded volume effect, closed loops are more likely the closer one gets to the chain end.<sup>12</sup> Following folding initiation, the remaining strands of a four-member,  $\beta$ -barrel (strands 2-5) zip-up *via* an "on site" construction mechanism; that is, the existing  $\beta$ -strand tertiary structure acts as scaffolding onto which subsequent  $\beta$ -strands are built. Assembly of the four-member  $\beta$ -barrel intermediate is a relatively rapid process, typically taking about

Continued on page 17

**Figure 1.** Representative denatured state conformation (1) in equilibrium with the native, Greek-key  $\beta$ -barrel shown in a side (2a) and top view (2b).



## Globular Protein Research

Continued from page 16

of the total folding time. The folding process is then punctuated by a very long pause, as the remaining random coil tail containing the long loop thrashes about while it hunts for the narrow entropic pass to the native state. Now, the partially assembled structure can hinder assembly. For example, the tail might be on the same side as

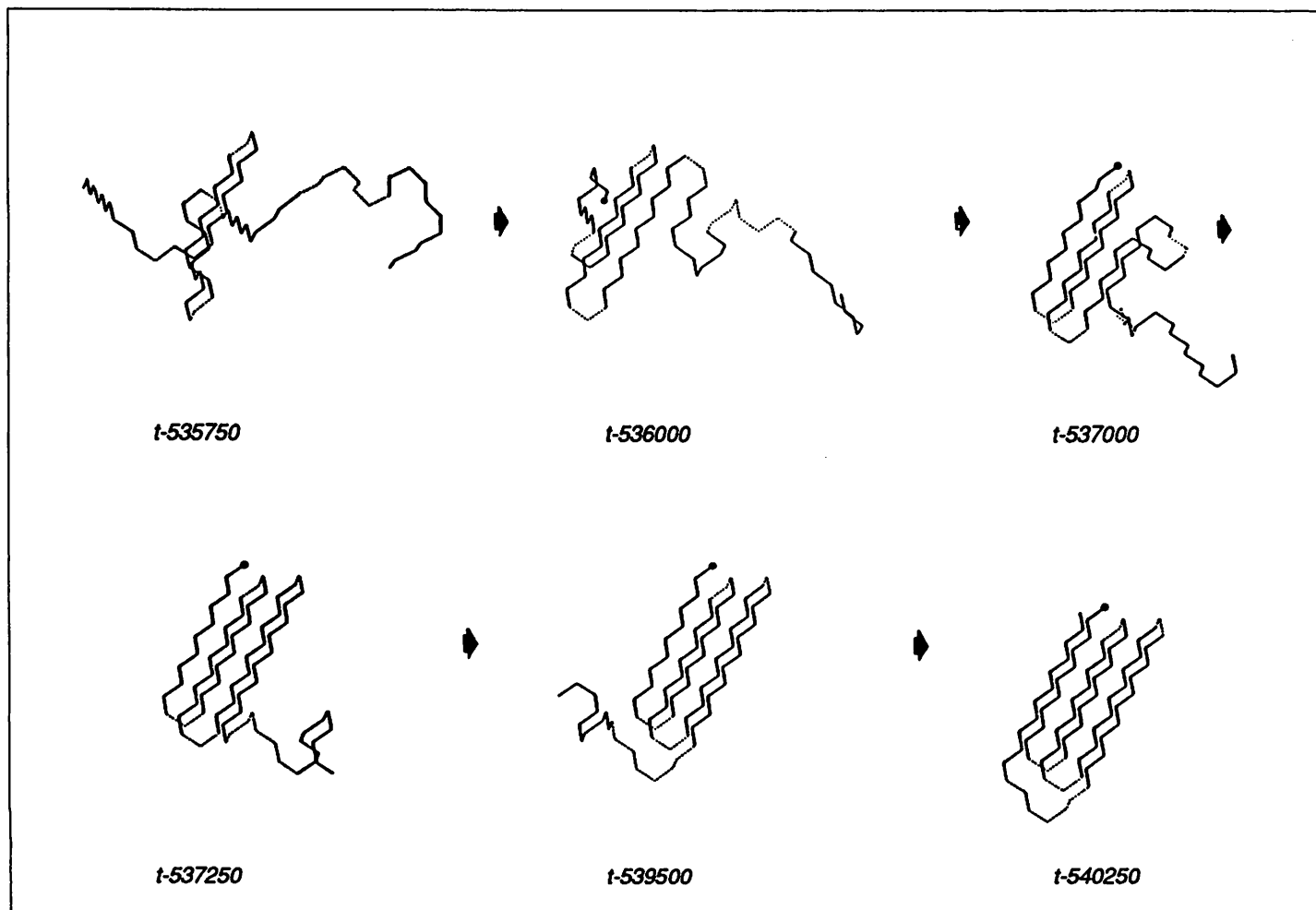
**Figure 2.** Representative folding trajectory of the Greek-key. The times indicated in the Figure are from the start of the simulation run. In every time step, on average, each bead is subject to all possible local micromodifications of the chain.

strands 1-2, and it must rattle about until it goes under or around the assembled portions of the protein. Eventually, the loop and strand 6 find themselves near strands 1 and 3, after which assembly is very rapid. Unfolding proceeds essentially in reverse.<sup>12</sup> Interestingly, this is the same pathway observed with the longer wavelength moves included, and more importantly, in a new simulation on a 24 nearest lattice model which contains sidechains.<sup>17</sup> That is, it appears that the folding pathway is universal, independent of both lattice representation and the choice of local elemental moves.

For the model  $\beta$ -proteins considered to date, although diffusion-collision adhesion assembly<sup>18</sup> involving preformed  $\beta$ -strands could in principle contribute

to the folding process, it has not been observed to do so. It is simply faster, and more probable to initiate at a turn and zip-up on-site, by pulling the pieces of randomly coiled chain from the denatured state, that to wait for the assembled, marginally stable, isolated  $\beta$ -strands to diffuse together and form a  $\beta$ -hairpin. Moreover, the folding pathways are defined in a general sense, with the degree of definition enhanced the closer one is to the native state. Folding can initiate at all the turns, but the attachment of the long loop is always the rate determining step for Greek-key assembly. While a coarse characterization of the folding pathway exists, there are a multiplicity of spatial trajectories consistent with this description.

Continued on page 18



## Globular Protein Research

Continued from page 17

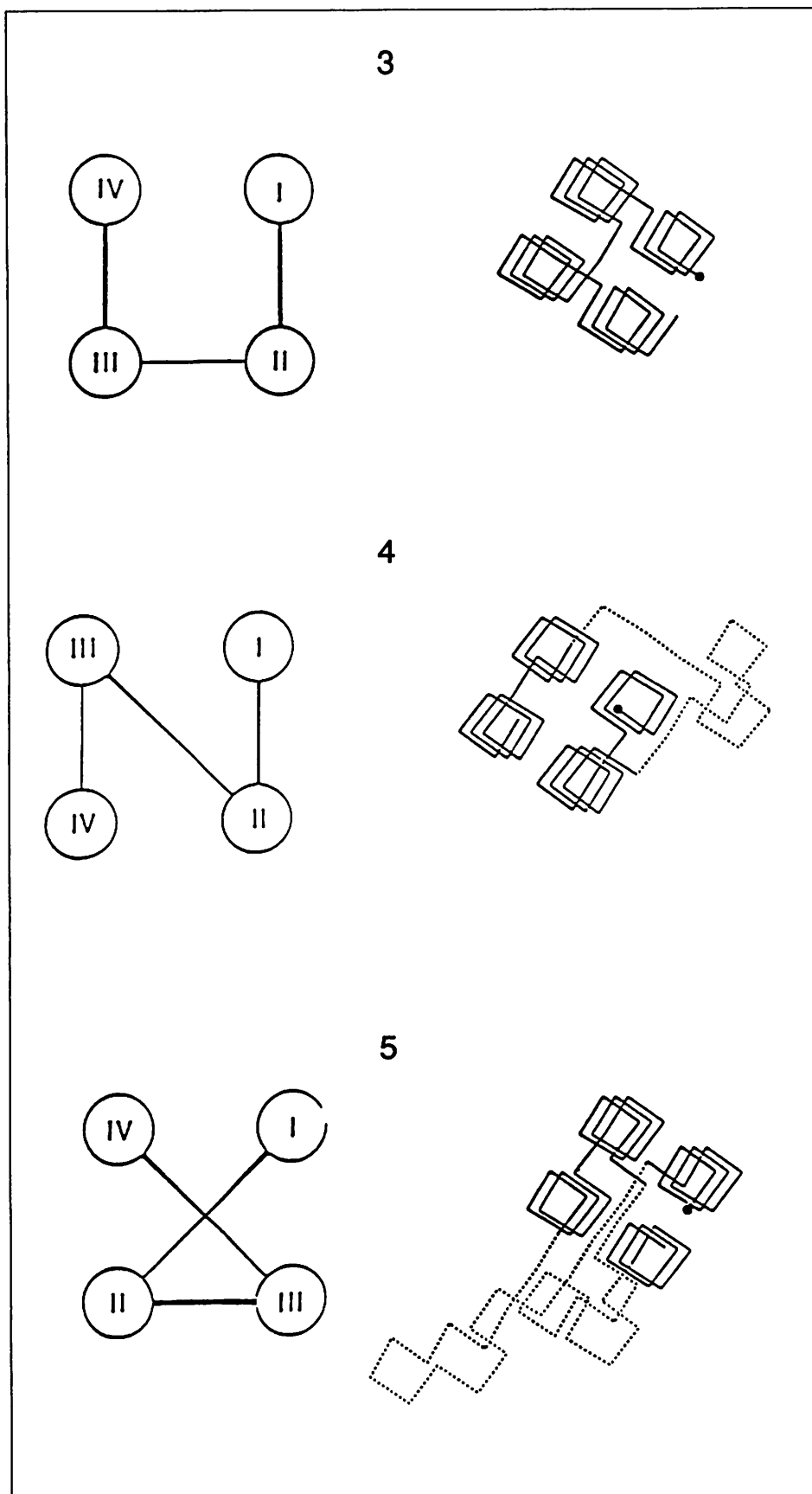
### $\alpha$ -Helical Protein Model

Next, we examined the equilibrium conformational transition between the denatured state and  $\alpha$ -helical folding motifs.<sup>10,11</sup> Helical wheel interactions are introduced to marginally stabilize  $\alpha$ -helical secondary structures, and the hydrophobic/hydrophilic primary sequence pattern is designed to distinguish exterior and interior surfaces of helices, *i.e.*, an amphipathic sequence is assumed. Initially, the folding of a left-handed, antiparallel four-helix bundle, having three tight turns, as shown in structure 3 of Figure 3, was investigated.<sup>11</sup> This is a crude model of the native conformation of cytochrome c' and myohemerythrin<sup>15</sup> and bears a close analogy to the synthetic four-helix bundle synthesized by Regan and DeGrado.<sup>19</sup> The sufficient conditions for an all-or-none collapse to the unique native structure are found to be an amphipathic pattern of hydrophilic-hydrophobic residues along with a central turn neutral region. However, unlike the  $\beta$ -barrel case, the external turns need not be turn neutral; this is because in the models considered, the stability of a helical hairpin is greater than the  $\beta$ -hairpin, and thus, out-of-register configurations of the four-helix bundle are unimportant.

We also folded an antiparallel four-helix bundle with one long loop between two of the helices (4 of Figure 3) which is the topology of apoferritin,<sup>20</sup> and four helix bundle with two long loops, which

Continued on page 19

**Figure 3.** Representative native state topologies (left-hand side of Figure) and actual folded native state conformations (right-hand side of Figure) for a four-helix bundle having three tight bends (3), one long loop between two of the helices (4) and two long loops between two pairs of helices (5).



## Globular Protein Research

Continued from page 18

has a topology close to that of methionyl porcine somatotropin<sup>21</sup> (5 of Figure 3). The long loops in each case are taken to be hydrophilic and based on short-range interactions, can *locally favor helix formation*. In addition, the tight turns need not be specified at all in the primary sequence. They can be part of an amphipathic amino acid pattern that locally favors helices, and the molecule

adopts a turn conformation there to minimize the global free energy. Thus, a given secondary structure is not automatically built into these models. The properties of the folding/unfolding transitions are qualitatively the same as for the topology with tight bends.

We, next, summarize the results on the folding/unfolding pathways of these  $\alpha$ -helical protein models. We began by investigating the folding of an  $\alpha$ -helical hairpin where the possibility of diffusion-collision-adhesion assembly<sup>18</sup> of pre-formed helices is explicitly implemented.<sup>13</sup> Again, as in the  $\beta$ -protein case, this mechanism doesn't successfully complete with the on-site mechanism. Folding typically initiates at or near the  $\alpha$ -helical turn, and then the

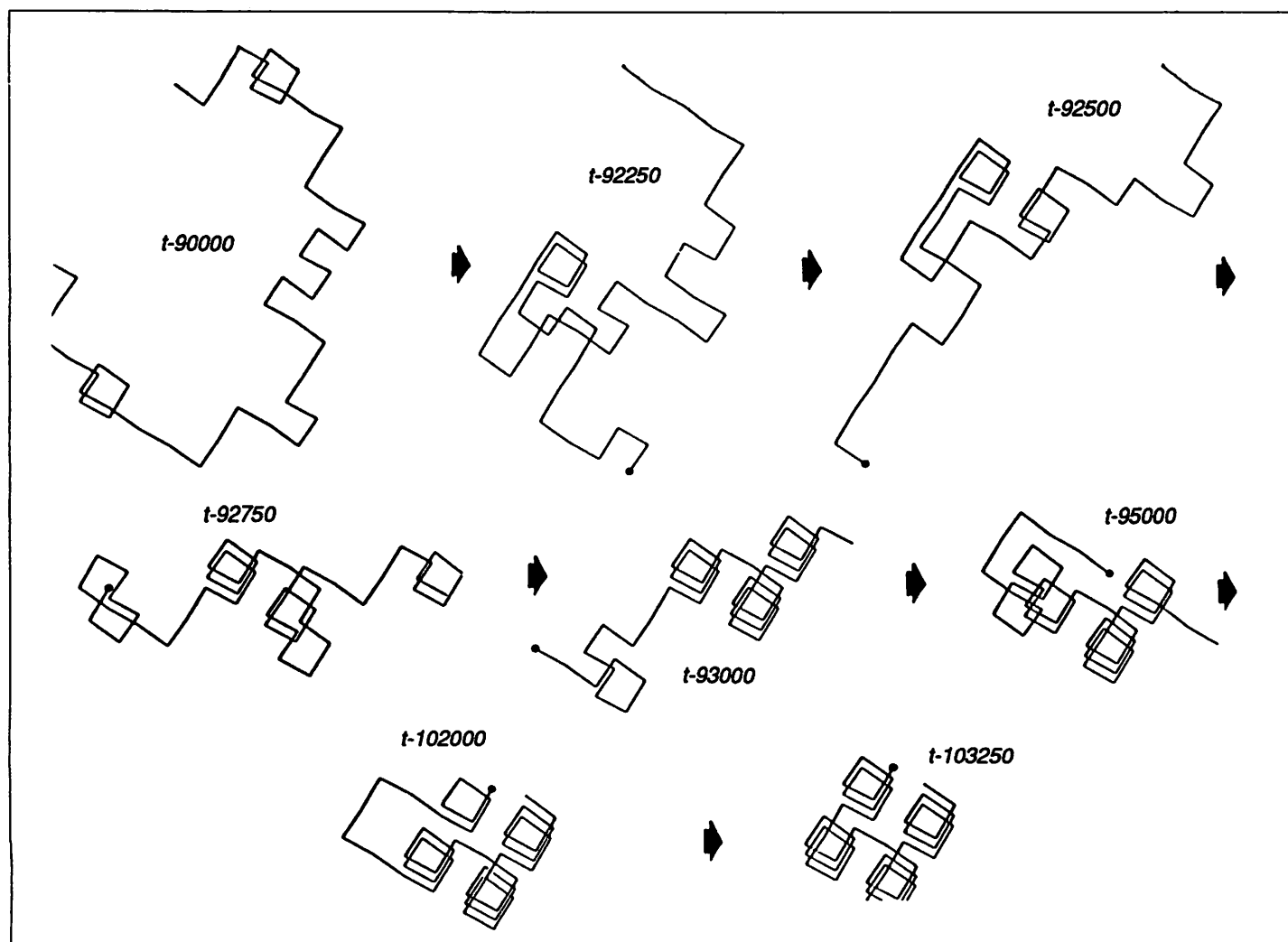
remaining portion of both helices zip-up on-site. Another possibility is single helix formation followed by subsequent on-site construction of the second helix. Turning to the four-helix bundle motif, Figure 4 shows a representative folding pathway of the four-helix bundle with tight bends. A punctuated on-site mechanism of folding is, once again, observed. Similar behavior is seen for the four-helix bundles with long loops and is described elsewhere.<sup>13</sup>

### Summary

In this brief review, highlights of results from a highly idealized series of models of globular proteins that reproduce all the qualitative aspects of the equi-

Continued on page 20

**Figure 4.** Representative folding trajectory of a four-helix bundle with three tight bends. The time indicated is from the start of the simulation run.



## Globular Protein Research

Continued from page 19

librium and dynamic features of the globular protein folding process have been presented. The fact that these models work as well as they do, indicates that the general rules of protein folding are rather robust, *i.e.*, while the local structure is dependent on the intimate details of the free energy surface, the approximate backbone folding topology, the thermodynamics and folding kinetics can emerge from a simplified model. The basic features that appear to dictate the overall folding topology are a general pattern of hydrophobic/hydrophilic residues, plus the presence of residues that are, at the very least, neutral toward turn formation at essential regions in the amino acid sequence. With respect to the pathways of assembly, for  $\alpha$ -helical,  $\beta$ -motifs and mixed  $\alpha/\beta$  motif proteins, the simulations indicate that assembly proceeds by an on-site framework construction mechanism where folding initiates at or near a native turn, followed by the sequential zipping-up of the secondary structure on-site. In the early folding stages, the already assembled secondary structure serves as scaffolding onto which subsequent secondary structure is built. In the later folding stages, the existing structure can hinder assembly, as the remaining unassembled parts of the chain thrash about trying to find the native state. Overall, there are a number of well-defined, but non-unique folding pathways. The pathways are independent of lattice description and level of detail (side chains vs. no sidechains). Recently, we have developed a more versatile lattice description including sidechains that allows for mixed  $\alpha/\beta$  proteins,  $\beta$ -sheet twists, realistic  $\beta$ -turn, etc.<sup>17</sup> Work is now in progress on these more sophisticated models. Helical hairpins, more realistic, four-helix bundles, plastocyanin models

and mixed  $\alpha/\beta$  motif proteins have already been folded, and the new model appears to be very promising. We have also developed algorithms that pull these chains off lattice and provide a full atom representation. Work on these latter stages of native conformation refinement is now in progress.

By Jeffrey Skolnick, Ph.D., Department of Molecular Biology, Scripps Clinic and Research Foundation, La Jolla, CA 92037 (619-554-4821) and Andrzej Kolinski, Ph.D. and Andrzej Sikorski, Ph.D., Department of Chemistry, University of Warsaw, Pasteura 1 02-093, Warsaw, Poland.

### Acknowledgement:

This research was supported in part by Grant No. GM-37408 of the National Institutes of Health. The authors have greatly benefitted from discussions with Drs. David Case, Alfred Holtzer, Lee Walters, Jane Richardson, David Richardson and Peter Wright.

### References

1. Creighton, T. E. (1985) *J. Phys. Chem.* **89**, 2452-2459.
2. Lewis, P. N.; Go, N.; Kotelchuck, D.; and Scheraga, H. A. (1970) *Proc. Natl. Acad. Sci., USA* **67**, 810-815.
3. Harrison, S. C. and Durbin, R. (1985) *Proc. Natl. Acad. Sci., USA* **82**, 4028-4030.
4. Tsong, T. Y. and Baldwin, R. L. (1978) *Biopolymers* **17**, 1669-1678.
5. Kolinski, A.; Skolnick, J. and Yaris, R. (1987) *Biopolymers* **26**, 937-962.
6. Kolinski, A. Skolnick, J. and Yaris, R. (1986) *Proc. Natl. Acad. Sci., USA* **83**, 7267-7271.
7. Skolnick, J.; Kolinski, A. and Yaris, R. (1988) *Proc. Natl. Acad. Sci., USA* **85**, 5057-5061.
8. Skolnick, J.; Kolinski, A. and Yaris, R. (1989) *Biopolymers* **28**, 1059-1095.
9. Skolnick, J.; Kolinski, A. and Yaris, R. (1989) *Proc. Natl. Acad. Sci., USA* **86**, 1229-1233.
10. Sikorski, A. and Skolnick, J. (1989) *J. Proc. Natl. Acad. Sci., USA* **86**, 2668-2672.
11. Sikorski, A. and Skolnick, J. (1989) *Biopolymers* **28**, 1097-1113.
12. Skolnick, J. and Kolinski, A. (1989) *J. Mol. Biol.*, in press.
13. Sikorski, A. & Skolnick, J. (1989) *J. Mol. Biol.*, in press.
14. Miyazawa, S. and Jernigan, R. L. (1985) *Macromolecules* **18**, 534-552.
15. Richardson, J. S. (1981) *Adv. Prot. Chem.* **34**, 167-339.
16. Baumgartner, A. (1984) *Ann. Rev. Phys. Chem.* **35**, 419-435.
17. Skolnick, J. and Kolinski, A. (1989) *J. Mol. Biol.*, submitted.
18. Karplus, M. and Weaver, D. L. (1976) *Nature* **260**, 404-406.
19. Regan, L. and DeGrado, W. F. (1988) *Science* **24**, 976-978.
20. Ford, G. C.; Harrison, P. M.; Rice, D. W.; Smith, J. M. A.; Treffry, A.; White, J. L. and Yariv, J. (1984) *J. Philos. Trans. R. Soc. Lond.* **B304**, 551-565.
21. Abdel-Meguid, S.; Shieh, H. S.; Smith, W. W.; Dayringer, H. E.; Violand, B. N. and Bentle, L. A. (1987) *Proc. Natl. Acad. Sci., USA* **84**, 6434-6437. ■