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.

an analogue of the antibiotic trichogin A IV. This is a small ($n = 38$ atoms) molecule on which we have been testing prototype algorithms. In Section 5 we conclude by identifying broader contexts in which having bounds on interatomic distances may be useful.

This report documents details of our research that have not been recorded elsewhere. It also provides other researchers with a plausible set of bounds on which to test their algorithms and—more importantly—with a methodology for generating other plausible data sets. Most of all, we hope that this report will provide computational scientists with a better understanding of the relation between certain mathematical problems and the chemical considerations that motivate them.

2. Background. The ultimate goal of structural molecular biology is to determine the unique 3-dimensional structure into which a given molecule folds. A natural place to begin this determination is with the structural information that is contained in the chemistry of the molecule. One way to represent such information is by lower and upper bounds, implied by the chemical structure, on the interatomic distances. The purpose of this report is to indicate the nature of such bounds; in this section, we establish a context for this representation by sketching how Cartesian coordinates can be extracted from bounds on interpoint distances.

A symmetric matrix $\Delta = (\delta_{ij})$ is a dissimilarity matrix if and only if $\delta_{ij} \geq 0$ and $\delta_{ii} = 0$. A dissimilarity matrix is a p-dimensional Euclidean distance matrix if and only if there exist $x_1, \ldots, x_n \in \Re^p$ such that $\delta_{ij} = \|x_i - x_j\|$. We denote the closed cone of p-dimensional Euclidean distance matrices by $\mathcal{D}_n(p)$. If $\Delta \in \mathcal{D}_n(p)$, then simple procedures for determining $x_1, \ldots, x_n \in \Re^p$ such that $\delta_{ij} = \|x_i - x_j\|$ are well-known.

Now suppose that the molecule in question has n atoms. Let $L = (\ell_{ij})$ and $U = (u_{ij})$ be $n \times n$ dissimilarity matrices that contain the specified lower and upper bounds on the interatomic distances. The rectangle $[L, U]$ is defined to be the set of dissimilarity matrices that satisfy the specified bounds, i.e. $\Delta \in [L, U]$ if and only if $\delta_{ij} \in [\ell_{ij}, u_{ij}]$. Glunt, Hayden and Raydan (1993) called this rectangle the *data box*. If $\Delta \in \mathcal{D}_n(3) \cap [L, U]$ and $x_1, \ldots, x_n \in \Re^3$ is such that $\delta_{ij} = \|x_i - x_j\|$, then $x_1, \ldots, x_n$ represent the Cartesian coordinates of a possible 3-dimensional structure of the molecule.

To find a dissimilarity matrix that is (approximately) contained in $\mathcal{D}_n(3) \cap [L, U]$, let ρ denote an error criterion for measuring the discrepancy between a given distance matrix and a given dissimilarity matrix. Then the problem of inferring a possible 3-dimensional structure from the specified interatomic

distance bounds can be formulated as the following optimization problem:

$$\text{(2.1)} \qquad \text{minimize } \rho(D, \Delta) \text{ subject to } D \in \mathcal{D}_n(3) \text{ and } \Delta \in [L, U].$$

Both the Data Box Algorithm proposed by Glunt, Hayden and Raydan (1993) and the embedding approach described by Trosset (1998) can be derived from special cases of this general formulation.

Problem (2.1) is an example of a problem in *distance geometry*. Other formulations are also possible; a recent example is Moré and Wu (1997). In statistics, techniques for inferring a p-dimensional configuration of points from information about interpoint distances are collectively known as *multidimensional scaling* (MDS). In analogy to Problem (2.1), these techniques can be conceived as algorithms for minimizing some measure of discrepancy between a set of distance matrices and a set of dissimilarity matrices. De Leeuw and Heiser (1982) and Trosset (1997) have surveyed a variety of MDS procedures from this perspective. Trosset (1998) described the relation between Problem (2.1) and nonmetric MDS.

This report is concerned with the reasoning by which a data box is derived from the chemical structure of a molecule. There are two reasons for wanting to make the data box as small as possible. First, we would like to eliminate as many distance matrices as possible in order to narrow the search for the correct interatomic distance matrix. Second, we would like to eliminate as many dissimilarity matrices as possible in order to facilitate solution of Problem (2.1). We now consider how to accomplish these objectives.

3. Bound derivation. We assume that the chemical structure of the molecule is known *a priori*, i.e. we assume knowledge of the atomic bonds within the molecule. To simplify our description of how bounds on interatomic distances can be inferred from such knowledge, we introduce some *ad hoc* notation and terminology.

Suppose that a particular pair of atoms has been specified. We denote each of these atoms by an upper case Roman letter, e.g. C for carbon, N for nitrogen, O for oxygen. If these atoms are bonded together, then they comprise a "1-2" pair, e.g. C-C. If they are not bonded together, but are each bonded to a common atom, which we denote by a lower case Roman letter, then they comprise a "1-3" pair, e.g. C-c-C. In similar fashion, we define and denote "1-4" pairs, "1-5" pairs, etc.

Atoms bond at distances that are approximately fixed by nature. These distances depend on identifiable chemical characteristics, e.g. the types of atoms (carbon, nitrogen, oxygen, etc.), the nature of the bond (covalent, double, etc.), and the type of structure within which the bond occurs (benzyl ring, ester,

peptide, etc.). Evidently, some knowledge of chemistry is required to identify comparable categories. Within such categories, there is a certain amount of apparently random variation in bond lengths for which lower and upper bounds can be obtained by inspecting data banks of structures whose bond lengths are known. Hence, it is essentially an empirical exercise to obtain fairly stringent upper and lower bounds on 1-2 distances.

It is also the case that atoms bond at angles that are approximately fixed by nature. Again, these angles depend on identifiable chemical characteristics. A guiding principle is that the bonds to a common atom will arrange themselves to maximize the minimum angle between any two bonds. For example, if an atom is bonded to four other atoms, then it is easily calculated that the minimum angle is maximized if each angle equals

$$\arccos(-1/3) \doteq 1.91 \doteq 109.5°.$$

Again, within the appropriate categories, there is a certain amount of apparently random variation in bond angles for which lower and upper bounds can be obtained by inspecting data banks of known structures.

If bond lengths and angles are both approximately fixed by nature, then 1-3 pairs are approximately rigid structures and 1-3 distances are approximately fixed. If we denote the bond lengths by a and b and the bond angle by τ, then the 1-3 distance x is given by

$$x^2 = a^2 - 2ab\cos(\tau) + b^2. \tag{3.2}$$

Given lower and upper bounds on a, b and τ, we can use equation (3.2) to derive lower and upper bounds on x. In practice, however, it seems preferable to infer these bounds directly by inspecting the appropriate 1-3 distances in data banks of known structures.

In fact, whenever the molecule contains an approximately rigid structure we eschew trigonometric calculation and directly infer bounds on all of the interatomic distances within that structure by inspecting known cases. For example, benzyl rings are approximately planar, hexagonal structures. The approximate rigidity of a benzyl ring can be described by specifying suitably stringent lower and upper bounds on the 1-2, 1-3 and 1-4 distances. All of these bounds can be obtained by inspecting examples of benzyl rings for which interatomic distances have been determined.

If rigidity cannot be assumed, then trigonometric calculation may be of considerable value. Consider the case of 1-4 distances. Let a, b and c denote the bond lengths; let τ_1 denote the angle between the a and b bonds; and let

τ_2 denote the angle between the b and c bonds. Let d denote the 1-3 distance defined by

$$d^2 = b^2 - 2bc\cos(\tau_2) + c^2$$

and let

$$\phi = \arcsin\left(\frac{c}{d}\sin(\tau_2)\right).$$

Then the 1-4 distance x lies between the lower $(+)$ and upper $(-)$ bounds defined by

$$x^2 = a^2 + 2ad\cos(\pi - \tau_1 \pm \phi) + d^2. \tag{3.3}$$

Given lower and upper bounds on a, b, c, τ_1 and τ_2, we can use equation (3.3) to derive lower and upper bounds on x.

Equations analogous to (3.3) can be derived for 1-5 distances, 1-6 distances, etc. Such equations, however, are of diminishing utility because the gap between the lower and upper bounds rapidly widens as the number of intervening bonds increases. It is easier (and often sharper) to impose lower bounds that approximate the repulsive forces that keep unbonded atoms apart. Reasonable upper bounds can be deduced by applying the triangle inequality to the upper bounds for the 1-2, 1-3 and 1-4 distances, an elementary example of *bound smoothing.* See Sections 5.3 (Triangle Inequality Bound Smoothing) and 5.4 (Tetrangle Inequality Bound Smoothing) of Crippen and Havel (1988) for an introduction to bound smoothing.

4. Example. We now apply the methods of Section 3 to a small molecule described by Crisma et al. (1994). This molecule, an analogue of the antibiotic trichogin A IV, contains 7 oxygen atoms, 4 nitrogen atoms, and 27 carbon atoms. Its chemical structure is diagrammed in Figure 1. Our goal is to infer plausible lower and upper bounds on the $703 = 38 \cdot 37/2$ interatomic distances associated with these $n = 38$ atoms.

The 3-dimensional structure of the molecule in Figure 1 is known. One coordinatization, measured in angstroms, is presented in Table 1. These coordinates were obtained from the Brookhaven Protein Data Bank (Bernstein et al., 1977), which can be accessed electronically at the web site `http://www.pdb.bnl.gov/`.

So that our example be self-contained, we will exploit replication within the molecule itself instead of replication in an external database of known molecules. This will cause us to underestimate the natural variability of the component structures, but the resulting bounds will be sufficiently plausible to be illustrative. The purpose of this report is to explicate the logic of how bounds are

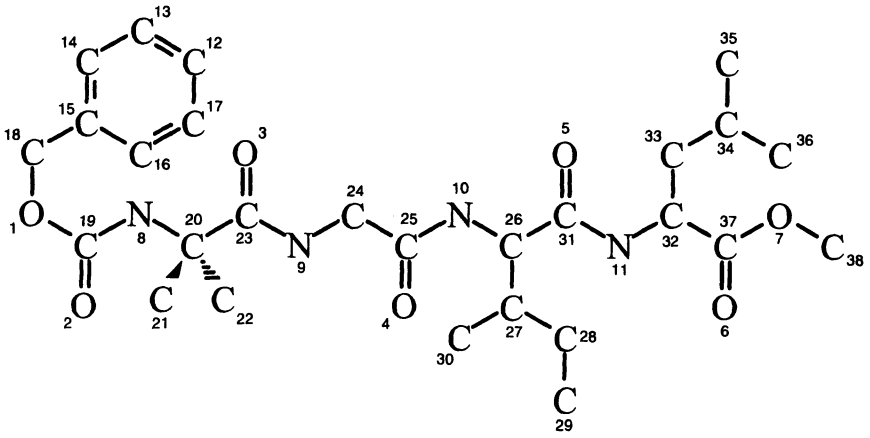

FIG. 1. *Chemical structure of an analogue of the antibiotic molecule trichogin A IV.*

derived and to reveal some of the qualitative features of a plausible data box, not to construct the most realistic data box imaginable for the molecule in question. We believe that using even crude data boxes constructed along the lines indicated in this report will advance the present state of numerical experimentation.

We illustrate the derivation of lower and upper bounds using a 10-atom fragment of the molecule in Figure 1. The fragment that we consider comprises a benzyl ring (atoms 12–17) and an ester (atoms 18,1,19,2).

We begin by examining the benzyl ring. The ring contains six 1-2 pairs, six 1-3 pairs, and three 1-4 pairs. The corresponding distances and bounds that include them are reported in Table 2. The 1-2 distances are listed clockwise from the 17-16 pair; the 1-3 distances are listed clockwise from the 17-16 pair; and the 1-4 distances are listed clockwise from the 17-14 pair.

Atom 18 is tetrahedral, i.e. it has bonds to four other atoms. (The bonds to two hydrogen atoms are not depicted in Figure 1.) Although the 18-15 bond is not replicated in the molecule, there are fifteen C-C bonds in which one (or both) of the atoms is tetrahedral. These 1-2 distances range from a minimum of 1.4687 to a maximum of 1.5456, so we adopt bounds of 1.46 and 1.55.

Because the benzyl ring is (approximately) rigid, each distance between atom 18 and an atom in the ring is (approximately) fixed. For the 18-16 and 18-14 distances we observe that

$$2.47 \leq 2.4769, 2.5212 \leq 2.53;$$

and for the 18-17 and 18-13 distances we observe that

$$3.76 \leq 3.7774, 3.7646 \leq 3.78.$$

TABLE 1
Atomic coordinates of an analogue of the antibiotic molecule trichogin A IV.

Atom		Coordinates			Atom		Coordinates		
1	O	1.629	1.310	4.250	20	C	3.277	4.529	4.289
2	O	2.633	2.357	2.508	21	C	2.422	5.328	3.320
3	O	5.199	5.284	3.098	22	C	3.418	5.277	5.619
4	O	7.777	2.147	1.545	23	C	4.682	4.358	3.714
5	O	5.088	2.230	-1.834	24	C	6.656	2.963	3.451
6	O	2.699	5.605	-1.867	25	C	6.701	2.307	2.089
7	O	3.967	5.011	-0.154	26	C	5.407	1.189	0.302
8	N	2.638	3.244	4.608	27	C	4.771	-0.186	0.520
9	N	5.327	3.216	3.970	28	C	5.660	-1.039	1.421
10	N	5.537	1.894	1.579	29	C	5.081	-2.325	1.831
11	N	3.443	2.477	-0.327	30	C	4.525	-0.901	-0.828
12	C	0.121	-3.130	5.716	31	C	4.643	2.031	-0.701
13	C	1.418	-2.846	5.531	32	C	2.627	3.347	-1.139
14	C	1.777	-1.783	4.797	33	C	1.163	3.290	-0.715
15	C	0.873	-0.950	4.212	34	C	0.576	1.866	-0.753
16	C	-0.453	-1.221	4.392	35	C	-0.919	2.004	-0.393
17	C	-0.830	-2.361	5.137	36	C	0.784	1.116	-2.012
18	C	1.272	0.224	3.357	37	C	3.094	4.785	-1.106
19	C	2.320	2.319	3.694	38	C	4.481	6.369	-0.016

The 18-12 pair is not replicated, but we might guess that

$$4.24 \leq 4.2590 \leq 4.28$$

are plausible bounds for our purposes.

Now we consider the ester C-O-C=O. This nonrigid structure appears twice: 18-1-19-2 and 38-7-37-6. Distances and bounds for the 1-2 and 1-3 pairs of atoms in this structure are reported in Table 3. Because the structure is not rigid, the 1-4 pair is omitted from Table 3. and bounds are obtained from equation (3.3). The bond angles observed for the C-O-C ester fragment are 115.7° and 117.6°,

TABLE 2
Distances between atoms in the benzyl ring.

	1-2 pairs	1-3 pairs	1-4 pairs
	1.4131	2.3972	2.6919
	1.3653	2.3351	2.7274
	1.3614	2.3731	2.7532
	1.3407	2.3241	
	1.3406	2.3332	
	1.3531	2.3931	
Lower bound	1.34	2.32	2.69
Upper bound	1.42	2.40	2.76

so we adopt bounds of 115° and 118°. The bond angles observed for the O-C=O ester fragment are 123.7° and 124.6°, so we adopt bounds of 123° and 125°. Using $a = 1.45$, $\tau_1 = 115°$, $b = 1.31$, $\tau_2 = 123°$ and $c = 1.18$, we obtain a lower bound on the 1-4 distance of 2.58. Using $a = 1.46$, $\tau_1 = 118°$, $b = 1.35$, $\tau_2 = 125°$ and $c = 1.23$, we obtain an upper bound on the 1-4 distance of 3.58.

TABLE 3
Distances between atoms in the C-O-C=O esters.

	1-2 pairs			1-3 pairs	
	C-O-c=o	c-O-C=o	c-o-C=O	C-o-C=o	c-O-c=O
	1.4506	1.3434	1.2272	2.3666	2.2669
	1.4586	1.3113	1.1864	2.3708	2.2125
Lower bound	1.45	1.31	1.18	2.36	2.21
Upper bound	1.46	1.35	1.23	2.38	2.27

Because atom 18 is a tetrahedral carbon, we can infer bounds on the angle between the 18-15 and 18-1 bonds by examining the twenty-two instances of such angles that occur in the molecule. They range from a minimum of 106.2° to a maximum of 115.8°, so we adopt bounds of 106° and 116°. Combining these bounds with the bounds previously obtained for the 18-15 and 18-1 distances, we can exploit equation (3.2) to obtain lower and upper bounds on the 15-1 distance. Using $a = 1.46$, $\tau = 106°$ and $b = 1.45$, we obtain a lower bound of 2.32. Using $a = 1.55$, $\tau = 116°$ and $b = 1.46$, we obtain an upper bound of 2.56. Similarly, we can exploit equation (3.3) to obtain lower and upper bounds on the 15-19 distance. Using $a = 1.46$, $\tau_1 = 106°$, $b = 1.45$, $\tau_2 = 115°$ and $c = 1.31$, we obtain a lower bound of 2.41. Using $a = 1.55$, $\tau_1 = 116°$, $b = 1.46$, $\tau_2 = 118°$ and $c = 1.35$, we obtain an upper bound of 3.80.

The angle between the 15-16 and 15-18 bonds is 119.1° and the angle between the 15-16 and 15-18 bonds is 123.0°, so we adopt bounds on these angles of 119° and 124°. We can now exploit equation (3.3) to obtain lower and upper bounds on the 16-1 and 14-1 distances. Using $a = 1.34$, $\tau_1 = 119°$, $b = 1.46$, $\tau_2 = 106°$ and $c = 1.45$, we obtain a lower bound of 2.51. Using $a = 1.42$, $\tau_1 = 124°$, $b = 1.55$, $\tau_2 = 116°$ and $c = 1.46$, we obtain an upper bound of 3.89.

The remaining lower bounds in the fragment approximate the repulsive forces that keep unbonded atoms apart. For N-O pairs, we suggest a lower bound of 2.80; for other pairs, we suggest a lower bound of 3.20.

Finally, the remaining upper bounds in the fragment are obtained by applying the triangle inequality to the upper bounds that we have already derived. This yields upper bounds of 5.32 for the 12-1 distance, 6.66 for the 12-19 distance, 7.59 for the 12-2 distance, 4.96 for the 17-1 and 13-1 distances, 6.16 for the 17-19 and 13-19 distances, 7.23 for the 17-2 and 13-2 distances, 4.91 for the

16-19 and 14-19 distances, 6.11 for the 16-2 and 14-2 distances, and 4.83 for the 15-2 distance.

We have derived lower and upper bounds for each of the $45 = 10 \cdot 9/2$ interatomic distances associated with a 10-atom fragment of a 38-atom molecule. The same techniques can be applied to the entire molecule. This methodology allows us to approximate plausible *a priori* lower and upper bounds on each of the molecule's 703 interatomic distances. A file containing a set of such bounds is available from the first author.

5. Discussion. The bounds derived in Section 4 define a rectangular feasible region whose features typify the feasible regions $[L, U]$ for Problem (2.1) that might actually arise in practice. As such, our derivations not only provide a useful example on which to test numerical algorithms, but also contribute to our understanding of what is involved in inferring 3-dimensional structure from information about interatomic distances.

We note that most of the distance information that can be derived from chemical structure pertains to atoms that are separated by a small number of bonds. It should be emphasized that this information will rarely—if ever—suffice to determine a unique 3-dimensional structure. Hence, solving Problem (2.1) will not produce the unique 3-dimensional structure that the molecule actually assumes, only one of many structures that are consistent with the specified bounds on the interatomic distances. Stated differently, we do not know how to infer Table 1 from Figure 1.

Of course, as remarked in Section 2, the ultimate goal of structural molecular biology is to find the solution of Problem (2.1) that corresponds to the actual 3-dimensional structure of the specified molecule. To do so, it is necessary to consider additional information about the molecule. We conclude this report by sketching some of the problems that arise in this manner.

5.1. *NMR spectroscopy.* Distances between nearby hydrogen atoms can at times be measured experimentally by Nuclear Magnetic Resonance (NMR) spectroscopy. Bounds on the measurement error resulting from this procedure can be interpreted as bounds on the interatomic distances themselves. This observation was the motivation for the Data Box Algorithm proposed by Glunt, Hayden, and Raydan (1993).

If we combine the distance bounds derived from chemical structure with the distance bounds determined by NMR spectroscopy, then we must again solve Problem (2.1). However, incorporating additional bounds restricts the original feasible region. Usually, the new feasible region is sufficiently restricted that, by solving Problem (2.1) repeatedly, one can obtain an ensemble of possible molec-

ular structures with common features of interest. General reviews of distance geometry in the context of using NMR spectroscopy to determine the structures of protein molecules have been provided by Havel and Wüthrich (1985), Braun (1987), Crippen and Havel (1988), Kuntz, Thomason and Oshiro (1989), Havel (1991), Brünger and Nilges (1993), and Havel, Hyberts and Naifeld (1997).

5.2. *Protein folding.* An alternative possibility is to introduce a function that quantifies Nature's preferences for certain molecular structures, thereby allowing us to anticipate which solutions of Problem (2.1) are most likely to occur. This possibility motivates us to consider the protein folding problem, which is to predict the 3-dimensional structure of a protein molecule from its amino acid sequence.

A self-contained introduction to the protein folding problem, together with many references, was recently provided by Neumaier [15]. Realistic models of protein folding, e.g. the CHARMM [3] potential, include terms corresponding to both bonded and unbonded interactions, the latter comprising both the long-range, slowly decaying electrostatic (Coulomb) interaction and the short-range, fast-decaying van der Waals interaction. Thus, a second possibility is to search for a unique 3-dimensional structure by minimizing a theoretical objective function subject to constraints imposed by the data box derived for Problem (2.1). It should be noted, however, that finding global solutions of such problems is an extremely difficult task.

5.3. *X-ray crystallography.* If the specified molecule can be crystallized, then a third possibility is to utilize data about its diffraction pattern obtained from a crystallography experiment. A general introduction to X-ray crystallography was provided by Glusker and Trueblood (1985); a more recent survey of the computational challenges associated with the phasing, model building, and refinement stages was provided by Brünger and Nilges (1993).

Mathematically, this possibility is similar to the preceding one in that the objective function for protein folding is replaced with a criterion for measuring the fit between the theoretical diffraction pattern of a 3-dimensional structure and the observed diffraction pattern of the molecule in question. Again, it should be noted that finding global solutions of such problems is an extremely difficult task.

Finally, we observe that there may be an interesting role for probability and statistics to play in solving the above problems. In each approach that we have described, it is necessary to sample from the data box. This is necessary in the first case in order to obtain a meaningful ensemble of possible structures; it is necessary in all three cases in order to search for global solutions for optimization

problems that are typically plagued with myriad nonglobal solutions. To the extent that additional "prior" information can be represented in the form of a probability distribution from which dissimilarity matrices in the data box are drawn, it may be possible to accelerate the search for meaningful 3-dimensional structures.

REFERENCES

BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, E., BRYCE, M. D., ROGERS, J. R., KENNARD, O., SHIKANOUCHI, T. and TASUMI, M. (1977). The protein data bank: A computer-based archival file for macromolecular structures. *Journal of Molecular Biology* **112** 535–542.

BRAUN, W. (1987). Distance geometry and related methods for protein structure determination from NMR data. *Quarterly Reviews of Biophysics* **19** 115–157.

BROOKS, B. R., BRUCCOLERI, R., OLAFSON, B., STATES, D., SWAMINATHAN, S. and KARPLUS, M. (1983). CHARMM: A program for macromolecular energy, minimization, and dynamics calculations. *Journal of Computational Chemistry* **4** 187–217.

BRÜNGER, A. T. and NILGES, M. (1993). Computational challenges for macromolecular structure determination by X-ray crystallography and solution NMR spectroscopy. *Quarterly Reviews of Biophysics* **26** 49–125.

CRIPPEN, G. M. and HAVEL, T. F. (1988). *Distance Geometry and Molecular Conformation.* John Wiley & Sons, New York.

CRISMA, M., VALLE, G., MONACO, V., FORMAGGIO, F. and TONIOLO, C. (1994). N alpha-benzyloxycarbonyl-alpha-aminoisobutyryl-glycyl-l-isoleucyl-l-leucine methyl ester monohydrate. *Acta Crystallographica* **50** 563–565.

DE LEEUW, J. and HEISER, W. (1982). Theory of multidimensional scaling. In Krishnaiah, P. R. and Kanal, I. N., editors, *Handbook of Statistics*, volume 2, chapter 13, pages 285–316. North-Holland Publishing Company, Amsterdam.

GLUNT, W., HAYDEN, T. L. and RAYDAN, M. (1993). Molecular conformations from distance matrices. *Journal of Computational Chemistry* **14** 114–120.

GLUSKER, J. P. and TRUEBLOOD, K. N. (1985). *Crystal Structure Analysis: A Primer.* Oxford University Press, New York.

HAVEL, T. F. (1991). An evaluation of computational strategies for use in the determination of protein structure from distance constraints obtained by nuclear magnetic resonance. *Progress in Biophysics and Molecular Biology* **56** 43–78.

HAVEL, T. F., HYBERTS, S. and NAIFELD, I. (1997). Recent advances in molecular distance geometry. In Hofestödt, R., editor, *Bioinformatics*, pages 62–71. Springer, Berlin. *Lecture Notes in Computer Science*, Number 1278.

HAVEL, T. F. and WÜTHRICH, K. (1985). An evaluation of the combined use of nuclear magnetic resonance and distance geometry for the determination of protein conformations in solution. *Journal of Molecular Biology* **182** 281–294.

KUNTZ, I. D., THOMASON, J. F. and OSHIRO, C. M. (1989). Distance geometry. In Oppenheimer, N. J. and James, T. L., editors, *Nuclear Magnetic Resonance, Part B: Structure and Mechanism*, pages 159–204. Academic Press, New York. Volume 177 of *Methods in Enzymology.*

MORÉ, J. J. and WU, Z. (1997). Global continuation for distance geometry problems. *SIAM Journal on Optimization* **7** 814–836.

NEUMAIER, A. (1997). Molecular modeling of proteins and mathematical prediction of protein structure. *SIAM Review* **39** 407–460.

TROSSET, M. W. (1997). Numerical algorithms for multidimensional scaling. In Klar, R. and Opitz, P., editors, *Classification and Knowledge Organization*, pages 80–92, Berlin. Springer. Proceedings of the 20th annual conference of the Gesellschaft für Klassifikation e.V., held March

6–8, 1996, in Freiburg, Germany.

TROSSET, M. W. (1998). Applications of multidimensional scaling to molecular conformation. *Computing Science and Statistics* **29**. To appear.

MICHAEL W. TROSSET
DEPARTMENT OF MATHEMATICS
COLLEGE OF WILLIAM AND MARY
P.O. BOX 8795
WILLIAMSBURG, VA 23187-8795
TROSSET@MATH.WM.EDU

GEORGE N. PHILLIPS, JR.
DEPARTMENT OF BIOCHEMISTRY AND CELL BIOLOGY
RICE UNIVERSITY—MS 140
HOUSTON, TX 77005-1892
GEORGEP@BIOC.RICE.EDU