Statistics in Molecular Biology
IMS Lecture Notes – Monograph Series (1999) Volume 33

DERIVING INTERATOMIC DISTANCE BOUNDS FROM CHEMICAL STRUCTURE

BY MICHAEL W. TROSSET[1] AND GEORGE N. PHILLIPS, JR.[2]

College of William and Mary and Rice University

Structural molecular biology is concerned with determining 3-dimensional representations of molecules. Various computational challenges arise in making such determinations, several of which have attracted some attention in the statistics and numerical optimization communities. One such problem is that of determining a 3-dimensional structure that is consistent with bounds on a molecule's interatomic distances; one source of such bounds is the molecule's chemical structure. Because realistic examples are not readily available to computational scientists hoping to test their algorithms, we provide a detailed description of how plausible bounds can be obained.

1. Introduction. Knowledge of 3-dimensional molecular structure can be of enormous value in a variety of scientific endeavors. In this report, we assume the importance of such knowledge and focus on one class of mathematical problems that are sometimes solved to obtain it.

The problem that motivates this report is that of calculating 3-dimensional Cartesian coordinates of atoms from information about interatomic distances. Such information usually assumes the form of lower and upper bounds on the distances. This report is concerned with the derivation of such bounds from a molecule's chemical structure, which we assume to be known. It is addressed to computational scientists who are interested in problems that involve determining 3-dimensional molecular structures that are consistent with specified bounds on interatomic distances. These researchers require sample problems on which to test new algorithms. The methods described herein provide an alternative to (1) inventing structures with no chemical plausiblity and (2) mastering specialized techniques that require considerable expertise in structural molecular biology.

In Section 2 we introduce some technical notation and provide some relevant background material. In Section 3 we provide detailed descriptions of several useful calculations for inferring lower and upper bounds on interatomic distances from chemical structure. In Section 4 we apply the techniques of Section 3 to

This research was supported by the W. M. Keck Center for Computational Biology under Medical Informatics Training Grant 1T1507093 from the National Library of Medicine.

[1]Also supported by Grant DMS-9622749 from the National Science Foundation.

[2]Also supported by Grant C-1142 from the Robert A. Welch Foundation.

AMS 1991 *subject classifications.* Primary 92E10, 92-08; secondary 62H99, 90C90.

Key words and phrases. Multidimensional scaling, distance geometry, computational biology.