Mathematical Modeling of Protein Structure Using Distance Geometry*

Jeong-Mi Yoon† Yash Gad ‡ Zhijun Wu §

July 24, 2000

Abstract. This paper reviews methods for structure determination with interatomic distances and explores possible improvement of the methods and ways of combining them with potential energy minimization.

Key words. Distance Geometry, Potential Energy Minimization, Nuclear Magnetic Resonance (NMR) Spectroscopy, Singular Value Decomposition (SVD), Graph Embedding, Global Optimization.

1 Introduction

Many of the research subjects in biology focus on properties and activities of cells that are primarily determined by proteins. Proteins are biopolymers made up of twenty different amino acids, each having an acid group, an amino group, and a side chain. The order of the amino acids and the properties of

*Technical Report TR00-24, Department of Computational and Applied Mathematics, Rice University, Houston, TX 77005

†Department of Computer & Mathematical Sciences, University of Houston, Downtown, Houston, TX 77002 (Yoonj@dt.uh.edu).

‡Center for Computational Sciences and Advanced Distributed Simulation, University of Houston, Downtown, Houston, TX 77002 (Yash@tiger-marmalade.com).

§Keck Center for Computational Biology, Rice University, Houston, TX 77005 (Zhijun@caam.rice.edu). Questions, comments, or corrections to this document may be directed to this email address.

their side chains in a protein determine a three-dimensional structure. The structure specifies the function of the protein [3].

1.1 Molecular Distance Geometry

The structure of a protein may be determined experimentally via NMR spectroscopy or X-ray crystallography or theoretically through potential energy minimization or molecular dynamics simulation. We study a problem related to the NMR approach to structure determination. More specifically, we consider the problem of determining the structure of a protein with a set of distances between pairs of atoms in the protein. The distances are either obtained with our knowledge on certain bond lengths and bond angles or estimated through NMR experiments. The problem is in general called a molecular distance geometry problem.

If the distances between all pairs of the atoms in the protein are given, a unique three-dimensional structure can be determined. However, in practice, only a subset of the distances are available and they often have experimental errors as well. The structure that can be determined by the given distances may not be unique and it may not even exist if the distances are not consistent. In any case, the solution to a distance geometry problem will be able to provide useful information about what the protein structure could be or where the inconsistency may occur in the data [5, 2, 10].

1.2 Potential Energy Minimization

A more general approach to protein structure determination is potential energy minimization, which assumes that the protein structure corresponds to the global minimum of the system's potential energy. Therefore, the structure may be found by minimizing a potential energy function for the protein to find a structure that has the least potential energy [17]. The global minimization problem is hard to solve in general. We would like to investigate an approach that can combine potential energy minimization with distance geometry. For example, we may use the solution to the distance geometry problem as an initial guess for potential energy minimization. Since the solution satisfies the NMR distance constraints, it must be a better starting point than arbitrarily generated structures. On the other hand, the NMR

distance data may be used as constraints for potential energy minimization to reduce the search space to more feasible regions.

1.3 Nuclear Magnetic Resonance (NMR) Spectroscopy

NMR [9] gives the labeled distances between the pairs of atoms to which they correspond. There are two kinds of two-dimensional NMR experiments which can yield information suitable for distance geometry calculations:

1. Correlated Spectroscopy (COSY) and
2. Nuclear Overhauser Enhancement Spectroscopy (NOESY)

The NMR experiments are based on the principle that hydrogen nuclei have two spin states. There exists a specific energy separation between the two states, and at a particular frequency, nuclei flip the spin. The structural information comes from "spin-spin coupling". If two nuclei are very close in space then their spins interact and the frequency required for a spin flip is shifted. Hence the peaks in the spectrum become shifted slightly. The intensity of this effect depends on the distance between the nuclei. The result of NMR analysis is a set of estimates of the distances between specific pairs of atoms, which can be used to determine the structure of the molecule.

2 Distance Geometry Problem

In this section, we discuss the distance geometry problem. We discuss problem formulations and complexity issues. Three classes of problems are described, including the problems with all exact distances, with sparse sets of distances, and with bounds on the distances.

2.1 History

According to Crippen and Havel [5], the general form of the distance geometry problem was given by Cayley in 1841. The problem was not systematically studied until 1928, when Menger showed how convexity and many other basic geometric properties could be defined and studied in terms of distances between pairs of points. In 1935, Schoenberg found an equivalent

characterization of Euclidean distances and realized the connection of the problem with bilinear forms. In 1953, Blumenthal published a monograph to introduce the subject, where he stated the fundamental problem in distance geometry as "When we have given a set of distances between pairs of points, the distance geometry can give a clue to find a correct set of coordinates for the points in three-dimensional Euclidean space satisfying the given distance constraints. [1] "

2.2 Problems with All Exact Distances

A particular case of the distance geometry problem is when exact distances between all pairs of points are given. Then, the problem can be solved by factorizing a distance matrix formed with the given distances. More specifically, we can define a special matrix with the given distances. If the distances are consistent in the sense that we can indeed find a set of feasible points in three-dimensional space, the distance matrix must be of rank three. If we can find the three nonzero eigenvalues of the matrix, we can use the eigenvectors to find the coordinates of the points immediately [1, 5].

If the number of arithmetic operations required for solving a problem is bounded by a polynomial function of the problem size, we say that the problem can be solved in polynomial time and it is a tractable problem. Note that computing the eigenvalues of a $n \times n$ matrix can be done in $\mathcal{O}(n^3)$ floating point operations. So, the distance geometry problem, when all exact distances are given, can be solved in polynomial time and is a tractable problem.

2.3 Problems with Sparse Sets of Distances

In practice, we are often given only a sparse set of distances. The distance geometry problem then becomes very hard to solve in general. Saxe [16] in 1979 showed that a one-dimensional distance geometry problem is equivalent to a set-partition problem which is known to be NP-hard. He also showed that a N-dimensional distance geometry problem is NP-hard for all N greater or equal to one. NP-hard is a computational complexity term. A simplified interpretation of NP-hard is that if a problem is NP-hard, it is not possible to find a polynomial time algorithm for solving the problem, and the problem is computationally intractable in general [6].

Figure 2.3 shows how a set partition problem can be reduced to a one-dimensional distance geometry problem. First, suppose that we have a set of integers $S = \{1, 2, 2, 4, 3\}$ and we want to partition the set to two subsets such that the sums of the integers in the subsets are equal. Given this problem, we can immediately construct a one-dimensional distance geometry problem: We consider 6 points in one-dimensional space and require the distance between the first and the second to be equal to the first integer in S, the distance between the second and the third to be equal to the second integer in S, and so on and so forth. Finally, we also require the distance between the last and the first points to be equal to zero. Suppose that we can solve this problem by folding the linear chain in a way that the distances between all adjacent points are satisfied and the first and last points also meet together. We can then obtain two subsets of integers, one corresponding to those distances pointing to the left and the other to the right: $S_1 = \{1, 2, 3\}$, and $S_2 = \{2, 4\}$. It is not difficult to verify that S_1 and S_2 solve the set partition problem for S.

Since the above construction can be used for arbitrary integer sets, a set partition problem can always be reduced in the same way to a one-dimensional distance geometry problem. The latter must be hard to solve, for otherwise, the set partition problem becomes easy to solve, which is not true because it is already known to be NP-hard.

2.4 Problems with Bounds on Distances

In molecular applications, for example, NMR molecular modeling, because of the experimental errors, we can only obtain some lower and upper bounds on the distances. Then the distance geometry problem becomes to find the coordinates for the atoms so that the distances between pairs of atoms are in between the corresponding lower and upper bounds. In general, this problem seems easier to solve since the distance constraints are relaxed and become easier to satisfy. However, in practice, some of the lower and upper bounds are tight, and the problem can still be hard to solve.

Moré and Wu [13] in 1996 considered a so-called ϵ-approximation problem similar to the problems with bounds on the distances. Given a set of distances, they allowed the distances to have a small error ϵ. If a solution to the distance geometry problem is found with distance errors within an ϵ tolerance, the solution is said to be ϵ-optimal, and an ϵ-approximation to the

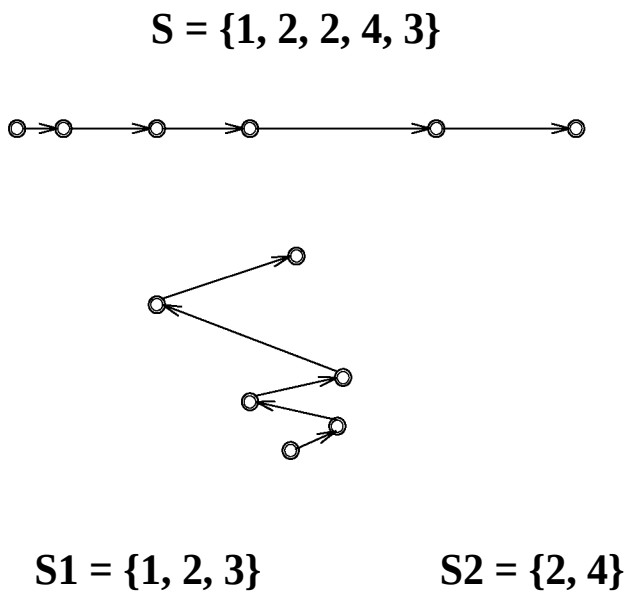

Figure 1: set partition and one-dimensional distance geometry

problem is obtained. Moré and Wu [13] showed that even obtaining such an ϵ-optimal solution to the distance geometry problem is still NP-hard if ϵ is small. There actually is a lower bound for ϵ. Once ϵ is smaller than this bound, the problem becomes equivalent to the set partition problem again.

3 Solution Techniques

In this section, we describe some techniques that can be used for solving the distance geometry problem. The first technique is concerned with the solution to the problems with all exact distances. The second one is used to reduce a given problem to smaller subproblems. The bound smoothing technique is used to estimate missing distances. The least-squares minimization technique is used to solve the distance geometry problem as a special class of optimization problems.

3.1 Decomposition of Distance Matrix

If all exact distances are given, they can be arranged into a matrix, $d = [d_{ij}]$, with d_{ij} corresponds to the distance between points i and j. Suppose that we have a set of points $x_0, x_1, \ldots, x_n$. Without loss of generality, we assume that the coordinates are translated so that $x_0 = (0, 0, 0)$. We can make this assumption since no matter what the coordinates are, we can always translate them without changing any distances between the points.

We consider the problem to find $x_1, \ldots, x_n$ so that the distance between points i and j are equal to given distance d_{ij} for all i and j. The distance constraints can be written in the following form,

$$\|x_i - x_j\| = d_{ij}, \qquad i, j = 0, 1, \ldots, n,$$

or equivalently,

$$\begin{aligned} \|x_i\|^2 &= d_{i0}^2, \\ \|x_i - x_j\|^2 &= d_{ij}^2, \qquad i, j = 1, \ldots, n. \end{aligned}$$

The second set of constraints are equivalent to

$$\|x_i\|^2 - 2x_i^T x_j + \|x_j\|^2 = d_{ij}^2, \qquad i, j = 1, \ldots, n.$$

We then obtain

$$d_{i0}^2 - d_{ij}^2 + d_{j0}^2 = 2x_i^T x_j, \qquad i, j = 1, \ldots, n.$$

Let $D_{ij} = (d_{i0}^2 - d_{ij}^2 + d_{j0}^2)/2$. We can then define a matrix $D = [D_{ij}]$. Let X be an $n \times 3$ matrix and $X = [x_1^T; \ldots; x_n^T]$. We then have

$$D = XX^T.$$

If a solution exists for this equation, matrix D must be of rank 3. Therefore, we can make a singular value decomposition for D to obtain

$$D = U\Sigma U^T,$$

where U is a $n \times 3$ orthogonal matrix and Σ a 3×3 diagonal matrix with the diagonal elements σ_1, σ_2, and σ_3 being three nonzero singular values of D. A solution for $D = XX^T$ can then be obtained with

$$X = U\Sigma^{1/2}.$$

Note that the singular value decomposition can be done in $\mathcal{O}(n^3)$ floating point operations. Therefore, the solution to the distance geometry problem can be obtained in polynomial time if given all exact distances between pairs of points. More details about factorizing a distance matrix can be found in [5].

3.2 Graph Reduction

If we consider the points as nodes and the distances as edges, the distance geometry problem can be described by a distance graph, and the solution to the problem is a realization of the distance graph in an Euclidean space. The problem is therefore also called graph embedding. The graph often is sparse. Therefore, the embedding may not be unique. In other words, there may be more than one ways to position the points so that the distance constraints can all be satisfied. Sometimes, some of the points can be moved continuously without violating any distance constraints. As a result, there may be infinitely many ways to embed the graph, or in other words, there may be infinitely many solutions to the distance geometry problems. The graph is then called flexible or otherwise rigid.

The rigidity of the distance graph can be important for the study of the distance geometry problem. A related issue is the uniqueness of embedding. A necessary condition for a graph to have a unique embedding is that it must be rigid. A rigid graph may still have multiple embeddings, for example, if the graph has partial reflections. Therefore, another necessary condition for a graph to have a unique embedding is that the graph does not have partial reflections. This can be assured if the graph is four-connected for a three-dimensional embedding. These conditions can be used to find graphs or subgraphs that have unique embeddings. The embedding problem for a given distance graph may then be solved by decomposing the graph into such subgraphs. Once the solutions for the subgraphs are found, they can be combined into a solution for the whole graph. For more details about this technique, readers are referred to Hendrickson [11].

3.3 Bound Smoothing

Bound smoothing can be used to estimate missing data or correct the inconsistencies in given distance bounds. The method is based on geometric rules

such as triangular inequalities as well as physical principles such as the van der Waal's spheres between pairs of atoms.

One of the geometric rules used in bound smoothing is the following. Suppose that the lower and upper bounds for two of the distances in between points i, j, and k are given. Let the bounds be denoted as l_{ij}, u_{ij}, l_{jk}, and u_{jk}. Then the lower and upper bounds for the distance between points i and k must agree with the following rules,

$$\begin{aligned} l_{ik} &= \max\{l_{ik},\ l_{ij} - u_{jk},\ l_{jk} - u_{ij}\} \\ u_{ik} &= \min\{u_{ik},\ u_{ij} + u_{jk}\}. \end{aligned}$$

Other rules can also be derived similarly for the distance bounds within more than three points [5, 10].

3.4 Least-Squares Formulation

The distance geometry problem can be formulated as a global least-squares problem. We first consider problems with exact distances. The problems can be defined by a set of equality constraints,

$$\|x_i - x_j\| = d_{ij}, \qquad (i,j) \in S,$$

where S may or may not be the whole set of distance pairs.

In order to solve this class of problems, we measure the following relative errors between calculated and given distances,

$$\frac{\|x_i - x_j\|^2 - d_{ij}^2}{d_{ij}^2}, \qquad (i,j) \in S,$$

and collect all of them to obtain an error function,

$$f(x_1, \ldots, x_n) = \sum_{(i,j)\in S} [\frac{\|x_i - x_j\|^2 - d_{ij}^2}{d_{ij}^2}]^2.$$

Here we see that if the distance constraints are satisfied, the error function is equal to zero.

Similarly, for problems with bounds on the distances, we have the following inequality constraints,

$$l_{ij} \le \|x_i - x_j\| \le u_{ij}, \qquad (i,j) \in S.$$

We can construct an error function,

$$f(x_1, \ldots, x_n) = \sum_{(i,j)\in S} \min^2\{\frac{\|x_i - x_j\|^2 - l_{ij}^2}{l_{ij}^2}, 0\} + \max^2\{\frac{\|x_i - x_j\|^2 - u_{ij}^2}{u_{ij}^2}, 0\}.$$

It is not difficult to verify that if all inequality constraints are satisfied, the error function is equal to zero.

Given the above error function f, it is easy to see that a set of coordinates $x_1, ..., x_n$ is a solution to the distance geometry problem if and only if it is the global minimizer of f with the global minimum equal to zero. Therefore, the distance geometry problem can be formulated as an optimization problem,

$$\min_{x_1,\ldots,x_n} \quad f(x_1, \ldots, x_n).$$

The challenge in solving this problem is to find a global minimizer of function f, which is an intractable problem in general. More details can be found in [5].

3.5 RMSD Calculation

The Root-Mean-Square-Deviation (RMSD) often is used to compare two molecular structures. In particular, when we obtain a structure with a distance geometry problem, we may want to compare it with an estimated structure by looking at the RMSD of the two structures. Let X be an $n \times 3$ coordinate matrix. Let Y be the target matrix. Suppose that the two structures X and Y have the same center of the mass after an appropriate translation of the structures. We define the RMSD of X and Y to be the minimum of the Frobenius norm of the difference of the two structures subject to an appropriate rotation of X, i.e.,

$$\text{RMSD}(X, Y) = \min_Q \|Y - XQ\|_F, \qquad Q^T Q = I,$$

where Q is a 3×3 orthonormal matrix.

Note that

$$\|Y - XQ\|^2 = tr(Y^T Y) + tr(X^T X) - 2\ tr(Q^T X^T Y).$$

Therefore, minimizing $\|Y - QX\|$ is equivalent to maximizing $tr(Q^T X^T Y)$. Let $C = X^T Y$ and $C = U\Sigma V^T$ be the singular value decomposition of C. Then

$$tr(Q^T X^T Y) = tr(Q^T C) = tr(Q^T U \Sigma V^T) = tr(V^T Q^T U \Sigma) \leq tr(\Sigma),$$

and $tr(Q^T X^T Y)$ is maximized when $Q = VU^T$.

It follows that RMSD(X, Y) can be computed by the following procedure:

1. Set $C = X^T Y$,
2. Obtain a singular value decomposition, $C = U\Sigma V^T$,
3. Set $Q = VU^T$, and then RMSD$(X, Y) = tr(\Sigma)$ [7].

4 Algorithms and Software

4.1 EMBED Algorithm

The EMBED algorithm was developed by Crippen and Havel [5, 10] who pioneered the work in distance geometry for molecular conformation. The algorithm can be used for solving the distance geometry problems arising in NMR molecular modeling and structure determination. In particular, the algorithms handles the bounds on the distances, since in practice, only a set of lower and upper bounds on the distances are available. Since not all distance bounds are given, the algorithm also uses some bound smoothing techniques to obtain some estimated bounds for the missing distances. Once all distance bounds are obtained, a set of coordinates for the atoms whose distances satisfy all the bounds are computed. Such a set of coordinates may not be found due to insufficient search or inconsistency in the data. In this case, the EMBED algorithm takes an approximate solution to the problem as a starting point for a least-squares optimization algorithm. Hopefully, a correct set of coordinates can be obtained by solving a least-squares problem as we described in previous sections.

More specifically, the EMBED algorithm follows three successive stages: The first stage takes the input distance bounds, which are usually quite sparse and imprecise, and attempts to estimate the range of values each distance can assume (bound smoothing). In the second step, a set of coordinates

for the atoms are computed such that the distances lie largely within the ranges from the previous step (embedding). The third stage uses standard numerical optimization methods to minimize an "error function" whose magnitude measures the deviations of the coordinates from the input distance constraints (least-squares optimization).

The EMBED algorithm also handles chirality constraints, which are very important in recognizing chemically valid structures among feasible solutions to a distance geometry problem.

4.2 ABBIE

ABBIE is a software package developed by Hendrickson [11] in 1991 for the determination of molecular structure with a given set of distances. The method used in the software is based on graph reduction. So, for a given distance graph, the program first decomposes the graph recursively into subgraphs with unique three-dimensional embeddings. The embedding problems for the subgraphs are then solved by minimizing a least-square error function. The method has several advantages. First, if there is not enough information to uniquely solve a given problem, the method will identify and solve unique subproblems. Second, for many applications, it is only a small portion of the molecule that is of interest, like a binding site. Therefore, the solution to a subproblem can be as important. Third, the method is able to determine whether or not there is sufficient data to the problem. Fourth, in any physical experiment, there can be some measurements that are in error. Inconsistent data would be indicated by the inability to solve a particular subproblem, which narrows the location of the erroneous data to the values in the subproblem.

4.3 Alternating Projection Algorithm

Glunt et al [8] developed an alternating projection algorithm for solving the distance geometry problem with a given set of bounds on the distances. The idea behind the algorithm is the following. First, a set of distances are generated from between the given distance bounds. Then a distance geometry problem with this set of distances is solved by minimizing an error function (optimization). If a solution is obtained, the program is done, otherwise, the

violated constraints are used to adjust the distances (projection), and the algorithm is repeated for a new set of distances.

This algorithm requires the bounds on all the distances available, or relies on a bound smoothing procedure to provide all the bounds. In every iteration, a least-squares problem is solved, which may require a large amount of computation. For example, if a Newton's algorithm is used, the total cost can be as much as $\mathcal{O}(n^3)$ floating point operations. When n is large (for example, 1,000) and the problem needs to be solved many times (for example, 1,000), the Newton's algorithm becomes too expensive to use. Therefore, a spectral gradient algorithm, which is much cheaper than conventional algorithms, has been used in the alternating projection algorithm instead.

4.4 DGSOL

DGSOL is a software package developed by Moré and Wu [14, 15] for solving molecular distance geometry problem by using a global smoothing and continuation method. The method considers the least-squares formulation of the distance geometry problem. It does not require all distances or bounds to be available.

The least-squares problem may have many local minimizers. In order to locate the global minimizer, the global smoothing and continuation method first transforms the least-squares function into a set of gradually deformed but smoother or easier functions with fewer local minimizers. The method then locate the minimizers of the transformed functions and trace their changes when the transformed functions are changed back to the original function. A global minimizer hopefully is located in the end.

This method has been applied to some small to medium-sized test problems with around 200 points or atoms. The results showed that the method was able to find the global minimizer of the least-squares function with a very high probability while a conventional multi-start random search algorithm failed to find a single global minimizer of the function.

One of the advantages of this method is that it does not need all the distances or bounds. The cost for solving a distance geometry problem is cheaper in the sense that the least-squares function contains a smaller number of terms. The function, the gradient, as well as the Hessian, if required, can all be computed with less cost than for all distances or bounds. The method is more practical as well since in practice only a sparse set of distances or bounds

are available. The bound smoothing technique may be helpful for providing some additional distance data, but they are not so reliable in general.

5 Future Research Directions

Distance geometry plays an important role in NMR protein structure determination. It has been a routine practice in many of the biochemistry labs to use distance geometry algorithms and software to predict protein structures with NMR experimental data. The algorithms need to be improved to obtain more efficient and accurate prediction, however. They may be combined with other technologies to make more powerful structural determination tools as well.

5.1 Improvement of the Algorithms

The distance geometry algorithms may be improved in two areas. The first one is related to the solution to the problems with all exact distances, and the second to the global least-squares problems. Both of them are subproblems to be solved in the EMBED and Alternating Projection Algorithms.

The current solution to the first class of problems requires computing three largest eigenvlues of a modified distance matrix, which is costly and therefore, the applicability and reliability of the algorithms for large-scale applications are limited. Improvement on the solution may be achieved by developing a combinatorial algorithm that can exploit the problem structure and find solutions to the problems in less computation than general eigenvalue algorithms.

For the global least-squares problems, a spectral gradient algorithm is used in the alternating projection algorithm, while a simulated annealing algorithm in the EMBED algorithm. The spectral gradient algorithm can only find a local minimizer for the least-squares function. A global optimization algorithm is required to guarantee a global or nearly global minimizer. The simulated-annealing algorithm may find a global minimizer only if sufficient sampling (slow cooling) of function values can be done, which usually are not possible for large-scale problems.

The global smoothing and continuation algorithm may be used to obtain better solutions to the global least-squares problems. Moré and Wu [14, 15]

showed promising results. More numerical work may be required to compare with other algorithms such as simulated-annealing algorithm. The algorithm may need to be improved to obtain good solutions for large-scale problems as well.

5.2 Energy Minimization with Distance Geometry

Possible combination of distance geometry with potential energy minimization has been studied by several research groups, including Hock and Stern [12], Brüger [4], and Smith-Brown et al [18]. Research in this area holds promises for better solutions to structure prediction and determination than those with only potential energy minimization or distance geometry.

The advantages in the two approaches are actually complementary. The potential energy minimization approach uses the knowledge of the potential field of given molecules. It is a powerful approach because it is based on more fundamental principles in molecular chemistry and physics. However, an accurate model for the potential field is hard to build, and the minimization problem is also hard to solve without using any structural heuristics.

On the other hand, the NMR distance geometry approach is more experimental. It is in some sense supported by the physical experiments. The disadvantage of the approach is that the experimental data usually is not adequate and accurate enough to fully determine a structure. Additional chemical or physical knowledge about the molecule will help.

Therefore, a novel combination of the two approaches may have a great advantage in solving a structure. One possible combination is to use the structure obtained from NMR distance geometry as a starting point for potential energy minimization. Another is to use the distance data obtained from the NMR experiments as constraints for potential energy minimization and develop a global constrained optimization algorithm to find the global energy minimum of the molecule. Work along these directions needs to be expanded in future.

References

[1] L. M. Blumenthal, *Theory and Applications of Distance Geometry*, Oxford, Clarendon Press, 1953.

[2] H. Bohr and S. Brunak, editors, *Protein Structure by Distance Analysis*, IOS Press, 1994.

[3] C. Branden and J. Tooze, *Introduction to Protein Structure*, Garland Publishing, Inc., 1991.

[4] A. T. Brüger and M. Niles, *Computational Challenges for Macromolecular Modeling*, in Reviews in Computational Chemistry, K. B. Lipkowitz and D. B. Boyd, eds., VCH Publishers, 1993, Vol. 5, pp. 299-335.

[5] G. M. Crippen and T. F. Havel , *Distance Geometry and Molecular Conformation*, John Wiley & Sons, 1988.

[6] M. R. Garey and D. S. Johnson, *Computers and Intractability*, W. H. Freeman, 1979.

[7] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Johns Hopkins University Press, 1989.

[8] W. Glunt and T. L. Hayden and M. Raydan, *Molecular Conformations from Distance Matrices*, J. Comput Chem, Vol.14, No.1, pp. 114–120, 1993.

[9] H. Gunther, *NMR Spectroscopy: Basic Principles, Concepts, and Applications in Chemistry*, John Wiley & Sons, 1995.

[10] T. F. Havel, *Distance Geometry*, in Encyclopedia of Nuclear Magnetic Resonance, D. M. Grant and R. K. Harris, eds., John Wiley & Sons, 1995, pp. 1701–1710.

[11] B. A. Hendrickson, *The Molecular Problem: Determining Conformation from Pairwise Distances*, Ph.D. thesis, Cornell University, 1991.

[12] J. C. Hock and A. S. Stern, *A Method for Determining Overall Protein Folding from NMR Distance Restraints*, J. Biomolecular NMR, 1992, Vol. 2, pp. 535–543.

[13] J. Moré and Z. Wu, *ϵ-Optimal Solutions to Distance Geometry Problems via Global Continuation*, in Global Minimization of Nonconvex Energy

Functions: Molecular Conformation and Protein Folding, P. M. Pardalos, D. Shalloway, and G. Xie, eds., American Mathematical Society, 1996, pp. 151–168.

[14] J. Moré and Z. Wu, *Global Continuation for Distance Geometry Problems*, in SIAM Journal on Optimization, Vol.7, No.3, 1997, pp. 814–836.

[15] J. Moré and Z. Wu, *Distance Geometry Optimization for Protein Structures*, Preprint MCS-P628-1296, Mathematics and Computer Science Division, Argonne National Laboratory, July 1997.

[16] J. B. Saxe, *Embeddability of Weighted Graphs in K-Space Is Strongly NP-Hard*, in Proc. 17th Allerton Conference in Communications, Control and Computing, 1979, pp. 480–489.

[17] H. A. Scheraga, *Predicting Three-Dimensional Structures of Oligopeptides*, in Reviews in Computational Chemistry, K. B. Lipkowitz and D. B. Boyd, eds., VCH Publishers, 1992, Vol. 3, pp. 73–142.

[18] M. J. Smith-Brown, D. Kominos, and R. M. Levy, *Global Folding of Proteins Using a limited Number of Distance Constraints*, Protein Eng., 1993, Vol. 6, pp. 605–614.