Journal of Theoretical Biology 262 (2010) 488–497

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

A novel approach for large-scale polypeptide folding based on elastic networks using continuous optimization

Sourav Rakshit *, G.K. Ananthasuresh

Mechanical Engineering, Indian Institute of Science, Bangalore 560012, India

ARTICLE INFO

Article history:
Received 20 March 2009
Received in revised form
23 September 2009
Accepted 8 October 2009
Available online 13 October 2009

Keywords:
MJ potential
Native state
Energy modeling
Folding funnel
Protein folding
Rigid body

ABSTRACT

We present a new computationally efficient method for large-scale polypeptide folding using coarse-grained elastic networks and gradient-based continuous optimization techniques. The folding is governed by minimization of energy based on Miyazawa–Jernigan contact potentials. Using this method we are able to substantially reduce the computation time on ordinary desktop computers for simulation of polypeptide folding starting from a fully unfolded state. We compare our results with available native state structures from Protein Data Bank (PDB) for a few *de-novo* proteins and two natural proteins, Ubiquitin and Lysozyme. Based on our simulations we are able to draw the energy landscape for a small *de-novo* protein, Chignolin. We also use two well known protein structure prediction software, MODELLER and GROMACS to compare our results. In the end, we show how a modification of normal elastic network model can lead to higher accuracy and lower time required for simulation.

1. Introduction

In this paper we present a novel approach for large-scale polypeptide folding from fully unfolded state to native state governed by the minimization of energy. The minimum energy is determined from MJ contact potentials (Miyazawa and Jernigan, 1996). It is generally accepted that folding of a polypeptide is governed by a funnel-like landscape (Dill, 1985; Leopold et al., 1992; Creighton, 1992; Dill and Chan, 1997). We intend to explore this funnel using continuous optimization techniques. Continuous optimization methods are more efficient in determining local minima than global stochastic search or combinatorial techniques. The simplicity and efficacy of our method is supported by our simple programs (written in MATLAB (www.mathworks.com)) which can run on personal computers taking insignificant time (generally 2–3 min) for small polypeptides; a moderate time of less than 1 h for larger polypeptides; and much less time for large proteins than other well-known methods.

While folding from an unfolded state to the native state, a polypeptide undergoes large changes in conformation. We have modeled these changes using the concept of an elastic network (EN). EN was first proposed by Bahar and Jernigan (1997) to provide a simple model for thermal fluctuations of proteins in their native crystal structures. It has been used for normal mode analysis (Haliloglu et al., 1997; Bahar and Rader, 2005) and for analyzing motions of proteins around their native state (Atilgan et al., 2001; Doruker et al.. 2001). Erman and Dill (2000) developed a dynamic model of EN and used it to predict native states. They showed that their model gave the same stable conformations as those of the corresponding lattice models. Kim et al. (2002) used EN to generate the intermediate states of proteins when their initial and final conformations are specified. Ball et al. (2002) framed protein folding problem as a variation of traveling salesman problem based on EN. They developed a fast combinatorial optimization method to predict native states of some small proteins. Our novelty lies in using continuous gradient based optimization on EN for predicting the native state, i.e., determining minimum energy (local minimum) conformations starting from a random and often a fully unfolded state. The deformations of the EN are expressed in terms of eigenvectors of its stiffness matrix. Eigenvectors are used for linear analysis, i.e., for calculating small deformations only. Here, we have formulated a novel algorithm by which we use eigenvectors for spanning the space of conformation change rather than normal Cartesian coordinates. By doing so we can produce large changes in conformation of the protein with much fewer variables than when the Cartesian coordinates themselves are taken as the conformation variables. One can view this change in conformation as similar to the deformation of an elastic structure under static internal loads. The forces that guide this folding are based on a novel potential function that uses MJ inter-residue contact

* Corresponding author.
E-mail addresses: srakshit@mecheng.iisc.ernet.in (S. Rakshit), suresh@mecheng.iisc.ernet.in (G.K. Ananthasuresh).

0022-5193/$ - see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2009.10.010

potentials. As these forces are static in nature, we neglect inertia or mass of the molecules but our method could consider this and other dynamic effects. We take initial conditions as unfolded states in which the protein is randomly oriented and fully stretched without violating bond-length and bond-angle constraints. We determine its final conformation by minimization of the energy using continuous optimization methods. We compare the final conformation with structures from Protein Data Bank (PDB; Bernstein et al., 1977) for a few *de-novo* proteins and two natural proteins, namely, Ubiquitin and Lysozyme. In case of these two naturally occurring proteins which are substantially larger than the *de-novo* proteins, we had to incorporate additional constraints to preserve the secondary structures.

The rest of the paper is organized as follows. In the next section we give a brief description of how we use EN for modeling large changes in conformation and formulate the energy minimization problem. Then, we apply our model to a few *de-novo* proteins and two natural proteins, namely Ubiquitin and Lysozyme. We compare the minimum energy conformation derived from our simulation with existing native conformations from PDB. We also generate the energy landscape for a particular *de-novo* protein, Chignolin (Honda et al., 2004). Next, we draw inferences from these simulations and discuss about the possible advantages and disadvantages of our technique. For Ubiquitin and Lysozyme, we make a modification of our method by considering the secondary structures as rigid bodies. We compare the results of our simulations with two well-known protein structure prediction software, namely MODELLER and GROMACS. The final section contains concluding remarks.

2. Elastic network model

Fig. 1 shows an EN model of a small 10-residue long *de-novo* protein called chigolin (PDB ID 1UAO). In this model all the C_α atoms are connected to one another by imaginary linear springs. Our EN model is a coarse-grained one as we assume all the amino acid residues to be spheres centered on their respective C_α atoms. As the bond-energies are much larger than the non-bonded interactions (due to which conformational changes occur), we take the length between the covalently linked residues to be fixed. Hence, the bonded C_α atoms are joined by springs of high stiffness. The non-bonded C_α atoms are joined by springs whose stiffness depends on the MJ contact potential of interacting residues. A lower MJ potential indicates lower stiffness and vice versa.

We derive the stiffness matrix K of the EN model (Cook et al., 2002). Any small deformed shape of the EN can be expressed as a linear combination of the eigenvectors of K. Thus, if $\{x_0\}$ is the position vector of all the residues in EN at an initial condition, we can express the position vector of all residues $\{x\}$ of a nearby conformation as

$$\{x\} = \{x_0\} + \sum_{i=1}^{3N} \alpha_i \{\omega_i\} \quad (1)$$

where N is the number of residues in the polypeptide, $\{x\}$ the column vector of Cartesian coordinates of all C_α atoms; so its dimension is $3N \times 1$, $\{\omega_i\}$ the ith eigenvector of K with dimension $3N \times 1$, α_i the scalar multiplier associated with ith eigenvector $\{\omega_i\}$.

The scalar multipliers $\{\alpha_i\}$ form the set of design variables in our optimization problem formulation. By varying these coefficients we change the conformation and with it the energy of the polypeptide. This method allows small changes in conformation only since eigenanalysis is essentially linear. To apply this method for large changes in conformation, we formulate a novel algorithm which updates the stiffness matrix of EN of the polypeptide from time to time as optimization progresses. This algorithm is shown in Fig. 2. As explained in this figure, the rate at which we update the stiffness matrix K is determined by the maximum number of iterations (*maxiter*) specified to the optimization program. In Fig. 3 we show for different values of *maxiter* how conformational energy of Chignolin varies with iteration as we minimize the energy from a fully unfolded state. Table 1 compares the energy of the final conformation, total number of iterations and actual CPU time for different values of *maxiter* . It is interesting to note that the energy of the optimal conformation shows insignificant change as *maxiter* is varied (Table 1). Even when we do not

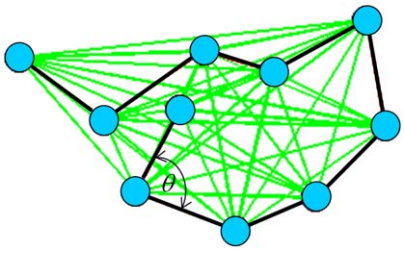

Fig. 1. Elastic network model of a small *de-novo* protein, Chignolin. The blue circles represent the amino acid residues centered on their respective C_α atoms. The covalently bonded residues are connected by black lines. The non-bonded residues are connected by green lines. (For interpretation of the references to the color in this figure legend, the reader is referred to the web version of this article.)

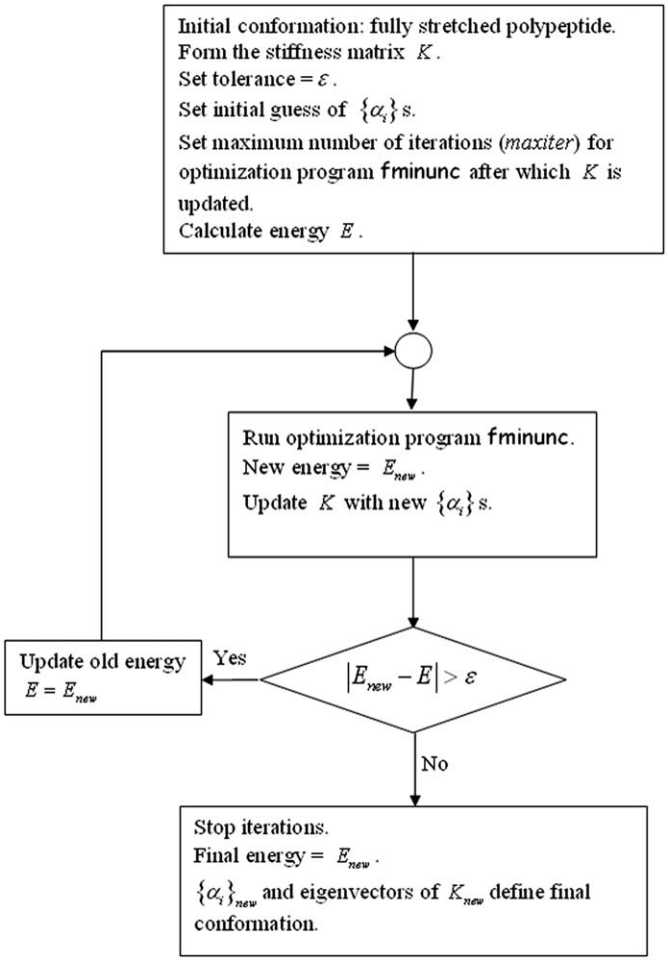

Fig. 2. Flowchart showing our algorithm for large change in conformation determined using eigenvectors of stiffness matrix K of EN.