John von Neumann Institute for Computing

An Improved Evolutionary Strategy for Protein Structure Prediction

Srinivasa Murthy Gopal, Kyu Hwan Lee, Wolfgang Wenzel

published in

NIC Workshop 2006,
From Computational Biophysics to Systems Biology,
Jan Meinke, Olav Zimmermann,
Sandipan Mohanty, Ulrich H.E. Hansmann (Editors)
John von Neumann Institute for Computing, Jülich,
NIC Series, Vol. **34**, ISBN-10: 3-9810843-0-6,
ISBN-13: 978-3-9810843-0-6, pp. 149-152 , 2006.

http://www.fz-juelich.de/nic-series/volume34

An Improved Evolutionary Strategy for Protein Structure Prediction

Srinivasa Murthy Gopal[1], **Kyu Hwan Lee**[2], **and Wolfgang Wenzel**[1]

[1] Institut für Nanotechnologie,
Forschungszentrum Karlsruhe, 76021 Karlsruhe, Germany
E-mail: {*gopal, wenzel*}*@int.fzk.de*

[2] Korea Institute of Science and Technology,
Supercomputational Materials Simulation lab,
130650 Seoul, South Korea
E-mail: khlee@fireworks.kist.re.kr

We propose a new evolutionary strategy for protein structure optimization in a free energy landscape. This strategy is an improved version of existing basin hoping technique where multiple independent trajectories are used. In the new method, these trajectories depend on each other. This technique is implemented as a simple master-client model for distributed computing. We demonstrate advantage of purposed method in terms of computational effort and structure diversity.

1 Introduction

Protein structure prediction(PSP) is regarded as one of the grand computational challenge. Ab initio protein structure prediction is one promising approach to obtain the native protein conformation from first principles. We use a free energy forcefield[1] with combination of several optimization methods[2] to predict the native structure of protein. We identify the native structure as the lowest energy conformation in our forcefield. The basin hopping technique(BHT)[3] has been our work horse for the structure optimization. We have predicted native structures of several proteins including 20 amino acid trp-cage protein, 40-amino acid headgroup of the HIV accessory protein using the BHT[4]. Though BHT proved to be a good optimization method, it has several drawbacks. Typically the independent BHT trajectories find identical structures corresponding to one local(global) minima of free energy landscape. There exists also a problem of BHT simulations going astray. We had previously identified this problem and proposed an evolutionary strategy[5] which eliminates the problems associated with BHT. In this current work we propose an improved version for protein structure prediction.

2 Method

We have generalized the BHT approach to a population of size N which is iteratively improved by P concurrent dynamical processes. The population is evolved towards a optimum of the free energy surface with a simple evolutionary strategy(ES). The strategy balances the energy improvement as well as the population diversity. The conformations are drawn from the population and subjected to an annealing cycle. At the end of each cycle the

resulting conformation is either integrated into the *active* population or discarded. The algorithm was implemented as a simple master-client model in which the idle clients request a task from the master. The master maintains the active conformation of the population and distributes the work to the clients. Each step in the algorithm has three phases.

- *Selection* A conformation is drawn randomly from the *active* population. We have used a uniform probability distribution with active population of 20 conformers.

- *Annealing cycle* We used a geometric cooling schedule with T_{start} drawn from a uniform or exponential distribution and T_{end} fixed at 2K. The number of steps per cycle is increased as $10^5 \times \sqrt{N_{cycle}}$.

- *Population Update* We have adjusted the acceptance criterion for newly generated conformations to balance the population diversity and energy enrichment. We define the close structures as conformation which have bRMSD(back bone RMSD) of 3Åto each other. The master performs one of the following operations.

  - *Add* The new conformation is not *close* to any structure in the population, it is added to the pool.
  - *Replace* If the new conformation is *closest* to some structure in the population, it replaces that structure provided its re-weighted energy (see below) is less than the closest one.
  - *Merge* If the new conformation has several *close* structures, it replaces this group of structures provided its re-weighted energy is less than the best one of the group.

We have used an energy criterion for the *Replace* and *Merge* operations during population update. We re-weigh the energy of the new conformation(Erew) as

$$Erew = A \times \tanh D \qquad where \qquad D = \frac{Ebest - Enew}{A}$$

Enew is the actual energy of the new conformation, Ebest is the current best low energy.
We have also optimized the number of concurrent processes with respect to the size of *active* population. We have investigated the folding of a small beta peptide, tryptophan zipper (PDBID:1LE0) for this purpose. We fixed the size of active population to 20 and used 10,20,30,40,50,60 concurrent processes(Fig 1). We found an optimal number of processes to be approximately equal to two times the active population. We confirmed this for population size of 5 and 10.

3 Results

We have used the improved evolutionary strategy for predicting the native structure of a 12 amino acid tryptophan zipper(PDBID:1LE0). The table below shows the top 10 structures, their energies, secondary structure and bRMSD with respect to experimental structure.

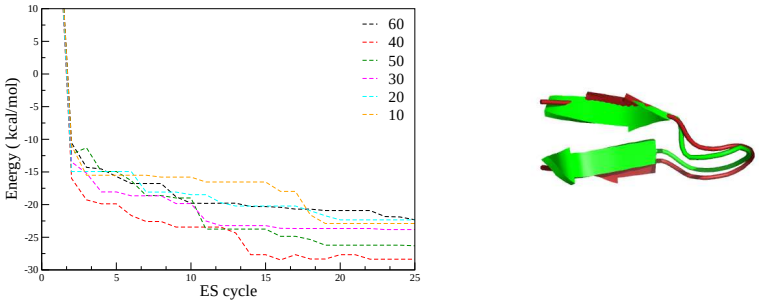

Figure 1. Left : The lowest Energy vs the ES cycle for different processes. Right : The predicted(red) and experimental(green) structures.

Name	Secondary structure	Energy (kcal/mol)	bRMSD Å
EXP	CEEECSSSEEEC		
1	CEEECSSCEEEC	-28.360	1.220
2	CCCEECSSCEEC	-21.350	3.800
3	CEEETTEEECCC	-19.470	3.790
4	CEESSSSSCEEC	-19.130	3.270
5	CCCEECSSCEEC	-19.040	3.630
6	CCCCCTTTTCCC	-18.820	6.170
7	CCCTTTTCCCCC	-18.450	4.510
8	CCCCBTTBCCCC	-18.120	3.360
9	CCCHHHHHHHHC	-17.850	6.880
10	CCCCCTTTTCCC	-17.600	6.370

4 Conclusions

We have developed and applied an improved evolutionary strategy which evolves a set of conformers to population of low energy diverse structures. We have optimized the algorithm and used it for tertiary structure prediction of a small beta hairpin. The method presented here is well suited for the distributed computational architecture.

Acknowledgments

We thank the BMBF, the Deutsche Forschungsgemeinschaft (grants WE 1863/10-2,WE 1863/14-1) and the Kurt Eberhard Bode Stiftung for financial support. We are grateful to S. Gregurick and J. Moult, for the contribution of part of code used in these simulations. We are thankful to KIST for computational resources.

References

1. T. Herges, W. Wenzel *An all-atom forcefield for tertiary structure prediction of helical proteins,* Biophysical Journal **87**, 3100 (2004) A. Verma and W. Wenzel (in preparation)
2. A. Schug, A. Verma, W. Wenzel and G. Schoen *Biomolecular structure prediction with stochastic optimization methods.*, Adv. Eng. Materials **7**, 1005 (2005)
3. A. Nayeem, J. Vila, and Harold A. Scheraga *A comparative study of the simulated-annealing and monte carlo-with-minimization approaches to the minimum-energy structures of polypeptides: [met]-enkephalin*, J. Comp. Chem **12**, 594 (1991)
4. A. Schug, A. Verma, K. H. Lee and W. Wenzel *Basin Hopping Simulations for All-Atom Protein Folding*, J. Chem. Phys. **(in press)**, (2006)
5. A. Schug and W. Wenzel *An Evolutionary Strategy for All-Atom Folding of the 60-Amino-Acid Bacterial Ribosomal Protein L20*, Biophysical Journal **90**, 4273 (2006)