

## The phase problem

**Garry Taylor**

Centre for Biomolecular Sciences, University of  
St Andrews, St Andrews, Fife KY16 9ST,  
Scotland

Correspondence e-mail: [glt2@st-andrews.ac.uk](mailto:glt2@st-andrews.ac.uk)

Given recent advances in phasing methods, those new to protein crystallography may be forgiven for asking 'what problem?'. As many of those attending the CCP4 meeting come from a biological background, struggling with expression and crystallization, this introductory paper aims to introduce some of the basics that will hopefully make the subsequent papers penetrable. What is the 'phase' in crystallography? What is 'the problem'? How can we overcome the problem? The paper will emphasize that the phase values can only be discovered through some prior knowledge of the structure. The paper will canter through direct methods, isomorphous replacement, anomalous scattering and molecular replacement. As phasing is the most acronymic realm of crystallography, MR, SIR, SIRAS, MIR, MIRAS, MAD and SAD will be expanded and explained in part. Along the way, we will meet some of the heroes of protein crystallography such as Perutz, Kendrew, Crick, Rossmann and Blow who established many of the phasing methods in the UK. It is inevitable that some basic mathematics is encountered, but this will be done as gently as possible.

Received 1 May 2003

Accepted 11 August 2003

### 1. Introduction

There are many excellent comprehensive texts on phasing methods (Blundell & Johnson, 1976; Drenth, 1994; Rossmann & Arnold, 2001; Blow, 2002) so this introduction to the CCP4 Study Weekend attempts to give an overview of phasing for those new to the field. Many entering protein crystallography are from a biological background, unfamiliar with the details of Fourier summation and complex numbers. The routine incorporation of selenomethionine into proteins and the wide availability of synchrotrons means that in many cases structure solution has become press-button. This is to be welcomed, but not all structure solutions are plain sailing and it is still useful to have some understanding of what phasing is. Here, we will emphasize the importance of phases, how phases are derived from some prior knowledge of structure and look briefly at phasing methods (direct, molecular replacement and heavy-atom isomorphous replacement). In most phasing methods the aim is to preserve isomorphism, such that the only structural change upon heavy-atom substitution is local and there are no changes in unit-cell parameters or orientation of the protein in the cell. Of course, single- and multi-wavelength anomalous diffraction (SAD/MAD) experiments achieve this. Where non-isomorphism does occur, then this can be used to provide phase information and we will look at an example where non-isomorphism was used to extend phases.

In the diffraction experiment (Fig. 1), we measure the intensities of waves scattered from planes (denoted by  $hkl$ ) in the crystal. The amplitude of the wave  $|F_{hkl}|$  is proportional to

the square root of the intensity measured on the detector. To calculate the electron density at a position  $(xyz)$  in the unit cell of a crystal requires us to perform the following summation over all the  $hkl$  planes, which in words we can express as: electron density at  $(xyz)$  = the sum of contributions to the point  $(xyz)$  of waves scattered from plane  $(hkl)$  whose amplitude depends on the number of electrons in the plane, added with the correct relative phase relationship or, mathematically,

$$\rho(xyz) = 1/V \sum |F_{hkl}| \exp(i\alpha_{hkl}) \exp(-2\pi ihx + ky + lz),$$

where  $V$  is the volume of the unit cell and  $\alpha_{hkl}$  is the phase associated with the structure-factor amplitude  $|F_{hkl}|$ . We can measure the amplitudes, but the phases are lost in the experiment. This is the phase problem.

## 1.1. The importance of phases

The importance of phases in producing the correct structure is illustrated in Figs. 2 and 3. In Fig. 2 three ‘electron-density waves’ are added in a unit cell, which shows the dramatically different electron density resulting from adding the third wave with a different phase angle. In Fig. 3, from Kevin Cowtan’s Book of Fourier (<http://www.ytbl.york.ac.uk/~cowtan/fourier/fourier.html>), the importance of phases in carrying structural information is beautifully illustrated. The calculation of an ‘electron-density map’ using amplitudes from the diffraction of a duck and phases from the diffraction of a cat results in a cat: a warning of model-bias problems in molecular replacement!

## 2. Recovering the phases

There is no formal relationship between the amplitudes and phases; the only relationship is *via* the molecular structure or electron density. Therefore, if we can assume some prior knowledge of the electron density or structure, this can lead to values for the phases. This is the basis for all phasing methods (Table 1).

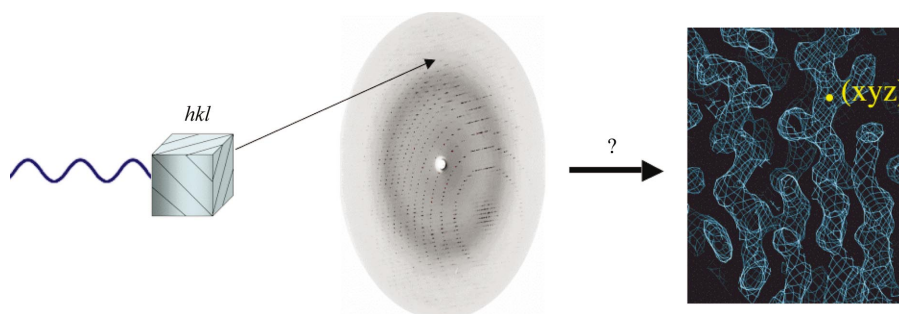
### 2.1. Direct methods

Direct methods are based on the positivity and atomicity of electron density that leads to phase relationships between the (normalized) structure factors, *e.g.*

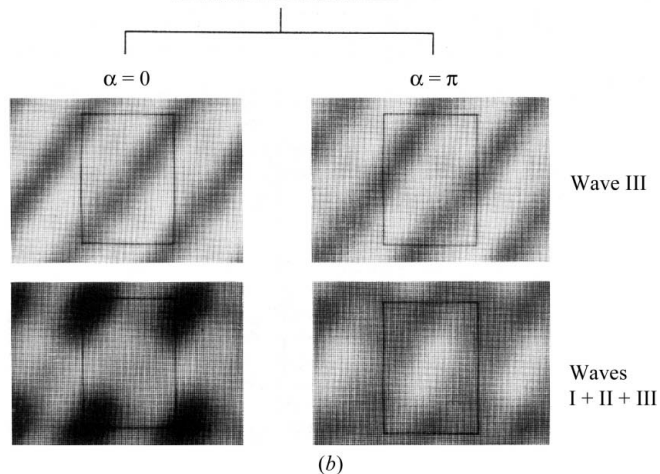
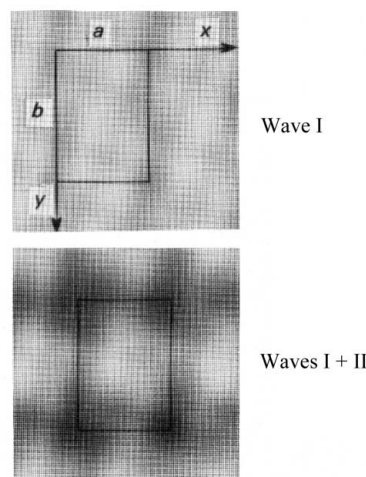
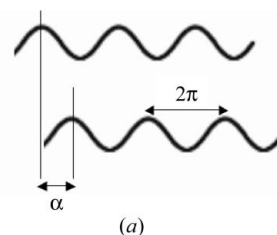
$$\alpha_{-h} + \alpha_h + \alpha_{h-h'} \simeq 0,$$

$$\tan \alpha_h = \frac{\langle E_{h'} E_{h-h'} \sin(\alpha_{h'} + \alpha_{h-h'}) \rangle_{h'}}{\langle E_{h'} E_{h-h'} \cos(\alpha_{h'} + \alpha_{h-h'}) \rangle_{h'}}$$

where  $E$  represents the normalized structure-factor amplitude; that is, the amplitude that would arise from point atoms at rest. Such equations imply that once the phases of some reflections are known, or can be given a variety of starting values, then

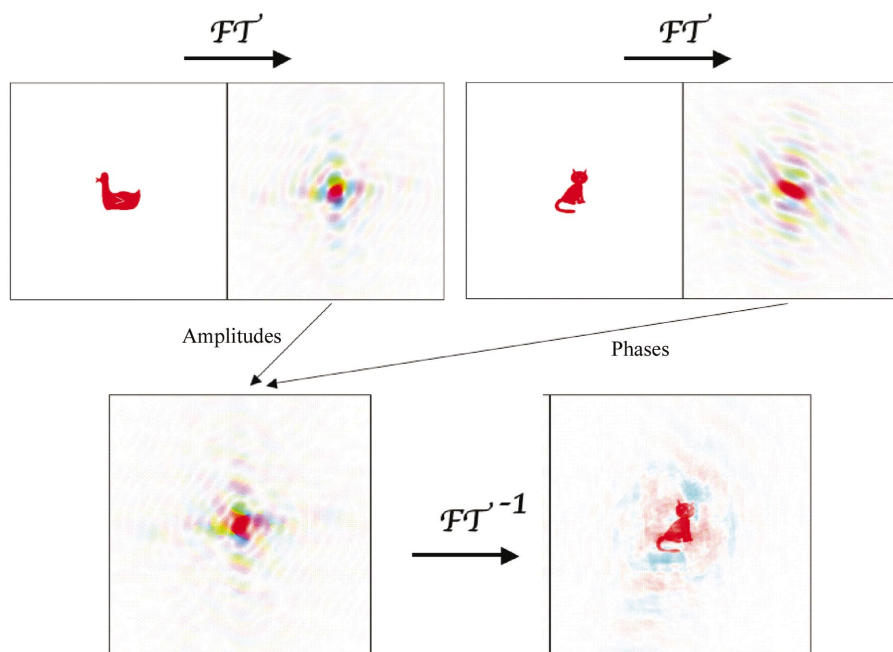


**Figure 1**  
The diffraction experiment.



**Figure 2**  
(a) The definition of a phase angle  $\alpha$ . (b) The result of adding three waves, where the third wave is added with two different phase angles.

the phases of other reflections can be deduced leading to a bootstrapping of phase values for all reflections. The

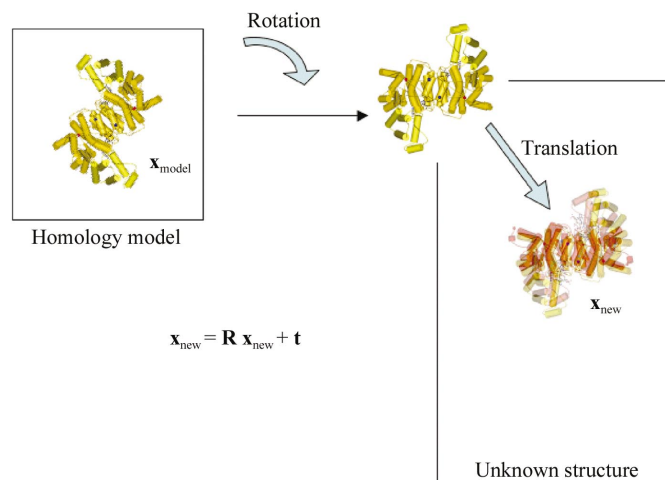

**Figure 3**

The importance of phases in carrying information. Top, the diffraction pattern, or Fourier transform (FT), of a duck and of a cat. Bottom left, a diffraction pattern derived by combining the amplitudes from the duck diffraction pattern with the phases from the cat diffraction pattern. Bottom right, the image that would give rise to this hybrid diffraction pattern. In the diffraction pattern, different colours show different phases and the brightness of the colour indicates the amplitude. Reproduced courtesy of Kevin Cowtan.

**Table 1**

Phasing methods.

Method	Prior knowledge
Direct methods	$\rho \geq 0$ , discrete atoms
Molecular replacement	Homology model
Isomorphous replacement	Heavy-atom substructure
Anomalous scattering	Anomalous atom substructure
Density modification (phase improvement)	Solvent flattening Histogram matching Non-crystallographic symmetry averaging Partial structure Phase extension


**Figure 4**

The process of molecular replacement.

requirement of, what is for proteins, very high resolution data ( $<1.2 \text{ \AA}$ ) has limited the usefulness of *ab initio* phase determination in protein crystallography, although direct methods have been used to phase proteins up to  $\sim 1000$  atoms. This so-called Sheldrick's rule (Sheldrick, 1990) has recently been given a structural basis with respect to proteins (Morris & Bricogne, 2003). However, direct methods are used routinely to find the heavy-atom substructure, such as in *Shake-and-Bake* (*SnB*; Miller *et al.*, 1994), *SHELXD* (Schneider & Sheldrick, 2002) and *SHARP* (de La Fortelle & Bricogne, 1997), and even subsequent phase determination from the substructure with programs such as *SHELXE* (Debreczeni *et al.*, 2003) and *ACORN* (Foadi *et al.*, 2000).

## 2.2. Molecular replacement (MR)

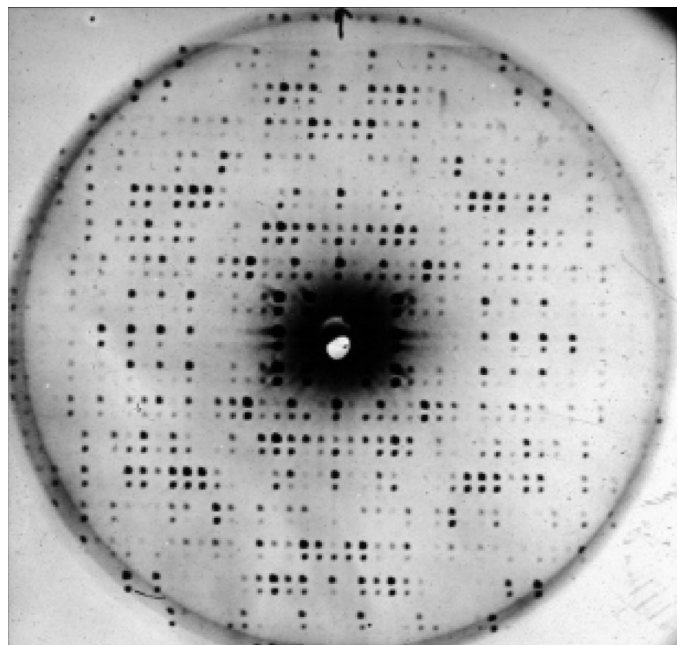
When a homology model is available, molecular replacement can be successful, using methods first described by Michael Rossmann and David Blow (Rossmann & Blow, 1962). As a rule of

thumb, a sequence identity  $>25\%$  is normally required and an r.m.s. deviation of  $<2.0 \text{ \AA}$  between the  $\alpha$  C atoms of the model and the final new structure, although there are exceptions to this. Patterson methods are usually used to obtain first the orientation of the model in the new unit cell and then the translation of the correctly oriented model relative to the origin of the new unit cell (Fig. 4).

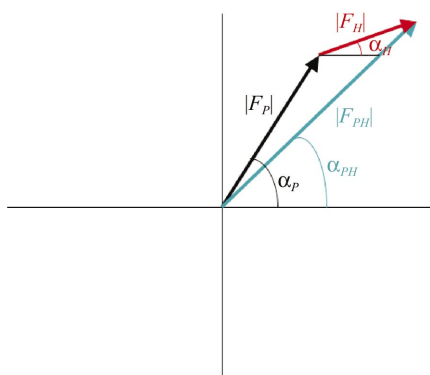
## 2.3. Isomorphous replacement

The use of heavy-atom substitution was invented very early on by small-molecule crystallographers to solve the phase problem; for example, the isomorphous crystals (same unit cells) of  $\text{CuSO}_4$  and  $\text{CuSeO}_4$  (Groth, 1908). The changes in intensities of some classes of reflections were used by Beevers & Lipson (1934) to locate the Cu and S atoms. It was Max Perutz and John Kendrew who first applied the methods to proteins (Perutz, 1956; Kendrew *et al.*, 1958) by soaking protein crystals in heavy-atom solutions to create isomorphous heavy-atom derivatives (same unit cell, same orientation of protein in cell) which gave rise to measurable intensity changes which could be used to deduce the positions of the heavy atoms (Fig. 5).

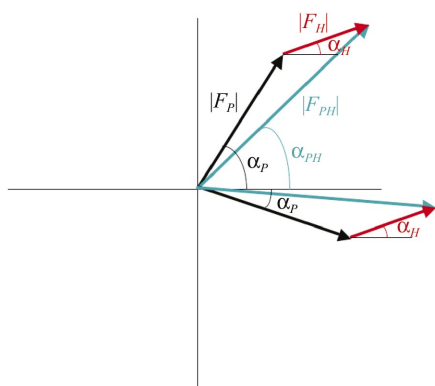
In the case of a single isomorphous replacement (SIR) experiment, the contribution of the heavy-atom replacement to the structure-factor amplitude and phases is best illustrated on an Argand diagram (Fig. 6). The amplitudes of a reflection are measured for the native crystal,  $|F_P|$ , and for the derivative crystal,  $|F_{PH}|$ . The isomorphous difference,  $|F_H| \simeq |F_{PH}| - |F_P|$ , can be used as an estimate of the heavy-atom structure-factor



**Figure 5**  
Two protein diffraction patterns superimposed and shifted vertically relative to one another. One is from the native bovine  $\beta$ -lactoglobulin, one from a crystal soaked in a mercury salt solution. Note the intensity changes for certain reflections and the identical unit cells suggesting isomorphism. (Photo courtesy of Dr Lindsay Sawyer).



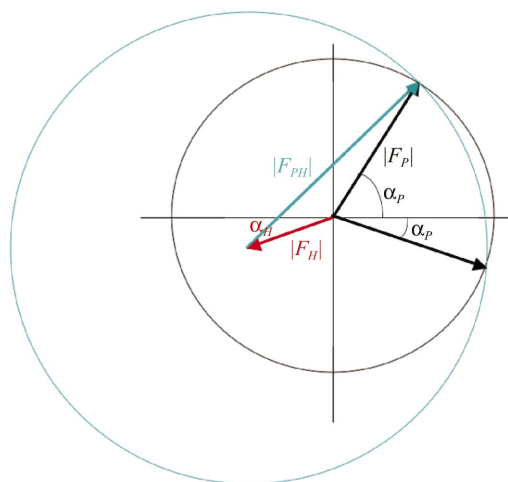
**Figure 6**  
Argand digram for SIR.  $|F_p|$  is the amplitude of a reflection for the native crystal and  $|F_{PH}|$  for the derivative crystal.



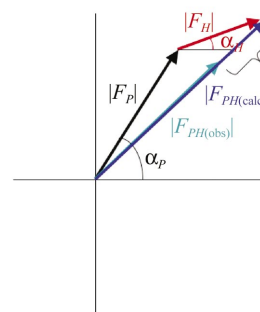
**Figure 7**  
Estimation of native protein phase for SIR.  $\alpha_p = \alpha_H + \cos^{-1} [(F_{PH}^2 - F_p^2 - F_H^2)/2F_p F_H]$ .

amplitude to determine the heavy-atom positions using Patterson or direct methods. Once located, the heavy-atom parameters ( $xyz$  positions, occupancies and Debye–Waller thermal factors  $B$ ) can be refined and used to calculate a more accurate  $|F_H|$  and its corresponding phase  $\alpha_H$ . The native protein phase,  $\alpha_p$ , can be estimated using the cosine rule (Fig. 7), leading to two possible solutions symmetrically distributed about the heavy-atom phase.

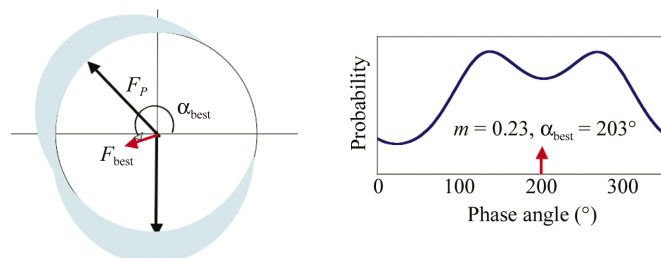
This phase ambiguity is better illustrated in the Harker construction (Fig. 8). The two possible phase values occur where the circles intersect. The problem then arises as to



**Figure 8**  
Harker construction for SIR.



**Figure 9**  
Phase probability (Blow & Crick, 1959). The lack of closure  $\epsilon = |F_{PH(obs)}| - |F_{PH(calc)}| = |F_{PH(obs)}| - [|F_p| \exp(i\alpha_p) + |F_H| \exp(i\alpha_H)]$ .



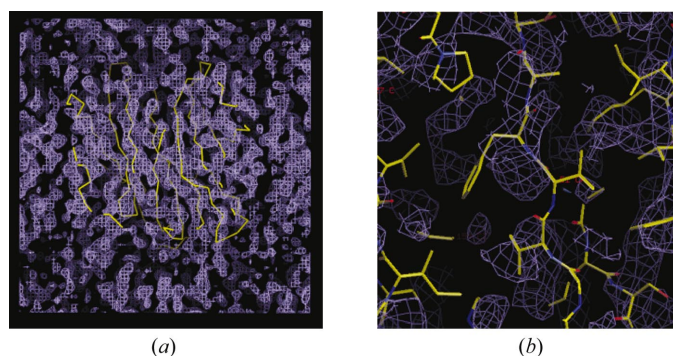
**Figure 10**  
Phase probability for one reflection in a SIR experiment.  $F_{best}$  is the centroid of the distribution. The map calculated with  $|F_{best}| \exp(i\alpha_{best})$  [or  $m|F_p| \exp(i\alpha_{best})$ , where  $m$  is the figure of merit,  $\langle \cos \Delta\alpha \rangle$  has least error.  $m = 0.23$  implies a 76° error.



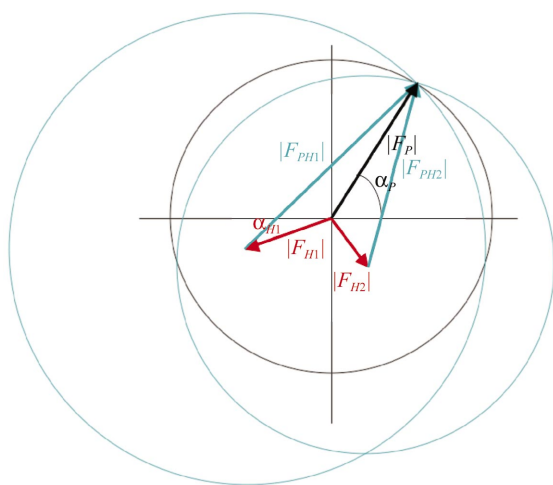
which phase to choose. This requires a consideration of phase probabilities.

### 3. Phase probability

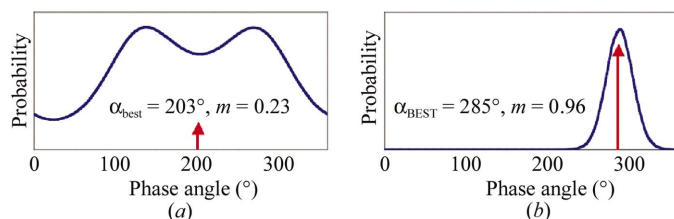
In reality, there are errors associated with the measurements of the structure factors and in the heavy-atom positions and their occupancies such that the vector triangle seldom closes. David Blow and Francis Crick introduced the concept of lack of closure ( $\epsilon$ ) and its use in defining a phase probability (Blow & Crick, 1959) (Fig. 9). Making the assumption that all the errors reside in  $F_{PH(\text{calc})}$  and that errors follow a Gaussian



**Figure 11**  
(a) A 2.6 Å SIR electron-density map with the final  $\alpha$ -carbon trace of the structure superimposed.  $\rho(\mathbf{x}) = (1/V)\sum m|F_p| \exp(i\alpha_{\text{best}}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x})$ .  
(b) A small section of the map with the final structure superimposed.



**Figure 12**  
Harker diagram for MIR with two heavy-atom derivatives.



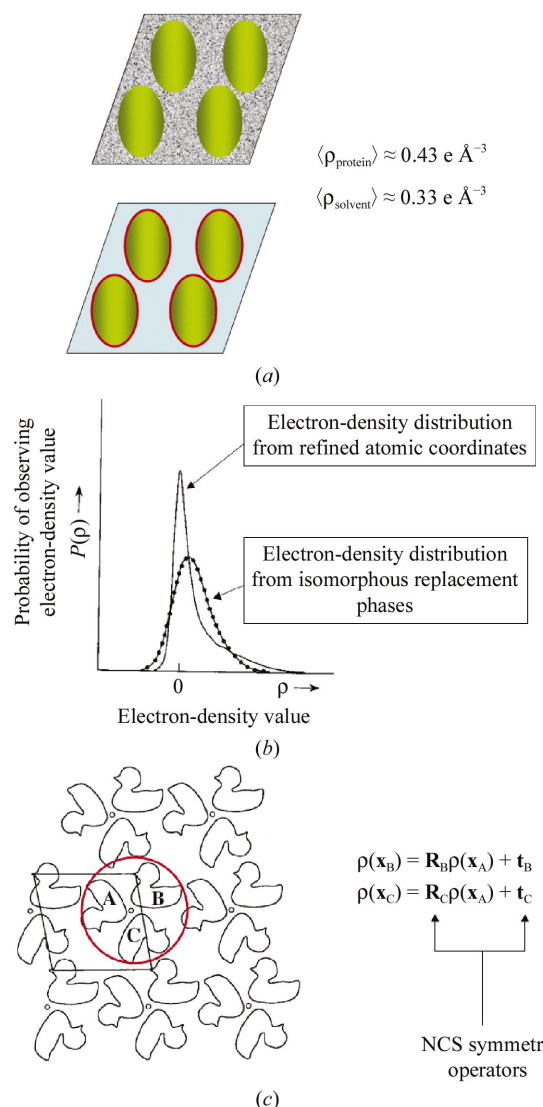
**Figure 13**  
Phase probability for one reflection in a MIR experiment. (a) One derivative. (b) Three derivatives.  $P(\alpha_p) \propto \prod_{i=1}^{\text{No. of derivatives}} \exp(-\epsilon_i^2/2E_i^2)$ .

distribution, the probability of a phase having a certain value is then

$$P(\alpha_p) \propto \exp(-\epsilon^2/2E^2),$$

where  $E = \langle [F_{PH(\text{obs})} - F_{PH(\text{calc})}]^2 \rangle$ .

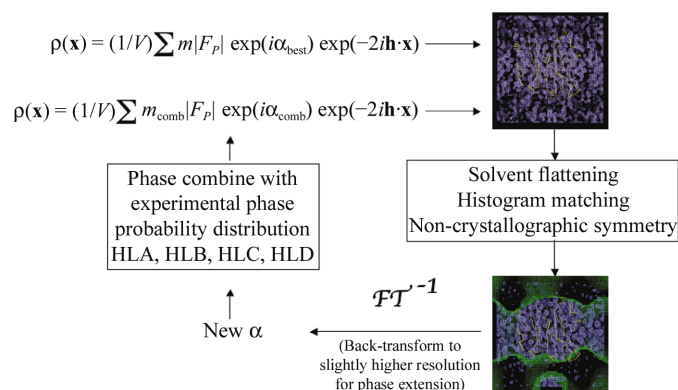
Most phasing programs calculate such a probability from 0 to 360° in 10° intervals, say, to produce a phase probability distribution whose shape can be represented by four coefficients of a polynomial, the so-called Hendrickson–Lattman coefficients HLA, HLB, HLC and HLD (Hendrickson & Lattman, 1970). Blow and Crick also showed that an electron-density map calculated with a weighted amplitude representing the centroid of the phase distribution gave the least



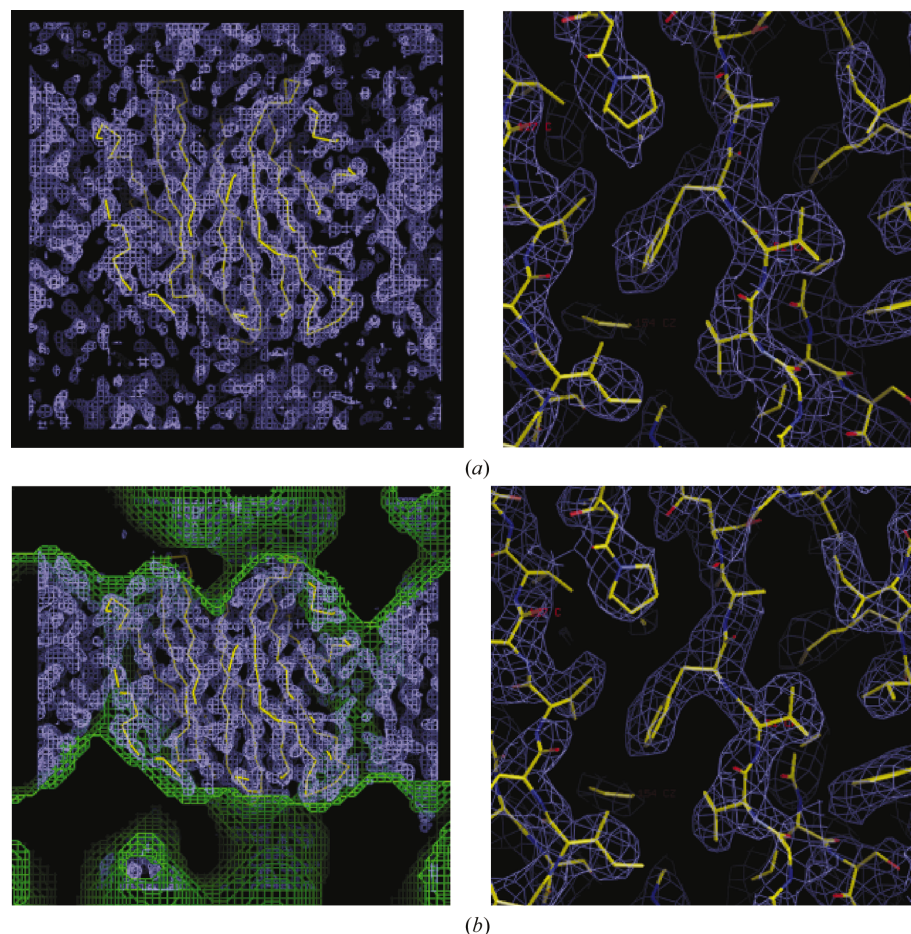
**Figure 14**  
Density-modification techniques. (a) Solvent flattening uses automated methods to define the protein–solvent boundary and then modify the solvent electron density to be of equal value. (b) Histogram matching redefines the values of electron-density points in a map to conform to an expected distribution of electron-density values. (c) Non-crystallographic symmetry averaging imposes identical electron-density values to points related by local symmetry, in this case a trimer of ducks that forms the asymmetric unit. The local NCS symmetry operators relating points in duck A to ducks B and C are shown.

error. Fig. 10 shows the phase probability distribution for one reflection from an SIR experiment. The centroid of the distribution is denoted by  $F_{\text{best}}$ , whose amplitude is the native amplitude  $|F_p|$  weighted by the figure of merit,  $m$ , which represents the cosine of the phase error. Modern phasing programs now use maximum-likelihood methods to derive phase probability distributions, as described in Read (2003).

Fig. 11 shows the electron density of part of the unit cell of the sialidase from *Salmonella typhimurium* (Crennell *et al.*,



**Figure 15**  
Phase improvement by density modification.



**Figure 16**  
(a) 2.6 Å MIR electron density. (b) Electron density after solvent flattening and histogram matching in *DM*. The solvent envelope determined by *DM* is shown in green.

1993) phased on a single mercury derivative. Although the protein–solvent boundary is partly evident, the electron density remains uninterpretable.

The use of more than one heavy-atom derivative in multiple isomorphous replacement (MIR) can break the phase ambiguity, as shown in Fig. 12. The phase probability is obtained by multiplying the individual phase probabilities, as shown in Fig. 13 for the same reflection as in Fig. 10, but this time three heavy-atom derivatives have resulted in a sharp unimodal distribution with a concomitantly high figure of merit.

## 4. Phase improvement

It is rare that experimentally determined phases are sufficiently accurate to give a completely interpretable electron-density map. Experimental phases are often only the starting point for phase improvement using a variety of methods of density modification, which are also based on some prior knowledge of structure. Solvent flattening, histogram matching and non-crystallographic averaging are the main techniques used to modify electron density and improve phases (Fig. 14). Solvent flattening is a powerful technique that removes negative electron density and sets the value of electron density in the solvent regions to a typical value of

$0.33 \text{ e} \text{ \AA}^{-3}$ , in contrast to a typical protein electron density of  $0.43 \text{ e} \text{ \AA}^{-3}$ . Automatic methods are used to define the protein–solvent boundary, first developed by Wang (1985) and then extended into reciprocal space by Leslie (1988). Histogram matching alters the values of electron-density points to concur with an expected distribution of electron-density values. Non-crystallographic symmetry averaging imposes equivalence on electron-density values when more than one copy of a molecule is present in the asymmetric unit. These methods are encoded into programs such as *DM* (Cowtan & Zhang, 1999), *RESOLVE* (Terwilliger, 2002) and *CNS* (Brünger *et al.*, 1998). Density-modification techniques will not turn a bad map into a good one, but they will certainly improve promising maps that show some interpretable features.

Density modification is often a cyclic procedure, involving back-transformation of the modified electron-density map to give modified phases, recombination of these phases with the experimental phases (so as not to throw away experimental reality) and calculation of a new map which is then modified and so the cycle continues until convergence. Such methods can also be used

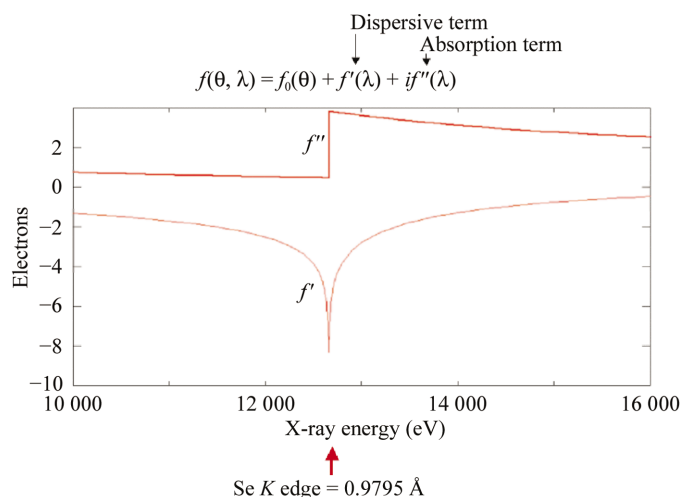
to provide phases beyond the resolution for which experimental phases information is available, assuming higher resolution native data have been collected. In such cases, the modified map is back-transformed to a slightly higher resolution on each cycle to provide new phases for higher resolution reflections. The process is illustrated in Fig. 15.

An example of the application of solvent flattening and histogram matching using *DM* is shown in Fig. 16 for the *S. typhimurium* sialidase phased using three derivatives.

#### 4.1. Anomalous scattering

The atomic scattering factor has three components: a normal scattering term that is dependent on the Bragg angle and two terms that are not dependent on scattering angle, but on wavelength. These latter two terms represent the anomalous scattering that occurs at the absorption edge when the X-ray photon energy is sufficient to promote an electron from an inner shell. The dispersive term reduces the normal scattering factor, whereas the absorption term is 90° advanced in phase. This leads to a breakdown in Friedel's law, giving rise to anomalous differences that can be used to locate the anomalous scatterers. Fig. 17 shows the variation in anomalous scattering at the *K* edge of selenium and Fig. 18 the breakdown of Friedel's law.

The anomalous or Bijvoet difference can be used in the same way as the isomorphous difference in Patterson or direct methods to locate the anomalous scatterers. Phases for the native structure factors can then be derived in a similar way to the SIR or MIR case. Anomalous scattering can be used to break the phase ambiguity in a single isomorphous replacement experiment, leading to SIRAS (single isomorphous replacement with anomalous scattering). Note that because of the 90° phase advance of the  $f'$  term, anomalous scattering provides orthogonal phase information to the isomorphous term. In Fig. 19, there are two possible phase values symmetrically located about  $f'$  and two possible phase values symmetrically located about  $F_H$ . For completeness, the use of

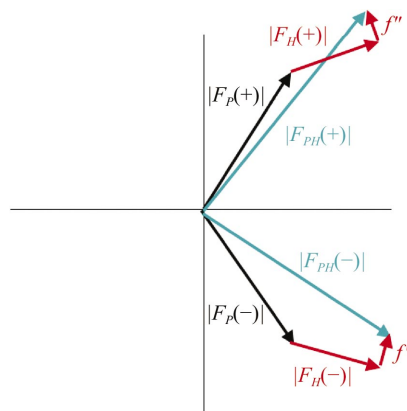


**Figure 17**  
Variation in anomalous scattering at the *K* edge of selenium.

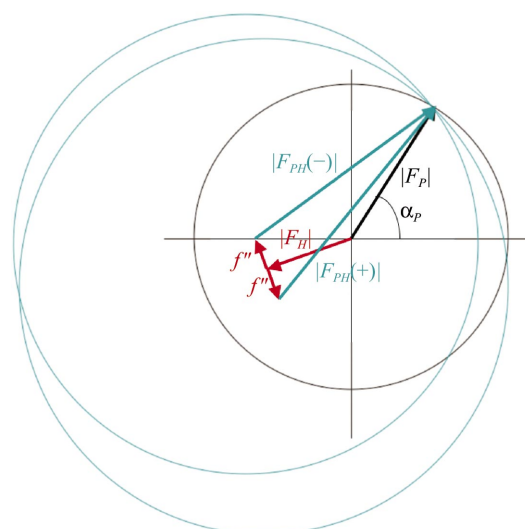
multiple isomorphous heavy-atom replacement using anomalous scattering is termed MIRAS.

#### 4.2. MAD

Isomorphous replacement has several problems: non-isomorphism between crystals (unit-cell changes, reorientation of the protein, conformational changes, changes in salt and solvent ions), problems in locating all the heavy atoms, problems in refining heavy-atom positions, occupancies and thermal parameters and errors in intensity measurements. The use of the multiwavelength anomalous diffraction (MAD) method overcomes the non-isomorphism problems. Data are collected at several wavelengths, typically three, in order to maximize the absorption and dispersive effects. Typically, wavelengths are chosen at the absorption,  $f''$ , peak ( $\lambda_1$ ), at the point of inflection on the absorption curve ( $\lambda_2$ ), where the dispersive term (which is the derivative of the  $f''$  curve) has its



**Figure 18**  
Breakdown of Friedel's law when an anomalous scatterer is present.  $f(\theta, \lambda) = f_0(\theta) + f'(\lambda) + if''(\lambda)$ .  $F_{hkl} \neq F_{\bar{h}\bar{k}\bar{l}}$ .  $\Delta F^\pm = |F_{PH}(+)| - |F_{PH}(-)|$  is the Bijvoet difference.

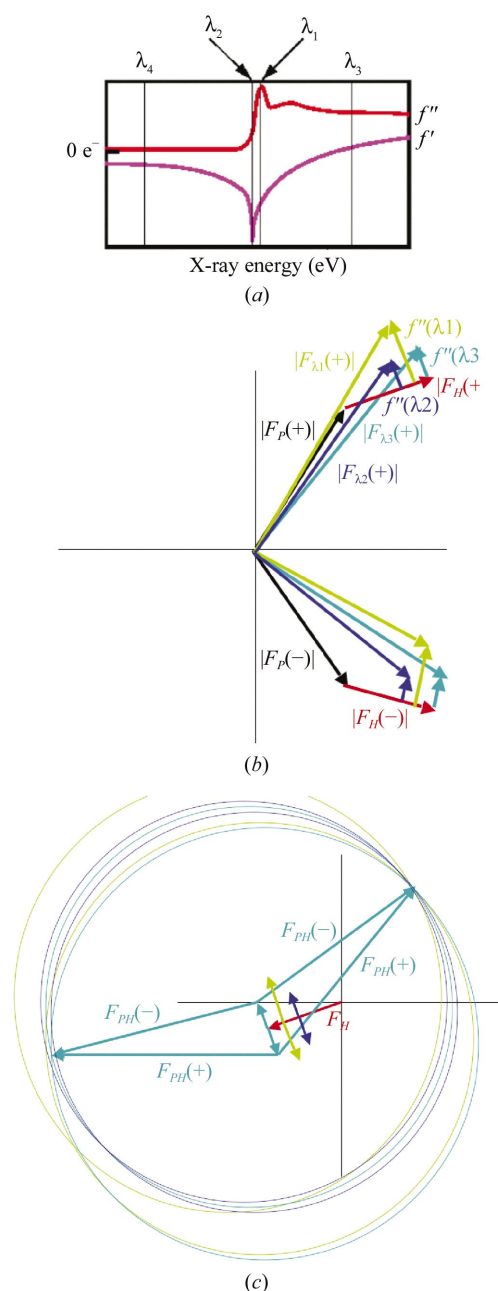


**Figure 19**  
Harker construction for SIRAS.



minimum, and at a remote wavelength ( $\lambda_3$  and/or  $\lambda_4$ ). Fig. 20 shows a typical absorption curve for an anomalous scatterer, together with the phase and Harker diagrams.

The changes in structure-factor amplitudes arising from anomalous scattering are generally small and require accurate measurement of intensities. The actual shape of the absorption curve must be determined experimentally by a fluorescence scan on the crystal at the synchrotron, as the environment of the anomalous scatterers can affect the details of the absorption. There is a need for excellent optics for accurate wavelength setting with minimum wavelength dispersion.



**Figure 20**  
MAD phasing. (a) Typical absorption curve for an anomalous scatterer. (b) Phase diagram.  $|F_p|$  is not measured, so one of the  $\lambda$ s is chosen as the 'native'. (c) Harker construction.

Generally, all data are collected from a single frozen crystal with high redundancy in order to increase the statistical significance of the measurements and data are collected with as high a completeness as possible. The signal size can be estimated using equations similar to those derived by Crick and Magdoff for isomorphous changes (Fig. 21), which also shows a predicted signal for the case of two Se atoms in 200 amino acids, calculated using Ethan Merritt's web-based calculator ([http://www.bmsc.washington.edu/scatter/AS\\_index.html](http://www.bmsc.washington.edu/scatter/AS_index.html)). Note that the signal increases with resolution owing to the fall-off of normal scattering with resolution.

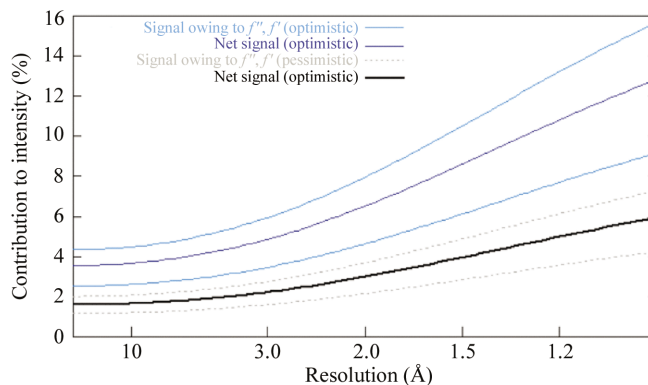
An example of MAD phasing is shown in Fig. 22. In this example of an archaeal chromatin modelling protein, Alba (Wardleworth *et al.*, 2002), the protein was expressed in a Met<sup>-</sup> strain of *Escherichia coli* and the single methionine was replaced with selenomethionine. Data were collected at three wavelengths around the Se K edge with a 12-fold redundancy to 3.0 Å on the ESRF beamline ID14-4. There were two monomers of 10 kDa in the asymmetric unit and SOLVE was used to determine the Se-atom positions and derive phases. RESOLVE was used to apply density modification to improve the phases.

#### 4.3. SAD

It is becoming increasingly possible to collect data at just a single wavelength, typically at the absorption peak, and use density-modification protocols to break the phase ambiguity and provide interpretable maps (Fig. 23). This so-called SAD (single-wavelength anomalous diffraction) method is described in Dodson (2003).

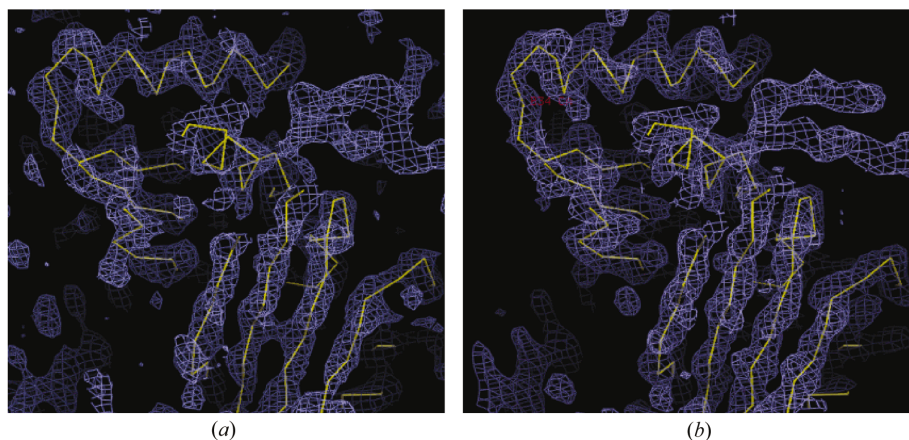
#### 5. Cross-crystal averaging

Protein crystallography is not a black-box technique for every protein; there are still challenges to be had in cases where MAD or SAD techniques cannot be used to derive a high-resolution map. On occasion, two or more crystal forms of a



**Figure 21**  
Estimation of signal size. The expected Bijvoet diffraction ratio  $r.m.s.(\Delta F^{\pm})/r.m.s.(|F|) \simeq (N_A/2N_T)^{1/2}(2f_A''/Z_{eff})$ . The expected dispersive ratio  $r.m.s.(\Delta F_{\Delta\lambda})/r.m.s.(|F|) \simeq (N_A/2N_T)^{1/2}[f_A''(\lambda_i) - f_A''(\lambda_j)]/Z_{eff}$ .  $N_A$  is the number of anomalous scatterers,  $N_T$  the total number of atoms in the structure and  $Z_{eff}$  is the normal scattering power for all atoms ( $6.7 e^-$  at  $2\theta = 0$ ).

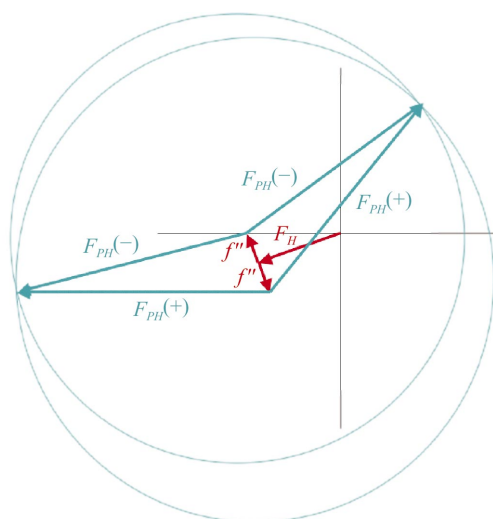




**Figure 22**  
(a) 3.0 Å electron-density map of Alba after *SOLVE*. (b) After *RESOLVE*, with the final  $\alpha$ -carbon trace of the structure superimposed.

protein are available: low-resolution phases may be known for one crystal form, but high-resolution data for another crystal form may be available. Cross-crystal averaging involves mapping the electron density from the one unit cell into the other; phases can then be derived for the new crystal form and through averaging of density between crystal forms and possibly phase extension as part of a density-modification procedure, one can bootstrap the phases to high resolution. The procedure is outlined in Fig. 24.

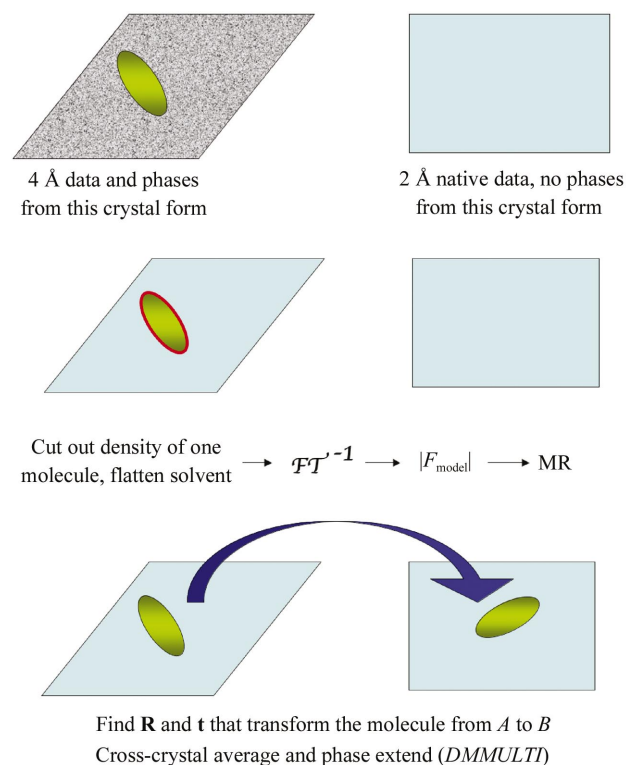
One example of the power of cross-crystal averaging is that of Newcastle disease virus haemagglutinin-neuraminidase (HN), whose structure solution was plagued with non-isomorphism problems (Crennell *et al.*, 2000). Native crystals from the same crystallization drop could have significantly different unit-cell parameters. The protein was derived from virus grown in embryonated chickens' eggs, so SeMet methods were out of the question. Most heavy-atom derivatives were non-isomorphous with native crystals and with one another. A platinum derivative was found that gave a clear peak in an



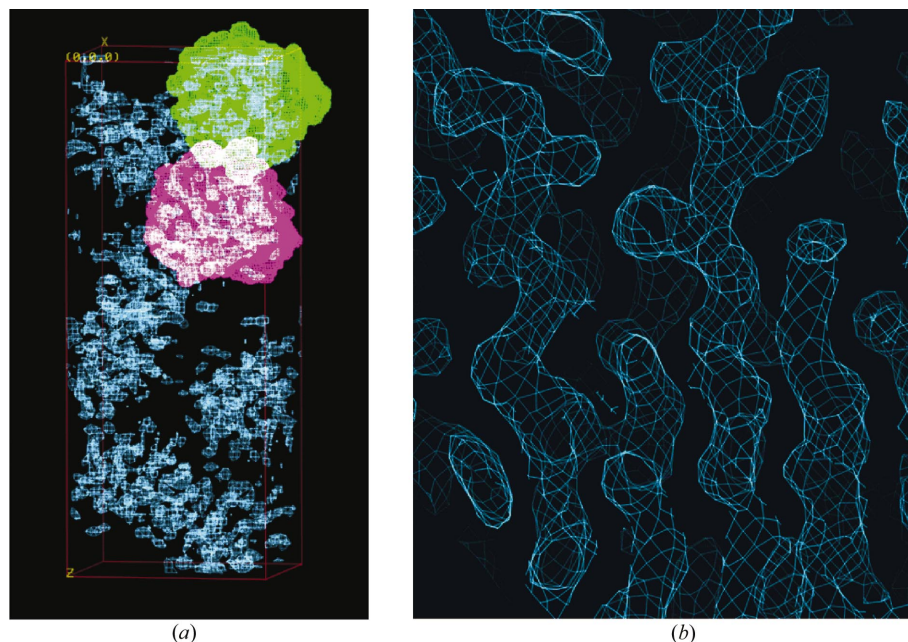
**Figure 23**  
Harker construction for SAD.  $\Delta F^{\pm}$  is used to find the substructure of anomalous scatterers, followed by phasing and phase improvement.

anomalous Patterson, which resulted in an attempt at MAD phasing, but the signal was just too small with one possibly not fully occupied Pt atom in 100 kDa. The  $P2_12_12_1$  unit cell had dimensions that varied as follows:  $a = 70.7\text{--}74.5$ ,  $b = 71.8\text{--}87.0$ ,  $c = 194.6\text{--}205.4$  Å. In the end, cross-crystal averaging was used to bootstrap from a poor uninterpretable 6.0 Å MIR map out to a clearly interpretable 2.0 Å map (Fig. 25). Four data sets were chosen for cross-crystal averaging in *DMMULTI*, chosen on the following criteria: (i) they were as non-isomorphous as possible to one another and (ii) they were to as high a resolution as possible. These were a pH 7.0 room-temperature data

set to 2.8 Å ( $a = 73.3$ ,  $b = 78.0$ ,  $c = 202.6$  Å), for which MIR phases were available to 6.0 Å, a pH 6 room-temperature data set to 3.0 Å ( $a = 72.0$ ,  $b = 83.9$ ,  $c = 201.6$  Å), a pH 4.6 frozen data set to 2.5 Å ( $a = 71.7$ ,  $b = 77.9$ ,  $c = 198.2$  Å) and a pH 4.6 frozen data set to 2.0 Å ( $a = 72.3$ ,  $b = 78.1$ ,  $c = 199.4$  Å). The power of the methods lies in the fact that the different unit cells are sampling the molecular transform in different places. Like most things, the idea is not new, and was indeed used by Bragg and Perutz in the early days of haemoglobin (Bragg &



**Figure 24**  
Cross-crystal averaging. Two crystal forms of the same protein for which phase information to low resolution is known for one form (left) and high-resolution data but no phase information is known for another form (right).



**Figure 25**

Cross-crystal averaging in haemagglutinin-neuraminidase (HN). (a) The unit cell showing the 6.0 Å MIR map derived from eight heavy-atom derivatives contoured at  $2.0\sigma$ , revealing two blobs corresponding to the two molecules in the asymmetric unit. (b) A section of the 2.0 Å map after phase extension and cross-crystal averaging over four non-isomorphous data sets.

Perutz, 1952), where they altered the unit cell of the crystals by controlled dehydration in order to sample the one-dimensional transform of the molecules in the unit cell. This paper is worth a read, if only for the wonderful inclusion of random test data in the form of train times between London and Cambridge!

## 6. Conclusion

The phase problem is fundamental and will never go away; however, its solution is now fairly routine thanks to SAD and MAD. The major problems in protein crystallography are now in the molecular biology, protein expression and crystallization, but perhaps most of all in interpreting the biological implications of structure which, after all, is where the fun starts.

I have been privileged to have received any understanding of phasing I possess from some excellent teachers. In particular, I would like to thank Stephen Neidle, Tom Blundell and Ian Tickle. I would like to thank Ethan Merritt for allowing me to reproduce graphs from his web site in Figs. 17, 20 and 21.

## References

- Beevers, C. A. & Lipson, H. (1934). *Proc. R. Soc. London Ser. A*, **146**, 570–582.
- Blow, D. M. (2002). *Protein Crystallography for Biologists*. Oxford University Press.
- Blow, D. M. & Crick, F. H. C. (1959). *Acta Cryst.* **12**, 794–802.
- Blundell, T. L. & Johnson, L. N. (1976). *Protein Crystallography*. London: Academic Press.
- Bragg, L. & Perutz, M. F. (1952). *Proc. R. Soc. London Ser. A*, **213**, 425–435.
- Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Cryst.* **D54**, 905–921.
- Cowtan, K. D. & Zhang, K. Y. (1999). *Prog. Biophys. Mol. Biol.* **72**, 245–270.
- Crennell, S. J., Garman, E. F., Laver, W. G., Vimr, E. R. & Taylor, G. L. (1993). *Proc. Natl Acad. Sci. USA*, **90**, 9852–9856.
- Crennell, S., Takimoto, T., Portner, A. & Taylor, G. (2000). *Nature Struct. Biol.* **7**, 1068–1074.
- Debreczeni, J. E., Bunkoczi, G., Ma, Q., Blaser, H. & Sheldrick, G. M. (2003). *Acta Cryst.* **D59**, 688–696.
- Dodson, E. (2003). *Acta Cryst.* **D59**, 1958–1966.
- Drenth, J. (1994). *Principles of Protein X-ray Crystallography*. Berlin: Springer-Verlag.
- Foadi, J., Woolfson, M. M., Dodson, E. J., Wilson, K. S., Jia-xing, Y. & Chao-de, Z. (2000). *Acta Cryst.* **D56**, 1137–1147.
- Groth, P. (1908). *Chemische Kristallographie*. Leipzig: Engelmann.
- Hendrickson, W. A. & Lattman, E. E. (1970). *Acta Cryst.* **B26**, 136–143.
- Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. & Phillips, D. C. (1958). *Nature (London)*, **181**, 662–666.
- La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
- Leslie, A. G. W. (1988). In *Proceedings of the CCP4 Daresbury Study Weekend*. Warrington: Daresbury Laboratory.
- Miller, R., Gallo, S. M., Khalak, H. G. & Weeks, C. M. (1994). *J. Appl. Cryst.* **27**, 613–621.
- Morris, R. J. & Bricogne, G. (2003). *Acta Cryst.* **D59**, 615–617.
- Perutz, M. F. (1956). *Acta Cryst.* **9**, 867–873.
- Read, R. J. (2003). *Acta Cryst.* **D59**, 1891–1902.
- Rossmann, M. G. & Arnold, E. (2001). Editors. *International Tables for Crystallography*, Vol. F. Dordrecht: Kluwer Academic Publishers.
- Rossmann, M. G. & Blow, D. M. (1962). *Acta Cryst.* **15**, 24–31.
- Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* **D58**, 1772–1779.
- Sheldrick, G. M. (1990). *Acta Cryst.* **A46**, 467–473.
- Terwilliger, T. C. (2002). *Acta Cryst.* **D58**, 1937–1940.
- Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
- Wardleworth, B. N., Russell, R. J., Bell, S. D., Taylor, G. L. & White, M. F. (2002). *EMBO J.* **21**, 4654–4662.