Multi-Objective Evolutionary Algorithm NSGA-II for Protein Structure Prediction using Structural and Energetic Properties

R. A. Faccioli, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, São Paulo, Brazil

L. O. Bortot, Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, São Paulo, Brazil

A. C. B. Delbem, Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos, São Paulo, Brazil

ABSTRACT

The Protein Structure Prediction (PSP) problem is concerned about the prediction of the native tertiary structure of a protein in respect to its amino acids sequence. PSP is a challenging and computationally open problem. Therefore, several researches and methodologies have been developed for it. In this way, developers are working to integrate frameworks in order to improve their capabilities and make their use more straightforward. This paper presents the application of NSGA-II algorithm using structural and energetic properties of protein. The implementation of this algorithm is based on ProtPred-GROMACS (2PG), an evolutionary framework for PSP. This framework is the integration between ProtPred and GROMACS. Six proteins were used to measure the capacity of ab initio predictions. The results were interesting since in all cases the native-like topology was obtained.

Keywords: Evolutionary Algorithms (EAs), Framework, Multi-Objective Optimization (MOP), Non-dominated Sorting Genetic Algorithm-II (NSGA-II), Protein Structure Prediction (PSP)

DOI: 10.4018/ijncr.2014010104
1. INTRODUCTION

Many essential functions for life are performed by proteins and the study of their structures yields the ability to elucidate their functionalities and properties in terms of a molecular view (Creighton, 1992). In many fields, there is great interest in discovering a methodology for Protein Structure Prediction (PSP), including drug design, diseases mechanisms and food industry. In this context, some experimental methods have been applied to determine the structure of proteins, such as X-ray crystallography and nuclear magnetic resonance. Despite their success, both methods have characteristic limitations. Conversely, the knowledge of the primary sequence of the amino acids of a protein is a simpler experimental procedure.

In protein molecular modeling, the determination of the native spatial arrangement of the molecule atoms, which corresponds to a global or local energy minimum state, is a fundamental task because it represents the protein functional conformation. However, there is not an efficient general computational method to achieve this purpose yet (Leach, 2001). Indeed, one of the main braches of the protein folding area is the computational problem of how to predict a protein native conformation solely from its amino acids sequence (Dill et al., 2008).

Although it is not possible assure that the global free energy minimum always corresponds to the protein bio-active conformation, at least it will be a local minimum (Leach, 2001). Several computational methods for PSP are semi ab inito methodologies, so they also use prior knowledge from both the sequence homology and the statistics found on protein databases (Miyazawa & Jernigan, 1985; Poole & Ranganathan, 2006). However, the use of this additional information restricts the search of protein structures that could be correctly predicted from the vast universe of proteins.

Evolutionary Algorithms (EAs) have been investigated because they have capacity to exploit extensively the search space in order to look for the best solutions (Michalewicz and Schoenauer, 1996). EAs are a metaheuristics inspired by evolutionary theory (Goldberg, 1989). These algorithms have been applied in mono and multi-objective optimization problems (Talbi, 2009).

In fact, macromolecular modeling and design are increasingly useful in basic research, such as biotechnology. However, the absence of a user-friendly modeling framework that provides access to a wide range of modeling capabilities discourages the wider adoption of computational methods by non-experts (Fleishman et al., 2011). Moreover, PSP requires efficient computational methods (Nair and Goodman, 1998).

All frameworks provide some information about protein but not all of them. Therefore, it is necessary to work with more than one, although it is an uncomfortable situation because frameworks can be operating with different force fields and/or units. Furthermore, PSP is an open problem, where researchers have applied different ideas and methods. Therefore, developers are working to integrate these frameworks in order to improve their capabilities and make their use more straightforward (Fleishman et al., 2011; Cock et al., 2009; Lund et al., 2008; Eswar et al., 2002). In this way, ProtPred-GROMACS (Faccioli et al., 2011; Faccioli et al., 2012) has been applied in PSP as mono-objective concept. Here, it is presented in multi-objective concept through implementation of NSGA-II algorithm.

1.1. Multi-Objective Optimization

Multi-Objective Optimization (MOP) discuss problems in which the number of objectives are not unique further they have to be treated simultaneously. Moreover, these objectives can be conflicting. In this type of problem there is a group of solutions that meet an “equilibrium” situation (solutions) (Coello Coello, 2006).

The objective space $Z$ is a multi-dimensional space, comprising the vector of objective functions $f(x)$. The difference between mono-objective and multi-objective is the search space where in the multi-objective, it is multidimensional and every solution $x$ in the space of decision has $f(x)$ in $Z$. On the other
A New Approach to Pattern Recognition in Fractal Ferns
Mamta Rani and Saurabh Goel (2010). *International Journal of Artificial Life Research* (pp. 21-28).
www.igi-global.com/article/new-approach-pattern-recognition-fractal/44668?camid=4v1a