References

COLLINS, D. M. (1982). *Nature (London)*, **298**, 49–51.

FOX, G. C. & HOLMES, K. C. (1966). *Acta Cryst.* **20**, 886–891.

HAMILTON, W. C., ROLLETT, J. S. & SPARKS, R. A. (1965). *Acta Cryst.* **18**, 129–130.

JAYNES, E. T. (1968). *IEEE Trans.* SSC-4, pp. 227–241.

JAYNES, E. T. (1979). In *The Maximum Entropy Formalism*, edited by R. D. LEVINE & M. TRIBUS, pp. 15–118. Cambridge: Massachusetts Institute of Technology.

MONAHAN, J. E., SCHIFFER, M. & SCHIFFER, J. P. (1967). *Acta Cryst.* **22**, 322.

NICHOLSON, W. L., PRINCE, E., BUCHANAN, J. & TUCKER, P. (1982). In *Crystallographic Statistics: Progress and Problems*, edited by S. RAMASESHAN, M. F. RICHARDSON & A. J. C. WILSON, pp. 230–363. Bangalore: Indian Academy of Science.

NIELSEN, K. (1977). *Acta Cryst.* **A33**, 1009–1010.

NORDMAN, C. E. (1960). *Acta Cryst.* **13**, 535–539.

RAE, A. D. (1965). *Acta Cryst.* **19**, 683–684.

RAE, A. D. & BLAKE, A. B. (1966). *Acta Cryst.* **20**, 586.

ROSSMANN, M. G., LESLIE, A. G. W., ABDEL-MEGUID, S. S. & TSUKIHARA, T. (1979). *J. Appl. Cryst.* **12**, 570–581.

Acta Cryst. (1984). A**40**, 708-712

Comparison of Methods of Matching Protein Structures

BY C. E. KENKNIGHT

Biochemistry Department, Biosciences West, University of Arizona, Tucson, AZ 85721, *USA*

(*Received* 24 *February* 1984; *accepted* 3 *July* 1984)

Abstract

Two fast methods of superposing two sets of atomic coordinates by least-squares refinement are described and related to two earlier fast methods. A Newton method is applied to rotations of a 3 × 3 outer product matrix used previously by Ferro & Hermans [*Acta Cryst.* (1977), A33, 345–347] and by McLachlan [*Acta Cryst.* (1972), A**28**, 656–657]. Three of the methods work better if one molecule has its inertial matrix aligned with *xyz*. A Newton–Gauss method that rotates the coordinates can converge rapidly after a rough orientation using three strategic atoms. The average superposition takes about 0·003 s on a Cyber 175 with the best method, rotations about the *xyz* axes in turn. Experience with reliability is reported for large residuals.

Introduction

This paper describes experience with several methods for calculating the rigid-body rotations that are needed for matching similar molecular structures. Two new methods are presented and compared with published methods. In applications to proteins there is a systematic search for a likeness between a fragment of structure anywhere in one protein, *A*, to any part of a second protein, *B* (Rao & Rossmann, 1973; Rossmann & Argos, 1976, 1977; Remington & Matthews, 1978; McLachlan, 1979). A typical search involves more than a million structure matches, so a fast method is essential. The frequency of matches that do not reach the global minimum for the least-squares search is also of concern for interpretation of supposed likenesses.

Three of the methods under discussion here are, in principle, equivalent. Thus, aspects of the theory given next are like that given by McLachlan (1972) and by Ferro & Hermans (1977). Performances of the associated algorithms are not equivalent and effort was directed to understanding why not. Running times were studied for a variety of conditions: magnitude of the residuals, magnitude of the relative rotations, orientation of one of the coordinate sets, and use of several tricks to speed or ensure convergence. The theory set forth here was helpful in understanding what gave minimal running times.

The problem and the Newton method

Let $\mathbf{a}_k$, $\mathbf{b}_k$ ($k = 1$ to N) be the position vectors of two sets of N atoms from the molecular fragments A and B. Let w_k be a weight for each atom. We want to minimize the residual E, an inner product,

$$E = \tfrac{1}{2}\sum w_k(R\mathbf{a}-\mathbf{b})'_k(R\mathbf{a}-\mathbf{b})_k. \tag{1}$$

Here the prime signifies a transpose so that $\mathbf{a}'$ is a row vector. If we were to interpret w_k as the strength of a linear spring joining the atoms numbered k (McLachlan, 1982), then E would have the interpretation of a potential energy. Such a system is static if the net force and torque on, say, A due to B vanishes. The vanishing force requires that the centroids of A and B coincide (McLachlan, 1972; Remington & Matthews, 1978) while the vanishing torque requires that the weighted vector cross product of the structures vanish:

$$\mathbf{g} = \sum w_k(\mathbf{b}_k \times \mathbf{a}_k) = 0. \tag{2}$$

Thus it is reasonable that a unique orthogonal proper rotation matrix R with determinant $+1$ exists (see McLachlan, 1979), which transforms $\mathbf{a}$ referred to the centroid as origin to $\mathbf{r} = R\mathbf{a}$ and minimizes the residual

0108-7673/84/060708-05$01.50

E. The only exception can be compared to finding a pencil balanced on its point.

The inner product (1) can be expanded (McLachlan, 1972; Ferro & Hermans, 1977). Using the invariance of the inner product $\mathbf{a}'\mathbf{a}$ under rigid-body rotation one gives attention to the coordinate outer product sum U, where

$$U_{ij} = \sum w_k (a_i b_j)_k. \tag{3}$$

The outer product U has the simple transformation $V = RU$, where

$$V_{ij} = \sum w_k (r_i b_j)_k. \tag{4}$$

Only the trace of V is involved in changes of the residual E. Both the first and second (and higher) derivatives of E can be expressed in terms of V. Then one may choose between updating the coordinates through $\mathbf{r} = R\mathbf{a}$ or updating the coordinate product through $V = RU$. The coordinate update requires $9N$ multiplications. The update of the outer product U requires 27 multiplications, so that we gain speed if N exceeds three atoms, other matters of overhead being equal. If the final coordinates are to be rotated for some purpose such as display, then a net rotation matrix must also be updated, another 27 multiplications. Then we gain speed for N greater than six atoms.

The search for a rotation matrix R is necessarily a series of approximations. The residual E is far from a parabolic function unless the fragments are closer than about 0·1 rad from the final solution. In fact, on most great circles of revolution the behavior of the residual function E is crudely sinusoidal. Thus, the second derivative of E on half of such a circle is apt to be negative, rather than positive, along the direction that the first derivative suggests is downward. In a Newton method we would estimate the step size to the solution from E'/E'', where the derivatives are evaluated along a great circle. The algorithm must resort to a restricted step method when E'' is negative; a reasonable step is 0·8 rad along the downward direction before evaluating a new gradient. The likelihood of finding a negative E'' can be made small by using a routine that precedes the least-squares matching and that accepts a rotation of 180 or 90° about an axis x, y or z if the trace of U increases. These operations require no multiplications, just sign changes and row or column permutations. Alternatively, a rough match can be done for three strategic atoms, as discussed later. In a conjugate gradient (CG) method (McLachlan, 1982) or a quasi-Newton method (*e.g.* Fletcher, 1980), the evaluation of E'' is normally done; it is the double inner product $\mathbf{s}'H\mathbf{s}$ of the Hessian (second derivative) matrix H with the path vector $\mathbf{s}$, and requires 12 multiplications. Attention to the sign of E'' with a Newton method greatly simplifies the logic of a line search and even permits the avoidance of a line search. I find that the shortest robust line search uses the current point and two new points, does a quadratic fit to the traces of U and interpolates for the maximum trace (minimum E).

Missing the global minimum with the Newton method is increasingly likely both when the molecules become unlike and when the molecule has few atoms. For definiteness I shall call small residuals those that are less than the r.m.s. separation along a strand, about 1·5 Å. I shall call large rotations those bigger than 1 rad. An occasional failure to find the global minimum owing to large residuals would have little significance in a search for cases of small differences between structures. Only experience with widely varied molecules can reveal the frequency of failures when the fragment residuals become large.

The search for the minimum of the residual E in a Newton method becomes a search for a zero of the gradient ∇E. In a Taylor series

$$\nabla E = \mathbf{g} + H\mathbf{s} + \tfrac{1}{2}\mathbf{s}'T\mathbf{s} + \ldots, \tag{5}$$

we set the left side to zero and evaluate the right-hand side derivatives at the current position. The third-derivative tensor T and higher orders are ignored. The solution for the step $\mathbf{s}$ becomes

$$H\mathbf{s} = -\mathbf{g}. \tag{6}$$

A large Hessian matrix often leads to handling of (6) by a CG method or a quasi-Newton method. But here the Hessian matrix is 3×3. Reduction to upper triangular form and solution by substitution requires only 11 multiplications and six divisions. The competing operation in the CG method is an update of the path direction, which requires three multiplications and one division, but the use of a relatively inexact path requires several additional steps.

Let us define a rotation matrix by successive plane rotations θ_3 about the z axis, θ_2 about the new y axis, and θ_1 about the final x axis (*e.g.* Remington & Matthews, 1978). Each positive rotation has the counter-clockwise sense of a polar angle when looking back along the named axis toward the origin. Defining the corresponding cos and sin values C_i and HHS_i, we find a finite rotation matrix R has the form

$$R = \begin{vmatrix} C_2C_3 & C_2S_3 & -S_2 \\ -C_1S_3 + S_1S_2C_3 & C_1C_3 + S_1S_2S_3 & S_1C_2 \\ S_1S_3 + C_1S_2C_3 & S_1C_3 + C_1S_2S_3 & C_1C_2 \end{vmatrix}. \tag{7}$$

Performing the three partial derivatives of R and evaluating them at $\theta_i = 0$, we find

$$D_1R = \begin{vmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{vmatrix}, \quad D_2R = \begin{vmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{vmatrix},$$

$$D_3R = \begin{vmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 0 \end{vmatrix}. \tag{8}$$

Forming the three matrix products with U, we find the three traces of the derivatives of RU becomes the gradient

$$\begin{aligned} g_i &= \mathrm{tr}\,(D_i RU) \\ &= \sum w_k(a_3b_2 - a_2b_3,\ a_1b_3 - a_3b_1,\ a_2b_1 - a_1b_2)_k \\ &= \sum w_k(\mathbf{b}\times\mathbf{a})_k. \end{aligned} \tag{9}$$

Proceeding similarly, the nine second partial derivatives of R are equally sparse when evaluated at $\theta_i = 0$ and yield, for a Hessian matrix H_1,

$$H_1 = \mathrm{tr}\,(D_{ij} RU)$$

$$= -\sum w_k \left|\begin{matrix} a_2b_2 + a_3b_3 & -a_1b_2 & -a_1b_3 \\ -a_1b_2 & a_1b_1 + a_3b_3 & -a_2b_3 \\ -a_1b_3 & -a_2b_3 & a_1b_1 + a_2b_2 \end{matrix}\right|_k . \tag{10}$$

On comparison with (8) we note that if $v = \mathrm{tr}\,(U)$ we can construct $-H_1$ from $v\delta_{ij} - U_{ij}$ provided that we use only the upper triangle of U, then make H_1 symmetric. Testing of the normal equations

$$-H_1\mathbf{s}_1 = \mathbf{g} \tag{11}$$

as described below revealed that the step vector $\mathbf{s}_1$ was usually about 10% wrong for random molecules, even for $s_1 < 0{\cdot}1$ rad.

The problem lies in the asymmetric use of information in U. Thus far we have considered rotating A into B. Had we rotated B into A, an independent problem, we would have used the transpose matrix R'. The normal equations become

$$-H_2\mathbf{s}_2 = -\mathbf{g}, \tag{12}$$

where H_2 involves the lower triangle of U. The overall change in sign from $\mathbf{s}_1$ to $\mathbf{s}_2$ is simply the change in sense of the rotations. Not surprisingly, the second kind of step is no better than the first. An order of magnitude in, but not complete, improvement (unless one of the molecules has its principal moments of inertia aligned with the coordinate system) is the averaging of $\mathbf{s}_1$ and $-\mathbf{s}_2$, which yields

$$H_{ij} = v\delta_{ij} - \tfrac{1}{2}(U_{ij} + U_{ij}) \tag{13}$$

in agreement with McLachlan (1982; his T_{ij}), who found an alternative algebraic derivation of (13). The present derivation is a reminder to use every symmetry in an optimization procedure.

The improvement of the symmetrized Hessian can be understood by consideration of the first neglected term in the Taylor series (5). The Hessian to higher order is $H + \mathbf{s}_1'({}_1\mathrm{T} - {}_2\mathrm{T})/4$. The difference of the tensors is 27 terms, of which all but six survive; they are components of g_i in (9) and become small near the solution. The six terms that subtract to zero involve the product a_2b_2, which becomes large near the solution in general; those terms caused the poor performance of H_1 and H_2.

The advantage of aligning one of the fragments so that its off-diagonal inertial matrix elements vanish requires consideration of the second neglected term in the Taylor series (5). The Hessian should properly be modified by terms quadratic in the angles needed to step to the solution, where the difference terms in the gradient (9) would vanish. When the angles are sizeable but one of the fragments is aligned with the coordinates, the perturbations to the Hessian are decreased. Even so, I found it imperative to check repeatedly whether the residual E could be decreased by flips of 180 or 90° during iterations of the Newton method. Such repeated flips were not needed by the CG method or the *xyz* rotations of Nyburg (1974) and Ferro & Hermans (1977).

The usefulness of aligning a fragment for the method of *xyz* rotations arises because the *xyz* axes otherwise have no special significance for the problem; they are not eigenvectors of the inertial matrix $\sum w_k|a\rangle\langle a|_k$, in the Dirac bracket notation. When fragment A is so aligned, its inertial matrix is diagonal. When residuals are small, the matching of B to A amounts to a procedure that diagonalizes $\sum w_k|b\rangle\langle b|_k$ in addition to $U = \sum w_k|a\rangle\langle b|_k$. In all the cases that posed difficulties of convergence I found that the determinant of U was negative, a circumstance never observed with $N = 50$ atoms but occurring with a frequency of a few percent with $N = 5$ and the 3 Å residuals of Table 1.

Tricks

A rough match of two fragments using three widely spaced atoms can be arranged as follows. Translate one atom pair to the origin. Put a second atom pair on the x axis by two rotations of the coordinate systems. Put a third atom pair in the $y = 0$ plane. A more efficient version of this copies the six coordinate triples, performs the operations only on the copies, and loads a suitable rotation matrix for multiplications on all the coordinates to be rotated, say, B. Schemes to select the three atoms include prior knowledge that they are widely separated or the calculation of the magnitude of cross products of the vectors defined by the three atoms. A less elaborate rough match was used by Nyburg (1974).

For the three special atoms I used the two atoms separated by the largest distance along one of the three original axes and that third atom having the largest cross product with those two. That procedure needs $6(N-2)$ multiplications. Even shorter is the use in the cross-product search of only the other four extreme atoms found during the distance search.

Alignment of A with *xyz* to adequate accuracy can be ensured in nine plane rotations. The rotation about each axis in turn is chosen to zero one of the off-

Table 1. *Execution times per* 250 *matches of random helices with residuals of about* 3Å

Four methods were used for $N = 5$ and $N = 50$ atoms per molecule. Each method could choose to use flips of 180° or of both 180 and 90°. Three methods could elect a line search (*LS*). The molecules were prepared with alignment of fragment *A* square with *xyz* (*SQ*), roughly with extreme atoms along *x* (*R*), or with random orientation (*N*). Columns repeat until a new entry appears.

Method	Atoms	Align	Flip	LS	Time(s)	Comments
xyz	5	*SQ*	Both	*N*	0·90	
			180		0·85	
			None		0·82	
	50		Both		0·82	
			180		0·82	
			None		0·82	
		N	Both		1·50	
			180		1·53	
			None		1·51	
CG	5	*SQ*	Both	*N*	1·68	
			180		1·49	
			None		1·62	
	50		Both		1·53	
			180		1·49	
			None		1·69	
	5	*N*	Both		1·53	
			180		1·51	
			None		1·59	
	50		Both		1·51	
			180		1·53	
			None		1·64	
Newton	5	*SQ*	Both	*N*	1·00	2 misses
				Y	1·74	1×
	50			*N*	1·05	0×
				Y	1·69	0×
	5	*N*		*N*	0·95	25×
				Y	2·77	7×
	50			*N*	1·02	11×
				Y	1·67	4×
Rotate coordinates	5	*SQ*	Both	*N*	3·14	7 slow
				Y	5·01	1 slow
			None		5·46	1 slow
	50		Both	*N*	7·69	
				Y	12·93	
			None	*N*	8·26	
				Y	13·66	
	5	*R*	Both	*N*	3·59	7 slow
				Y	5·08	1 slow
	50			*N*	8·03	
				Y	12·99	
	5	*N*		*N*	3·42	10 slow
				Y	5·14	1 slow
	50			*N*	7·41	
				Y	13·77	

diagonal elements of the inertial matrix. The calculation requires $6N$ multiplications to obtain the matrix elements and $4N$ multiplications per plane rotation, up to $90N$ multiplications for one of the two molecules. Alignment of A with $N = 50$ takes 0·013 s on the Cyber. However, the rotation of fragment A is done only once for many matches with fragments B.

The coordinate alignment about z, say, minimizes the weighted distance to the plane of z and the rotated x axis:

$$T(\theta_3) = \sum w_k (Cy - Sx)_k^2. \tag{14}$$

The required rotation by θ satisfies

$$\tan 2\theta_3 = 2 \sum (wxy)_k / \sum w_k (x^2 - y^2)_k. \tag{15}$$

The numerator and denominator can be treated like sine and cosine components of a phasor in assigning the quadrant for $2\theta_3$. Similar relations hold for each axis. Three cycles through the three axes ensure that the off-diagonal elements of the inertial matrix vanish for the molecular fragment.

The *xyz* rotation method

Nyburg (1974) noted that the plane rotation, say, about z that best matches the xy coordinates has a closed form; *i.e.* if

$$\begin{vmatrix} r_1 \\ r_2 \end{vmatrix} = \begin{vmatrix} C & S \\ -S & C \end{vmatrix} \begin{vmatrix} a_1 \\ a_2 \end{vmatrix} \tag{16}$$

and

$$\begin{aligned} T(\theta_3) &= \sum w_k [(r_1 - b_1)^2 + (r_2 - b_2)^2] \\ &= \sum w_k [(Ca_1 + Sa_2 - b_1)^2 + (-Sa_1 + Ca_2 - b_2)^2], \end{aligned} \tag{17}$$

then the derivative of T vanishes for

$$\tan \theta_3 = \sum w_k (a_2 b_1 - a_1 b_2)_k / \sum w_k (a_1 b_1 + a_2 b_2)_k. \tag{18}$$

We recognize a cross product in the numerator and a dot product in the denominator, so, as in the last section, the phase of θ_3 can be assigned to a quadrant. There is no need for concern about large rotations. Nyburg rotated coordinates as in (16), but Ferro & Hermans (1977) noted that the quantities in (18) are in the coordinate outer product sum U and rotated U. Table 1 shows that sign flips that avoid large rotations save some time, but the saving is not significant. Three iterations through the three axes will usually ensure 100 μrad matching when the residuals are small and one of the fragments is squared up, just as the squaring up of that fragment converges in three tries in most cases.

The Newton–Gauss method

I found a method that is rapidly convergent for small residuals and for any orientation after a rough match is arranged. However, it almost certainly requires rotating the coordinates, rather than updating the coordinate outer product U. This time we ask for a vector $\mathbf{c}$ that minimizes the estimated residual

$$E = \tfrac{1}{2} \sum w_k \| (\mathbf{a} - \mathbf{c} \times \mathbf{a}) - (\mathbf{b} + \mathbf{c} \times \mathbf{b}) \|_k^2. \tag{19}$$

Unlike the correct circular arcs that the atoms follow in the previous method, we here use tangent line segments $\mathbf{c} \times \mathbf{a}$ and $\mathbf{c} \times \mathbf{b}$ in *estimating* the residual. For brevity we introduce the vectors $\mathbf{x} = \mathbf{a} + \mathbf{b}$ and

$\mathbf{q}=\mathbf{a}-\mathbf{b}$. The normal equations become

$$H\mathbf{c}=\sum w_k(\mathbf{x}\times\mathbf{q})_k, \tag{20}$$

where the Hessian

$$H=\sum w_k \begin{vmatrix} x_2^2+x_3^2 & -x_1x_2 & -x_1x_3 \\ -x_1x_2 & x_1^2+x_3^2 & -x_2x_3 \\ -x_1x_3 & -x_2x_3 & x_1^2+x_2^2 \end{vmatrix}_k \tag{21}$$

is manifestly symmetric in the coordinates of the fragments.

This new method, however, is not like each of the others reviewed here. Some of the terms in the Hessian do not occur in the outer product U, although some do. Expanding the vector cross product quickly shows that the term

$$\mathbf{x}\times\mathbf{q}=2\mathbf{b}\times\mathbf{a} \tag{22}$$

transforms under rotations like $\mathbf{g}$ in (9), which are off-diagonal terms in U. On the diagonal of (21) terms like

$$x_2^2+x_3^2=2(a_2b_2+a_3b_3)+(a_2^2+b_2^2+a_3^2+b_3^2) \tag{23}$$

correspond to previous cross terms plus unchanging terms. But off the diagonal we find terms

$$x_1x_2=(a_1b_2+a_2b_1)+(a_1a_2+b_1b_2) \tag{24}$$

that are in the outer product U plus other terms whose changing is complex unless we rotate the coordinates. I find that convergence is completely reliable with large residuals but strugglingly slow near the solution.

It can be shown that the method of (19) is a small-angle Newton–Gauss treatment of (1), which ensures that the Hessian (21) is positive definite, unlike the indefinite Hessian of (11), (12), or the average of those equations. The Newton–Gauss approximation ignores the curvature of the paths taken by the atoms in rotating into a match. That also suppresses information and causes the slow convergence with large residuals, *i.e.* with a U having a negative determinant. In a Newton method starting with (1) one adds the second-order corrections and obtains (11) or (12). Thus, the simple-appearing rotation problem is a useful tutorial exercise for any introduction to numerical methods.

Testing the algorithms

The 'random molecules' used for test purposes were based on a linear random walk and were of two types. 'Random helices' used consecutive thirds of the walk for x coordinates of the atoms, then the y and z coordinates. The r.m.s. spatial separation of adjacent atoms was scaled to a constant value. The molecule B was generated by adding Gaussian noise of zero mean to each of the coordinates of fragment A. Then some rotation of B was added. 'Random pencils' used three consecutive numbers of the walk for x, y and z of an atom, and so on. The resultant high correlations between x, y and z left the 'pencil' aligned along a space diagonal passing through the origin. The latter type afforded the more severe test for the algorithms because one of the three real eigenvalues of the Hessian is very near to zero while the other two are almost coincident. The advantage of using computer-generated coordinate sets was that the magnitude of the residuals could be varied as desired, unlike the situation in the final application.

The result of comparisons of various methods with rather unlike molecules is given in Table 1. The independent methods sometimes gave unequal residuals after apparently valid stops. The minimum residual was accepted as correct. Major conclusions are as follows. The *xyz* rotations are fastest, simplest and completely reliable. The Newton method is as fast as the CG method, but not reliable at a few misses of the global minimum per thousand and rather more complex because a line search is imperative. The CG method is fully reliable but takes a bit more code and memory than *xyz* rotations. It gains almost nothing from the tricks used by the other methods. The rotation of the entire molecule coordinate set puts the Newton–Gauss method at a disadvantage in general. It is more suited for special problems where small rotations suffice. Slow convergence means 30 iterations or more to reach 100 μrad steps. The residuals were unchanging in the fourth decimal place during most of these iterations.

Fortran listings are available from the author. This work was supported by NIH grant 30393 and by the Computer Center of the University of Arizona.

References

FERRO, D. R. & HERMANS, J. (1977). *Acta Cryst.* **A33**, 345–347.
FLETCHER, R. (1980). *Practical Methods of Optimization.* Vol. 1. *Unconstrained Optimization*, ch. 3. Chichester: Wiley.
MCLACHLAN, A. D. (1972). *Acta Cryst.* **A28**, 656–657.
MCLACHLAN, A. D. (1979). *J. Mol. Biol.* **128**, 49–79.
MCLACHLAN, A. D. (1982). *Acta Cryst.* **A38**, 871–873.
NYBURG, S. C. (1974). *Acta Cryst.* **B30**, 251–253.
RAO, S. T. & ROSSMANN, M. G. (1973). *J. Mol. Biol.* **105**, 241–256.
REMINGTON, S. J. & MATTHEWS, B. W. (1978). *Proc. Natl Acad. Sci. USA*, **75**, 2180–2184.
ROSSMANN, M. G. & ARGOS, P. (1976). *J. Mol. Biol.* **105**, 75–96.
ROSSMANN, M. G. & ARGOS, P. (1977). *J. Mol. Biol.* **109**, 99–129.