Recent Improvements in the Automatic System for Protein Tertiary Structure Prediction GAX

Carlos Adriel Del Carpio Munoz
carlos@translell.eco.tut.ac.jp
Department of Ecological Engineering, Toyohashi University of Technology, Tempaku, Toyohashi 441-8580, Japan

1 Introduction

In recent years the author has been concentrated in the development of the automatic system for protein 3-D structure prediction GAX [2,3]. The system, having as backbone a configurational search engine based on the genetic algorithm (GA) paradigm, optimizes a potential function which expresses the internal energy of hundreds of conformers at successive stages of the GA process. According to previous reports, structures of the last generation of evolved structures contain conformers highly similar to the reported crystallographic data.

The system however has been improved in several aspects, leading to an overall improvement of performance. Especially improvements both in the quality of the results and the processing times required for an specific job is remarkable.

Here I report some aspects related to those improvements, and present some examples to illustrate them. The main aspects discussed here are 1) the prediction of regions of defined secondary structure, 2) Improvement of the potential function by explicitly taking into account the solvent effect, 3) prediction of turns using distance geometry techniques, 4) Improvement of the final coordinates by molecular dynamic simulations.

2 Methodology

2.1 Prediction of regions of defined secondary structure

Several works to predict regions of defined secondary structure such as α helices or β sheets have been reported in the literature. On the other hand the number of protein crystallographic data is increasing steadily allowing the use of experimental data in more efficient ways. Here we describe the incorporation of the prediction of α helices for regions of proteins of high amphipatic amino acid contents. Similarly, we discuss the retrieval of fragments in the data base which under different environments adopt similar secondary structure characteristics.

2.2 Incorporation of the solvent effect into GAX potential function

I report here on the improvement of the potential energy function of GAX by introducing a solvent effect model which takes explicit account of the solvent protein interactions, with an emphasis of the hydrophobic effect on the folding process of a protein. The quantification of this effect, always qualitatively mentioned is one of the aspects of relevance in this solvent effect model.

2.3 Prediction of turns using distance geometry

Distance Geometry (DG) has been a technique used for prediction of 3D structures of small molecules, as well as in ligand docking processes [1]. Here we adopt the technique to rationally predict regions

of not well defined secondary structure such as turns or completely undefined secondary structures. I believe that these type of structures are consequence of the interaction within the molecule of well defined substructures such as α helices, sheets and coils in order to optimize their positions inside the protein. Similarly, I believe that cooperativity among these types of substructures leads also to disruption of part of the internal structure during the folding process.

2.4 Output refinement by molecular dynamics

Refinement of the fittest conformers of the last generation of the GA is performed with the help of a system for molecular dynamic simulations DYNAX developed in our laboratories [4]. The refined structures are closer to the crystallographic data.

3 Results and Discussion

I report here the prediction of the 3D structures for several medium sized proteins using the modified potential function (i.e. including explicitly the solvent effect), and then refining the fittest conformers of the GA by performing a molecular dynamics simulation. The results are shown in Fig. 1, where the structure predicted by GAX is shown for the protein VEGF(PDB-CODE:2VGH) of 55 amino acids long (left). The Dynax refined structure is shown in the middle, while the native is illustrated in the right. Superposition of the refined structure with the experimentally reported yields an RMS of 9.6 Å.

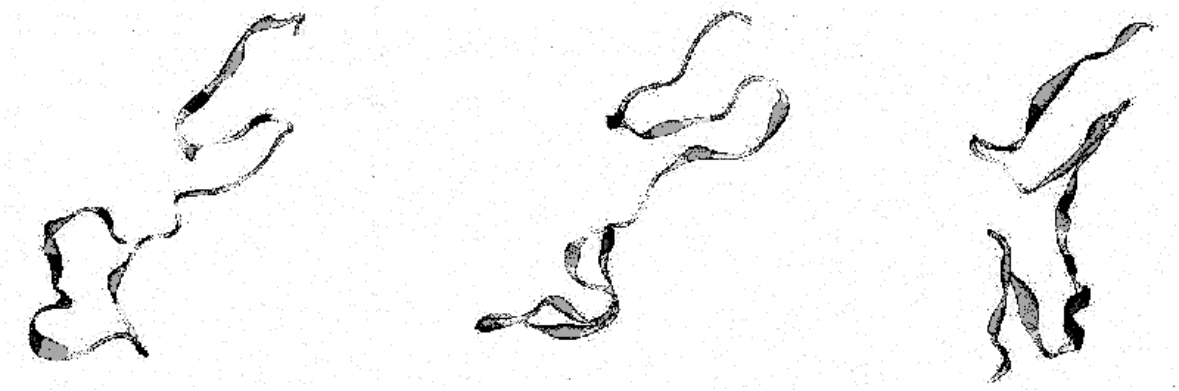

Figure 1: Predicted and observed 3D structures for VEGF. GAX (left), Dynax (middle), Observed (right).

References

[1] Crippen, G.M. andHavel, T.F., *Distance Geometry and Molecular Conformation*, Research Studies Press Ltd., John Wiley and Sons Inc., 1988.

[2] Del Carpio, C.A., A parallel genetic algorithm for polypeptide three dimensional structure prediction, a transputer implementation, *J. Chem. Inf. Comput. Sci.*, 36(2):258–269, 1996.

[3] Del Carpio, C.A. and Gogonea, V., Modelling proteins conformation in solution, Part I: a parallel GA engine for protein conformational space mapping, *Genome Informatics 1996*, Universal Academy Press, 108–118, 1996.

[4] Yoshimori A., and Del Carpio, C.A., DYNAX: a new system for biomolecular dynamics simulation in solution, *Genome Informatics 1998*, Universal Academy Press, 374–375, 1998.