Computers and Mathematics with Applications 57 (2009) 1855–1861

Contents lists available at ScienceDirect

Computers and Mathematics with Applications

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

Extremal Optimization for protein folding simulations on the lattice☆

Hengyun Lu, Genke Yang*

School of Electronic Information and Electrical Engineering, Shanghai Jiaotong University Shanghai, 200240, Shanghai, China

ARTICLE INFO

Keywords:
Protein folding
HP model
Extremal Optimization

ABSTRACT

This paper presents a novel guided search strategy Extremal Optimization (EO) with constrained structure for protein folding. In the proposed algorithm, evaluating the fitness of each monomer in an amino-acid sequence is introduced to guide the improvement of the conformation. In addition, a constrained structure is proposed to reduce the complexity of algorithm. We demonstrate that EO can be applied successfully to the protein folding problem. The results show that the algorithm can find the best solutions so far for the listed benchmarks. Within the achieved results, the search converged rapidly and efficiently.

1. Introduction

The prediction of the native structure of a protein from its amino-acid sequence is one of the most attractive problems in computational biology. Proteins are polymer chains of amino acid residues of twenty different kinds. They are essential to the structures and functions of all living cells and perform functions including acting as a catalyst for biomedical reactions, as structures of cells and receptors for hormones. Furthermore, the biological functions and properties of proteins are crucially determined by their structures, so understanding how to predict protein structure from sequence information is of immense significance. Famous folding experiments [1] by Anfinsen et al. showed that a protein in its natural environment folds into, i.e. vibrates around, a unique three dimensional structure, the native conformation. At present, experimental determination of native conformation by techniques such as magnetic resonance imaging (MRI) and X-ray crystallography is expensive and time consuming, so much work has been done to predict the native conformation computationally [2–5].

The difficulties in predicting the native conformation of a protein molecule from the amino acid sequence mainly stem from two sources: (1) finding proper measurements for the quality of candidate structures (e.g., energy models), (2) upon such measurements, determining optimal or close-to-optimal structures for a given amino-acid sequence. While the first issue is usually addressed by biochemists who study and model the protein folding processes, the second attracts many experts who working on computational problems in local and global optimization. In this paper, we address to the latter which can be investigated on a particular simplified model: the 2-dimensional hydrophobic-polar (2D HP) model.

Recently, for 2D HP models, many heuristic algorithms have been explored to find the minimum energy configuration for small protein. Since Genetic Algorithm (GA) proposed by Unger opens a door for applying evolutionary computation in Protein folding [2], the subsequent variation comes with the other stochastic search approaches, such as varieties of Monte Carlo(MC) [6], Evolutionary MC(EMC) [5], Simulated Annealing (SA), Tabu Search with GA(GTB) [3] and a guided GA(GGA) [7]. Meanwhile, statistical approaches such as Contact Interaction (CI), Chain Growth (CG), pruned-enriched-Rosenbluth method (PERM) and its variations nPERMis [8] have also been applied to 2D HP models.

In this paper, Extremal Optimization with constrained structure is proposed for the protein folding problem. Extremal Optimization (EO) is a general purpose local search heuristic method based on recent progress in understanding far-from-equilibrium phenomena in terms of Self-Organized Criticality (SOC), a concept introduced to describe emergent complexity

☆ This work was supported in part by National Natural Science Foundation under grant no. 60574063.
* Corresponding author.
E-mail addresses: luhy@mail.sjtu.edu.cn (H. Lu), gkyang@mail.sjtu.edu.cn (G. Yang).

0898-1221/$ – see front matter © 2009 Elsevier Ltd. All rights reserved.
doi:10.1016/j.camwa.2008.10.061

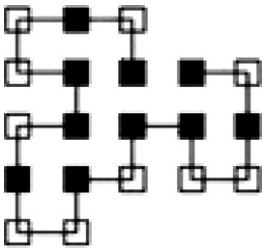

Fig. 1. A sample of the native conformation for a protein in the 2D HP model, showing a sequence of amino acids connected by solid line. '■': Hydrophobic '□': Hydrophilic.

in many physical systems [9,10]. There are two crucial pivots in EO: the definition of fitness for every monomer in the sequence and the strategy for neighborhood search. The former leads to the improvement of current conformation, while the latter, a variation of PERM, determines the convergent speed of the whole algorithm. In addition, a constrained structure is used to reduce the complexity and runtime of the algorithm, motivated by the assumption that the main driving force of the formation of the tertiary structure is the hydrophobic interaction.

The rest of this paper proceeds as follows: In Section 2, the 2D HP model and the energy function are introduced. In Section 3, we give a detailed description of EO with constrained structure for protein folding problem. In Section 4, we define the fitness for every monomer in the sequence. In Section 5, we introduce the neighbourhood search inspired by advanced PERM. Section 6 is the experimental results and comparisons between the proposed approach and other state-of-the-art algorithms. Finally, a short discussion and conclusion are given in Section 7.

2. The 2D HP model

The HP model proposed first by Dill in 1985 [11] is a free energy model based on the belief that interactions between hydrophobic amino acids give a major contribution to the free energy of the natural conformation of a protein. In this model, the amino acid sequence of a protein is abstracted as a binary sequence comprising hydrophobic and hydrophilic amino acids. Even though some amino acids cannot be classified clearly as being either hydrophobic or hydrophilic, the model disregards this fact to achieve simplicity. The model is usually referred to as the HP model where H stands for hydrophobic and P stands for polar(hydrophilic).

In the 2D HP model, the sequence is folded on a 2-dimensional square lattice. Here we will use s to denote a protein sequence with length N. That is, $s = s_1 s_2 \dots s_N$ is represented by a string in a binary alphabet: $s_i \in \{H, P\}$, $i = 1, \dots, N$, where s_i is H if the ith amino acid in the sequence is hydrophobic and P if it is hydrophilic. Then a folding of a protein in this model can be regarded as an embedding of s in the square lattice such that adjacent characters in s occupy adjacent grid points in the lattice, and no grid point is occupied by more than one character. Two amino acids are namely topological adjacent if they are neighbors in the lattice. Then a topological H–H bond is formed between two topological adjacent hydrophobic amino-acids which are not neighbors in sequence. The free energy of a conformation depends on the number of such H–H bonds. In other words, if a conformation denoted as c has exactly n such H–H bonds, its free energy $E(c) = n \cdot (-1)$. Fig. 1 shows a conformation in the 2D HP model with energy -9.

Then the 2D HP protein folding problem can be formally defined as follows: Given an amino-acid sequence $s = s_1 s_2 \dots s_N$, find an energy-minimizing conformation of s, i.e., find $c^* \in C$ such that $E(c^*) = \min\{E(c) | c \in C\}$, where C is a set of all possible conformations of sequence s. It has been proven that this problem and its several variations are NP-complete [12].

3. Extremal Optimization with constrained structure for 2D HP model

Extremal Optimization, proposed by Boettcher and Percus [9], is inspired by a simplified model of natural selection. It is developed to show the emergence of Self-Organized Criticality (SOC), a concept introduced to describe the tendency of large driven dissipate systems to self-organize to a peculiar stationary state. Benefited from its generality and ability to explore complicated configuration spaces efficiently, EO and its variations have been widely applied in many combinational optimization problems such as graph bi-partitioning problem [9], graph coloring [13] and Satisfiability Problem [14] etc.

The same as GA, EO is based on the principles of natural selection, but it does not use the GA's framework of population reproduction. For example, rather than by expressly breeding those species best adapted to their environment, evolution in EO is driven by successively eliminating extremely undesirable components of sub-optimal solutions, and generating their replacements stochastically. The dynamics of this extremal process shows characteristics of SOC, such as punctuated equilibrium, that is also observed in natural ecosystems.

To avoid a search based only on the forced mutation of the weakest species and get out of local optima, the basic EO algorithm has been modified by introducing an adjustable parameter. This variation of EO algorithm is called τ-EO and shows superior performance over the EO even in the cases where the later works well.