Effective Search of the Energy Landscape for Protein Folding

Eugene Santos Jr.[1], Keum Joo Kim[1], and Eunice E. Santos[2]

[1] University of Connecticut, Storrs, CT 06269
{eugene,keumjoo}@engr.uconn.edu,
[2] Virginia Polytechnic Institute and State University, Blacksburg, VA 24061
santos@cs.vt.edu

Abstract. We propose a new algorithmic approach for global optimization in protein folding. We use the information found in various local minima to direct the search for the global minimum. In this way, we explore the energy landscape efficiently by considering only the space of local minima instead of the whole feasible space of conformations.

Our fundamental approach is to sample only the space of local minima and guide the sampling process by exploring protein structure building blocks found in sampled local minima. These building blocks form the basis of information in searching for the global minimum. In particular we employ an iterative algorithm that begins with an initial pool of local minima; construct a new pool of solutions by combining the various building blocks found in the original pool; take each solution and map them to their representative local minima; and, repeat the process. Our procedure seems to share a great deal of commonality with evolutionary computing techniques. Indeed, we even employ genetic operators in our algorithm. However, unlike existing hybrid evolutionary computing algorithms where local minimization algorithms are simply used to "fine-tune" the solutions, we focus primarily on constructing local minima from previously explored minima and only use genetic operators to assist in diversification. Hence, our total number of iterations/generations were demonstrated (empirically) to be quite low (≈ 50) whereas standard genetic algorithms and Monte Carlo are very high ranging from 150,000 to nearly 20,000,000 generations in order to provide sufficient opportunity for these methods to converge and achieve their best solution. We applied our idea to several proteins from the Protein Data Bank (PDB) using the UNRES model[1]. We compared against Standard Genetic Algorithms(SGA) and Metropolis Monte Carlo(MMC) approaches. In all cases, our new approach computed the lowest energy conformation.

Procedure LMBE
begin
$t = 0$*;*
initialize $P(t)$ *with* ***local minima****;*
while termination condition not satisfied do
    begin
    select individuals $P^{sub}_{new}(t)$ *from current pool* $P(t)$*;*
    recombine structures with selected individuals $P^{sub}_{new}(t)$*;*

E. Cantú-Paz et al. (Eds.): GECCO 2003, LNCS 2723, pp. 642–643, 2003.

determine local minima corresponding to $P_{new}(t)$
replace local minima in $P_{new}(t)$***;***
evaluate structures $P_{new}(t)$*;*
end
end.

Although LMBE is clearly derived from standard genetic algorithm approaches, our emphasis is on exploring the local minima space and exploits the genetic operators for diversification of the population. Furthermore, this is potentially more systematic in local minimization than memetic algorithms.

Given the prohibitive amount of time to conduct multiple runs of each method over all 100 proteins, each method was run exactly once using the parameter settings determined from pre-trial runs. Hence, the weaknesses and strengths of each method is averaged over the testbed.

For each protein, we initially constructed 100 random conformations. Next, we found the local minimum for each conformation with the gradient descent algorithm [2]. The initial pool consists of these 100 random minimized conformations. The same initial pool was used for LMBE, SGA and MMC for algorithm comparison. The computation time of LMBE varied from 10 mins to 13 hrs, depending on the protein length, amino acid sequence and the genetic parameters (i.e. crossover rate, mutation rate). For MMC, the time was between 21 mins and 16 hours. For SGA, the time varied between 13 mins to 14 hours. Table 1 shows the average energy improvement of LMBE compared with SGA, MMC, and the baseline from PDB. For all 100 proteins, LMBE computed the best energy conformation. Finally, it is interesting to observe that the improvement using LMBE seems to improve significantly for longer proteins on comparison to the existing baseline.

Table 1. Percentage improvement of LMBE over SGA, MMC, and the baseline

Protein Group	SGA(%)	MMC(%)	baseline(%)
Group A (11–20 res.)	8.75	8.82	25.81
Group B (21–30 res.)	11.94	12.50	40.45
Group C (31–40 res.)	13.67	14.05	44.95
Group D (41–50 res.)	13.93	14.30	56.47

References

1. Liwo, A., Kazmierkiewicz, R., Oldziej, S., Pincus, M. R., Wawak, R. J., Rackovsky, S., and Scheraga, H. A.: A United-Residue Force Field for Off-Lattice Protein-Structure Simulations: III. Origin of Backbone Hydrogen-Bonding Cooperativity in United-Residue Potentials. J. Com. Chem. (1998) 19, 259–276
2. Gay, David M.: Algorithm 611: Subroutines for Unconstrained Minimization Using a Model/Trust-Region Approach. ACM ToMS(1983) 9, 503–524