graphite monochromator and $\omega/2\theta$ scan mode in the range $1 \leq \theta \leq 25°$. Table 2 gives the figures of merit of the best sets for all three methods; that of *MAGEX*89 gave an E map showing 22 atoms and is reproduced in Fig. 1. Fourier calculations revealed the remaining atoms and least-squares refinement gave a final residual of 0·055 for the observed reflexions.

Discussion

It will be seen from Table 1 that for the trial structures the *MAGEX*89 method performed somewhat better than the other two and the computer resources used by *MAGEX*89 were 9% less than those of *RANTAN*. Not too much should be made of that since by modifying parameters all three methods are probably capable of solving all the structures. What we do say is that it is worthwhile having *MAGEX*89 available. While any individual method may not succeed for a particular structure, the probability of failure is far lower with many methods available.

The version of *MAGEX*89 we have used has been programmed for a PDP11/44 but should be able to run on most standard personal computers. It can handle all 230 space groups in the standard orientations, including alternative settings, as given in *International Tables for Crystallography* (1987).

We are most grateful to the National Natural Science Foundation of China for its support of this project and to Professor M. M. Woolfson for advice and help with the preparation of this manuscript.

References

Debaerdemaeker, T., Tate, C. & Woolfson, M. M. (1985). *Acta Cryst.* **A41**, 286-290.

Debaerdemaeker, T., Tate, C. & Woolfson, M. M. (1988). *Acta Cryst.* **A44**, 353-357.

Hull, S. E. & Irwin, M. J. (1978). *Acta Cryst.* **A34**, 863-870.

Hull, S. E., Viterbo, D., Woolfson, M. M. & Zhang, S. H. (1981). *Acta Cryst.* **A37**, 566-572.

International Tables for Crystallography (1987). Vol. A. Dordrecht: Kluwer.

White, P. S. & Woolfson, M. M. (1975). *Acta Cryst.* **A31**, 53-56.

Yao, J. X. (1981). *Acta Cryst.* **A37**, 642-644.

Zhang, S. H., Luo, B. S., Chen, S. K. & Yao, J. W. (1989). *Acta Phys. Chim. Sin.* **5**, 536-540.

Zhang, S. H. & Woolfson, M. M. (1982). *Acta Cryst.* **A38**, 683-685.

Acta Cryst. (1992). **A48**, 11-14

A Method for Multiple Superposition of Structures

By A. Shapiro and J. D. Botha

Department of Mathematics, Applied Mathematics and Astronomy, University of South Africa, PO Box 392,
Pretoria 0001, *South Africa*

A. Pastore

European Molecular Biology Laboratory, Meyerhofstrasse 1, *Postfach* 1022.09, 6900 *Heidelberg, Germany*

and A. M. Lesk

Department of Haematology, University of Cambridge Clinical School, MRC Centre, Hills Road,
*Cambridge CB*2 2*QH, England*

(*Received* 5 *April* 1991; *accepted* 15 *July* 1991)

Abstract

The study of families of protein structures is important in analysing the results of NMR structure determinations and in investigating mechanisms of molecular evolution at the level of conformation. A method is discussed for finding the transformations that mutually superpose an arbitrary number of structures in the least-squares sense given specified atom-to-atom correspondence.

0108-7673/92/010011-04$03.00

1. Introduction

Superposition has become an important tool for comparing protein structures and for deriving an 'average' structure from a family of conformations of a protein. The problem arises regularly in the determination of the three-dimensional structure of a protein in solution using nuclear magnetic resonance, which typically produces an ensemble of conformations (Wuethrich, 1986). Sets of interproton distances obtained from measuring cross relaxation between nuclei less than 5 Å apart can be converted into a 3D structure by several methods, but all suffer from underdetermination as the number of observable distances is far smaller than the number of degrees of freedom. As a result, a family of structures, all consistent with experimental data, is produced (Kaptein, Zuiderweg, Scheek, Boelens & van Gunsteren, 1985; Clore, Gronenborn, Bruenger & Karplus, 1985; Havel, Kuntz & Crippen, 1983; Braun & Gō, 1985; Bassolino *et al.*, 1988; Nilges, Clore & Gronenborn, 1988).

Given a family of conformations, a method to measure their similarity and dispersion is necessary. Techniques in common use rely on calculations of minimal root-mean-square deviations of atomic positions to achieve optimal superposition. For a pair of structures, techniques for finding the global minimum root-mean-square deviation in atomic position, with respect to rigid-body relative motions, are known (von Neumann, 1937; Kabsch, 1976, 1978; Golub & Van Loan, 1983). However, for more than two structures, most methods so far described involve combinations of pairwise superpositions. In practice, NMR spectroscopists usually either (*a*) calculate an average structure and then superpose all the structures on it or (*b*) superpose all the structures on one member of the family chosen as a reference.

Three previous reports address directly the problem of multiple structural superposition (Gerber & Müller, 1987; Sutcliffe, Haneef, Carney & Blundell, 1987; Kearsley, 1990). These methods provide solutions of multiple superposition problems in many practical cases, especially if the structures are quite similar. However, it is of interest to note that another general solution, free of simplifying assumptions and heuristic devices, exists. This is discussed in § 2.

2. Problem formulation and iterative procedures

In this paper we discuss an optimization problem of simultaneous rotation of M rigid molecules to maximal similarity. That is, let $(\mathbf{x}_{ik})$ be a set of 3×1 vectors such that for every $k=1,\ldots,M$ the set $(\mathbf{x}_{ik})$, $i=1,\ldots,N$, represents atomic coordinates of a given molecule. Following Gerber & Müller (1987) we take the following criterion for a match (similarity) between rotated molecules:

$$E=\sum_{k<l}^{M} v_{kl}\sum_{i=1}^{N} w_i\|\mathbf{T}_k\mathbf{x}_{ik}-\mathbf{T}_l\mathbf{x}_{il}\|^2, \qquad (1)$$

where $\mathbf{T}_1,\ldots,\mathbf{T}_M$ are rotation (orthogonal) matrices and v_{kl} and w_i are given positive weights. (Here $\|\mathbf{x}\|^2=\mathbf{x}'\mathbf{x}$ denotes the squared length of a vector $\mathbf{x}$ and $\mathbf{x}'$ stands for the transpose of $\mathbf{x}$.) An optimal similarity is achieved by minimization of the criterion E as function of the orthogonal matrices $\mathbf{T}_1,\ldots,\mathbf{T}_M$.

It will be convenient to formulate the obtained optimization problem in a matrix form. Consider $3\times N$ data matrices $\mathbf{X}_k=(\mathbf{x}_{1k},\ldots,\mathbf{x}_{Nk})$, $k=1,\ldots,M$, the $N\times N$ diagonal matrix $\mathbf{W}=\operatorname{diag}(w_i)$ and the 3×3 matrices $\mathbf{S}_{kl}=v_{kl}\mathbf{X}_k\mathbf{W}\mathbf{X}'_l$, $k, l=1,\ldots,M$. Then

$$\begin{aligned}E&=\sum_{k<l}^{M} v_{kl}\operatorname{tr}(\mathbf{T}_k\mathbf{X}_k-\mathbf{T}_l\mathbf{X}_l)\mathbf{W}(\mathbf{T}_k\mathbf{X}_k-\mathbf{T}_l\mathbf{X}_l)'\\&=-2\sum_{k<l}^{M}\operatorname{tr}\mathbf{T}_k\,\mathbf{S}_{kl}\mathbf{T}'_l\\&\quad+\text{a sum independent of }\mathbf{T}_j.\end{aligned}$$

Consequently, the problem of minimization of E is equivalent to maximization of

$$G=\sum_{k<l}^{M}\operatorname{tr}\mathbf{T}_k\mathbf{S}_{kl}\mathbf{T}'_l. \qquad (2)$$

For $M=2$ this optimization problem has a closed-form solution. Namely, one has to choose $\mathbf{T}_1$ and $\mathbf{T}_2$ in such a way that the matrix $\mathbf{T}_1\mathbf{S}_{12}\mathbf{T}'_2$ is diagonal and positive semidefinite. That is, $\mathbf{T}_1$ and $\mathbf{T}'_2$ are formed by orthonormal eigenvectors of the matrices $\mathbf{S}_{12}\mathbf{S}'_{12}$ and $\mathbf{S}'_{12}\mathbf{S}_{12}$, respectively. This result has a long history. It was derived by von Neumann (1937) and has been rediscovered many times since. Notice that the optimal solution is not unique. We can always premultiply $\mathbf{T}_1$ and $\mathbf{T}_2$ by an orthogonal matrix without changing the trace of $\mathbf{T}_1\mathbf{S}_{12}\mathbf{T}'_2$. In particular, we can replace $\mathbf{T}_1\mathbf{S}_{12}\mathbf{T}'_2$ by $\mathbf{T}'_2\mathbf{T}_1\mathbf{S}_{12}$. The orthogonal matrix $\mathbf{T}=\mathbf{T}'_2\mathbf{T}_1$ is then optimal if and only if $\mathbf{T}\mathbf{S}_{12}$ is symmetric and positive semidefinite (*cf.* Ten Berge, 1977).

For $M\geq 3$ a closed-form solution is not available and an iterative procedure is required. A simple and surprisingly efficient algorithm was proposed by Ten Berge (1977). The idea is to maximize G with respect to one orthogonal matrix at a time keeping other orthogonal matrices fixed. That is, let matrices $\mathbf{T}_2,\ldots,\mathbf{T}_M$ be fixed and consider G as a function of $\mathbf{T}_1$ alone,

$$G=\operatorname{tr}\mathbf{T}_1\mathbf{S}_{1.}+\text{a sum independent of }\mathbf{T}_1,$$

where

$$\mathbf{S}_{k.}=\sum_{l\neq k}\mathbf{S}_{kl}\mathbf{T}'_l,\qquad k=1,\ldots,M.$$

We can now maximize G with respect to $\mathbf{T}_1$ by the diagonalization procedure and similarly for $\mathbf{T}_2$ *etc.*

Ten Berge's algorithm can be described as follows. Choose initial values of the matrices $\mathbf{T}_k$, $k=1,\ldots,M$. Usually these initial values are taken to be the identity matrices. Maximize G with respect to $\mathbf{T}_1$ and substitute the obtained optimal value. Do the same, in turn, for $\mathbf{T}_2,\ldots,\mathbf{T}_M$. Repeat the procedure until the increment in G becomes less than a prescribed precision given by a positive number ε, say $\varepsilon=10^{-4}$.

Ten Berge's algorithm converges to a stationary point where all matrices $\mathbf{S}_{k\cdot}$, $k=1,\ldots,M$, are symmetric and positive semidefinite. This is a necessary condition for optimality. Unfortunately, this necessary condition may be not sufficient (see Ten Berge, 1977, p. 270). Therefore a dual procedure was proposed by Shapiro & Botha (1988) in order to verify optimality of an obtained stationary point. Let

$$\mathbf{S}(\mathbf{Z})=\begin{bmatrix} \mathbf{Z}_1 & \mathbf{S}_{12} & \ldots & \mathbf{S}_{1M} \\ \mathbf{S}_{21} & \mathbf{Z}_2 & \ldots & \mathbf{S}_{2M} \\ \vdots & \vdots & \ldots & \vdots \\ \mathbf{S}_{M1} & \mathbf{S}_{M2} & \ldots & \mathbf{Z}_M \end{bmatrix}$$

be a $3M\times 3M$ symmetric matrix considered as a function of the symmetric block-diagonal matrix $\mathbf{Z}=\operatorname{diag}(\mathbf{Z}_1,\ldots,\mathbf{Z}_M)$. Denote by $\lambda_1(\mathbf{Z})\geq\ldots\geq\lambda_{3M}(\mathbf{Z})$ the eigenvalues of $\mathbf{S}(\mathbf{Z})$. Then for every $\mathbf{Z}$ the number

$$f(\mathbf{Z})=\tfrac{1}{2}\left[M\sum_{i=1}^{3}\lambda_i(\mathbf{Z})-\operatorname{tr}\mathbf{S}(\mathbf{Z})\right] \tag{3}$$

gives an upper bound for the maximum of G (Shapiro & Botha, 1988, theorem 1). By minimizing $f(\mathbf{Z})$ over all block-diagonal matrices, one obtains the corresponding least upper bound. It can be shown that this least upper bound is equal to the maximum of G if $\lambda_3(\mathbf{Z}_0)\neq\lambda_4(\mathbf{Z}_0)$, where $\mathbf{Z}_0$ is the minimizer of $f(\mathbf{Z})$.

It was found by Shapiro & Botha (1988) that a very good starting value for minimization of $f(\mathbf{Z})$ is constructed as follows. Let $\mathbf{P}_{kl}\mathbf{D}_{kl}\mathbf{Q}'_{kl}$ be singular-value decompositions of the matrices $\mathbf{S}_{kl}$, *i.e.* $\mathbf{D}_{kl}$ are diagonal and positive semidefinite and $\mathbf{P}_{kl}$ and $\mathbf{Q}_{kl}$ are orthogonal. Then take

$$\mathbf{Z}_k^*=-\sum_{\substack{l=1\\ l\neq k}}^{M}\mathbf{P}_{kl}\mathbf{D}_{kl}\mathbf{P}'_{kl}, \tag{4}$$

$$k=1,\ldots,M \text{ and } \mathbf{Z}^*=\operatorname{diag}(\mathbf{Z}_1^*,\ldots,\mathbf{Z}_M^*).$$

It was found that for this value of $\mathbf{Z}$ the corresponding upper bound $f(\mathbf{Z}^*)$ was usually very close to the maximum value of G produced by Ten Berge's algorithm. Therefore, for practical purposes, it was usually sufficient to evaluate the upper bound $f(\mathbf{Z}^*)$ in order to confirm optimality of the calculated stationary point.

It may happen that some of the optimal orthogonal matrices calculated by Ten Berge's algorithm have negative determinant (equal to -1). This will be the case if the determinant of the corresponding matrices $\mathbf{S}_{k\cdot}=\sum_{l\neq k}\mathbf{S}_{kl}$ is negative. Then the obtained optimal solution suggests reflections as well as rotations of the considered structures. Although such a situation is unlikely to happen in practice, it is possible from the theoretical point of view. In the remainder of this section we briefly discuss how to deal with this case.

Let us study first the case of $M=2$. Again, in this case the problem has a closed-form solution. That is, consider the matrix $\mathbf{S}_{12}$ and suppose that $\det\mathbf{S}_{12}<0$. Let $\bar{\mathbf{T}}_1$ and $\bar{\mathbf{T}}_2$ be orthogonal matrices such that the matrix $\mathbf{S}_{12}^*=\bar{\mathbf{T}}_1\mathbf{S}_{12}\bar{\mathbf{T}}'_2$ is diagonal, $\mathbf{S}_{12}^*=\operatorname{diag}(d_1,d_2,d_3)$, and positive semidefinite. Since $\det\mathbf{S}_{12}<0$, we have here that the determinants of $\bar{\mathbf{T}}_1$ and $\bar{\mathbf{T}}_2$ have different signs, say $\det\bar{\mathbf{T}}_1=1$ and $\det\bar{\mathbf{T}}_2=-1$. We can choose matrices $\bar{\mathbf{T}}_1$ and $\bar{\mathbf{T}}_2$ in such a way that the diagonal elements of $\mathbf{S}_{12}^*$ are arranged in decreasing order, *i.e.* $d_1\geq d_2\geq d_3$.

We have to find rotation matrices $\mathbf{T}_1$ and $\mathbf{T}_2$ such that the trace of $\mathbf{T}_1\mathbf{S}_{12}\mathbf{T}'_2$ is maximized. It is claimed then that the optimal rotation matrices are given by the matrices $\bar{\mathbf{T}}_1$ and $\mathbf{D}\bar{\mathbf{T}}_2$, where $\mathbf{D}$ is the diagonal matrix with diagonal elements 1, 1 and -1, *i.e.* $\mathbf{D}=\operatorname{diag}(1,1,-1)$. That is, the maximum of the trace of $\mathbf{T}_1\mathbf{S}_{12}\mathbf{T}'_2$, subject to $\mathbf{T}_1$ and $\mathbf{T}_2$ being orthogonal and $\det\mathbf{T}_1=\det\mathbf{T}_2=1$, is given by $d_1+d_2-d_3$.

Indeed, consider a 3×3 orthogonal matrix $\mathbf{T}$ with $\det\mathbf{T}=-1$. Such a matrix can be represented in the form $\mathbf{T}=\mathbf{QAQ}'$, where $\mathbf{Q}$ is an orthogonal matrix and $\mathbf{A}$ is a block-diagonal matrix with the first block given by a 2×2 rotation matrix $\mathbf{B}$ and the second block consisting of the element -1 (*e.g.* Curtis, 1979). We have then

$$\operatorname{tr}\mathbf{S}_{12}^*\mathbf{T}=\operatorname{tr}\mathbf{Q}'\mathbf{S}_{12}^*\mathbf{QA}=\operatorname{tr}\mathbf{Q}'_1\mathbf{S}_{12}^*\mathbf{Q}_1\mathbf{B}-\mathbf{q}'_3\mathbf{S}_{12}^*\mathbf{q}_3,$$

where $\mathbf{q}_1$, $\mathbf{q}_2$ and $\mathbf{q}_3$ are column vectors of the matrix $\mathbf{Q}$ and $\mathbf{Q}_1$ is the 3×2 matrix formed by the first two columns of $\mathbf{Q}$, *i.e.* $\mathbf{Q}_1=[\mathbf{q}_1,\mathbf{q}_2]$. Since $\mathbf{Q}'_1\mathbf{S}_{12}^*\mathbf{Q}_1$ is symmetric and positive semidefinite, we know that the maximum of $\operatorname{tr}\mathbf{Q}'_1\mathbf{S}_{12}^*\mathbf{Q}_1\mathbf{B}$ is attained when $\mathbf{B}$ is the identity matrix. Also by Ky Fan's inequality (Ky Fan, 1949) we have $\operatorname{tr}\mathbf{Q}'_1\mathbf{S}_{12}^*\mathbf{Q}_1$ is less than or equal to the sum of the two largest eigenvalues of $\mathbf{S}_{12}^*$, that is d_1+d_2. Similarly, $\mathbf{q}'_3\mathbf{S}_{12}^*\mathbf{q}_3\geq d_3$. All this implies

$$\operatorname{tr}\mathbf{T}_1\mathbf{S}_{12}\mathbf{T}'_2=\operatorname{tr}\mathbf{S}_{12}^*\mathbf{T}\leq d_1+d_2-d_3$$

where

$$\mathbf{T}=\bar{\mathbf{T}}_2\mathbf{T}'_2\mathbf{T}_1\bar{\mathbf{T}}'_1 \quad\text{and}\quad \det\mathbf{T}=-1.$$

For $M\geq 3$ an obvious analogue of Ten Berge's iterative algorithm can then be derived. Unfortunately, it appears that in the case of negative determinants the Shapiro–Botha upper-bound procedure cannot be applied for verification of optimality of the calculated solution.

3. Numerical experimentation

Ten Berge's algorithm together with the Shapiro–Botha dual procedure was applied to five representative sets of coordinates of the protein molecule acyl phosphatase in an attempt to superpose them optimally. For each representative set there are 392 triples of coordinates that represent the positions of the 392 atoms that make up the whole structure. Only the unweighted optimization problem was considered; that is, minimization of E in (1) with respect to $\mathbf{T}_1, \ldots, \mathbf{T}_M$, with $v_{kl} = 1 (1 \le k < l \le M)$ and $w_i = 1 (1 \le i \le N)$.

First the representative sets were translated so that the centres of all lie at the origin. Hence for each set $\mathbf{X}_k = (\mathbf{x}_{1k}, \ldots, \mathbf{x}_{Nk})$, $1 \le k \le M$, $\sum_{i=1}^{N} \mathbf{x}_{ik} = \mathbf{0}$. Then the maximal value (denoted by G_0) of the function G in (2) was obtained by means of the Ten Berge algorithm. Recall that the maximum of G corresponds to the minimum of the criterion E given in (1).

Note that this algorithm differs from algorithm 1 proposed by Sutcliffe *et al.* (1987, p. 378), which incidentally is identical to an algorithm proposed earlier by Kristof & Wingersky (1971). Both the Ten Berge and Kristof–Wingersky algorithms yield upon convergence stationary points which satisfy some necessary conditions for the optimality of G. However, as was indicated by Ten Berge (1977, pp. 270–272), the necessary condition satisfied by a stationary point obtained from the Ten Berge algorithm is stronger than that of Kristof & Wingersky.

The value of G_0, which was obtained in two iterations, is 605 606·0. This turned out to be very close to the optimal value of G. Evaluation of the upper bound $f(\mathbf{Z}^*)$ of G, with f and $\mathbf{Z}^*$ defined as in (3) and (4) respectively, yields the value 605 607·0.

The program used for the above calculation also provides for an attempted minimization of the upper bound $f(\mathbf{Z})$ of G by means of a Newton-like differentiation method. It uses $\mathbf{Z}^*$ as defined in (4) as starting value. However, this is not always possible, since the function f is convex, but not everywhere differentiable. In fact, it is differentiable at a point $\mathbf{Z} = \operatorname{diag}(\mathbf{Z}_1, \ldots, \mathbf{Z}_M)$ if and only if $\lambda_3(\mathbf{Z}) \ne \lambda_4(\mathbf{Z})$ [see Shapiro & Botha (1988) for a more detailed discussion of the properties of f and the relevant minimization technique].

In this particular case $f(\mathbf{Z})$ was minimized successfully. The difference $\lambda_3(\mathbf{Z}) - \lambda_4(\mathbf{Z})$ was sufficiently large throughout the minimization process with $|\lambda_4(\mathbf{Z})/\lambda_3(\mathbf{Z})| \ge 10^2$ (the eigenvalues were all negative). This means that $f(\mathbf{Z})$ was differentiable at each step and that the stationary point is a global minimizer, since $f(\mathbf{Z})$ is convex. Furthermore, the minimal value of f turned out to be equal to G_0, which implies that the Ten Berge algorithm converged to the optimal solution. Hence, in this case the orthogonal matrices obtained from the Ten Berge algorithm optimally superpose the original five representative sets of coordinates. Also, it was found that these orthogonal matrices are all rotation matrices.

Finally, note that in this case the upper bound $f(\mathbf{Z}^*)$ is very close to the optimal value of both f and G. This was also borne out by other experiments. This may therefore be a good way to check the performance of the Ten Berge algorithm, especially if one does not want to go through the process of minimizing f.

AML thanks the Kay Kendall Foundation for generous support.

References

Bassolino, D. A., Hirata, F., Kitchen, D. B., Kominos, D., Pardi, A. & Levy, R. M. (1988). *Int. J. Supercomput. Appl.* **2**, 41–61.

Braun, W. & Gō, N. (1985). *J. Mol. Biol.* **186**, 611–626.

Clore, G. M., Gronenborn, A. M., Bruenger, A. T. & Karplus, M. (1985). *J. Mol. Biol.* **186**, 435–455.

Curtis, C. W. (1979). *Linear Algebra*, 3rd ed., p. 270. Boston: Allyn and Bacon.

Gerber, P. R. & Müller, K. (1987). *Acta Cryst.* **A43**, 426–428.

Golub, G. H. & Van Loan, C. F. (1983). *Matrix Computations*, pp. 425–426. Baltimore, MD: The Johns Hopkins Univ. Press.

Havel, T., Kuntz, I. D. & Crippen, G. M. (1983). *Bull. Math. Biol.* **45**, 655–720.

Kabsch, W. (1976). *Acta Cryst.* **A32**, 922–923.

Kabsch, W. (1978). *Acta Cryst.* **A34**, 827–828.

Kaptein, R., Zuiderweg, E. R. P., Scheek, R. M., Boelens, R. & van Gunsteren, W. F. (1985). *J. Mol. Biol.* **182**, 179–182.

Kearsley, S. K. (1990). *J. Comput. Chem.* **11**, 1187–1192.

Kristof, W. & Wingersky, B. (1971). Proc. 79th Annual Convention of the American Psychological Association, pp. 89–90.

Ky Fan (1949). *Proc. Natl Acad. Sci. USA*, **35**, 652–655.

Neumann, J. von (1937). *Tomsk Univ. Rev.* **1**, 286–300.

Nilges, M., Clore, G. M. & Gronenborn, A. M. (1988). *FEBS Lett.* **229**, 317–324.

Shapiro, A. & Botha, J. D. (1988). *SIAM* (*Soc. Ind. Appl. Math.*) *J. Matrix Anal. Appl.* **9**, 378–383.

Sutcliffe, M. J., Haneef, I., Carney, D. & Blundell, T. L. (1987). *Protein Eng.* **1**, 377–384.

Ten Berge, J. M. F. (1977). *Psychometrika*, **42**, 267–276.

Wuethrich, K. (1986). *NMR of Proteins and Nucleic Acids.* New York: Wiley.