A Discrete Approach for Finding the Conformation of Molecules from NMR Data

Antonio Mucherino[1], Carlile Lavor[2], Leo Liberti[3], Nelson Maculan[4]

[1] INRIA, Lille Nord Europe, antonio.mucherino@inria.fr
[2] Dept. of Applied Mathematics, State University of Campinas, Brazil, clavor@ime.unicamp.br
[3] LIX, École Polytechnique, Palaiseau, France, liberti@lix.polytechnique.fr
[4] Systems Engineering, Federal University of Rio de Janeiro, Brazil, maculan@cos.ufrj.br

Mots-Clés : *distance geometry, combinatorial optimization, molecules, NMR experiments*

Experiments of Nuclear Magnetic Resonance (NMR) spectroscopy are able to provide information from which some distances between pairs of atoms of a molecule can be estimated. These distances, together with some geometric constraints that the molecule must satisfy, can be used for finding the coordinates of all its atoms, i.e. its three-dimensional conformation. This problem is known in the literature as the Molecular Distance Geometry Problem (MDGP), which is NP-hard.

The MDGP is usually formulated as a global continuous optimization problem, where the considered *potential energy* only encodes geometric structure information, and possible artifacts introduced by using molecular force fields are avoided. As a consequence, the objective function is usually the combination of a given number of penalty terms, where each of them provides a measure of how much a certain geometric constraint is violated. The minimization of all such violations can bring to the identification of the three-dimensional conformation of the molecule.

Many methods have been proposed for the solution of this optimization problem. As an example, the package Xplor-NIH [8] has been particularly designed for solving MDGPs arising from NMR experiments. It makes use of heuristic methods (such as Simulated Annealing) and local search methods (such as Conjugate Gradient Minimization) for finding possible solutions to the optimization problem. Other methods have been also proposed, which are based on more complex techniques, but they have been tested on theoretical instances only. Some recent examples are [1, 2].

The information provided by NMR experiments is usually limited. Only distances shorter than 6Å are usually found, and the majority of such distances regard hydrogen atoms only. Therefore, the task of positioning all the atoms of a molecules (which may contain also carbons, nitrogens, oxygens, and so on) can be very difficult. Nowadays, molecular conformations obtained through NMR experiments are not precisely identified, but more than one possible model is usually found. In general, from 2 to over 100 models can be presented for the same molecule (the most common number of provided models is 20). The variability among the different models can be due to the uncertainty in positioning some or even all the atoms of the molecule. Moreover, a recent deep study on models obtained by NMR showed that they are very likely to contain major errors [7].

The aim of our work is to apply a discrete approach for the MDGP in order to find good-quality solutions of instances related to NMR experiments. To this purpose, we consider the DMDGP [5], which consists in a subclass of instances of the MDGP for which the problem can be discretized. After the discretization, the MDGP is formulated as a combinatorial optimization problem, that

is NP-hard. However, this combinatorial problem can be solved efficiently by an algorithm called *Branch & Prune* (BP) [6], which is able to find better-quality solutions with respect to the ones provided by methods based on a continuous formulation.

The main problem to be solved is to build instances of the MDGP which are related to NMR data that satisfy the assumptions of the DMDGP. This is not trivial, because of the limited information available from NMR experiments. We propose the following strategy for finding the three-dimensional conformation of molecules from NMR data.

Since the majority of the distances obtained through NMR are distances between pairs of hydrogens, we firstly focus our attention on the problem of finding the coordinates of all the hydrogen atoms forming a given molecule, where only data from NMR experiments are considered. A suitable ordering (if it exists) can be searched for the hydrogens of the molecules, in order to have the assumptions of the DMDGP satisfied, so that we can apply the BP algorithm for the identification of their coordinates. In [3], a particular ordering has been presented that allows the combinatorial reformulation for instances related to the hydrogen atoms of *protein backbones*.

Moreover, in [4], it has been shown that all the atoms of a protein backbone can be reconstructed by using the coordinates of the hydrogens (found by the BP algorithm) and information on bond lengths and angles. In order to compute all the coordinates of these atoms, only a sequence of small linear systems needs to be solved.

We are currently working for generalizing this strategy to entire protein conformations and to molecules in general. The main difficulty is that NMR data can be affected by noise and the measurements may contain some errors. However, we believe this discrete approach to the problem could be able to provide three-dimensional conformations of molecules which are more accurate than the models that are currently available. We plan to reach this aim in future works.

Références

[1] P. Biswas, K.-C. Toh, Y. Ye, *A Distributed SDP Approach for Large-Scale Noisy Anchor-Free Graph Realization with Applications to Molecular Conformation*, SIAM Journal on Scientific Computing **30** (3), 1251–1277, 2008.

[2] A. Grosso, M. Locatelli, F. Schoen, *Solving Molecular Distance Geometry Problems by Global Optimization Algorithms*, Computational Optimization and Applications **43**, 23–27, 2009.

[3] C. Lavor, A. Mucherino, L. Liberti, and N. Maculan, *Computing Artificial Backbones of Hydrogen Atoms in order to Discover Protein Backbones*, IEEE Conference Proceedings, IMCSIT09 (WCO09), Poland, 751-756, 2009.

[4] C. Lavor, A. Mucherino, L. Liberti, and N. Maculan, *An Artificial Backbone of Hydrogens for Finding the Conformation of Protein Molecules*, Proceedings of the Conference CSBW09, Washington D.C., USA, 152–155, 2009.

[5] L. Liberti, C. Lavor, N. Maculan, *Discretizable Molecular Distance Geometry Problem*, Tech. Rep. q-bio.BM/0608012, arXiv, 2006.

[6] L. Liberti, C. Lavor, and N. Maculan, *A Branch-and-Prune Algorithm for the Molecular Distance Geometry Problem*, International Transactions in Operational Research **15** (1), 1–17, 2008.

[7] S.B. Nabuurs, C.A.E.M. Spronk, G.W. Vuister, G. Vriend, *Traditional Biomolecular Structure Determination by NMR Spectroscopy Allows for Major Errors*, PLoS Computational Biology **2**(2), 71–79, 2006.

[8] C.D. Schwieters, J.J. Kuszewski, G.M. Clore, *Using Xplor-NIH for NMR Molecular Structure Determination*, Progress in Nuclear Magnetic Resonance Spectroscopy **48**, 47–62, 2006.