position is 24 E_1:24 E_2:12 E_3; where E_1 and E_3 are dimeric proteins, no other subunit arrangement is possible.

This research was supported by grant AM28607 for P. A. Frey from the National Institute of Arthritis, Diabetes and Digestive and Kidney Diseases, and grant RR0177 to J. S. Wall from the National Institutes of Health Biotechnology Resource Branch.

Received for publication 9 May 1985.

REFERENCES

1. Eley, M .H., G. Namihira, L. Hamilton, P. Munk, and L. J. Reed. 1972. α-Keto acid dehydrogenase complexes. Subunit composition of the *Escherichia coli* pyruvate dehydrogenase complex. *Arch. Biochem. Biophys.* 152:655–669.
2. Bleile, D. M., P. Munk, R. M. Oliver, and L. J. Reed. 1979. Subunit structure of dihydrolipoyl transacetylase component of pyruvate dehydrogenase complex from *Escherichia coli. Proc. Natl. Acad. Sci. USA.* 76:4385–4389.
3. Angelides, K. J., S. K. Akiyama, and G. G. Hammes. 1979. Subunit stoichiometry and molecular weight of the pyruvate dehydrogenase multienzyme complex from *Escherichia coli. Proc. Natl. Acad. Sci. USA.* 76:3279–3283.
4. CaJacob, C. A., P. A. Frey, J. F. Hainfeld, J. S. Wall, and H. Yang. 1985. *Escherichia coli* pyruvate dehydrogenase complex. Particle masses of the complex and component enzymes measured by scanning transmission electron microscopy. *Biochemistry.* 24:2425–2431.
5. Willms, C. R., R. M. Oliver, H. R. Henney, Jr., B. B. Mukherjee, and L. J. Reed. 1967. α-Keto acid dehydrogenase complexes. Dissociation and reconstitution of the dihydrolipoyl transacetylase of *Escherichia coli. J. Biol. Chem.* 242:889–897.
6. Reed, L. J. 1974. Multienzyme complexes. *Accts. Chem. Res.* 7:40–46.
7. Reed, L. J., F. H. Pettit, M. H. Eley, L. Hamilton, J. H. Collins, and R. M. Oliver. 1975. Reconstitution of the *Escherichia coli* pyruvate dehydrogenase complex. *Proc. Natl. Acad. Sci. USA.* 72:3068–3072.
8. Reed, L. J., and C. R. Willms. 1966. Purification and resolution of the pyruvate dehydrogenase complex (*Escherichia coli*). *Methods Enzymol.* 9:247–265.
9. Speckhard, D. C., and P. A. Frey. 1975. *Escherichia coli* pyruvate dehydrogenase complex: improved purification and the flavin content. *Biochem. Biophys. Res. Commun.* 62:614–620.
10. Steven, A. C., J. F. Hainfeld, B. L. Trus, P. M. Steinert, and J. S. Wall. 1984. Radial distributions of density within macromolecular complexes determined from dark-field electron micrographs. *Proc. Natl. Acad. Sci. USA.* 81:6363–6367.

APPLICATION OF RESTRAINED LEAST-SQUARES REFINEMENT TO FIBER DIFFRACTION FROM MACROMOLECULAR ASSEMBLIES

GERALD STUBBS,* KEIICHI NAMBA,* AND LEE MAKOWSKI‡
**Department of Molecular Biology, Vanderbilt University, Nashville, Tennessee 37235; and*
‡Department of Biochemistry and Molecular Biophysics, College of Physicians and Surgeons of Columbia University, New York, New York 10032

The principal concern in the refinement of the structure of a macromolecular assembly against fiber diffraction data is the relatively small number of independent data available compared with the number of adjustable molecular parameters. At the resolution limit of the best fiber diffraction patterns from helical viruses (~3 Å) there are too few measurable intensities per model parameter for stable refinement of the structure using diffraction data alone. This is because the intensities observed in a fiber diffraction pattern are the cylindrical average of the intensity distribution on layer planes in reciprocal space. Thus, far fewer data are available as restraints on the values of the molecular parameters. For example, cylindrical averaging reduces the effective number of observable diffraction data for tobacco mosaic virus (TMV) at 3 Å resolution by a factor of ~2.5 and for the bacteriophage Pf1 at the same resolution by a factor of 1.7 (Makowski, 1982). Consequently, stereochemical information must be incorporated into the refinement process to increase the ratio of observations to parameters. This can be done in two ways. Stereochemical information can be used in the form of constraints that fix the values of selected bond lengths and angles, reducing the number of parameters to be refined. This is the approach taken in the linked-atom least-squares (LALS) method of Arnott and his collaborators (Arnott and Wonacott, 1966; Smith and Arnott, 1978), in which dihedral bond angles are refined. This has been particularly effective in refining the structure of helical polymers with relatively small repeating units such as nucleic acids and polysaccharides using diffraction data from crystalline fibers. An alternative approach is one in which stereochemical information is used in the form of restraints. This method as developed by Hendrickson and Konnert (1980) has been widely used in the refinement of protein structures against crystallographic data. In this approach, the ideal values of bond lengths and angles are treated as additional observational equations and the refinement of atomic coordinates attempts to minimize deviations from these values while simultaneously minimizing the difference between calculated and observed diffraction data. Additional restraints can readily be incorporated into the refinement. Because restrained least-squares methods offer some advantages in speed and in final agreement with the observed diffraction data (Dod-

BIOPHYS. J. © Biophysical Society · 0006–3495/86/01/58/03 $1.00
Volume 49 January 1986 58–60

son et al., 1976), as well as allowing a realistic flexibility in the molecular structure, we have chosen to adapt this method for use in the refinement of macromolecular assemblies against fiber diffraction data.

ADAPTATION TO DIFFRACTION FROM HELICAL ASSEMBLIES

Two fundamental changes in the method are required to make it appropriate for use with helical assemblies. First, the structure factor calculation must be replaced with a Fourier-Bessel transform because fiber diffraction data are calculated in cylindrical coordinates. Second, intermolecular interactions must be considered in solving the structure of a macromolecular assembly because these are usually extensive in such assemblies.

Computer programs provided by Dr. W. Hendrickson were used as a starting point and some modifications written by Dr. E. Westhof to handle nucleic acid structures were incorporated. Adaptation of these programs was greatly facilitated by their modular form. The structure factor calculation was replaced by a Fourier-Bessel transform using extensive Bessel function look-up tables. Optimization of this calculation was essential because fiber diffraction structure factors usually take much longer to calculate than crystallographic structure factors and the structure factor calculation is the most time-consuming part. Routines were added to generate molecules related to each other by helical symmetry. Van der Waals interactions between molecules related by crystallographic symmetry were not part of the original programs, because these are not usually important in crystalline structures. They are, however, significant in close-packed structures such as tobacco mosiac virus (TMV) and Pf1, so they were included. Also included were restraints from covalent bonds between helical subunits, because such bonds are extremely common in fiber diffraction systems. For example, TMV contains a single continuous strand of RNA that follows the basic helix of the virus structure, with three nucleic acid residues in each repeating unit of the helix.

APPLICATION TO TOBACCO MOSIAC VIRUS

Restrained least-squares refinement was applied to the structure of TMV. The starting structure was determined from fiber diffraction data at 3.6 Å resolution by a combination of multi-dimensional isomorphous replacement, layer-line splitting, and electron-density modification (Namba and Stubbs, 1985). The starting model had a crystallographic *R*-factor for data between 10 Å and 3.6 Å resolution of 0.31. It should be noted that, because of the cylindrical averaging of the data, fiber diffraction *R*-factors are inherently lower than crystallographic *R*-factors. 30 cycles of refinement, allowing only atomic coordinates to vary, reduced the *R*-factor to 0.17. Fifteen additional cycles, during which isotropic temperature factors of the atoms were also allowed to vary, reduced the *R*-factor further to 0.14. In the refined model, the rms deviation from ideal bond lengths was 0.017 Å, compared with 0.027 Å in the starting model.

The refinement appears to behave well, despite the small number of data available in fiber diffraction. One indication of the quality of the refinement is the set of refined temperature factors. Several regions of the protein chain have higher than average temperature factors, and these correlate well with the mobile regions described by Westhof et al. (1984) for the TMV coat protein disk structure, which has a very similar backbone fold to that found in the virus.

Restrained least-squares refinement is now being used to extend the resolution of the TMV structure. This work is still in progress, but the *R*-factor against the x-ray diffraction data between 10 Å and 3 Å resolution is currently 0.126.

DISCUSSION

Fiber diffraction patterns measurable to 3 Å resolution or better can now be obtained from several helical viruses, and the use of magnetic orientation and other new methods for producing specimens is likely to provide high-resolution fiber diffraction patterns from other assemblies in the future. Further work is required to develop the methods of refinement discussed in this paper; in particular, the nature of the *R*-factor in fiber diffraction must be investigated in detail to understand fully the significance of the improvements obtained during the refinement. Nonetheless, our application of least-squares refinement to fiber diffraction from TMV has demonstrated the feasibility of using a combination of fiber diffraction data and stereochemical restraints for the refinement of the structure of a macromolecular assembly.

We would like to thank Dr. W. Hendrickson for programs and valuable discussions.

This work was supported by National Institutes of Health grants GM33265 and GM29829.

Received for publication 22 May 1985.

REFERENCES

Arnott, S., and A. J. Wonacott. 1966. The refinement of the crystal and molecular structures of polymers using x-ray data and stereochemical constraints. *Polymer.* 7:157–166.

Dodson, E. J., N. W. Isaacs, and J. S. Rollett. 1976. A method for fitting satisfactory models to sets of atomic positions in protein structure refinements. *Acta Crystallogr. (Sect. A).* 32:311–315.

Hendrickson, W. A,. and J. H. Konnert. 1980. Incorporation of stereochemical information into crystallographic refinement. *In* Computing in Crystallography. R. Diamond, S. Ramaseshan, and K. Venkatesan, editors. The Indian Academy of Science, Bangalore. 13.01–13.26.

Konnert, J. H., and W. A. Hendrickson. 1980. A restrained-parameter thermal-factor refinement procedure. *Acta Crystallogr. (Sect. A).* 36:344–350.

Makowski, L. 1982. The use of continuous diffraction a phase constraint. II. Application to fiber diffraction data. *J. Appl. Cryst.* 15:546–557.

Namba, K., and G. Stubbs. 1985. Solving the phase problem in fiber diffraction. Application to tobacco mosaic virus at 3.6 Å resolution. *Acta Crystallogr. (Sect. A)*. 41:252–262.

Smith, P. J. C., and S. Arnott. 1978. LALS: A linked-atom least-squares reciprocal-space refinement system incorporating stereochemical restraints to supplement sparse diffraction data. *Acta Crystallogr. (Sect. A)*. 34:3–11.

Stubbs, G., S. Warren, and K. Holmes. 1977. Structure of RNA and RNA binding site in tobacco mosaic virus from 4-Å map calculated from x-ray fibre diagrams. *Nature (Lond.)*. 267:216–221.

Westhof, E., D. Altschuh, D. Moras, A. C. Bloomer, A. Mondragon, A. Klug, and M. H. V. Van Regenmortel. 1984. Correlation between segmental mobility and the location of antigenic determinants in proteins. *Nature (Lond.)*. 311:123–126.

TOWARDS DIRECT PHASE RETRIEVAL IN MACROMOLECULAR CRYSTALLOGRAPHY

R. P. MILLANE,* R. H. T. BATES,‡ W. R. FRIGHT,‡ AND W. A. NORTON§

**Department of Biological Sciences, Purdue University, West Lafayette, Indiana 47907; ‡Electrical and Electronic Engineering Department, University of Canterbury, Christchurch, New Zealand; and §Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Silver Street, Cambridge, CB3 9EW, United Kingdom*

Classical direct methods for x-ray structure determination (1) have made solution of the phase problem for small organic molecules relatively straightforward in most cases. However, these techniques (at least in their present forms) are not useful for the solution of large structures for two reasons. First, they are based on an atomicity property of the electron density that does not apply to macromolecules when diffraction data are not available at atomic resolution. Second, for large molecules, direct methods are unwieldy and unreliable.

Although phase problems in one dimension are inherently nonunique (2), recent theoretical work has shown that in two or more dimensions, the phase of the Fourier transform of a localized, positive function is uniquely determined by the amplitude (2–4). Furthermore, phase retrieval algorithms have been developed which are capable of reconstructing images from the magnitudes of their Fourier transforms with no phase information (5–7). Both the uniqueness properties and the phase retrieval algorithms depend on the amplitude being available continuously (in effect) in reciprocal space. Although this is the case in most imaging applications, in crystallography the amplitude is measured only at the reciprocal lattice points, so these results cannot be applied directly to crystallographic phase retrieval. To distinguish the former case from the later, we refer to them as "optical" (although this is not restricted to optics) and "crystallographic" phase problems. We describe here implications of the optical results for crystallographic phase retrieval and show how they may be incorporated into existing macromolecular phase retrieval algorithms to improve the convergence properties. This is a deterministic approach that is distinct from techniques based on maximum entropy.

THEORY

For ease of exposition, our discussion is restricted to two dimensions (as is usual in optical applications) and we consider crystal structures in the plane group P1 with a rectangular unit cell. Extension of the analysis and algorithms to three dimensions and to arbitary space groups is strightforward. The measured diffraction amplitudes are equal to the amplitude of the continuous Fourier transform of the electron density in a single unit cell, sampled at the reciprocal lattice points. The continuous intensity is the Fourier transform of the autocorrelation (8) of a single unit cell, which is identical to the Patterson of the density in an isolated unit cell. The linear extent of the autocorrelation is twice that of the unit cell so that, as a result of the sampling theorem (8), the continuous intensity (or amplitude) can be constructed from its samples only if they are separated by no more than half the spacing of the reciprocal lattice points. Hence, to make use of the optical results, amplitude measurements must be available on a grid with a spacing no greater than this. We call the amplitudes at the reciprocal lattic points the "ordinary" structure amplitudes and others, which are on a grid with half the spacing, the "inbetween" structure amplitudes (Fig. 1). Knowledge of both the ordinary and inbetween structure amplitudes then is equivalent to knowing the continuous amplitude.

Using the continuous amplitude, optical *ab initio* phase retrieval involves two steps. First, approximate phases are generated from the ordinary and inbetween amplitudes using a noniterative procedure (3) that we call "crude phase estimation. " The second step involves iteratively improving these phases by iterating between real space and reciprocal space, forcing the image to conform to any a priori information in the former, and forcing the calculated amplitudes to be equal to those measured in the latter. The iterative procedure is often called a "Fienup algorithm" (5, 7) in optics and typical constraints on the image are positivity and known extent. Equivalent procedures (referred to as density modification) are used in crystallography, and typical constraints in real space are positivity, known molecular boundaries, and equivalence of identical subunits (9, 10). Application of the Fienup algorithm in

BIOPHYS. J. © Biophysical Society · 0006-3495/86/01/60/03 $1.00
Volume 49 January 1986 60–62