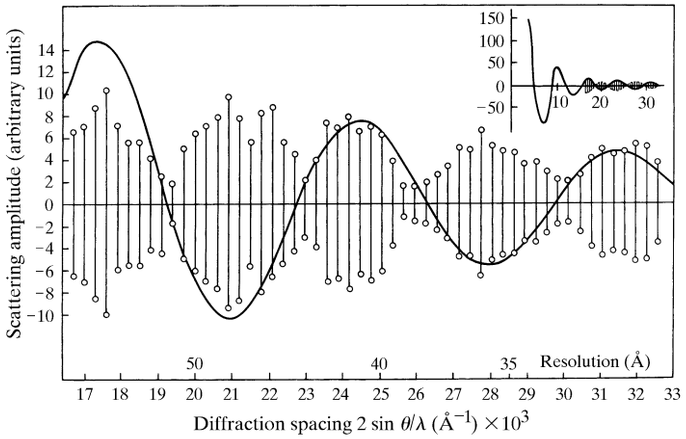

Fig. 13.4.12.1. Structure amplitudes of the type II crystals of southern bean mosaic virus, averaged within shells of reciprocal space, shown in relation to the Fourier transform of a 284 Å diameter sphere. The inset shows the complete spherical transform from infinity to 30 Å resolution. [Reproduced with permission from Johnson *et al.* (1976). Copyright (1976) Academic Press.]

resolution. These often show the anticipated distribution of a Fourier transform of a uniform sphere (Fig. 13.4.12.1). Thus, initiating phasing using a spherical model does require the prior determination of the average radius of the spherical virus. This can be done either by using an *R*-factor search (Tsao, Chapman & Rossmann, 1992) or by using low-angle X-ray scattering data (Chapman *et al.*, 1992). A minimal model would be to estimate the value of $F(000)$ on the same relative scale as the observed amplitudes. This structure factor must always have a positive value. Such a limited initial start was first explored by Rossmann & Blow (1963).

In surprisingly many cases (Valegård *et al.*, 1990; Chapman *et al.*, 1992; McKenna, Xia, Willingmann, Ilag, Krishnaswamy *et al.*, 1992; McKenna, Xia, Willingmann, Ilag & Rossmann, 1992; Tsao, Chapman & Rossmann, 1992; Tsao, Chapman, Wu *et al.*, 1992), it has been found that initiating phasing by using a very low resolution model results in a phase solution of the Babinet inverted structure ($\alpha \rightarrow \alpha + \pi$), where the desired density is negative instead of positive. Presumably, this is the result of phase convergence in a region where the assumed spherical transform is π out of step with reality. As long as this possibility is kept in mind with a watchful eye, such an inversion does not hamper good phase determination. In the case of phase extension, stepping too far in resolution can also lead to analogous problems (Arnold *et al.*, 1987).

Similar errors can occur due to lack of information on the correct enantiomorph in the initial phasing model. In some cases, where spherical envelopes are used and the distribution of NCS elements is also centric, there will be no decision on hand, and the phases will remain centric (Johnson *et al.*, 1975). However, in general, the enantiomorphic ambiguity (hand assignment) can be resolved by providing a model that has some asymmetry or by arbitrarily selecting the phase of a large-amplitude structure factor away from its centric value.

The progress of phase refinement away from false solutions has been the subject of 'post mortem' examinations (Valegård *et al.*, 1990; Chapman *et al.*, 1992; McKenna, Xia, Willingmann, Ilag, Krishnaswamy *et al.*, 1992; McKenna, Xia, Willingmann, Ilag & Rossmann, 1992; Tsao, Chapman & Rossmann, 1992; Tsao, Chapman, Wu *et al.*, 1992; Dokland *et al.*, 1998). The main lesson learned from these observations is that phase determination using NCS is amazingly powerful. Most initial errors in phasing gradually work themselves out with subsequent iterations and phase extension.

Perhaps the power of NCS phase determination should not be overly surprising. When phases are determined by multiple isomorphous replacement, the amount of data collected for the given molecular weight is $(N + 1)$, where N is the number of derivatives and is usually 3 or 4. Similarly, for multiwavelength anomalous-dispersion data collection, there might be measurements at four different wavelengths, essentially giving $N = 8$ data points for each reflection. However, icosahedral virus determination frequently provides $N = 60$ data points for the equivalent resolution.

13.4.13. Recent salient examples in low-symmetry cases: multidomain averaging and systematic applications of multiple-crystal-form averaging

When averaging molecules that have segmental flexibility, it is essential to be able to define the extents of and noncrystallographic relationships among multiple segments which can flexibly reorient. No general protocol has been described for determining the minimum size or optimal number of segments to use in such cases. If the number of segments used for averaging is too small, then the NCS parameters cannot accurately superpose the entirety of the related segments. If too many segments are used for averaging, the segments may become too small for accurate determination of the NCS parameters. The use of too many segments may also become awkward and somewhat inefficient, since in some program systems the total number of maps that must be stored in a given cycle of averaging is proportional to the number of segments used for averaging. Comparison of atomic models for related segments that have been built or refined independently may provide convenient definitions of envelopes for averaging. In practice, a radius of 2 Å or more (depending upon the stage of structure solution and completeness and expected reliability of the model) may be added around the atoms used to define a molecular mask or envelope used in averaging. As with other averaging procedures, multidomain and multiple-crystal-form averaging approaches generally benefit from updating the molecular masks as structure determination progresses.

Often, a macromolecule can be crystallized in multiple crystal forms. Advances in crystallization technology leading to the frequent occurrence of multiple crystal forms, coupled with the availability of convenient programs, have led to increasing frequency of application of multiple-crystal-form averaging for structure solution.

Proteins, especially those containing more than one folded domain, often contain flexible hinges. As long as the boundaries of and noncrystallographic relationships among the related domains in multiple copies can be determined, then density averaging can be used to improve phasing. Programs such as *O* can be conveniently used to obtain the initial transformations necessary for correct superposition of related segments. NCS parameters can be refined using routines that either minimize the density differences among related copies or that perform rigid-body refinements of atomic models.

A number of experimental techniques have been described that may permit more widespread application of multiple-domain and multiple-crystal-form averaging. Freezing of macromolecular crystals to liquid-nitrogen temperatures has become a routine approach for enhancing the resolution and quality of macromolecular X-ray diffraction data. With most macromolecular crystals, there is a shrinkage of the 'frozen' unit cell relative to the lattice of the 'unfrozen' crystals. In many cases, significantly different cell dimensions can also be obtained by using different cryo-protective

buffer and salt conditions. These variations can be exploited in a systematic fashion for phasing by electron-density averaging, so long as (1) the shrinkage relationships among the different crystals are not merely isotropic and (2) the boundaries and NCS parameters among related segments can be determined. Perutz (Perutz, 1946; Bragg & Perutz, 1952) recognized the potential utility of such shrinkage stages for crystallographic phasing in studies of haemoglobin crystals with varying degrees of hydration.

Recent examples of structure solutions involving multidomain and multiple-crystal-form averaging include studies of HIV reverse transcriptase (RT) (Ren *et al.*, 1995; Ding *et al.*, 1995). Studies of HIV RT by Stuart and coworkers involved multidomain and multiple-crystal-form averaging using different soaking solutions (Esnouf *et al.*, 1995; Ren *et al.*, 1995), in some cases with dramatically improved diffraction resolution. Arnold and co-workers have applied multidomain and multiple-crystal-form averaging to studies of HIV RT, including a systematic application of averaging electron density between 'frozen' and 'unfrozen' crystal forms (Ding *et al.*, 1995; Das *et al.*, 1996). Tong *et al.* (1997) recently described electron-density averaging among multiple closely related crystal forms of the human cytomegalovirus protease that were obtained by treatment of the crystals with different soaking buffers containing differing levels of precipitants, such as salt and polyethylene glycol.

13.4.14. Programs

This review hopefully covers most aspects encountered when employing electron-density averaging, yet the authors have drawn liberally from their own experience. There are now a large number of averaging programs and procedures available, some more suitable for structure determinations of proteins with low NCS redundancy and improper relationships (Jones, 1992) and others particularly suitable for high NCS redundancy, such as is encountered in the study of icosahedral viruses. For large structures, phase determination can be a very time-consuming computer operation. Therefore, attempts have been made to parallelize some programs (Cornea-Hasegan *et al.*, 1995), although this may lead to difficulties in exporting the programs to new and different computers.

Recently described program packages for symmetry averaging have been successfully applied to a number of cases. General program systems for averaging that are well suited to cases with high NCS include *ENVelope* (Rossmann *et al.*, 1992) and *GAP* (Jonathan Grimes and David Stuart, unpublished results); these same packages have also been used for multiple-crystal-form averaging and problems with low symmetry. A number of the program packages have been conveniently integrated with interactive computer-graphics programs such as *O* (Jones *et al.*, 1991) and most permit molecular-envelope definition by a number of possible approaches. *RAVE* and *MAVE* (Kleywegt & Jones, 1994), programs for graphics-assisted averaging within and between crystal forms, also come with an array of tools for flexible map handling and envelope definition (Kleywegt & Jones, 1996). The program systems *DMMULTI* (Cowtan & Main, 1993) and *MAGICSQUASH* (Schuller, 1996), which both derive from the program *SQUASH* (Zhang, 1993), can simultaneously apply real-space (symmetry averaging and solvent levelling with or without histogram matching) and reciprocal-space (phase refinement by the Sayre equation) constraints for phase improvement and extension. The advantage of adding phasing by the Sayre equation is greater at higher resolution, but appears to be significant in some cases, even at relatively low resolution (Cowtan & Main, 1993). *MAGICSQUASH* has been used to determine a number of structures which required multiple-domain and multiple-crystal-form averaging (Schuller, 1996). The *DEMON/ANGEL* package allows noncrystallographic averaging among multiple crystal forms together with solvent flattening and histogram matching (Vellieux *et al.*, 1995). Other versatile programs for electron-density averaging include *AVGSYS* (Bolin *et al.*, 1993) and *PHASES* (Furey & Swaminathan, 1990, 1997), both of which have features for facilitating definition and refinement of NCS parameters.

Acknowledgements

We are most grateful to Sharon Wilder and Cheryl Towell for extensive help in creating this manuscript. We are also grateful for decades of financial support by the National Science Foundation and the National Institutes of Health during the development of the techniques reported here.

References

13.1

Abrahams, J. P. & Leslie, A. G. W. (1996). *Methods used in the structure determination of bovine mitochondrial F_1 ATPase. Acta Cryst.* D**52**, 30–42.

Argos, P., Ford, G. C. & Rossmann, M. G. (1975). *An application of the molecular replacement technique in direct space to a known protein structure. Acta Cryst.* A**31**, 499–506.

Blow, D. M., Rossmann, M. G. & Jeffery, B. A. (1964). *The arrangement of α-chymotrypsin molecules in the monoclinic crystal form. J. Mol. Biol.* **8**, 65–78.

Brünger, A. T. (1997). *Patterson correlation searches and refinement. Methods Enzymol.* **276**, 558–580.

Crowther, R. A. (1967). *A linear analysis of the non-crystallographic symmetry problem. Acta Cryst.* **22**, 758–764.

Crowther, R. A. (1969). *The use of non-crystallographic symmetry for phase determination. Acta Cryst.* B**25**, 2571–2580.

Crowther, R. A. & Blow, D. M. (1967). *A method of positioning a known molecule in an unknown crystal structure. Acta Cryst.* **23**, 544–548.

Cruickshank, D. W. J. (1999). *Remarks about protein structure precision. Acta Cryst.* D**55**, 583–601; erratum D**55**, 1108.

Dodson, E., Gover, S. & Wolf, W. (1992). Editors. *Proceedings of the CCP4 study weekend. Molecular replacement.* Warrington: Daresbury Laboratory.

Fehlhammer, H. & Bode, W. (1975). *The refined crystal structure of bovine β-trypsin at 1.8 Å resolution. I. Crystallization, data collection and application of Patterson search techniques. J. Mol. Biol.* **98**, 683–692.

Hoppe, W. (1957). *Die 'Faltmolekülmethode' – eine neue Methode zur Bestimmung der Kristallstruktur bei ganz oder teilweise bekannter Molekülstruktur. Acta Cryst.* **10**, 750–751.

Hoppe, W. & Gassmann, J. (1968). *Phase correction, a new method to solve partially known structures. Acta Cryst.* B**24**, 97–107.

Huber, R. (1965). *Die automatisierte Faltmolekülmethode. Acta Cryst.* **19**, 353–356.

Huber, R., Kukla, D., Bode, W., Schwager, P., Bartels, K., Deisenhofer, J. & Steigemann, W. (1974). *Structure of the complex formed by bovine trypsin and bovine pancreatic trypsin inhibitor. II. Crystallographic refinement at 1.9 Å resolution. J. Mol. Biol.* **89**, 73–101.