An Evolutionary Search Framework to Efficiently Sample Local Minima in the Protein Conformational Space

Brian Olson[1] and Amarda Shehu[1,2,3]

[1]Department of Computer Science
[2]Department of Bioinformatics and Computational Biology
[3]Department of Bioengineering
George Mason University, Fairfax, VA 22030

http://www.cs.gmu.edu/~ashehu
amarda@gmu.edu | bolson3@gmu.edu

Abstract

The quest to characterize the protein native state from primarily its amino-acid sequence remains a central challenge in computational biology. The problem is computationally intractable due to the continuous and high-dimensional conformational space and the corresponding rugged energy surface [1]. A plethora of conformational search algorithms are devoted to this problem, employing diverse strategies to navigate the search space with the goal of converging to the native energy basin. Our previous work shows that an explicit focus on sampling local minima in the conformational space is an effective approach for sampling conformations near the experimentally-determined native state[2]. Here we extend the explicit local minima sampling in basin hopping to a population-based evolutionary search framework to more effectively cover the breadth of the search space.

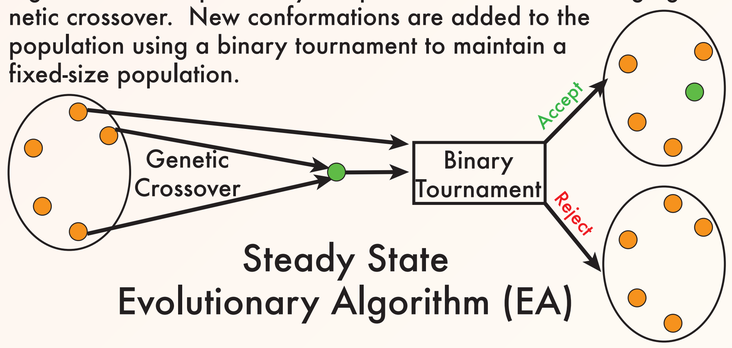
The framework unifies strategies from evolutionary computation and computational biology to more effectively navigate the subspace of conformations corresponding to local minima. We employ a hybrid approach which combines a perturbation step to make larger global moves with a minimization step to map each conformation sampled at the global level to a nearby local minimum. The minimization step draws on computational biology, consisting of a coarse-grained greedy local search employing molecular fragment replacement. An evolutionary-inspired perturbation operator escapes each mapped local minimum to effectively explore the breadth of the search space. A multi-objective approach to potential energy allows the framework to maintain a representative population of "interesting" low-energy local minima to effectively guide the search towards unexplored regions of the search space.

Our experiments show that the proposed framework samples conformations near the experimentally-determined native structure more effectively or as well as state-of-the-art methods in ab-initio protein structure prediction [4]. On more complex α/β topologies, employing the Pareto front as the population effectively guides the search closer to the native state.

Grant No. 1016995

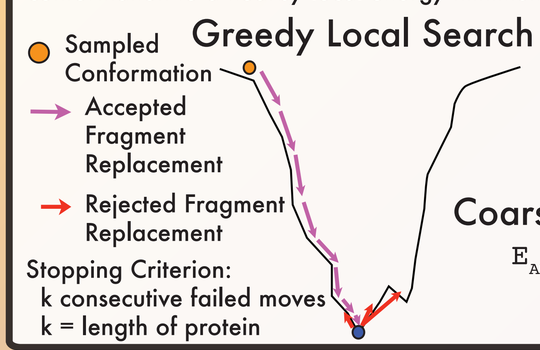
Evolutionary Algorithm (EA)

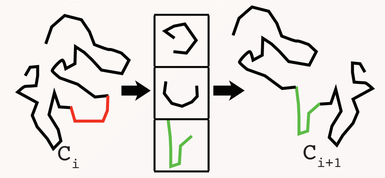
A population of conformations is iteratively evolved by combining features from previously sampled conformations through genetic crossover. New conformations are added to the population using a binary tournament to maintain a fixed-size population.



Local Search For Biologically Realistic Structures

A coarse-grained greedy local search employing molecular fragment replacement maps a sampled conformation to a nearby local energy minimum.





Coarse-grained Energy Function

$$E_{AMW} = E_{electro\text{-}static} + E_{H\text{-}Bond} + E_{water} + E_{bond} + E_{Rg}$$

Evolutionary Computation

-How to sample biologically relevant conformations with crossover?

-How to maintain a diverse population of conformations?

Interdisciplinary Crossover

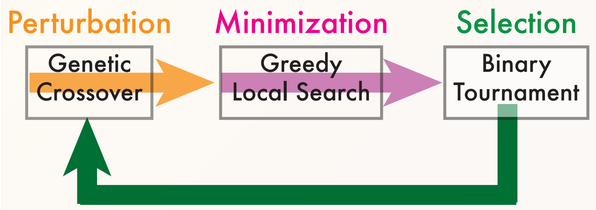
Computational Biology

-How to explore the breadth of the protein conformational space?

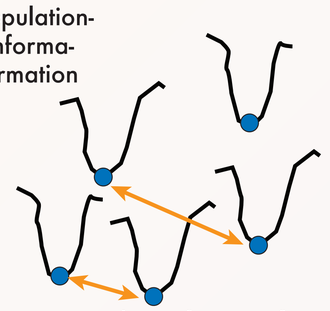
-How to distinguish between two low-energy local minima?

Hybrid Global-Local Search

A Hybrid Evolutionary Algorithm (HEA) combines the population-based global search for exploring the breadth of the conformational space with a minimization step to map each conformation sampled at the global level to a nearby local minimum.



Conformations sampled through crossover are optimized to biologically relevant structures through the local search.

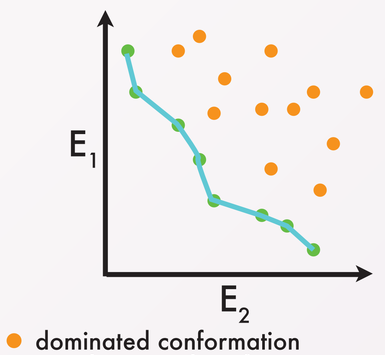


Crossover combines features from different locally optimized structures to effectively explore the conformational space.

Multi-Objective Conformational Ensemble

The error resulting from the weighted linear combination of energy terms in current energy functions make it difficult to distinguish between two low-energy conformations based on total energy. This error is avoided by a multi-objective approach where energy terms are considered individually.

In a Multi-Objective Hybrid EA (MOHEA), conformations in the non-dominated Pareto front are retained in the population (rather than using binary tournament with total energy). The result is an unbounded population retaining only an "interesting" sub-set of sampled local minima.



References

1. K. A. Dill, B. Ozkan, M. S. Shell, and T. R. Weikl. "The protein folding problem." Annu. Rev Biophys., 37:289–316, 2008.

2. Brian Olson and Amarda Shehu. Evolutionary-inspired Probabilistic Search for Enhancing Sampling of Local Minima in the Protein Energy Surface. Proteome Science 2012, 10(Suppl1): S5.

3.G. A. Papoian, J. Ulander, M. P. Eastwood, Z. Luthey- Schulten, and P. G. Wolynes, "Water in protein structure prediction," Proc. Natl. Acad. Sci. USA, vol. 101, no. 10, pp. 3352–3357, 2004.

4. J. DeBartolo, A. Colubri, A. K. Jha, J. E. Fitzgerald, K. F. Freed, and T. R. Sosnick, "Mimicking the folding pathway to improve homology-free protein structure prediction," Proc. Natl. Acad. Sci. USA, vol. 106, no. 10, pp. 3734–3739, 2009.

Results

We test the effectiveness of the HEA and MOHEA to sample coarse-grained decoy conformations near the protein native state. The table shows the RMSD to the experimentally-determined native structure of the best decoy conformations sampled for 15 target protein systems. While the HEA tends to sample lower RMSD structures for the larger α-helix proteins, the addition of the multi-objective analysis in the MOEA improves results on the smaller α/β proteins.

Results for 11 targets are compared to published results from the Sosnick group [4].