Modeling protein flexibility by distance geometry

Mathilde Le Boudic-Jamin, Antonio Mucherino, Rumen Andonov

Modeling protein flexibility by distance geometry

Mathilde Le Boudic-Jamin, Antonio Mucherino, Rumen Andonov

IRISA, University of Rennes 1, Rennes, France.
{mathilde.le_boudic-jamin,antonio.mucherino,rumen.andonov}@irisa.fr

Mots-clés : *protein conformation, protein flexibility, distance geometry, combinatorial optimization, protein docking.*

1 Introduction

The Molecular Distance Geometry Problem (MDGP) is the problem of finding the possible three-dimensional conformations of a molecule for which a certain subset of its atoms satisfies some distance constraints. A great interest is given to this problem because experimental techniques such as the Nuclear Magnetic Resonance (NMR) are able to estimate some distances between pairs of atoms of the molecule, which can then be exploited for reconstructing the possible conformations associated to a given molecule. Much attention is given to proteins, because they are biological molecules that are able to perform several important functions in the cells of living beings.

Due to the great interest given to the MDGP, many approaches for the solution of this problem have been proposed over the last years. By its definition, the MDGP is a constraint satisfation problem, but it is usually reformulated as a global optimization problem in a continuous space, where a penalty function is used for measuring the satisfation of the constraints. The interested reader is referred to a recent survey for a wide discussion on the different approaches to the MDGP [2].

In this context, we are particularly interested in the approach to the MDGP recently proposed in [1], the Discretizable MDGP (DMDGP). The search space of the optimization problem is here transformed in a discrete space, so that the problem becomes combinatorial. Even if the NP-hardness of the problem is not lost while performing the reformulation, the combinatorial structure of the new search domain makes it possible to use an efficient Branch & Prune (BP) algorithm for the solution of MDGPs that can be discretized. The BP algorithm is the first algorithm potentially able to identify all possible solutions to the problem. As explained in the following, we can exploit this property of BP for modeling the natural flexibility of proteins.

2 Protein flexibility

Even if the BP algorithm was initially tested on artificially generated instances, where the assumptions needed for the discretization were always satisfied, it was recently proved in [3] that its extension *i*BP (*interval* BP) can be used for solving DMDGPs containing real NMR data. This is possible because of a special ordering that have been identified for the atoms forming protein backbones. In this special ordering, atoms are re-arranged so that the distance between each atom and its three immediate predecessors in the ordering is always known. Moreover, the ordering ensures that only one of these three distances may be imprecise, while all others are exact (distances between chemically bonded atoms, or between pairs of atoms bound to a common atom). In this situation, a finite set of possible coordinates can be computed for each atom of the molecule.

We artificially generated some instances of the DMDGP with the technique detailed in [3]. These instances simulate NMR data for short peptides containing 3 amino acids, and use

intervals to define distances between hydrogen atoms (which are likely to be estimated by NMR). We defined different instances with different lengths $\ell \in \{1.0\text{Å}, 0.8\text{Å}, 0.5\text{Å}\}$ for the intervals corresponding to distances between hydrogens. As expected, when we tried to solve these instances with the iBP algorithm, the total number of solutions we found was different for each instance, and it increased with the length ℓ of the intervals. For a peptide formed by 3 amino acids, we obtained 8 solutions when ℓ was 0.5, 155 solutions when ℓ was 0.8, and 2368 solutions when ℓ was 1.0.

This simple experiment shows that modifying the lengths of the intervals allows to model the flexibility of the molecule. When iBP is invoked for solving the problem, the flexibility is reflected by the increased total number of solutions, while the solutions corresponding to shorter intervals keep being contained in the final solution set. Naturally, for longer peptides, the number of solutions should not increase exponentially with ℓ, because of the presence of short range distances between pairs of amino acids which are far in the sequence. Nevertheless, we believe that DMDGPs can be used as a valid model for the flexibility of compact molecules such as proteins.

We are starting to work in order to exploit this property of the iBP algorithm in the context of protein docking. Given a protein A and a protein B with known conformations, the conformation of the two proteins $A+B$, during their interaction, is searched [5]. During the interaction, it is evident that the surfaces of the two molecules may slightly modify their conformations. We will attempt the generation of all suitable three-dimensional conformations for A+B by finding all solutions to a DMDGP where the flexibility of the two molecules is regulated by the intervals used to replace computed distances. This is one of our main research directions.

3 Conclusions

We discussed about the possibility of modeling the natural protein flexibility by an MDGP where known exact distances between atoms are substituted by suitable intervals. Among the approaches to the MDGP proposed over the years, we considered a recent combinatorial approach, where the search domain of the formulated optimization problem is discrete. Differently from other methods for the MDGP which are based on other approaches, this combinatorial approach allows for enumerating all possible conformations for a given instance of the problem. As a consequence, when intervals representing distances are enlarged for modeling flexibility, the total number of solutions for the instance increases, because the molecule is actually free to assume a larger range of conformations. One possible application of this idea is to protein docking, where the interacting surfaces of the two molecules are in fact flexible.

Références

[1] C. Lavor, L. Liberti, N. Maculan, A. Mucherino, *The Discretizable Molecular Distance Geometry Problem*, to appear in Computational Optimization and Applications, 2012. *Available on the Internet on the "Online First Articles" section of the journal.*

[2] C. Lavor, L. Liberti, N. Maculan, A. Mucherino, *Recent Advances on the Discretizable Molecular Distance Geometry Problem*, to appear in European Journal of Operational Research, 2012.

[3] C. Lavor, L. Liberti, A. Mucherino, *The interval Branch-and-Prune Algorithm for the Discretizable Molecular Distance Geometry Problem with Inexact Distances*, to appear in Journal of Global Optimization, 2012.

[4] N. Tuncbag, G. Kar, O. Keskin, A. Gursoy, R. Nussinov, *A Survey of Available Tools and Web Servers for Analysis of Protein-Protein Interactions and Interfaces*, Briefings in Bioinformatics **10**(3), 217–232, 2009.