FORTIER, S., DE TITTA, G. F., FRONCKOWIAK, M., SMITH, G.
    D. & HAUPTMAN, H. (1979). *Acta Cryst.* B35, 2062-2066.
FORTIER, S., FRONCKOWIAK, M., SMITH, G. D., HAUPTMAN,
    H. & DE TITTA, G. (1978). International School of Crystallogra-
    phy, Erice, Italy, Supplement XXXIII-5-20.
GERMAIN, G., MAIN, P. & WOOLFSON, M. M. (1970). *Acta Cryst.*
    B26, 274-285.
KANTERS, J. A. & VAN BOMMEL, M. (1987). In preparation.
KANTERS, J. A., JANSMA, S. & PONTENAGEL, W. M. G. F. (1986).
    To be published.

KOTEN, G. VAN, JASTRZEBSKI, J. T. B. H., NOLTES, J. G.,
    PONTENAGEL, W. M. G. F., KROON, J. & SPEK, A. L. (1978).
    J. Am. Chem. Soc. 100, 5021-5028.
KROON, J., DUISENBERG, A. J. M. & PEERDEMAN, A. F. (1984).
    Acta Cryst. C40, 645-647.
PONTENAGEL, W. M. G. F. (1984). *Acta Cryst.* A40, 314-323.
SPEK, A. L., DE RUITER, E. T. & PONTENAGEL, W. M. G. F.
    (1982). *Cryst. Struct. Commun.* 11, 1869-1872.
VAND, V. & PEPINSKY, R. (1956). *Z. Kristallogr.* 107, 202-
    224.

Isomorphous Replacement in Fiber Diffraction Using Limited Numbers of Heavy-Atom Derivatives

BY KEIICHI NAMBA AND GERALD STUBBS

Department of Molecular Biology, Vanderbilt University, Nashville, TN 37235, USA

Abstract

Multi-dimensional isomorphous replacement, analogous to isomorphous replacement in protein crystallography, can be used in fiber diffraction analysis to overcome the problems caused by the cylindrical averaging of the intensity data. Large numbers of heavy-atom derivatives are needed, however. A method is presented by which molecular structures may be determined using data from only one or two derivatives, similar to crystallographic single isomorphous replacement. Partial structure information may also be incorporated. Examples are given using data from oriented gels of tobacco mosaic virus, and possibilities for further applications are discussed.

Introduction

Many important biological macromolecules, for example actin, myosin, tubulin, flagellin, and the coat proteins of some viruses, naturally form filamentous assemblies and have functions specific to those assemblies. Even in cases where it is possible to crystallize such molecules as monomers or small aggregates, it is important to know the molecular structure of the intact assembly in order to understand the function of the molecule. It is therefore necessary to use the methods of fiber diffraction.

Fiber diffraction from macromolecules has many aspects in common with protein crystallography, but there are also a number of major differences. The most important of these stems from the fact that although the filamentous particles in a fiber diffraction specimen are oriented with their long axes

approximately parallel, they are randomly oriented about those axes. As a consequence, the observed diffraction pattern is the cylindrical average of the diffraction pattern to be expected from one particle (in the absence of interference effects) or from a fully ordered array of particles (in the case of a crystalline fiber). Considerable information is lost in this averaging; for example, the effective number of observable diffraction data for tobaco mosaic virus (TMV) at 3 Å resolution is reduced by a factor about 2·5, and for the bacteriophage Pfl at the same resolution by 1·7 (Makowski, 1982). These factors are much higher for lower-symmetry systems such as microtubules.

The intensity of fiber diffraction can be written (Waser, 1955; Franklin & Klug, 1955)

$$I(R,l) = \sum_n G_{n,l}(R)G^*_{n,l}(R) \qquad (1)$$

where l is the layer-line number, R is the reciprocal-space radius and n is the order of the Bessel functions that contribute to the complex Fourier-Bessel structure factor G (Klug, Crick & Wyckoff, 1958). Equation (1) can be compared with the crystallographic equation

$$I(h,k,l) = F_{hkl}F^*_{hkl}.$$

If each G is known, a Fourier-Bessel transform of the G terms will give an electron density map, just as a Fourier transform of the F terms will give an electron density map for a crystal. Instead of two unknowns (the real and imaginary parts of F), there are $2N$ unknowns, where N is the number of terms (G terms) contributing to the sum in (1). N depends on the size and symmetry of the diffracting particle

and increases with resolution. For TMV at 4 Å, $N = 5$ (Namba & Stubbs, 1985) but for microtubules at only 18 Å resolution, N is already 7 (Beese, 1984).

It was shown by Stubbs & Diamond (1975) that this problem, the phase problem for fiber diffraction, could be solved by multi-dimensional isomorphous replacement, a technique analogous to the isomorphous replacement method of protein crystallography. This technique, however, requires $2N$ heavy-atom derivatives. To determine the structure of TMV at 4 Å would require ten derivatives. In favorable cases, the number of derivatives required can be reduced by making use of the fine splitting of layer lines (Stubbs & Makowski, 1982), which occurs when an integral number of turns in the diffracting helical structure contains very close to, but not exactly, an integral number of identical subunits. For example, TMV has 49·02 subunits in three turns. Layer-line splitting may in theory reduce the number of derivatives required by a factor of two, but its use requires accurate measurements of layer-line splitting (manifested as very small shifts in the layer-line position), and these can only be made with particularly well oriented specimens.

It is evidently desirable to develop methods which will allow us to determine structures with as few heavy-atom derivatives as possible, particularly when layer-line splitting data are not available. In this paper, we consider what may be done with data from a native specimen and either one or two derivatives. If the relative magnitudes of the terms contributing to the sum (1) are known for both native and derivative data (colloquially, if the separation of the Bessel orders is known), the phase of each term may be determined separately, without considering the other terms. We consider the effect of assuming that all terms in the sum (1) contribute equally, and the effect of assuming that their contributions may be approximated by the expected contributions from a related structure. We use TMV as a test case, comparing the electron density maps we obtain with a model built from a 3·6 Å resolution map in which intensity data from six derivatives and layer-line splitting data from four were used (Namba & Stubbs, 1985, 1986).

Data collection

The data used were part of the data set described by Namba & Stubbs (1985). Briefly, X-ray fiber diffraction patterns were obtained from oriented gels of TMV and its derivatives, and diffracted intensities were determined using the angular deconvolution method of Makowski (1978). Intensity data were sampled along layer lines at intervals of 0·001 Å^{-1}. Data to 4 Å resolution were used. Derivative data were scaled to native data using a Fourier surface whose coefficients (usually for five orders) were determined by a least-squares method, taking into con-

sideration the heavy-atom contribution (Namba & Stubbs, 1985). The derivative data used were the methyl mercury nitrate derivative of TMV *vulgare* and the methyl mercury nitrate derivative of the mutant Ni-2068 (Tyr 139 → Cys). Heavy-atom parameters are given in Table 1 of Namba & Stubbs (1985).

Double isomorphous replacement

With data available from a native and two heavy-atom derivatives, we have three observations with which to determine as many as ten unknowns. While this is clearly an extremely under-determined situation, we can make certain assumptions which will allow us to estimate the phases even in this case.

Initial phasing

We begin by assuming that all the significant **G** terms in (1) contribute equally to the observed intensity, both in the native and the derivative diffraction patterns. With this assumption, each term can be phased separately, using the conventional two-dimensional phasing procedure of protein crystallography described in such texts as Blundell & Johnson (1976). The mean-square lack-of-closure error can be estimated from the low-resolution region (lower than 10 Å) of the equator, a centric zone where only zero-order Bessel functions contribute. A phase probability distribution is calculated using this mean-square error, and the 'best' phase and figure of merit (Blow & Crick, 1959) are determined. The mean-square error could be re-estimated at this point using all the data, although this was not done in this work. An electron density map is calculated by combining the 'best' phases with the evenly divided structure amplitudes, weighted by the figures of merit. It is extremely important to use appropriate values for the mean-square error, because although this does not affect the 'best' phase angle very much, the quality and interpretability of the electron density map depend very strongly on the figures of merit, which are directly coupled with the mean-square error. While this is always true in protein crystallography, it is particularly important here, because the figures of merit partly reflect the quality of the assumption that the Bessel order terms contribute equally to the intensity.

A map of TMV was calculated in this way. The mean figure of merit was relatively low (0·44) as would be expected. [This may be compared with the value of 0·56, which was obtained using six derivatives and with the contributions of the Bessel order terms calculated by multi-dimensional isomorphous replacement with layer-line splitting (Namba & Stubbs, 1985).] The electron density in this map (Fig. 1a) was very weak, but comparing it with the model built from the six-derivative map (Namba & Stubbs,

1985), the principal features of the virus structure could be seen. The four α helices that run approximately radially could be recognized, as could the RNA backbone running along the virus helix. Although the peptide chain was disconnected in several places, most parts of the molecule were in density. Even though this map was calculated using the obviously wrong assumption that all Bessel order terms contribute equally to the intensity, it clearly contains useful information, and would be well worth calculating if no other source of phase information were available.

Refinement by density modification

This map was improved by a procedure similar to solvent flattening refinement. Solvent flattening as
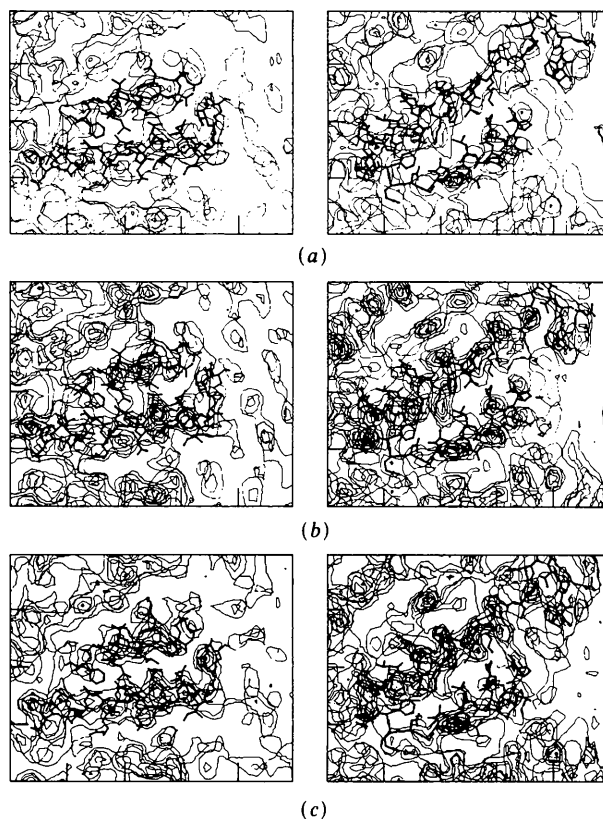

(a)

(b)

(c)

Fig. 1. Sections of the TMV map calculated using two heavy-atom derivatives. Left side: three superimposed sections, 1·4 Å apart, containing the left and right radial α helices of the coat protein. Residues 73 to 87 and 114 to 138, taken from the six-derivative multi-dimensional isomorphous replacement structure (Namba & Stubbs, 1986), are superimposed as a skeletal model. Right side: four sections containing the left and right slewed helices, with residues 19 to 63 superimposed. (a) Calculated assuming equal contributions from all significant Bessel order terms in both native and derivative data. (b) Calculated after one cycle of solvent flattening refinement of (a), followed by separation of the Bessel order terms based on the electron density. (c) Calculated by separating the Bessel order terms as predicted from the atomic coordinates of a partial subunit structure from the crystalline protein disk, rotated into map (b).

such, where density outside an envelope presumed to exclude only solvent regions is set to zero (Namba & Stubbs, 1985), was not sufficiently powerful to correct the poorly determined multi-dimensional phases. Instead, a contour level that appeared to contain most of the structure was selected, and all density below this contour was set to zero. This contour was set high enough to effectively form a much tighter envelope than is usually used in solvent flattening, penetrating all parts of the molecule. It appears that the errors introduced by excluding small areas of real structure are more than offset by the errors removed. Such a procedure is only appropriate for a poorly determined map, where the level of error is expected to be higher than the level of density in the low-density regions around the edge of the structure.

This modified density distribution was treated as a model to be used in refinement of the phases. The separation of Bessel orders was calculated from the modified density for each data point, for native and derivative data. This was done by deriving the relative magnitudes of the terms for the native data directly from the transform of the density, while the relative magnitudes of the terms for the derivative data were calculated using the heavy-atom parameters and the calculated native phases, as described by Namba & Stubbs (1985). The magnitudes were normalized so that the calculated intensity agreed with the observed intensity. Double isomorphous replacement phases were calculated using these separated Bessel terms. The mean figure of merit at this point was 0·57, a
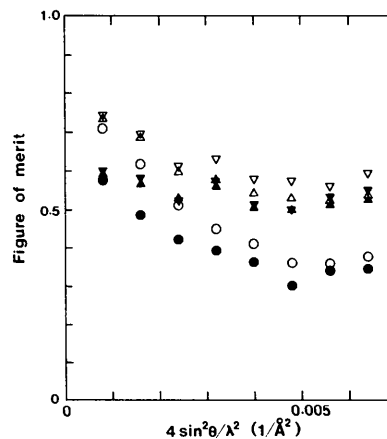

Fig. 2. Figures of merit (m) as a function of resolution. Open symbols: phases calculated using two heavy-atom derivatives (DIR). Closed symbols: phases calculated using one heavy-atom derivative (SIR).

	$<m>$	
	DIR	SIR
○ Assuming equal contributions to intensity from contributing Bessel orders:	0·44	0·39
△ After one cycle of solvent-flattening refinement:	0·57	0·53
▽ After two cycles of solvent-flattening refinement:	0·61	0·54

definite improvement over the initial 0·44. The improvement was particularly evident at high resolution (Fig. 2). Further cycles of refinement did not significantly improve the figures of merit or the interpretability of the electron density map.

Parts of the electron map from one cycle of density modification refinement are shown in Fig. 1(*b*). This map has considerably higher electron density than the unrefined map in Fig. 1(*a*), and is somewhat more interpretable. Both of the regions shown contain two α helices running approximately parallel to the sections, and models of these helices, as determined using six derivatives (Stubbs, Warren & Holmes, 1977; Namba & Stubbs, 1986) are shown superimposed on the maps. Although there are some areas of disconnected density and unexplained extra density, most of the model fits the map well, and the general molecular structure is easy to recognize.

Refinement by model building

It might be possible, but it would probably be difficult, to build an atomic model solely on the basis of the map shown in Fig. 1(*b*). If an atomic model of a similar structure or a partial structure were available, however, it would be quite easy to orient such
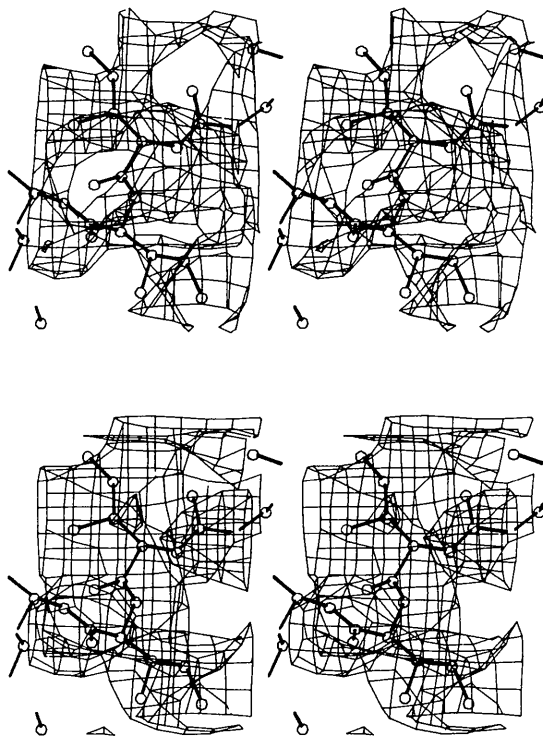

Fig. 3. Improvement in the electron density map of TMV using a partial structure model to separate Bessel orders. Stereo pairs with superimposed model of –Ile 93–Ile 94–. Top: Bessel order separation assumed equal, then refined for one cycle of solvent flattening. Bottom: Bessel order separation on the basis of a molecular model that did not include the residues shown.

a model in the map. The separation of the Bessel order terms derived from such a model could then be used as a basis for a further round of double isomorphous replacement phasing, and the model could be corrected or extended. Such models may be available when the monomer or some other fraction of the assembly under study has been crystallized, as is the case for actin (Kabsch, Mannherz & Suck, 1985) and TMV (Bloomer, Champness, Bricogne, Staden & Klug, 1978) respectively. Such a procedure would be analogous to the alternating rounds of model-building and two-dimensional isomorphous replacement phase determination used by Namba & Stubbs (1985).

The feasibility of this approach was investigated, using unrefined coordinates from the structure of the protein disk of TMV. The disk crystallizes, and its structure has been determined at 2·8 Å resolution by Bloomer *et al.* (1978). It does not contain RNA, and 24 of its amino acid residues are part of a disordered loop in a region corresponding to the inner surface of the virus, but coordinates were available for 134 residues. The rigid-body transformation needed to locate the disk subunit in the map was first determined by eye in projection, and then adjusted using an Evans and Sutherland graphics system. The transformed coordinates were used to calculate structure factors and thus separate Bessel order terms, and new phases were determined by double isomorphous replacement.

The resulting map, shown in Fig. 1(*c*), was significantly better than the density modification map, and structure could be recognized in many areas outside the region defined by the partial structure. The density corresponding to Ile 93 and Ile 94, which are both part of the disordered loop in the disk but are ordered in the virus, is shown with an atomic model fitted in Fig. 3.

Single isomorphous replacement

Even when only one heavy-atom derivative is available, the procedures described above may still be used. Given the assumption of equal contributions to the intensity from all the Bessel order terms, it is still possible to determine the centroid of the phase probability and obtain a 'best' phase and a figure of merit. Since the single isomorphous replacement phase probability is always a bimodal function, the mean figure of merit is somewhat lower than it is in double isomorphous replacement (Fig. 2). The electron density map thus calculated (Fig. 4*a*) is, as expected, of lower quality, and the density is weaker. However, there are some recognizable features such as α helices in this map.

Refinement by density modification was applied as in the case of double isomorphous replacement, although the lower overall density meant that a lower

contour level had to be used as a molecular envelope. Once again there was a marked improvement in the mean figure of merit (Fig. 2), from 0·39 to 0·53. The map was significantly improved by the refinement (Fig. 4b), both in density and in interpretability. Further cycles of refinement did not result in any appreciable change. Although this map is not as good as the double isomorphous replacement map shown in Fig. 1(b), it is certainly good enough to recognize the general features of the structure and to allow molecular replacement if a related or partial structure is available.

Bessel order separations calculated using the transformed coordinates from the disk subunit partial structure were used to obtain the single isomorphous replacement map shown in Fig. 4(c). This map is almost as good as the corresponding double isomorphous replacement map (Fig. 1c) in the region shown, but this is the part of the molecule whose coordinates were included in the partial structure. The density in the part of the molecule corresponding to the flexible

(a)

(b)

(c)

Fig. 4. Sections of the TMV map calculated using one heavy-atom derivative. Sections and superimposed models as for Fig. 1. (a) Calculated assuming equal contributions from all significant Bessel order terms in both native and derivative data. (b) Calculated after one cycle of solvent flattening refinement of (a), followed by separation of the Bessel order terms based on the electron density. (c) Calculated by separating the Bessel order terms as predicted from the atomic coordinates of the disk subunit partial structure.

loop of the disk, whose coordinates were not available for use in the Bessel order separation, is not as good as, for example, that shown in Fig. 3.

Potential applications

The procedures we have described appear to work well for determining a structure of the degree of difficulty of TMV, that is, one where the data include five overlapping Bessel orders. This is particularly so if two heavy-atom derivatives are available, but even the maps derived from one derivative are informative. Thus it appears that the assumption that the Bessel order terms contribute equally to the intensity is good enough to allow isomorphous replacement to be used for phase determination. Qualitatively, this is because, for this number of terms, ΔI between the native and the derivative correlates reasonably well with $\Delta |G|^2$. This correlation will decrease progressively with an increasing number of terms, so that although there is no limitation inherent in the algorithm on the number of terms that can be separated, there will come a point where the maps calculated will be uninterpretable. Nonetheless, there are many structures having sufficiently high symmetry to be accessible to these methods, particularly at low resolution. Viruses and most crystalline fibers are particularly good potential candidates.

If an initial molecular model can be built, either directly from a map or with the aid of a related partial structure, further refinement with the goal of a complete detailed model can be considered. In fiber diffraction, care must always be taken during such a refinement to ensure that the model does not completely dominate the refinement procedure, and thus prevent convergence to any structure different from itself; this is an obvious consequence of the reduced number of observational data. It is thus advisable not to discard the isomorphous replacement data in favor of, for example, a restrained least-squares procedure (Hendrickson & Konnert, 1980; Stubbs, Namba & Makowski, 1986) during the early stages of refinement.

Ideally, one would calculate maps using a method that combined information from the partial structure with isomorphous replacement information. Such methods are available for crystallography (Sim, 1959; Hendrickson & Lattman, 1970), but have not yet been developed in a full multi-dimensional form for fiber diffraction. Our experience with the structure determination of TMV by multi-dimensional isomorphous replacement (Namba & Stubbs, 1985) suggests, however, that a satisfactory expedient is the calculation of Bessel order separations from the model, followed by determination of phases by combining model and isomorphous replacement information, using a weighting scheme analogous to that of Sim (1959). Combination of phase information in this way
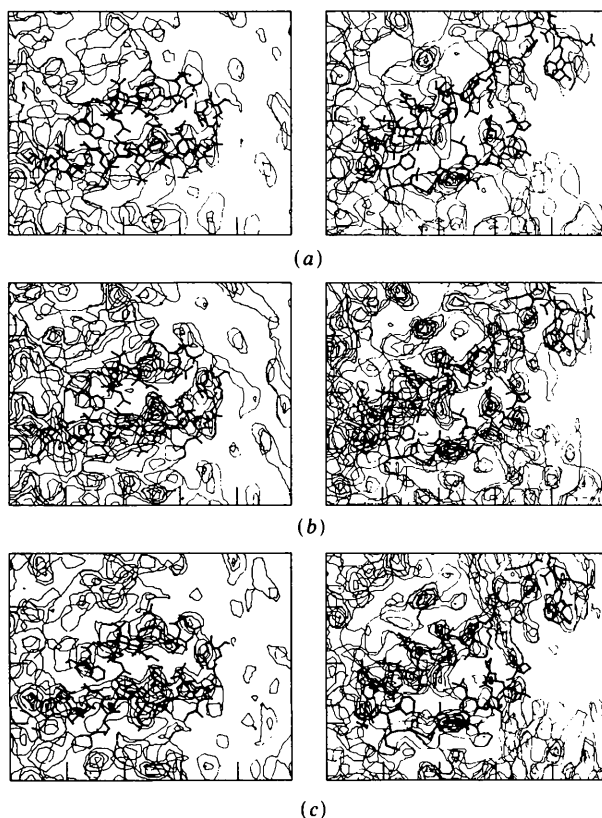
would be particularly important in the case of single isomorphous replacement, where the isomorphous replacement phase distributions are bimodal, and could not be expected to give highly interpretable maps. Further rounds of density modification might also be useful.

The value of these methods is greatly increased when a partial structure is known, and here our results are very encouraging. The crystalline TMV protein disk is not a particularly good model for the intact virus, not only because it lacks RNA and provides no information at all for the inner loop, about 15% of the protein, but because virtually all of the amino acid side chains in the subunit interfaces (a high proportion in such a long thin molecule) have different conformations in the two assemblies (Namba & Stubbs, 1986). Even so, interpretable maps were obtained. In a case where a better starting model was available, one might expect to be able to determine a structure relatively easily.

An immediately obvious application of these methods is to other strains of a virus whose structure is already known. A number of strains of TMV have yielded good diffraction patterns, but we have not been able to determine their structures by simply applying difference Fourier methods. Obtaining the number of derivatives required for a complete *de novo* structure determination would be a long and difficult process but, in several cases, one or two derivatives are available. The virus U2, with 70% coat protein homology to TMV, appears to have a very similar structure to TMV except in a region near the inner wall (Holmes & Franklin, 1958), and two derivatives of this virus have been made (Mandelkow & Holmes, 1974). Cucumber green mottle mosaic virus (watermelon strain), with 40% homology to TMV, has more significant differences, judging from a radial density distribution, but presumably still has essentially the same protein fold as TMV. At least one derivative of this virus is available (Lobert, Heil, Namba & Stubbs, 1986). Both of these viruses appear to differ in structure from TMV in regions that are important in the control of viral assembly; thus, their detailed molecular structures are of considerable interest. Application of the single or double isomorphous replacement phasing methods described here should permit elucidation of these structures.

References

BEESE, L. (1984). PhD Thesis, Brandeis Univ.

BLOOMER, A. C., CHAMPNESS, J. N., BRICOGNE, G., STADEN, R. & KLUG, A. (1978). *Nature (London)*, **276**, 362-368.

BLOW, D. M. & CRICK, F. H. C. (1959). *Acta Cryst.* **12**, 794-802.

BLUNDELL, T. L. & JOHNSON, L. N. (1976). *Protein Crystallography.* New York: Academic Press.

FRANKLIN, R. E. & KLUG, A. (1955). *Acta Cryst.* **8**, 777-780.

HENDRICKSON, W. A. & KONNERT, J. H. (1980). *Computing in Crystallography*, edited by R. DIAMOND, S. RAMASESHAN & K. VENKATESAN, pp. 13.01-13.26. Bangalore: Indian Academy of Science.

HENDRICKSON, W. A. & LATTMAN, E. E. (1970). *Acta Cryst.* **B26**, 136-143.

HOLMES, K. C. & FRANKLIN, R. E. (1958). *Virology*, **6**, 328-336.

KABSCH, W., MANNHERZ, H. G. & SUCK, D. (1985). *EMBO J.* **4**, 2113-2118.

KLUG, A., CRICK, F. H. C. & WYCKOFF, H. W. (1958). *Acta Cryst.* **11**, 199-213.

LOBERT, S., HEIL, P. D., NAMBA, K. & STUBBS, G. (1986). *Biophys. J.* **49**, 528a.

MAKOWSKI, L. (1978). *J. Appl. Cryst.* **11**, 273-283.

MAKOWSKI, L. (1982). *J. Appl. Cryst.* **14**, 160-168.

MANDELKOW, E. & HOLMES, K. C. (1974). *J. Mol. Biol.* **87**, 265-273.

NAMBA, K. & STUBBS, G. (1985). *Acta Cryst.* **A41**, 252-262.

NAMBA, K. & STUBBS, G. (1986). *Science*, **231**, 1401-1406.

SIM, G. A. (1959). *Acta Cryst.* **12**, 813-815.

STUBBS, G. & DIAMOND, R. (1975). *Acta Cryst.* **A31**, 709-718.

STUBBS, G. & MAKOWSKI, L. (1982). *Acta Cryst.* **A38**, 417-425.

STUBBS, G., NAMBA, K. & MAKOWSKI, L. (1986). *Biophys. J.* **49**, 58-60.

STUBBS, G., WARREN, S. & HOLMES, K. (1977). *Nature (London)*, **267**, 216-221.

WASER, J. (1955). *Acta Cryst.* **8**, 142-150.