

# De Novo Simulations of the Folding of GCN4 and Its Mutants

J. SKOLNICK,<sup>a\*</sup> M. VIETH,<sup>a</sup> A. KOLINSKI,<sup>a,b</sup> and C. L. BROOKS, III<sup>c</sup>

<sup>a</sup>Departments of Molecular Biology and Chemistry, The Scripps Research Institute, 10666 N Torrey Pines RD, La Jolla, CA 92037, U.S.A.

<sup>b</sup>Department of Chemistry, University of Warsaw, 02-093 Warsaw, Poland

<sup>c</sup>Department of Chemistry, Carnegie Mellon University, Pittsburgh, PA 15213, U.S.A.

**Abstract.** A hierarchical approach to protein folding is employed to examine the folding pathway and predict the quaternary structure of the GCN4 leucine zipper. Structures comparable in quality to experiment have been predicted. In addition, the equilibrium between dimers, trimers and tetramers of a number of GCN4 mutants has been examined. In most of the cases, the simulation results are in accord with the experimental studies of Harbury *et al.* [1].

**Key words.** Protein folding/GCN4, leucine zipper/computer, simulations/lattice, protein models.

## Introduction

One of the most important unsolved problems in molecular biology is the prediction of protein structure from amino acid sequence [2–4]. As a step in this direction, we have examined the folding pathway and predicted quaternary structure of the GCN4 leucine zipper [5]. In addition, the equilibrium between various GCN4 mutants [1] has been explored. The method uses a hierarchical approach to protein folding [6,7]. The basic idea is to employ a reduced model to simulate the early stages of folding and to obtain the native state topology. Restrained molecular dynamics are then used to produce full atom models [5,7]. An overview of the entire procedure is presented in Figure 1.

Folding of the model chains commences on a very high coordination lattice, from a pair of random unfolded chains. Following dimer assembly to a parallel, left handed in register coiled coil, the resulting structures are refined on the lattice to produce a family of native structures whose rms from native [8] ranges from 2.3 to 3.7 Å [5]. Then, full atom models are built, the structures are solvated and refined using molecular dynamics with the CHARMM potential [9]. The resulting family of structures is indistinguishable from the native one when subjected to the molecular dynamics refinement protocol. The average structure from this family has a backbone heavy atom rms of 0.8 Å, all heavy atoms in the dimerization interface differ by 1.31 Å rms from native, and all heavy atoms differ from the crystal structure by 2.29 Å rms. These studies marked the first time that protein quaternary structures of this quality have been obtained from random unfolded conformations.

In a recent paper in *Science* by Harbury *et al.* [1], they examined the shift in equilibrium between parallel, coiled coil dimers, trimers and tetramers associated

★ Author for correspondence.

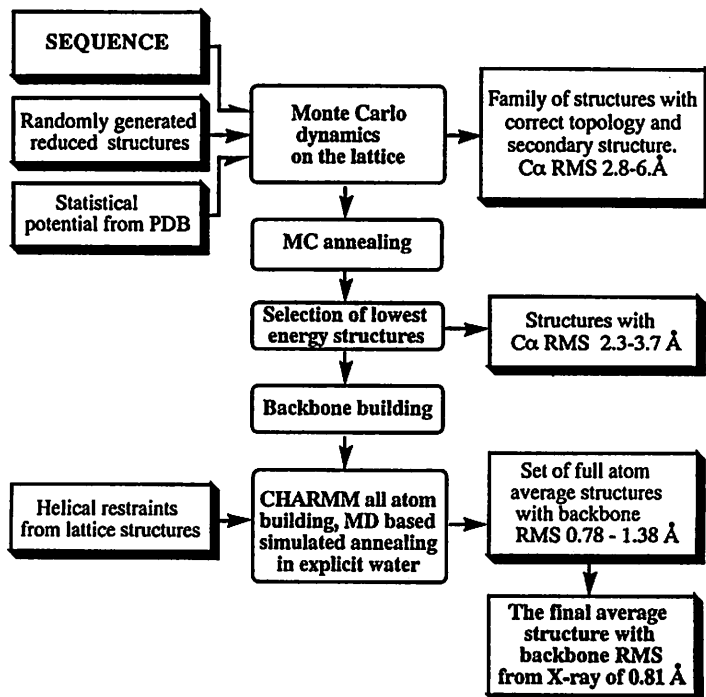


Fig. 1. An overview of the hierarchical approach. The left hand column displays the input information, the central column displays the protocol, and the right-hand column shows the resulting output.

with varying the identity of the residues of the two hydrophobic residues in the coiled coil heptad repeat. To further test the validity of the model, we prepared a series of dimers, trimers and tetramers for each of the eight species that have been experimentally studied. In order to compare with experiment, we have to calculate the equilibrium constants associated with the dimer,  $D$  trimer,  $T$ , and tetramer,  $R$ , species.



with  $\{L\}$ , the number of chains in species  $L = D, T, R$ . The equilibrium constants are

$$K_{DT} = \{T\}^2 / \{D\}^3 = VZ_T^2 / Z_D^3 \quad (2a)$$

$$K_{DR} = \{R\}^3 / \{D\}^2 = VZ_R / Z_D^2 \quad (2b)$$

with  $V$  the total accessible volume of the system. Finally, the total concentration of chains,  $C$ , independent of their degree of association is

$$C = 2\{D\} + 3\{T\} + 4\{R\}. \quad (3)$$

Since Equations 2a, 2b, and 3 comprise three equations in three unknowns

Table I.  
Comparison or simulation with experiment on GCN4 mutants

Residues at positions		Experimental results	Simulation results	
a	d	Dominant species	Dominant species	Concentration dependence
Wild	Type	2	2	99:1:0 *low 94:6:0 high
I	L	2	3 !	13:87:0 low 4:95:1 high
I	I	3	3	0:100:0 low 0:100:0 high
L	I	4	4	0:4:96 low 0:0:100 high
V	I	?	3	0:100:0 low 0:100:0 high
L	V	3	3	1:99:0 low 0.2:99.8:0 high
V	L	(2,3)	3 !	2:98:0 low 0.1:99.9:0 high
L	L	3	3	0.1:99.9:0 low 0:100:0 high

Low and high indicate 2  $\mu\text{M}$  and 200  $\mu\text{M}$  concentration of a peptide respectively. ! indicates those results which apparently disagree with experiment. In the VI case, the degree of association could not be assigned experimentally.

( $\{D\}$ ,  $\{T\}$ ,  $\{R\}$ ), these equations can be solved numerically. The key to the analysis is to employ a method based on Mayer and Mayer [10] which requires that in each species,  $\gamma$  the system be in a local energy minimum, with energy  $E_{\text{avg}}$ . Then, the internal partition function is given by

$$Z_{\text{int},\gamma} = V_{\text{exp}}(-E_{\text{avg}}/kT) \left\{ \frac{\prod_N dv_i}{P(R=0)} \right\}^{1-1/N} \quad (4)$$

with  $dv_i$  the infinitesimal volume element centered about the origin which corresponds to the minimum energy conformation,  $N$  is the number of groups in the species  $\gamma$ , and  $P(R=0)$  is the probability that all the  $N$  groups in the molecule are in the minimum energy state. In our case, the groups correspond to the alpha carbons and side chains' center of mass positions.

The results of our simulations are shown in Table I, where rather good agreement with experiment is demonstrated. We found that antiparallel species do not contribute at all. Finally, to verify that our results are not limited to the GCN4 and its mutants, we examined the equilibrium degree of association of a sequence designed by DeGrado, Eisenberg and coworkers, coil-Ser [11]. In agreement with experiment, we predict that the dominant species is an antiparallel, three helix bundle.

## Conclusions

These simulations on the folding of coiled coils demonstrate that it is possible to predict quaternary structure of simple folding motifs from amino acid sequence alone. Furthermore, the ability to predict the dominant species provides confidence that the phenomenological potentials have captured some aspects of the experimental situation. Future work will entail extension to larger coiled coils and extension of the model to predict the quaternary structure of some small globular proteins.

## Acknowledgment

This research was supported in part by NIH Grant No. GM48835 and NIH Fogarty International Center Grant No. 1 R03 TW00418.

## References

1. P. B. Harbury, T. Zhang, P. S. Kim, and T. Alber: *Science* **262**, 1401 (1993).
2. T. E. Creighton: *Biochem. J.* **270**, 131 (1990).
3. M. Levitt: *Current Opinion in Structural Biology* **1**, (1991).
4. O. B. Ptitsyn: *Journal of Protein Chemistry* **6**, 273 (1987).
5. M. Vieth, A. Kolinski, C. L. Brooks, III and J. Skolnick: *J. Mol. Biol.* **237**, 361 (1994).
6. A. Kolinski, A. Godzik and J. Skolnick: *J. Chem. Phys.* **98**, 7420 (1993).
7. J. Skolnick, A. Kolinski, C. L. Brooks and A. Godzik: *Current Biol.* **3**, 414 (1993).
8. E. K. O'Shea, J. D. Klemm, P. S. Kim, and T. Alber: *Science* **254**, 539 (1991).
9. B. R. Brooks, R. Brucoleri, B. Olafson, D. States, S. Swaminathan, and M. Karplus: *J. Comp. Chem.* **4**, 187 (1983).
10. J. E. Mayer and M. G. Mayer: *Statistical Mechanics*, New York: John Wiley & Sons, Inc. (1963).
11. B. Lovejoy, C. Seunghyan, D. Cascio, D. K. McRorie, W. F. DeGrado, and D. Eisenberg: *Science* **259**, 1288 (1993).