Molecular Dynamics Simulation of Protein Folding with Supersecondary Structure Constraints

Zhi Rong Sun,1,3 Yan Cui,2 Lun Jiang Ling,2 Qing Guo,1 and Run Sheng Chen2,3

Received May 7, 1998

We integrate molecular dynamics simulation methods with a newly developed supersecondary structure prediction method and compute the structure of a protein molecule, crambin. The computed structure is similar to the crystal structure with an rms error of 3.94 Å.

KEY WORDS: Molecular dynamics simulation; supersecondary structure; protein structure simulation.

1. INTRODUCTION

Molecular dynamics (MD) has been widely used in protein structure simulation. However, it is impossible to simulate the whole process of protein folding only with MD methods. This is mainly because that the time scale of protein folding is from $10^{-3}$ sec to several seconds, while the time scale of a typical MD simulation of a protein molecule is from $10^{-11}$ to $10^{-10}$ sec with present computational resources. In this paper, we integrate MD methods with a newly developed supersecondary structure prediction method (Sun et al., 1997). Supersecondary structure was defined as the combination of two secondary structural elements with a short connecting peptide of length between one and five residues. Many methods have been developed and improved for the secondary structure prediction (such as Chou and Fasman, 1974; Garner et al., 1978; Cohen et al., 1986; Qian and Sejnowski, 1988; Holley and Karplus, 1989). One important step toward building a tertiary structure from the specified secondary structures is to identify how secondary structures as building blocks arrange themselves in space. The conformation of the connecting peptide is the key issue in identifying a supersecondary motif. We used an artificial neural network method to predict supersecondary motifs from protein sequences. The predicted supersecondary structures were used as constraints in the MD simulation. We computed the structure of crambin, a 46-residue protein molecule with this method. The root mean square (rms) error of the computed structure is 3.94 Å.

2. METHOD

2.1. Prediction of Supersecondary Structures

Supersecondary structure is defined as the combination of two secondary structural elements with a short connecting peptide of length between one and five residues. A short connecting peptide can have a large number of conformations. They play an important role in defining the protein structure. A connecting peptide usually changes the trend of the protein backbone so as to form an antiparallel turn, a vertical corner, a twist, or just a slight bend in a peptide chain. The conformations of the residues in the short connecting peptides are classified into five major types, namely a, b, e, l, or t, each represented by a region on the $\Phi$-$\Psi$ map (Fig. 1). Supersecondary structures are classified according to their component secondary structural elements, the length of the connecting peptide, and the type of residues in the connecting peptide. In a survey of 240 proteins (Sun and Jiang, 1996), it was found that there are 34 types of
supersecondary structures which occur more than five times. Of these 34 types, 11 occur more than 25 times. These 11 types of supersecondary structures are called frequently occurring supersecondary structures. The 34 types of supersecondary structures occurred altogether 766 times, of which the 11 frequently occurring supersecondary structures occurred 568 times. This result shows that about 75% of the short connecting peptides belong to the 11 types of frequently occurring supersecondary structures. An artificial neural network method was developed to predict the 11 frequently occurring supersecondary structures; H-b-H, H-\(\beta\)-H, H-\(\beta\)b-H, H-\(\beta\)l-E, E-aa-E, E-ea-E, H-\(\beta\)b-H, H-\(\beta\)l-a-E, E-aal-E, E-aalal-E, and H-\(\beta\)l-E, where H and E represent \(\alpha\)-helix and \(\beta\)-strand, respectively. Each of these corresponds to a well-defined three-dimensional motif. Examples of the 11 motifs are given in Fig. 2. Study of the sequence, of the 11 types of frequently occurring protein supersecondary motifs lead to a classification scheme. A database of protein supersecondary motifs is first set up for further prediction. Then an artificial neural network method, i.e., the backpropagation neural network, is applied to the supersecondary motifs from protein sequences. The prediction correctness ratio is higher than 70% often from 75% to 82% (Sun et al., 1997). We applied the trained neural networks to predict the protein crambin. The results are compared with the native supersecondary structures in (Fig. 3) and the predicted supersecondary structures.