

The Solution of the Distance Geometry Problem for Protein Modeling Using Geometric Buildup

Zhijun Wu

Department of Mathematics
Program on Bioinformatics and Computational Biology
Iowa State University

A well-known problem in protein modeling is the determination of the structure of a protein with a given set of inter-atomic or inter-residue distances obtained from either physical experiments or theoretical estimates. A general form of the problem is known as the distance geometry problem in mathematics, the graph embedding problem in computer science, and the multidimensional scaling problem in statistics. The problem has applications in many other scientific and engineering fields as well such as sensor network localization, image recognition, and protein classification. We describe the formulations and complexities of the problem in its various forms, and introduce a geometric buildup approach to the problem. Central to this approach is the idea that the coordinates of the atoms in a protein can be determined one atom at a time, with the distances from the determined atoms to the undetermined ones. The determination of each atom requires the solution of a small system of distance equations, which can usually be obtained in constant time. Therefore, in ideal cases, the coordinates of n atoms can be determined by a geometric buildup algorithm with $O(n)$ distances in $O(n)$ computing time instead of $O(n^2)$ distances in $O(n^2)$ computing time as required by a conventional singular-value decomposition algorithm. We present the general algorithm and discuss the methods for controlling the propagation of the numerical errors in the buildup process, for determining rigid vs. unique structures, and for handling problems with inexact distances (distances with errors). We show the results from applying the algorithm to a set of model protein problems with varying degrees of availability and accuracy of the distances and justify the potential use of the algorithm in protein modeling practice.