Poster Presentation: Protein Folding Simulation in CCP

Alessandro Dal Palù[1], Agostino Dovier[1], and Federico Fogolari[2]

[1] Dip. di Matematica e Informatica, Univ. di Udine
Via delle Scienze 206, 33100 Udine (Italy). (dalpalu|dovier)@dimi.uniud.it
[2] Dip. Scientifico-Tecnologico, Univ. di Verona
Strada Le Grazie 15, 37134 Verona (Italy). fogolari@sci.univr.it

Background. A protein is a list of linked units called aminoacids. There are 20 different kinds of aminoacids and the typical length of a protein is 100–500 units. The Protein Structure Prediction Problem (PSP) is the problem of predicting the 3D native conformation of a protein, when its aminoacid sequence is known. The process for reaching this state is called the *protein folding*. This work deals with *ab-initio* prediction, based on *Anfinsen thermodynamic hypothesis* [1] which states that the conformation adopted by a protein (also known as the *native* conformation) is the one with minimum free energy. We can identify two main problems: the first is to choose a representation of the protein and an energy function, which must be at minimum for native-like conformations. The second is, given the representation and the energy function, to find the 3D conformation that minimizes the function. Due to intrinsic computational limits, no general solution to the latter problem is currently available. In particular, simulation-based techniques that take into account all atoms constituting the aminoacids (and the solvent) and simulate the folding process approximating atom interactions, run extremely slow due to the huge number of calculations.

Results. In this work we adopt a simplified representation of a protein where each aminoacid is represented by a center of interaction. This simplification, while loosing some details, has the advantage of being computationally tractable and of having smoother potential landscapes. A recently developed empirical contact energy function [3], already used in constraint-based approach to the problem [4], is modified and augmented by local terms which describe bond lengths, bend angles, and torsion angles. The energy term related to bond lengths is designed to keep the distance of two consecutive aminoacids in the chain fixed. The bend angle and the torsion angle potentials are a statistical mean of the behaviour of the aminoacids, extracted from analysis performed on the Protein Data Bank [2].

With this energy function the problem is approached by a high-level simulation method which makes use of concurrent constraint programming. Basically, each aminoacid of the protein is viewed as an independent process that moves in the space and communicates with other aminoacids. Each process waits for a communication of the modification of other processes' position; after receiving a message, it stores the information in a list and performs a move. The new position is computed using a Montecarlo simulation, based on the spatial information

available to the aminoacid, which may not be the current dislocation of the protein, due to asynchrony in the communication. Once the move is performed, the aminoacid communicates its new position to all the others. The code has been implemented in Mozart [5], where it is natural to model concurrent programs in a simple notation mixing classes, constraints, and logic variables. During the computation each process keeps track of the whole history of the folding. There are various parameters to be correctly set to properly model the system; with the currently computed values we are able to properly obtain helices (see figure on the right). There are also difficulties that arise due to concurrency; among them there is the asyncronization between the communication and the moves of the aminoacids, which lead to modification of the energy landscape.

Conclusions. We have presented a preliminary concurrent constraint implementation of protein folding simulation. The results are promising. As far as we know, all current parallel approaches to the Protein Folding are simply a division of the computational load between the processors, while here we are starting to model the problem in a real concurrent setting. We wish to concentrate our efforts in this direction, developing more sophisticated communication strategies and especially some cooperative approach, in which the cooperation strategy adapts to the current configuration. The target is to derive a model that is able to represent the dynamical evolution of the system and that exploits these features to perform faster searches for the minimum energy point in the space of conformations. However, the actual energy function is too simple to capture all the complex interactions between the aminoacids; for this reason we would like to introduce the representation of the so-called *side chain* of an aminoacids and to use some already tested potentials. Moreover, we are planning to integrate this software with other tools, both from the constraint minimization side and from the molecular dynamics side.

References

1. C. B. Anfinsen. Principles that govern the folding of protein chains. *Science*, 181:223–230, 1973.
2. H. M. Berman et al. The Protein Data Bank. *Nucleic Acids Research*, 28:235–242, 2000. `http://www.rcsb.org/pdb/`.
3. M. Berrera, H. Molinari, and F. Fogolari. Amino acid empirical contact energy definitions for fold recognition in the space of contact maps. *BMC Bioinformatics*, 4(8), 2003.
4. A. Dal Palù, A. Dovier, and F. Fogolari. Protein folding in $CLP(\mathcal{FD})$ with empirical contact energies. In *Recent Advances in Constraints*, vol. 3010 of *Lecture Notes in Artificial Intelligence*, pp. 250–265, 2004.
5. Univ. des Saarlandes, Sweedish Inst. of Computer Science, and Univ. Catholique de Louvain. The Mozart Programming System. `www.mozart-oz.org`.