



Doctoral Thesis

**Software for determination and visual display of NMR structures of proteins
the distance geometry program DGPLAY and the computer graphics programs CONFOR and XAM**

Author(s):

Xia, Tai-he

Publication Date:

1992

Permanent Link:

<https://doi.org/10.3929/ethz-a-000654743> →

Rights / License:

[In Copyright - Non-Commercial Use Permitted](#) →

This page was generated automatically upon download from the [ETH Zurich Research Collection](#). For more information please consult the [Terms of use](#).

Diss. ETH Nr. 9831

**Software for Determination and Visual Display of NMR Structures of Proteins:
the Distance Geometry Program DGPLAY and the Computer Graphics Programs
CONFOR and XAM**

A dissertation submitted to the
SWISS FEDERAL INSTITUTE OF TECHNOLOGY ZÜRICH

for the degree of
Doctor of Natural Sciences

presented by

Tai-he Xia

Master of Science, Institute of Biophysics, Academia Sinica, Beijing, China

born in Pingchang, Sichuan, the People's Republic of China

accepted on the recommendation of

Prof. Dr. K. Wüthrich, examiner

Prof. Dr. W.F. van Gunsteren, co-examiner

Dr. M. Billeter, co-examiner

1992

Abstract

A new distance geometry algorithm for three-dimensional protein structure calculations from upper and lower distance limits as well as dihedral angle constraints derived from NMR spectra was implemented in the program DGPLAY. The algorithm, which is based on structure randomization, was tested with two proteins, using experimental NMR constraints. One is the 69-residue N-terminal domain of the 434 repressor, 434 repressor(1-69), which is an α protein with five helices. The second protein, reduced *E.Coli* glutaredoxin, has 85 residues, and its structure consists of a four-stranded β -sheet and three α -helices. The starting structures of the two proteins are produced by the embedding algorithm, which is also implemented in DGPLAY. Test calculations showed that the program behaves similarly when applied to the two proteins, and that three rounds of randomization are generally sufficient to achieve good sampling and good quality of the structures.

The new algorithm was compared with two other widely used distance geometry algorithms for structure calculations with NMR data, namely, the variable target function method in torsion angle space (DIANA) and the simulated annealing with a simplified force field in which nonbonded interactions are replaced by a steric repulsion term (X-PLOR). The aforementioned two proteins were used in the comparison study, the default protocols for the three programs were employed. The global folds defined by each of three sets of structures are the same, the residual violations are comparable among the three sets of structures, the variations of local r.m.s.d. along the polypeptide chain are similar. However, overall the DGPLAY conformers have a wider spread than the DIANA conformers, while the XPLOR conformers have the smallest spatial spread.

The calculations of the n.m.r. structures of reduced *E. coli* glutaredoxin and the oxidized form of the protein in aqueous solution is described. With the input derived

from the NMR spectra the distance geometry program DIANA was used followed by simulated annealing with the program X-PLOR. The molecular architectures of the two redox states of the protein are similar, they are made up of three helices and a four-stranded β -sheet. Comparison was made between the two redox states of the protein.

A new implementation of the molecular graphics program CONFOR on an Evans & Sutherland PS 390 was presented. It served both as a molecular modeling package and as a suitable tool for displaying the solution structures of macromolecules. The implementation involved retaining the original 70 commands, building 20 new commands for additional facilities and convenience, creating FUNCTION NETWORK for supporting local operations on molecules, and reconstructing the library.

The last step in the structure determination with nuclear magnetic technique is to analyze the obtained structures. The program XAM provides a platform for examining the various features of the structures, viewing them on a SUN sparc station, and efficient preparation of hardcopy illustrations.

Kurzfassung

Ein neuer Distanzgeometriealgorithmus, der auf zufälligen Verschiebungen der Atompositionen beruht, wird vorgestellt für die Berechnung dreidimensionaler Proteinstrukturen aufgrund vom experimentellen Kernspinresonanzdaten, die sowohl obere und untere Distanzschranken als auch Torsionswinkeleinschränkungen umfassen. Das Programm DGPLAY, das diesen Algorithmus verwendet, wird mit den experimentellen Datensätzen von zwei Proteinen getestet. Die N-terminale Domäne des 434 Repressor, 434 Repressors(1-69) ist ein rein α -helikales Protein, das fünf Helices enthält; die reduzierte Form von *E.Coli* Glutaredoxin umfasst 85 Aminosäurereste und seine Struktur besteht aus einem 4-strängigen β -Faltblatt und drei α -Helices. Die Startstrukturen für beide Proteine wurden mit dem Einbettungsalgorithmus, der die metrische Matrix verwendet und auch im Programm DGPLAY verfügbar ist, erhalten. Testrechnungen zeigen, dass das Programm sich für beide Proteine ähnlich verhält und im Allgemeinen drei Iterationen zufälliger Atomverschiebungen ausreichen, um sowohl eine genügende Streuung im erlaubten Konformationsraum als auch die verlangte Qualität der Strukturen zu erreichen.

Der neue Algorithmus wird mit zwei anderen weitverbreiteten Distanzgeometriealgorithmen für die Strukturberechnung aufgrund vom Kernresonanzdaten verglichen: mit dem Verfahren der variablen Zielfunktion im Torsionswinkelraum (Programm DIANA) und mit Molekuldynamikrechnungen, die eine vereinfachte Energiefunktion verwenden, bei der die nicht-bindende Wechselwirkung durch eine bloße sterische Abstossung ersetzt wird (Programm X-PLOR). Die oben erwähnten zwei Proteine wurden für diesen Vergleich herangezogen und mit allen drei Programmen wurde das jeweilige Standardrechenverfahren durchgeführt. Die drei Programme liefern jeweils die gleiche globale Faltung und ähnliche Änderungen der lokalen RMSD Werte entlang der Sequenz. Allerdings

streuen die DGPLAY Strukturen mehr als die DIANA Strukturen, während die Streuung der X-PLOR Strukturen am geringsten ist.

Die Berechnung der NMR Strukturen von reduziertem und oxidiertem *E.Coli* Glutaredoxin in wässriger Lösung wird beschrieben. Aus der Auswertung der NMR Spektren stammende Daten wurden im Distanzgeometrieprogramm DIANA und danach im Moleküldynamikprogramm X-PLOR verwendet. Die dreidimensionalen Strukturen der oxidierten und der reduzierten Form des Glutaredoxins sind ähnlich und enthalten drei Helices und ein viersträngiges β -Blatt. Die beiden Formen des Proteins werden miteinander verglichen.

Eine Neuimplementation des Molekülgraphikprogramms CONFOR auf einer Evans & Sutherland PS 390 Graphikstation wird vorgestellt. Dieses Programm dient sowohl der Modellierung von Molekülen als auch als ein geeignetes Werkzeug zur Darstellung von makromolekularen Lösungsstrukturen. Die Implementation beinhaltet die ursprünglichen 70 Befehle, 20 weitere Kommandos für zusätzliche Funktionen und erleichterte Bedienung, ein Funktionsnetzwerk zur Unterstützung lokaler Operationen mit Molekülen und eine neu erstellte Bibliothek.

Der letzte Schritt in einer NMR Strukturbestimmung ist die Analyse der erhaltenen Strukturen. Das Programm XAM bietet die Möglichkeit, verschiedene Eigenschaften der Strukturen zu untersuchen, Strukturen auf einer SUN SPARC Station darzustellen und auf effiziente Weise Zeichnungen der Struktur herzustellen.