A PARALLEL HYBRID GA FOR PEPTIDE 3-D STRUCTURE PREDICTION

Carlos A. DEL CARPIO and Shin-ichi SASAKI[1] Lajos BARANYI and Hidechika OKADA [2]
carlos@translell.tutkie.tut.ac.jp Compuserve:100210.1512

[1] Department of Knowledge-based Infofrmation Engineering
Toyohashi University of Technology
Tempaku. Toyohashi, 441 Japan

[2] Department of Molecular Biology Nagoya City University School of Medicine
Mizuho-ku, Mizuho-cho. Nagoya 467, Japan

Abstract

The present work describes recent advances made in the system for 3-D structure prediction of polypeptides being developed in our laboratory. The system was originally conceived as a conformational space search procedure based on a simple genetic algorithm. However, the complexity of the system and the need to produce better fit conformers as artificial evolution proceeds, compelled us to improve the algorithm in two substantial aspects. The first is a parallelization of the original algorithm to enrich the diversity of conformers in the population and the second a hybridization of the original GA in order to process the atoms of the side chains.
The results are exemplified with the prediction of the 3D structure for CRAMBIN.

1 Introduction

An evolutionary algorithm for peptide 3-D structure prediction [1,2,3] has been under development in our laboratories during the last two years. The characteristics of the algorithm are essentially the coding of each 3-D conformer as a list of the dihedral angles of amino acids constituting the primary sequence of the peptide (chromosome), and the fitness or function to examine each chromosome and decide on its survival, reproduction or mutation in following generations. This evaluation function takes into consideration several factors in protein folding.
Parallelizing genetic algorithm (GA) and the implementation of the parallel version on a network of transputers increased the variety of conformers within the population at each generation of the GA. The second aspect of the improvement is concerned with the conformation of the side chains within the polypeptide. The original algorithm leaves the configuration of the side chains unchanged through the genetic process. We propose a hybrid GA to relax side chains after a GA operation is performed.

2 THE HYBRID GENETIC ALGORITHM

The initial population of chromosomes is created linking amino acid structures following the order of the primary sequence using pairs of dihedral angles taken from a steric contour diagram for every amino acid. At the initial stage, each amino acid is represented by an energetically optimized structure extracted from a file containing the 20 basic optimized amino acid structures.
In the present study, we suggest a new methodology for side chain packaging by hybridizing the tools of the GA used so far (rotation, translation, etc.) by optimization of the geometry of the particular conformer. This is

[1] Del Carpio C.A.

performed by minimization of its steric energy. This minimization is performed by the Newton-Raphson second derivative method[4]. However a restriction is imposed on the type of atoms allowed to move in the course of this local minimization process. All the atoms of the side chains are allowed to move while those of the main chain are kept static. This minimization leads to a local energy minima. With the local minimization the protein conformers constituting the population of a determined generation are refined both, energetically and structurally.

3 PARALLELIZATION OF THE HYBRID GA

The evaluation of the potential function, requires intensive computation [5]. The terms which are especially computer power consuming are the local minimization of the potential function, and the calculation of the solvent accessible surface area.
To speed up calculations, we have developed a parallel version of GA and implemented it on a network of transputers.
The architecture of a transputer, consists of a processor with 4 hardware channels that allows it to transfer and receive data to or from any other four processors. The algorithm developed here, in its simplest fashion runs on a net of five transputers. The population of the GA is divided among the processors.
The root processor where the master process runs, reads data from the host, performs assignment of parameters to the molecule, and transmits data to the slave processes to perform potential function evaluations. The slave processor on the other hand, performs the task of sending data to any of its four neighbour transputer, or receive data from any of them.
A generation is built in the network selecting the best individuals from its subpopulation of individuals, and sends it to the neighboring transputers. Similarly, it gets the signal to receive data from any of its neighbors .

4 RESULTS AND CONCLUSIONS

We illustrate the performance of the new version of the system predicting the structure of the protein CRAMBIN, and we compare the automatically predicted structure with the native crystal structure recorded in the PDB. CRAMBIN is composed of 36 amino acids.
A comparison of the best fit individual after the 500th generation obtained by the hybrid GA reveals more accurately the level of evolution of the process. Superposition of both conformers results in a RMS of 6.69 A, which although still high, shows the tendency of the molecule to approach that of the crystal structure. A further inspection of the conformers shows patterns of substructures of high similarity.
Fragments as large as 30 atoms of the main chain when superimposed yield a RMS of 1.796 A. Comparison of other less similar conformers reveal that these substructural patterns are also being formed and that the GA as a whole progresses in the direction of more stable conformers, i.e. conformers closer to a minimum.
In conclusion, the local optimization of conformers generated by the GA, proposed here, seems to improve the performance of the program significantly. We hope that upon further paramenter optimization, it will yield a new tool for predicting the 3-D structure of peptides.

References

[1] Uchida M., Del Carpio C., Sasaki S.; Proceedings of the 17th Symposium on Chemical Information(JCS); 1994 (82-85).

[2] Del Carpio C., Uchida M., Sasaki S., Baranyi L., Campbell W., Okada H.; Pr oceedings of the 17th Symposium on Chemical Information(JCS); 1994(126-129).

[3] Del Carpio C., Sasaki S., Proceedings of the XIth International Conference on Computers in Chemical Research and Education. Paris-France 1995

[4] Allinger N.L, Tribble M.T., Miller D.H.; J. Am. Chem. Soc.; 1971; 93(1637)

[5] Del Carpio C. A., Sasaki S., Baranyi L., Okada H.; J. of Chem. Inf. and Co mp. Sci. (In press)