

# Local Minima-Based Exploration for Off-Lattice Protein Folding

Eugene Santos Jr., Keum Joo Kim  
University of Connecticut, Storrs, CT 06269  
eugene@enr.uconn.edu, keumjoo@enr.uconn.edu

Eunice E. Santos  
Virginia Polytechnic Institute and State University, Blacksburg, VA 24061  
santos@cs.vt.edu

## Abstract

*We present a new and simple algorithmic approach to help predict protein structures from amino acid sequences based on energy minimization. In the search for the minimal energy conformation, we analyze and exploit the protein structures found at the various local minima to direct the search the global minimum. As such, we explore the energy landscape efficiently by considering only the space of local minima instead of the whole feasible space of conformations. Our specific algorithmic approach is comprised of two different elements: local minimization and operators from genetic algorithms. Unlike existing hybrid approaches where the local optimization is used to fine-tune the solutions, we focus primarily on the local optimization and employ stochastic sampling through genetic operators for diversification. Our empirical results indicate that each local minimum is representative of the substructures contained in the set of solutions surrounding the local minima. We applied our approach to determining the minimal energy conformation of proteins from the Protein Data Bank (PDB) using the CHARMM and UNRES energy model. We compared against standard genetic algorithms and Monte Carlo approaches as well as the conformations found in the PDB as the baseline. In all cases, our new approach computed the lowest energy conformation.*

Proteins have specific native folding structures, but without proper folding, these proteins function incorrectly and lead to disease. The major challenge lies in effectively predicting such native conformations. Particularly we employ an iterative algorithm in searching for the global minimum; begin with an initial pool of local minima; construct a new pool of solutions by combining the various building blocks found in the original pool; map all combined solutions to their representative local minima; and, repeat the process. (For specific details on our algorithm, see

[4][5].) Our procedure seems to share a great deal of commonality with evolutionary computing techniques because we employ genetic operators for diversification. However, most hybrid evolutionary computing algorithms used local minimization algorithms as the solutions of “fine-tune”, we concentrate primarily on constructing local minima from previously explored minima and only use genetic operators to assist our search. For our baseline comparisons, we referenced the structure from the Protein Data Bank (PDB) for each protein sequence. These structures represent the predicted native structures for each protein known to date. Furthermore, because these structures were determined using various molecule models, their energy values are not directly related to CHARMM[1] or UNRES[3] energy function. Therefore, we computed the UNRES energy value from the given PDB structure and further minimized it using our gradient descent algorithm[2] for UNRES. For CHARMM, we computed the CHARMM energy value from the given PDB structure and further minimized it using our gradient descent algorithm[2] for CHARMM. Given the debates regarding the efficacy of different energy models such as CHARMM and UNRES in accurately predicting the native structure of a protein, our focus for this testbed is to determine the effectiveness and efficiency of our algorithmic approach for CHARMM and UNRES and ultimately any other realistic energy model. Hence, assuming that the molecular energy model adequately captures the necessary physical properties and real-world constraints, the energy value itself is our primary measurement criteria. We picked up several proteins arbitrarily from the Protein Data Bank (PDB) and minimized using the CHARMM and UNRES model. We compared against Standard Genetic Algorithms(SGA) and Metropolis Monte Carlo with Simulated Annealing(MC/SA) approaches. In all cases, our new approach computed the lowest energy conformation. 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. With our experimental results our total number of iterations/generations were demonstrated (empirically) to be quite low ( $\approx 50$ ) whereas Standard Genetic Algorithms and Monte Carlo with Simulated Annealing are very high number of generations in order to provide sufficient opportunity for these methods to achieve their best solution. The computation time of LMBE for UNRES 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 MC/SA, the time was between 21 mins and 16 hours. Table 1 shows the obtained energy values of LMBE compared with SGA, MC/SA, and the baseline from PDB (due to space limitations, we have only presented the first 50 proteins from our experiments). For all 100 proteins, LMBE computed the best energy conformation with the UNRES model. Moreover, it is interesting to observe that the improvement using LMBE seems to improve significantly for longer proteins on comparison to the existing baseline. Furthermore, our results on the CHARMM energy model have been equally promising. Finally, our goal is to ultimately obtain a better understanding of the computational requirements and complexities posed by different molecular energy models. With such an understanding, we can provide critical input during the design of future realistic energy models that avoid unnecessary algorithmic complexities with a guide towards efficiency for the computational biologist.

## References

- [1] B. R. Brooks, R. E. Brucoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. Charmm: A program for macromolecular energy, minimization and dynamics calculations. *Journal of Computational Chemistry*, 4:187–217, 1983.
- [2] D. M. Gay. Algorithm 611: Subroutines for unconstrained minimization using a model/trust-region approach. *ACM ToMS*, 9:503–524, 1983.
- [3] A. Liwo, R. Kazmierkiewicz, S. Oldziej, M. R. Pincus, R. J. Wawak, S. Rackovsky, and H. A. Scheraga. A united-residue force field for off-lattice protein-structure simulations: Iii. origin of backbone hydrogen-bonding cooperativity in united-residue potentials. *Journal of Computational Chemistry*, 19:259–276, 1998.
- [4] E. Santos, Jr., K. Kim, and E. E. Santos. Effective search of the energy landscape for protein folding. In *Proceedings of the Genetic and Evolutionary Computation Conference. GECCO*, 2003.
- [5] E. Santos, Jr., K. Kim, and E. E. Santos. Efficient global optimization for protein folding. IDIS Technical Report 104, University of Connecticut, Intelligent Distributed Information Systems Laboratory, 2003.

**Table 1. Performance of LMBE versus SGA, MMC, and the baseline for 50 proteins with UNRES model**

PDB ID	LMBE	SGA	MMC	Baseline
1A11	-98.53	-85.84	-86.02	-95.03
1AGG	-190.08	-162.31	-161.29	-143.22
1AGT	-140.35	-118.49	-119.81	-117.99
1AHL	-178.63	-166.4	-165.89	-136.15
1ALC	-127.19	-107.83	-107.02	-113.19
1ALE	-39.98	-36.68	-36.12	-39.6
1ALF	-48.48	-41.37	-41.02	-45.13
1ALG	-78.17	-69.33	-68.93	-71.04
1AMB	-80.12	-74.79	-74.02	-76.63
1AML	-153.51	-132.46	-132.46	-126.62
1ANP	-94.05	-84.52	-84.52	-78.57
1ANS	-97.99	-95.09	-95.09	-80.68
1A00	-114.26	-98.12	-97.5	-99.05
1APF	-186.96	-175.99	-175.45	-137.3
1APO	-126.08	-113.18	-112.95	-100.04
1AQG	-26.88	-25.69	-25.47	-25.89
1ARD	-84.49	-79.97	-79.97	-70.85
1ARE	-88.02	-77.73	-77.47	-75.53
1ARF	-90	-80.33	-79.85	-75.31
1ATX	-178.23	-153.76	-153.32	-134.86
1AW6	-145.18	-123.74	-123.52	-123.94
1AXH	-115.79	-99.22	-98	-86.89
1AZ6	-128.92	-123.33	-122.58	-93.32
1AZJ	-129.02	-116.98	-116.67	-86.1
1AZK	-131.58	-117.6	-117.6	-94.26
1B03	-49.29	-43.57	-45.63	-35.29
1B1V	-70.29	-62.65	-62.95	-57.91
1B8W	-145.9	-128.63	-128.63	-103.1
1B9P	-90.86	-88.26	-86.47	-87.07
1BBA	-116.73	-100.44	-100.32	-106.05
1BBG	-141.87	-135.73	-135.16	-114.6
1BDS	-164.32	-152.01	-152	-141.56
1BGK	-129.01	-107.94	-107.47	-104.29
1BKT	-130.98	-112.84	-112.8	-113.99
1BL1	-116.4	-98.26	-96.25	-101.36
1BOE	-164.46	-149.73	-149.47	-126.74
1BWX	-145.28	-131.57	-131.57	-98.84
1BXJ	-102.58	-91.54	-91.54	-54.61
1BY0	-49.47	-44.04	-43.79	-48.14
1BZG	-113.43	-96.76	-96.76	-81.51
1C49	-96.67	-85.94	-85.94	-55.12
1C55	-148.63	-136.3	-136.3	-98.8
1C6W	-91.7	-79.76	-79.58	-34.04
1CBN	-199.7	-171.97	-174.54	-169.57
1CCM	-183.63	-162.4	-161.83	-146.02
1CE4	-101.1	-89.18	-88.68	-91.93
1CFG	-80.19	-70.51	-69.78	-58.67
1CIX	-161.74	-144.58	-143.97	-82.45
1CKW	-75.7	-67.63	-67.63	-65.48
1CKX	-89.64	-80.99	-80.98	-84.73