Systematic Search of Constrained Molecular Conformations

Lisa Tucker-Kellogg, Bruce Tidor & Tomás Lozano-Pérez

Artificial Intelligence Laboratory
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: Accurate determination of the three-dimensional atomic structures of molecules is currently a bottleneck in drug design research and is a crucial part of understanding the molecular mechanisms of biological systems. The goal of this project is to systematically enumerate feasible conformations (three-dimensional configurations) of a flexible molecule, subject to a given set of constraints on the distances between atoms and on the torsional angles of rotatable bonds.

Without any constraints, the conformation space of a flexible molecule is astronomical (exponential in the number of degrees of freedom). Thus any naive, exhaustive search would be intractable. However, depending on the given set of constraints, the region of conformation space that satisfies the constraints may be small enough to search systematically.

Motivation: There are many molecular modeling goals and medically important problems involving molecular structure determination that can be expressed as determining the set of structures that satisfy a given set of constraints.[3] For example, NMR spectroscopy data can provide bounds on certain interatomic distances in a protein.

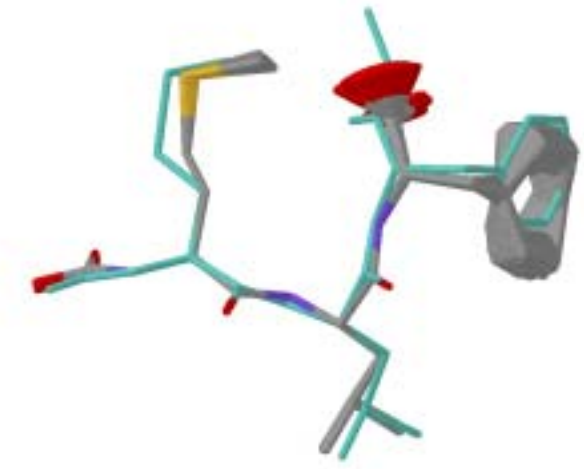

Figure 1: A family of 56,975 fMLF structures consistent with the constraint data. For comparison, the crystal structure[2] of the methyl ester analogue (fMLF-OMe) is also shown in aqua.

Approach: Assume that the large, high-dimensional space of conformations is subdivided into small hypercubes at some predetermined level of resolution. One possible approach for a systematic search would be to output the set of all hypercubes that contain at least some satisfying conformation(s). A similar but more useful goal is to output one satisfying conformation from each hypercube in that set. Throughout this work, single conformations are used to represent continuous regions of conformation space.

There are three conceptual layers to our search algorithm. The outer layer implements a divide-and-conquer scheme: it splits the molecule into small subchains, calls the main conformational search engine on the small subchains, and combines those results to create the set of conformations for intermediate subchains. Eventually, larger subchains are combined to create the set of conformations for the whole molecule.

The main, middle layer is a conformational search engine that iterates over possible ranges of values, for each rotatable bond in a molecule or sub-molecule, thus dividing the search space of each subchain into hypercubes.

The lowest-level layer takes each combination (a hypercube) generated by the other layers and evaluates that small region of space by heuristically searching for satisfying conformations. We have no *systematic* method for evaluating hypercubes, but our heuristic search uses minimization and randomization to explore the space. For sufficiently small and low-dimensional spaces it appears to find any satisfying conformations that exist. However, performance degrades with increasing size and number of dimensions. If a hypercube contains some very narrow region that satisfies the constraints, our search might, if it never finds that region, conclude incorrectly that the hypercube is infeasible.

The advantage our our multi-layered approach is that many infeasible solutions can be ruled out by considering only small subchains, thus reducing the amount of computation performed on the whole molecule.

When ruling out branches of the space as infeasible, efficiency is improved when large regions are excluded *early* during the search. Therefore, at the divide-and-conquer layer, we attempt to precompute which choice of subunits would be most highly constrained, which pairs of subunits would be most highly constrained, and so on, to create the best *merge tree*.

Impact: Experimental NMR constraints are typically interpreted using one of a family of stochastic "distance geometry" algorithms.[1] Although empirically they are very successful, they are at best accompanied by only probabilistic guarantees of completeness or accuracy. If we can perform our systematic searches at high enough resolution, our method may reveal alternative interpretations of NMR data that the popular, stochastic search algorithms have overlooked. Even when existing methods are correct, our method's enumeration will be more complete and more thorough.

Future Work: Having succeeded with the systematic search of a small peptide, we must investigate how the method scales to larger systems and assess the possibility of using the algorithm for full-scale protein structure determination.

Research Support: This work is supported by the Singapore-MIT Alliance and by the Cecil and Ida Green Chair Fund.

References:

[1] G.M. Crippen and T.F. Havel. *Distance Geometry and Molecular Conformation*. Wiley, 1988.

[2] E. Gavuzzo, F. Mazza, G. Pochetti, and A. Scatturin. Crystal structure, conformation, and potential energy calculations of the chemotactic peptide N-formyl-L-Met-L-Leu-L-Phe-OMe. *Int. J. Pept. Protein Res.*, 5:409–415, 1989.

[3] G.P. Gippert, P.E. Wright, and D.A. Case. Distributed torsion angle grid search in high dimensions: a systematic approach to nmr structure determination. *J. Biomolec. NMR*, 11:241–263, 1998.

[4] C.M. Rienstra, L. Tucker-Kellogg, C.P. Jaroniec, M. Hohwy, B. Reif, M.T. McMahon, B. Tidor, T. Lozano-Perez, and R.G. Griffin. De novo determination of peptide structure with solid-state magic-angle spinning nmr spectroscopy. *Proc. Natl. Acad. Sci. USA*, 99(16):10260–65, Aug 2002.