Systematic Conformational Search with Constraint Satisfaction

Lisa Tucker-Kellogg

Artificial Intelligence Laboratory
Massachusetts Institue Of Technology
Cambridge, Massachusetts 02139

http://www.ai.mit.edu

The Problem: 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: NMR (nuclear magnetic resonance) spectroscopy experiments yield constraints on the conformations of biologically interesting molecules. For example, NMR data might provide bounds on certain interatomic distances in a protein. Our most immediate motivation is to use systematic conformational search for interpreting NMR data about a small peptide called fMLF. Solid-state NMR experiments performed by collaborators are yielding constraints on interatomic distances, torsion angles, and pairs of torsion angles for fMLF. More generally, there are many molecular modeling goals and structure determination problems that can be expressed as determining the set of structures that satisfy a given set of constraints.

Previous Work: A number of systematic conformational search methods have been developed as greater computational resources have become available. Pincus, Klausner & Scheraga[1] proposed a divide-and-conquer approach wherein a molecule is divided into small subunits, the conformations of each subunit are optimized individually, and then conformations for the whole molecule are constructed by combining optimized conformations for the subunits. Rather than using all feasible conformations for their subunits, they choose only the conformations with optimized solvent stabilization and other empirical features.

Gippert *et al.* [2, 3] created a set of systematic search methods aimed at interpreting experimental constraints from NMR spectroscopy. One very efficient method is only applicable to linear chains of torsional angles. Another method takes an arbitrary molecule, applies the linear-chain method to each linear subsegment, and then combines the linear search results at each branching point to create a systematic exploration of the whole molecule.

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 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. The outer layer implements a divide-and-conquer scheme: it splits the molecule into small subchains, calls the main conformational search engine on each subchain, then combines those results to create the set of conformations for a larger subchain. Eventually, larger subchains are combined to create the set of conformations for the whole molecule. The lowest-level layer takes each combination generated by the other layers and evaluates that small region of space by heuristically searching for satisfying conformations. If this layer determines a region or subregion to be infeasible, the other layers may be able to avoid searching larger regions of conformation space that depend on the existence of a solution in that subregion. Thus, branches of the search can be pruned in advance.

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*.

Difficulty: We have no systematic method to evaluate a small continuous region of conformation space. 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, but performance degrades with increasing size and number of dimensions. When a hypercube contains some very narrow region that satisfies the constraints, our search might, if it never finds that region, conclude that the hypercube is simply infeasible.

Another obstacle to this research is the amount of time required to perform the portions of the search that are still exponential. Even if a set of constraints overdetermines the placement of a molecule's "core" or backbone bonds, the exterior sidechain or "loop" bonds could be underdetermined, and their degrees of freedom would still create an exponentially large search space.

Impact: Experimental NMR constraints are typically interpreted using one of a family of stochastic "distance geometry" algorithms.[4] 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.

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.

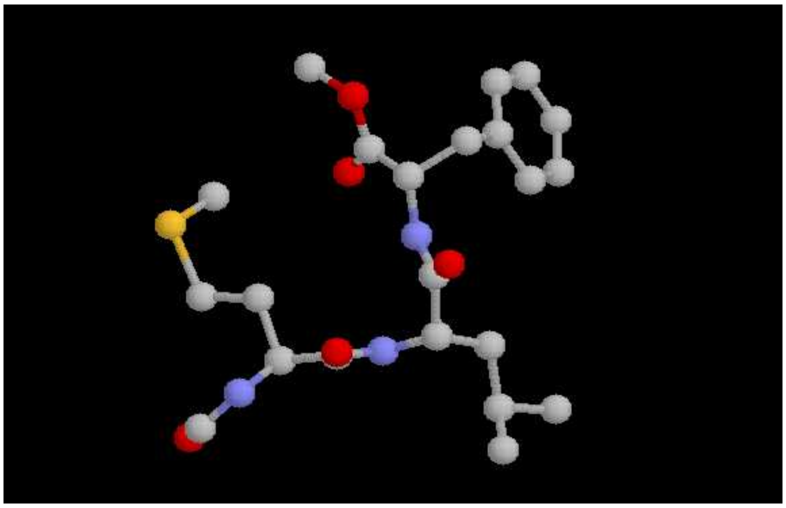

Figure 1: The conformation of fMLF-OMe, an analogue of fMLF, as determined by X-ray crystallography[5].

Future Work: We are primarily working on improving the low-level heuristic search and on demonstrating the benefit of optimized merge-trees. We will also demonstrate that our results do not depend on our choice of simplifying assumptions about molecular geometry.

Research Support: This work was supported by the Cecil and Ida Green Chair Fund.

References:

[1] M.R. Pincus, R.D. Klausner and H.A. Scheraga. Calculation of the three-dimensional structure of the membrane-bound portion of melittin from its amino acid sequence. *Proc. Natl. Acad. Sci. USA* (1982) 79: 5107–5110.

[2] 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* (1998) 11: 241–263.

[3] G.P. Gippert, P.F. Yip, P.E. Wright and D.A. Case. Computational methods for determining protein structures from NMR data. *Biochem. Pharmacol.* (1990) 40: 15–22.

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

[5] 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.* (1989) 5:409–415.