DIMACS Series in Discrete Mathematics
and Theoretical Computer Science
Volume 23, 1996

A hierarchical approach to the prediction of the quaternary structure of GCN4 and its mutants

Michal Vieth, Andrzej Kolinski, C.L. Brooks, III, and Jeffrey Skolnick*

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 five out of eight cases, the simulation results are in accordance with the experimental studies of Harbury, et al [1].

1. Introduction

An outstanding unsolved problem in molecular biology is the prediction of protein structure from amino acid sequence. As a step in this direction, we have examined the folding pathway and the native structure GCN4 leucine zipper. In addition, the equilibrium between various GCN4 mutants has been explored. The basic idea is to employ a reduced model to stimulate the early stages of folding and to obtain the native state topology. Restrained molecular dynamics are then used to produce full atom models [2]. We have also developed a general formalism to estimate the equilibrium between various multimeric species. An overview of the procedure to predict the native conformation of a protein from sequence is presented in Figure 1 [3, 4]. 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 ranges from 2.3 to 3.7 Å. Subsequently, full atom models are built, the structures are solvated and refined, using molecular dynamics with the CHARMM potential [5]. The resulting family of structures is indistinguishable from the native one when subjected to the molecular dyanmics refinement protocol. The average structure from this family has a backbone heavy atom rms 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 were the first time that protein quaternary structures of this quality have been obtained from random

1991 *Mathematics Subject Classification.* Primary 65K10, 90C30; Secondary 92C05, 92E10.
Key words and phrases. Protein folding, GCN4 Leucine zipper, Computer simulations, Lattice protein models.

TABLE 1. Comparison of simulation results with experiment on GCN4 mutants. *indicates those results which apparently disagree with experiment. In the VI case, the degree of association could not be assigned experimentally.

	Experiment	Simulation
Residues at positions a d	Dominant species	Dominant species
wild type	2	2
I L	2	2
I I	3	3
L I	4	3,4!
V I	?	3!
L V	3	3
V L	(2,3)	2!
L L	3	3

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 with various residue identities of the two hydrophobic residues in the coiled 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. To compare with the experiment, we have to calculate the equilibrium constants associated with the dimer, D, trimer, T, and tetramer, R, species

$$3D \leftrightarrow 2T \tag{1}$$

$$2D \leftrightarrow R \tag{2}$$

with (L), the number of chains in species $L = D, T, R$. The equilibrium constants are

$$K_{DT} = \frac{(T)^2}{(D)^3} = \frac{VZ_T^2}{Z_D^3} \tag{3}$$

$$K_{DR} = \frac{(R)}{(D)^2} = \frac{VZ_R}{Z_D^2} \tag{4}$$

with V the total accessible volume of the system. The key to the calculation of the partition function is to use a generalization of a rotational isomeric approach originally applied to 24 neighbor lattice by Skolnick and Kolinski [6] to the present lattice. With the neglect of excluded volume between residues that are more than 4 neighbors down the chain (a fairly minor approximation in the denatured state of proteins), we can calculate the configurational partition function using a transfer matrix approach that is logically equivalent to Zimm-Bragg helix coil theory [7]. For the native molecule, we approximate the partition function by the Boltzmann factor of the energy of the most probable conformation times the product of lattice accessible lattice states. The latter are decomposed into the product of the probabilities of accessible three bond states. (Sampling problems preclude the inclusion of higher order correlations.) The comparison with experiment is shown in Table 1.

2. 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 some confidence that the phenomenological potentials have captured aspects of the experimental situation. Future work will entail extension to larger coiled coils and extension of the model to predict the quarternary structure of some small globular proteins.

References

[1] P. Harbury, T. Zhang, P. Kim S., T. Alber, *Science*, **262**, 1401–1407 (1993).
[2] M. Vieth, A. Kolinski, C.L. Brooks III, J. Skolnick, *J. Mol. Biol.*, (submitted), (1995).
[3] A. Kolinski, J. Skolnick, *Proteins*, **18**, 338–352 (1994).
[4] M. Vieth, A. Kolinski, C.L. Brooks III, J. Skolnick, *J. Mol. Biol.*, **237**, 361–367 (1994).
[5] B.R. Brooks et al., *J. Comp. Chem.*, **4**, 187–217 (1983).
[6] J. Skolnick, A. Kolinski, *J. Mol. Biol.*, **221**, 499–531 (1991).
[7] D. Poland, H.A. Scheraga, *Theory of helix-coil transitions in biopolymers*, (Academic Press, New York, 1970).

THE SCRIPPS RESEARCH INSTITUTE, 10666 NORTH TORREY PINES ROAD, LA JOLLA, CA 92037

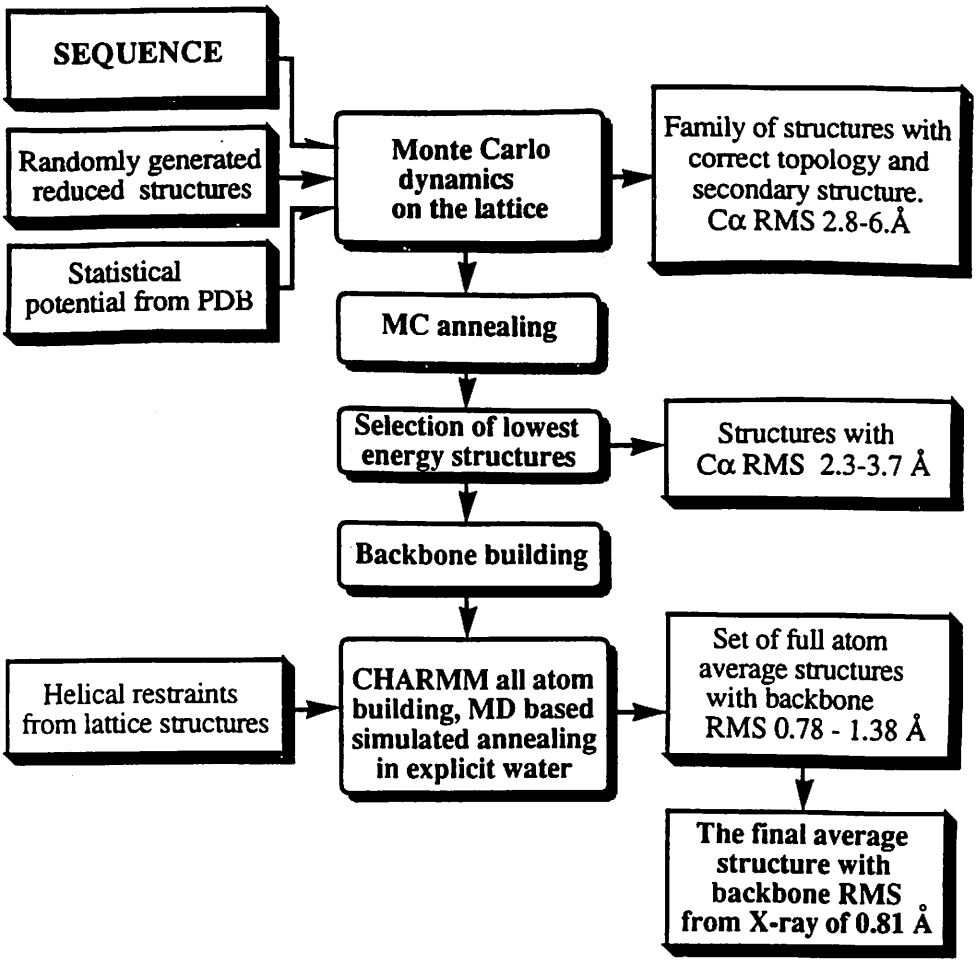

FIGURE 1. An overview of the hierarchical approach. The left hand column displays the input information, the central column displays the protocol, and the right-hand column shows the resulting output.