Reduced Pareto Set Genetic Algorithm: Application to Protein Structure Prediction

R. A. Faccioli[1] , A. Gaspar-Cunha[2], A. Defelicibus[3], L. O. Bortot[4] and A. C. B. Delbem[1]

1. Instituto de Ciência Matemática e da Computação, Universidade de São Paulo, São Carlos, Brasil.
rodrigo.faccioli@gmail.com
acbd@icmc.usp.br

2. Instituto de Polímeros e Compósitos – IPC/I3N, Universidade de Minho, Guimarães, Portugal.
agc@dep.uminho.pt

3. Programa de Pós-Graduação em Bioengenharia EESC/FMRP/IQSC, Universidade de São Paulo, São Carlos, Brasil.
adefelicibus@usp.br

4. Laboratório de Física Biológica, Faculdade de Ciências Farmacêuticas de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, Brasil.
leandro.bortot@usp.br

Keywords: multi-objective optimisation, bioinformatics.

1 Introduction

The determination of the native spatial arrangement of the molecule proteins, which corresponds to a global or local energy minimum state, is a fundamental task because it represents the protein functional conformation. In spite of this importance, there is no efficient computational method that achieves such purpose for any protein.

The process of finding such conformation is called *folding.* There are some physical properties that guide this process [1]:

- Flexibility of the backbone of the sequence;
- Interactions among amino acids, including electrostatic interactions and van der Waals forces;
- Volume constraints;
- Specific bonds: salt bridges , Hydrogen and disulfide bounds ;
- Interactions of amino acids with water.

Due to the difficulty to model and simulate the folding process, it has been formulated as an optimization problem. Indeed, by some research groups, one of the main branches of the protein folding area is the computational problem of how to predict a protein native conformation solely from its amino acids sequence. Furthermore, different protein conformations may involve a tradeoff among different physical properties. Therefore, an optimum solution (protein conformation) with respect to one physical property usually are not the optimum with respect to another physical property.

Multi-Objective Evolutionary Algorithms (MOEAs) have been investigated in protein structure prediction since they can efficiently exploit the search space finding solutions that present the best tradeoff between two criteria or more. Those papers shared a conclusion: those MOEAs are promised techniques for protein structure prediction.

In this way, developers have worked to create frameworks that integrate optimization algorithms with molecular modeling for the purpose of applying them to protein structure prediction. ProtPred-Gromacs (2PG) [2] has been developed for this purpose. 2PG implements optimization algorithms such as Monte

Carlo Metropolis, a Mono-Objective Evolutionary Algorithm and NSGA-II (Non-Sorting Genetic Algorithm-II). All optimization algorithms from 2PG utilize Gromacs (a framework for calculating molecule potentials and simulation) [3] to compute the physical properties of each candidate to a protein conformation (point in the search space).

In this paper will be presented the integration of Reduced Pareto Set Genetic Algorithm (RPSGA) with Elitism [4] in 2PG. The performance of 2PG using RPSGA will be compared with the result from 2PG using NSGA-II for a set of benchmark proteins.

References

[1] M. Karplus and E. Shakhnovich. Protein folding: Theorical studies of thermodynamics and dynamics. Protein Folding (Creighton, T.,ed), 125–195, 1992.

[2] R. A. Faccioli, L. O. Bortot and A. C. B. Delbem. Multi-Objective Evolutionary Algorithm NSGA-II for Protein Structure Prediction using Structural and Energetic Properties. IJNCR, 4(1), 43-53, 2014

[3] HESS, B.; KUTZNER, C.; SPOEL, D. van der; LINDAHL, E. GROMACS 4: algorithms for highly e cient, load-balanced, and scalable molecular simulation. Journal of Chemical Theory and Computation, v.4, n.3, p.435 - 447, 2008.

[4] A. Gaspar-Cunha and J. A. Covas, "RPSGAe--Reduced Pareto Set Genetic Algorithm: Application to Polymer Extrusion," in Metaheuristics for Multiobjective Optimisation, X. Gandibleux, M. Sevaux, K. Sörensen, and V. T'kindt, Eds. Berlin Heidelberg: Springer. Lecture Notes in Economics and Mathematical Systems Vol. 535, 2004, pp. 221–249.