Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Low-resolution phase information in multiple-wavelength anomalous solvent contrast variation experiments

W. Shepard,[a]* R. Kahn,[b] M. Ramin[c] and R. Fourme[a]

[a]LURE, Bâtiment 209D, BP 34, Centre Université Paris-Sud, 91898 Orsay, France, [b]Institut de Biologie Structurale, 41 Rue Jules Horowitz, 38027 Grenoble, France, and [c]Université d'Evry-Val-d'Essonne, Boulevard François Mitterrand, 91025 Evry CEDEX, France

Correspondence e-mail: shepard@lure.u-psud.fr

Received 3 February 2000
Accepted 16 July 2000

The basic theory and principles of the multiple-wavelength anomalous solvent-contrast (MASC) method are introduced as a contrast-variation technique for generating low-resolution crystallographic phase information on the envelope of a macromolecule. Experimental techniques and practical considerations concerning the choice of anomalous scatterer, sample preparation and data acquisition are discussed. Test cases of crystals of three proteins of differing molecular weights from 14 kDa through to 173 kDa are illustrated. Methods for extracting the moduli of the anomalous structure factors from the MASC data are briefly discussed and the experimental results are compared with the known macromolecular envelopes. In all cases, the lowest resolution shells exhibit very large anomalous signals which diminish at higher resolution, as expected by theory. However, in each case the anomalous signal persists at high resolution, which is strong evidence for ordered sites of the anomalous scatterers. For the smaller two of these proteins the heavy-atom parameters could be refined for some of these sites. Finally, a novel method for phasing the envelope structure-factor moduli is presented. This method takes into account the relatively low number of observations at low resolution and describes the macromolecular envelope with a small number of parameters by presuming that the envelope is a compact domain of known volume. The parameterized envelope is expressed as a linear combination of independent functions such as spherical harmonics. Phasing starts from solutions of a sphere in the unit cell after positional refinement from random trials and the parameters describing the envelope are then refined against the data of structure-factor moduli. The preliminary results using simulated data show that the method can be used to reconstruct the correct macromolecular envelope and is able to discriminate against some false solutions.

1. Introduction

Contrast variation is a means of gleaning information from a scattering experiment on the form of a dispersed particle (a solute) by altering the level of its surrounding solvent. In experimental terms, this requires the measurement of scattering intensities of the particle in several different solvent compositions. The level of the solvent may be electron density which scatters X-rays, H/D isotope mixtures which scatter neutrons or any other physical density which interacts with the incident photons in a scattering experiment (for a review, see Williams *et al.*, 1994). The difference between the solute and the solvent density levels is defined as the contrast (Stuhrmann & Kirste, 1965; Ibel & Stuhrmann, 1975).

© 2000 International Union of Crystallography
Printed in Denmark – all rights reserved

This method is experiencing a renaissance in small-angle scattering studies on biological samples (*e.g.* Junemann *et al.*, 1998; Stuhrmann & Nierhaus, 1996; Zhao *et al.*, 1999). However, since 30–70% of the volume of macromolecular crystals constitutes solvent molecules, contrast-variation techniques can also be extended to macromolecular crystallography. Such was the case when Bragg & Perutz (1952) applied these methods to a haemoglobin crystal by altering the electronic density of the mother liquor and then observing changes in the intensities of X-ray reflections at low resolution. The changes derived from a contrast-variation series provide information on the 'solvent'-accessible region of the crystal, a region which is closely related to the 'negative' image of the macromolecular envelope. A number of contrast-variation experiments on macromolecular crystals have been demonstrated using either X-rays or neutrons to extract the low-resolution structures (Harrison, 1969; Jack *et al.*, 1975; Moras *et al.*, 1983; Roth *et al.*, 1984; Bentley *et al.*, 1984; Podjarny *et al.*, 1987 and references cited therein; Carter *et al.*, 1990; Badger, 1996) or to locate the lipid phase in membrane protein crystals (Roth *et al.*, 1991; Timmins *et al.*, 1992; Pebay-Peyroula *et al.*, 1995; Timmins & Pebay-Peyroula, 1996; Penel *et al.*, 1998; Pignol *et al.*, 1998). These experiments highlight certain disadvantages in the contrast-variation methods used, notably in the X-ray diffraction experiments, where there is a risk of lack of isomorphism when the unit-cell parameters change after soaking the crystals in different mother liquors, and in the H/D isotope-exchange experiments, which suffer from the low flux of neutron sources. Although anomalous scattering effects have been employed in small-angle scattering experiments at the Fe *K*-edge (Stuhrmann, 1980) and at the P *K*-edge (Hütsch, 1993), the possibility of using anomalous scattering from the solvent in crystals has only been considered as a supplement to a standard contrast-variation series at a single wavelength by Wyckoff and others (Dumas, 1988; Crumley, 1989; Carter *et al.*, 1990).

Anomalous scattering effects used at multiple wavelengths, however, can offer an excellent way of generating contrast variation by tuning the X-ray wavelength about an absorption edge of the anomalous scatterer, which is dispersed in the solvent channels of the macromolecular crystal. This method avoids the difficulties associated with the lack of isomorphism often encountered in X-ray chemical contrast-variation experiments because the changes in X-ray wavelength are physical changes applied *externally* to the sample. Consequently, isomorphism is conserved. The possibility of exploiting anomalous scattering at multiple wavelengths for contrast-variation experiments was originally suggested by Bricogne (1993), but its full potential was realised and demonstrated in the theoretical formalism and preliminary results presented by Fourme *et al.* (1995).

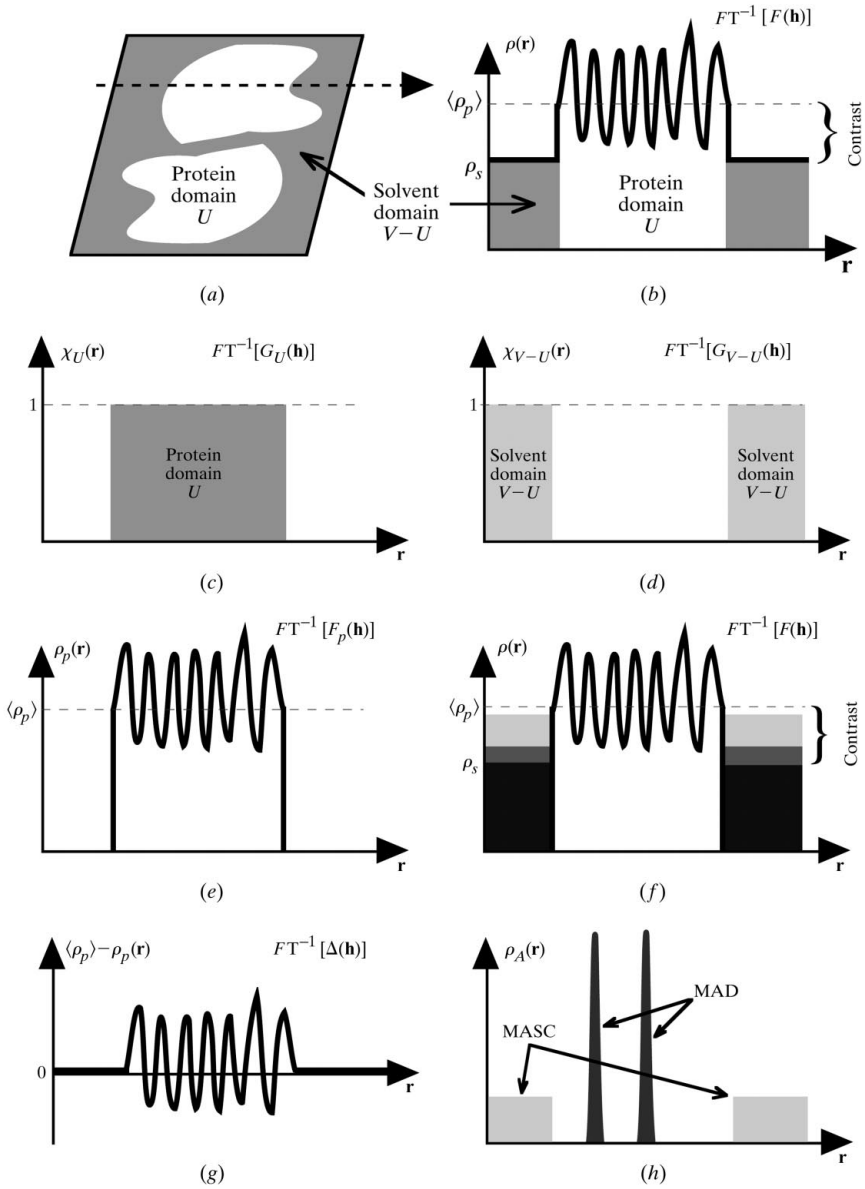


Figure 1
(*a*) A two-dimensional representation of the unit cell of volume V of a macromolecular crystal illustrating the ordered domain U which contains the macromolecule and the disordered domain $V - U$ which contains the solvent. One-dimensional slices portray the different components in contrast-variation theory. (*b*) The variation of the electron density, $\rho(\mathbf{r})$, in the macromolecular and solvent domains. (*c*) Indicator function of the ordered macromolecular domain U. (*d*) Indicator function of the disordered solvent domain $V - U$. (*e*) The electronic density of only the ordered domain U. This corresponds to the macromolecule in a vacuum. (*f*) The electronic density for both the macromolecule and solvent regions. Three different electronic densities of the solvent are represented by the three shades of grey. The contrast is shown for one of these. (*g*) The internal electronic density fluctuations inside the macromolecule. (*h*) The anomalous electronic density for both the MAD and MASC cases.

Here, we will present the principles, experimental aspects and recent advances that have been made in what is known as the multiple-wavelength anomalous solvent contrast method or MASC.

2. Theoretical principles

The theoretical principles of MASC (Fourme *et al.*, 1995) start at the basic principles of contrast variation, where the macromolecular crystal unit cell (volume V) is assumed to be separated into two phases: the domain U occupied by the macromolecule and the domain $V - U$ occupied by the solvent, which is in a liquid-like state of rapid exchange (Fig. 1*a*). The domain U containing the macromolecule is presumed to be ordered and this also includes the shell of ordered solvent molecules, whereas the domain $V - U$ is presumed to be the solvent region, which is completely disordered (Fig. 1*b*). It will be seen later that this model is oversimplified because the solvent region overlaps with the ordered zone.

A formalism derived by Bricogne (unpublished work) and demonstrated in Carter *et al.* (1990) separates the diffraction effects of the ordered and disordered domains by defining $G_U(\mathbf{h})$ as the Fourier transform of the indicator function $\chi_U(\mathbf{h})$, which is 1 inside the domain U and 0 elsewhere (Fig. 1*c*). Likewise, $G_{V-U}(\mathbf{h})$ is the FT of the complementary indicator function $\chi_{V-U}(\mathbf{h})$ which corresponds to the disordered solvent zones (Fig. 1*d*). It should be noted that $G_U(\mathbf{h}) = G_{V-U}(\mathbf{h})$ for all $\mathbf{h} \neq 0$. The overall structure factor $F(\mathbf{h})$ can then be written as two components, one as a term for the ordered structure domain, $F_P(\mathbf{h})$, and the other as the term for the disordered solvent domains, $\rho_s G_{V-U}(\mathbf{h})$ or $\rho_s G_U(\mathbf{h})$ for $\mathbf{h} \neq 0$, where ρ_s is the constant electron density of this flat and featureless domain,

$$F(\mathbf{h}) = F_P(\mathbf{h}) - \rho_s G_U(\mathbf{h}) \text{ for } \mathbf{h} \neq 0.$$

Note that $F_P(\mathbf{h})$ is the Fourier transform of the protein in a vacuum (Fig. 1*e*) and may be rewritten in terms of the average electron density inside the domain U, $\langle\rho_P\rangle$, and the internal fluctuations from this average (Fig. 1*g*), $\langle\rho_P\rangle - \rho_P(\mathbf{r})$, such that their Fourier transforms become $\langle\rho_P\rangle G_U(\mathbf{h})$ and $\Delta(\mathbf{h})$, respectively. Substituting this into the overall expression one obtains

$$F(\mathbf{h}) = (\langle\rho_P\rangle - \rho_s)G_U(\mathbf{h}) + \Delta(\mathbf{h}) \text{ for } \mathbf{h} \neq 0.$$

The term $\langle\rho_P\rangle - \rho_s$ is defined as the contrast as illustrated in Fig. 1(*f*) (Stuhrmann & Kirste, 1965). When $\langle\rho_P\rangle = \rho_s$ the system is said to be at the contrast matching point, and only the Fourier transform of the internal electron density, $\Delta(\mathbf{h})$, contributes to the overall structure factor $F(\mathbf{h})$.

In order to account for anomalous scatterers included in the solvent, we use the seminal idea of Karle (1980) where the structure factors of an anomalously scattering atom, A, are separated into wavelength-independent (f°) and wavelength-dependent ($^\lambda f' + i^\lambda f''$) parts,

$$^\lambda f = f^\circ + {}^\lambda f' + i^\lambda f''.$$

We then consider that the structure factor $^\lambda f$ is constant to a first-order approximation with respect to scattering angle at low resolution. The density of the anomalous scatterers in the solvent can be treated as a complex quantity, $^\lambda\rho_{sA}$, which is dependent upon wavelength and can be separated into wavelength-independent and wavelength-dependent parts, *i.e.*

$$^\lambda\rho_{sA} = {}^\circ\rho_{sA}(1 + {}^\lambda f'/f^\circ + i^\lambda f''/f^\circ).$$

The total electronic density of the solvent, $^\lambda\rho_s$, becomes a function of wavelength and separable into wavelength-independent and wavelength-dependent parts,

$$^\lambda\rho_s = {}^\circ\rho_s + {}^\circ\rho_{sA}({}^\lambda f'/f^\circ + i^\lambda f''/f^\circ).$$

Note that the term $^\circ\rho_s$ includes the normal scattering part of the anomalous scatterer. Thus, one obtains for the overall structure factor $^\lambda F(\mathbf{h})$,

$$\begin{aligned} {}^\lambda F(\mathbf{h}) &= (\langle\rho_p\rangle - {}^\lambda\rho_s)G_U(\mathbf{h}) + \Delta(\mathbf{h}) \\ {}^\lambda F(\pm\mathbf{h}) &= [(\langle\rho_p\rangle - {}^\circ\rho_s)G_U(\mathbf{h}) + \Delta(\mathbf{h})] \\ &\quad - [{}^\circ\rho_{sA}({}^\lambda f'/f^\circ \pm i^\lambda f''/f^\circ)G_U(\mathbf{h})]. \end{aligned}$$

The terms between the first set of brackets represent the wavelength-independent part of the overall structure factor, denoted $^\circ F(\mathbf{h})$. It includes the envelope, the contrast and the fluctuation terms. The second set of brackets is wavelength-dependent and incorporates the envelope and the anomalous structure factors of A, $^\lambda f'$ and $^\lambda f''$. Note that the wavelength-dependent contribution is subtracted from the normal scattering part, indicating that the anomalous and dispersive structure factors of A are applied to the Fourier transform of the indicator function of the solvent-accessible domain $-G_U(\mathbf{h})$. The overall contrast becomes $\langle\rho_p\rangle - {}^\lambda\rho_s$ and contrast variation is generated by tuning the wavelength to different values of $^\lambda f'$ and $^\lambda f''$.

By defining $\Gamma(\mathbf{h}) = -{}^\circ\rho_{sA}G_U(\mathbf{h})$, one generates a expression of the overall structure factor similar to the starting point used for the algebraic MAD method (Hendrickson, 1985), where $\Gamma(\mathbf{h})$ replaces the normal scattering component of the partial structure A, $^\circ F_A(\mathbf{h})$,

$$\begin{aligned} {}^\lambda F(\pm\mathbf{h}) &= {}^\circ F(\mathbf{h}) + ({}^\lambda f'/f^\circ \pm i^\lambda f''/f^\circ){}^\circ F_A(\mathbf{h}) \quad \text{(MAD)} \\ {}^\lambda F(\pm\mathbf{h}) &= {}^\circ F(\mathbf{h}) + ({}^\lambda f'/f^\circ \pm i^\lambda f''/f^\circ)\Gamma(\mathbf{h}) \quad \text{(MASC)}. \end{aligned}$$

The substitution of $\Gamma(\mathbf{h})$ for $^\circ F_A(\mathbf{h})$ has an obvious physical meaning. The anomalous partial structure, A, which is a set of a few punctual and ordered scatterers in a MAD experiment, is exchanged for an extended uniform electron density in a MASC experiment (see Fig. 1*h*). The separation of the effects of the anomalous partial structure A (and hence the Fourier transform of the solvent-accessible volume) from the overall diffraction effects can be applied using a set of equations analogous to the *MADLSQ* equations, where they are solved for $|{}^\circ F_T(\mathbf{h})|$, $|\Gamma(\mathbf{h})|$ and the phase difference between $^\circ F_T(\mathbf{h})$ and $\Gamma(\mathbf{h})$, $\Delta\varphi = (\varphi_T - \varphi_\Gamma)$, *i.e.*

$$|^{\lambda}F(\pm\mathbf{h})|^2 = |^{\circ}F_T(\mathbf{h})|^2 + a(\lambda)|\Gamma(\mathbf{h})|^2 + b(\lambda)|^{\circ}F_T(\mathbf{h})||\Gamma(\mathbf{h})|\cos(\Delta\varphi) \pm c(\lambda)|^{\circ}F_T(\mathbf{h})||\Gamma(\mathbf{h})|\sin(\Delta\varphi),$$

where

$$a(\lambda) = ({}^{\lambda}f'^2 + {}^{\lambda}f''^2)/f^{\circ 2}$$
$$b(\lambda) = 2{}^{\lambda}f'/f^{\circ}$$
$$c(\lambda) = 2{}^{\lambda}f''/f^{\circ}.$$

3. Strength of the anomalous signal in MASC

The strength of the anomalous signal in a MASC experiment can be estimated as in a MAD experiment by measuring differences between Bijvoet pairs (anomalous or ${}^{\lambda}f''$ contribution) and wavelength pairs (dispersive or $\Delta f'$ contribution). Intuitively, the magnitude of the anomalous signal in a MASC experiment is expected to vary considerably with resolution, being very large in the lowest resolution shells and then diminishing rapidly with increasing resolution. One also expects the anomalous signal to be directly proportional to the concentration of the anomalous scatterer in the solvent-accessible volume. Furthermore, the signal will be maximized at the point of contrast matching. By making a certain number of approximations (globular protein, Porod's law, smooth interface *etc*), it is possible to derive expressions for and calculate the expected anomalous and dispersive ratios (Fourme *et al.*, 1995), but to be concise only the final expressions will be given here. Thus, for anomalous and dispersive differences, respectively, one obtains[1]

$$\langle|{}^{\lambda}\Delta F(\pm\mathbf{h})|\rangle/\langle|{}^{\lambda}F(\pm\mathbf{h})|\rangle = 3.44\times10^{-4}[A](2{}^{\lambda}f''/f_{\rm eff})(M_w^{1/12}s)^{-2}\exp(-Bs^2/4)$$

and

$$\langle|{}^{\Delta\lambda}\Delta F(\mathbf{h})|\rangle/\langle|{}^{\Delta\lambda}F(\mathbf{h})|\rangle = 3.44\times10^{-4}[A](\Delta f'/f_{\rm eff})(M_w^{1/12}s)^{-2}\exp(-Bs^2/4).$$

M_w is the molecular weight and $f_{\rm eff}$ is the r.m.s. of the structure-factor moduli of the atoms in the macromolecule ($f_{\rm eff}$ = 6.7 electrons for proteins at s = 0). Obviously, the anomalous signal is dependent on the molar concentration of the anomalous scatterer, $[A]$, and the magnitudes of $\Delta f'$ and ${}^{\lambda}f''$, but the strongest effect on the anomalous signal is a result of the resolution, $s = 2\sin\theta/\lambda$, diminishing as a function of $1/s^2$ and $\exp(-Bs^2/4)$. In the term $\exp(-Bs^2/4)$, B is a pseudo-temperature factor which models the combined effects of the smearing of the envelope boundary as well as the temperature factor of the solvent. However, even with a perfectly sharp interface (*i.e.* B = 0 Å^2), the anomalous signal still drops away as $1/s^2$ owing to Porod's law (Porod, 1951). The loose dependence upon the molecular weight arises from the assumptions of a globular protein and Porod's law. Hypothetical MASC signals are plotted against resolution for different molecular weights in Fig. 2. As is clearly illustrated, one expects very large signals in the lowest resolution shells that will diminish sharply with increasing resolution. Thus, to obtain a measurable anomalous signal (>0.020) out to 10 Å resolution, either multimolar quantities of a K-edge scatterer or molar quantities of an L-edge scatterer are required.

[1] Where $s = 2\sin\theta/\lambda$, $|{}^{\lambda}\Delta F(\pm\mathbf{h})| = \big||{}^{\lambda}F(+\mathbf{h})| - |{}^{\lambda}F(-\mathbf{h})|\big|/\langle|{}^{\lambda}F(\pm\mathbf{h})|\rangle$, $\langle|{}^{\lambda}F(\pm\mathbf{h})|\rangle = \big||{}^{\lambda}F(+\mathbf{h})| + |{}^{\lambda}F(-\mathbf{h})|\big|/2$, $|{}^{\Delta\lambda}\Delta F(\mathbf{h})| = \big||{}^{\lambda_i}F(\mathbf{h})| - |{}^{\lambda_j}F(\mathbf{h})|\big|/\langle|{}^{\Delta\lambda}F(\mathbf{h})|\rangle$, $\langle|{}^{\Delta\lambda}F(\mathbf{h})|\rangle = [|{}^{\lambda_i}F(\mathbf{h})| + |{}^{\lambda_j}F(\mathbf{h})|]/2$ and $\Delta f' = |{}^{\lambda_i}f' - {}^{\lambda_j}f'|$.

4. Experimental considerations

4.1. The experimental setup

Since a MASC experiment is analogous to a MAD experiment but at low resolution, the data collection should ideally be carried out at X-ray wavelengths near an absorption edge of the anomalous scatterer where the ${}^{\lambda}f'$ and ${}^{\lambda}f''$ vary abruptly. Thus, a MASC experiment requires, in addition to the usual equipment of a crystallography experiment, a tuneable source of X-rays with a narrow bandpass ($\Delta\lambda/\lambda \simeq 10^{-4}$), an X-ray fluorescence detector to determine precisely the wavelengths of ${}^{\lambda}f''_{\rm max}$ and $|{}^{\lambda}f'|_{\rm max}$, and an experimental setup designed to collect reflections at the lowest possible resolution. This last aspect is not as trivial as it first appears (also see Evans *et al.*, 2000), since small-angle scattering from different components of the setup can severely add to the background and deteriorate the quality of the data, especially if the solvent electron density is at the contrast matching where the low-resolution reflections are weakest. The X-ray background around the beamstop originates from several sources, in particular from the slits, the air path and the windows of the He cone, as well as the sample and its holder. The mounting of a low-resolution beamstop requires special attention and a schematic layout is shown in Fig. 3. Typically, a MASC experiment requires the mounting of a small beamstop just in front of the detector entrance window inside the He cone. However, a very small beamstop which is only slightly larger than the beam size and is accurately placed just behind

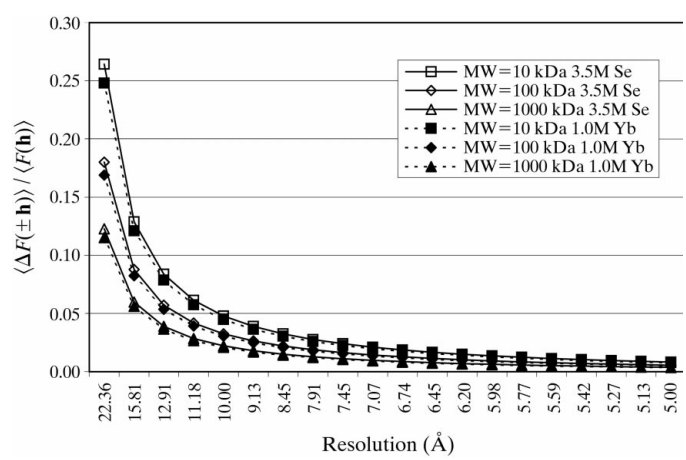


Figure 2
Hypothetical Bijvoet signals plotted against resolution (as a function of s^2) for three different molecular weights (10, 100 and 1000 kDa) and in two different solutions of anomalous scatterers (3.5 M SeO_4^{2-} K-edge, ${}^{\lambda}f''$ = 7 e$^-$ and 1.0 M Yb^{3+} $L_{\rm III}$-edge, ${}^{\lambda}f''$ = 23 e$^-$).

Table 1
Selected crystallographic data for MASC experiments on HEWL, P64k and XI crystals (taken from Ramin *et al.*, 1999).

Protein crystal	HEWL				P64k			XI			
Space group and unit cell (Å)	$P4_32_12$; $a = b = 78.68$, $c = 37.05$				$P4_32_12$; $a = b = 140.62$, $c = 77.02$			$P3221$; $a = b = 141.91$, $c = 227.48$			
Fractional volume of protein (%)	70.7				39.5			37.4			
Wavelengths (Å)	1.3928	1.3859	1.3852	1.3776	0.9919	0.9795	0.9793	0.9919	0.9795	0.9793	0.9791
Number of measured reflections	10257	10427	10436	10617	44222	45453	45310	110358	114299	114009	114302
Number of unique reflections	1092	1112	1111	1129	6008	6107	6106	19684	20430	19429	20456
Resolution limits (Å)	33.5–3.95	33.5–3.93	33.5–3.92	33.5–3.91	100–4.19	100–4.18	100–4.18	105.4–4.18	105.4–4.12	105.4–4.13	105.4–4.12
Overall completeness (%)	84.7	86.4	86.3	87.8	98.2	99.9	99.9	97.3	95.4	96.2	95.5
R_{sym}, $d = 20$ Å (%)	3.7	3.1	3.5	2.7	4.2	4.4	4.1	3.7	4.3	8.5	6.0
R_{sym}, $d = 10$ Å (%)	3.2	3.5	3.3	2.8	3.4	3.6	3.6	2.7	3.6	7.8	5.1
R_{sym}, global (%)	2.8	2.5	2.7	2.3	5.5	5.1	5.6	3.7	7.0	19.4	10.6
$\langle I\rangle/\langle\sigma(I)\rangle$, $d = 20$ Å	11.6	14.9	12.6	18.0	12.3	7.9	14.1	12.2	11.5	6.3	7.7
$\langle I\rangle/\langle\sigma(I)\rangle$, $d = 10$ Å	15.5	16.4	15.7	20.6	15.9	14.4	16.1	18.9	16.2	8.2	11.9
$\langle I\rangle/\langle\sigma(I)\rangle$, global	18.3	20.9	19.6	23.3	5.2	9.4	8.0	17.3	11.8	3.9	6.9
Completeness ($d = 10$ Å) (N_{obs}/N_{poss}) (%)	73.86 (65/88)				100 (501/501)			99.38 (1591/1601)			
Completeness ($d = 20$ Å) (N_{obs}/N_{poss}) (%)	53.8 (7/13)				100 (73/73)			96.9 (217/224)			
R factor, *MADLSQ* ($d = 20$ Å) (%)	32.9 (16.3)†				32.4			23.7			
R factor, *GFROMF* ($d = 20$ Å) (%)	44.5 (27.3)†				33.9			26.9			

† *R* factors for HEWL MASC data after accounting for six Yb sites.

the crystal will minimize considerably the background arising from the air scatter by cutting down on the length of the air path travelled by the X-ray beam. If a long crystal-to-detector distance is used, then it is important that the beam divergence is low or that the beam is carefully focused onto the detector rather than the crystal. This helps to avoid an enlargement or elongation of the reflection profiles.

4.2. Choice of the anomalous scatterer and preparation of crystals

A variety of anomalous scatterers may be used in a MASC experiment and the most suitable ones will depend on the crystallization conditions of the macromolecule. Analogues of the precipitating agent are good choices, as such compounds are less likely to perturb the crystalline lattice (*e.g.* selenate for sulfate, bromide for chloride, tribromoacetate for acetate *etc*). In general, the MASC compound used should have a high solubility in the mother liquor, should not influence the pH and should be inert to avoid binding to and denaturing the protein. Crystals may be prepared either by co-crystallization with the MASC compound or by soaking techniques. In the latter case, special attention is often needed to prevent crystals from cracking owing to the osmotic shock upon addition of the high concentration of the MASC compound. Small amounts of the MASC compound can be added progressively to the mother liquor over a period of several hours, or vapour-diffusion techniques may be used to slowly increase the ionic strength of the mother liquor before attaining the desired concentration. Dialysis buttons could also be used, although their rate of exchange is often too fast for many crystals unless the reservoirs are incremented in a stepwise fashion. Another factor to consider is that the macromolecular crystals of interest may dissolve at high ionic strengths and this could be understood as a form of 'dehydration' of the crystal, which suffers from over-competition with the MASC compound for water molecules. Finally, it can be difficult to find a MASC compound suitable for those crystals which grow under low ionic strength conditions, such as polyethylene glycol precipitating agents. One possibility is to use neutral polar compounds, such as aurothioglucose; however, these compounds remain to be examined in detail.

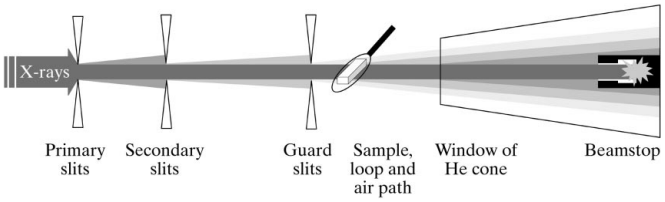


Figure 3
The primary slits are closed down to the size of the crystal, while the secondary slits are opened slightly larger. This shields the detector from the X-ray scatter of the primary slits and avoids the secondary slits contributing more X-ray scatter. A third series of guard slits, which are opened wider than the previous two, are inserted just prior to the crystal to further screen out the X-ray scatter from the slits and air upstream of the crystal. When the slits are correctly aligned, the largest contribution to low-angle X-ray scatter will come from the transmitted X-ray beam traveling through air. If the crystal-to-detector distance must be large, then a He cone should be set up to minimize the scattering from air. Some scattering will arise from the entrance window of the He cone and ideally the entire assembly (crystal, cryostat, slits *etc.*) should be encased in a He atmosphere and a He cryostat should be used.

4.3. X-ray fluorescence and absorption effects

The combination of working with high concentrations of an anomalous scatterer in the crystal and with X-ray wavelengths at one of its absorption edges will provoke two drawbacks: severe absorption of the Bragg reflection intensities of the crystal and an increased background from X-ray fluorescence.

Little can be done to minimize severe absorption effects except to optimize the crystal size such that the reflection intensities are maximized with respect to the crystal volume and its linear absorption coefficient μ. Fluorescence is the emission of a photon after the absorption of an incident photon with an energy greater than that of the absorption edge. X-ray fluorescence effects can be very strong in MASC experiments because of the high concentrations of anomalous scatterers in the crystal. The fluorescence yield is at a maximum at X-ray wavelengths where ${}^{\lambda}f''$ is a maximum. The effects can be very detrimental at the sharp white-line resonances, where the emitted radiation floods out the diffraction image (see Fig. 4). This is especially true for K edges, which have higher fluorescence yields than L edges. Remedies include increasing the crystal-to-detector distance, collecting images with small rotation slices or maximizing the dispersive differences by collecting diffraction data at the inflection point and then at two other wavelengths as remote as possible from the absorption edge. However, perhaps the best solution is to switch the anomalous scatterer from a K-edge element to an L-edge element. For L-edge white-line resonances, ${}^{\lambda}f''$ can be over 20 electrons, so that only a fraction of the concentration is needed to generate an anomalous signal; near 1 Å the X-ray fluorescence yield can be halved (Kortright, 1986).

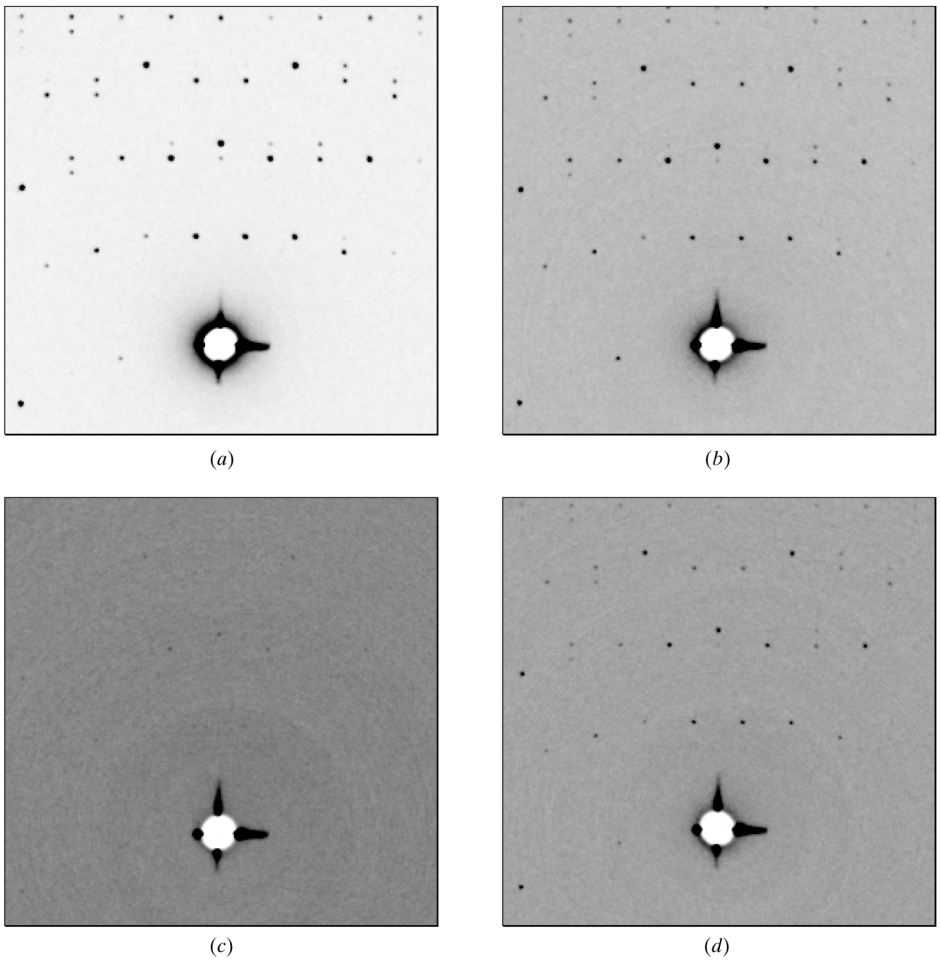


Figure 4
Fluorescence and absorption effects of the diffraction pattern from an XI crystal in 2 M $(NH_4)_2SeO_4$ recorded at four different wavelengths about the Se K-edge and of the same region of reciprocal space. (a) Low-energy remote, 12 500 eV; (b) inflection point, 12 657.5 eV; (c) peak fluorescence, 12 660 eV; (d) high-energy remote, 12 663 eV.

5. Experimental test-case studies

To date, MASC data have been collected and analysed on crystals of three proteins of differing molecular weights (14, 54 and 173 kDa) and using a variety of anomalous scatterers (see Table 1). All of the cases are known crystal structures, which have allowed the experimental results to be compared with their known envelope structure-factor moduli and phases (Ramin *et al.*, 1999). In each of the experiments, the X-ray diffraction data were recorded at the wavelengths corresponding to the $|{}^{\lambda}f'|_{\max}$ and ${}^{\lambda}f''_{\max}$, which were determined from the X-ray fluorescence spectra from a solution of the anomalous scatterer, as well as for at least one wavelength remote from the absorption edge. A small beamstop (~2.5 mm) was mounted and aligned just in front of the entrance window of the detector. Where possible, the crystallographic axes were aligned so that Bijvoet pairs could be measured on the same image. The full details of these experiments have been reported in Ramin *et al.* (1999) and here we will present each case study briefly.

5.1. Hen egg-white lysozyme co-crystallized in $YbCl_3$

As a first test case, hen egg-white lysozyme (HEWL) crystals were grown directly from $YbCl_3$ solutions to obtain robust HEWL crystals, which normally crystallize in NaCl, and to exploit the white-line structure of the Yb L_{III}-edge. X-ray diffraction data were collected for the first ever MASC experiment in 1993 on station D23 at LURE-DCI (Kahn *et al.*, 1986) at the Yb L_{III}-edge from a single crystal of HEWL containing 0.8 M $YbCl_3$ and then later with 0.5 M $YbCl_3$ crystals on station DW21b at LURE-DCI. In both experiments, the results confirmed a large anomalous signal at low resolution as expected by theory (see Fig. 5a). The internal agreement between true equivalent reflections is within ~1–3%, implying that the anomalous signal is reproducible and not an artefact of either absorption effects, data processing or reflections hidden behind the beamstop. In all cases, the

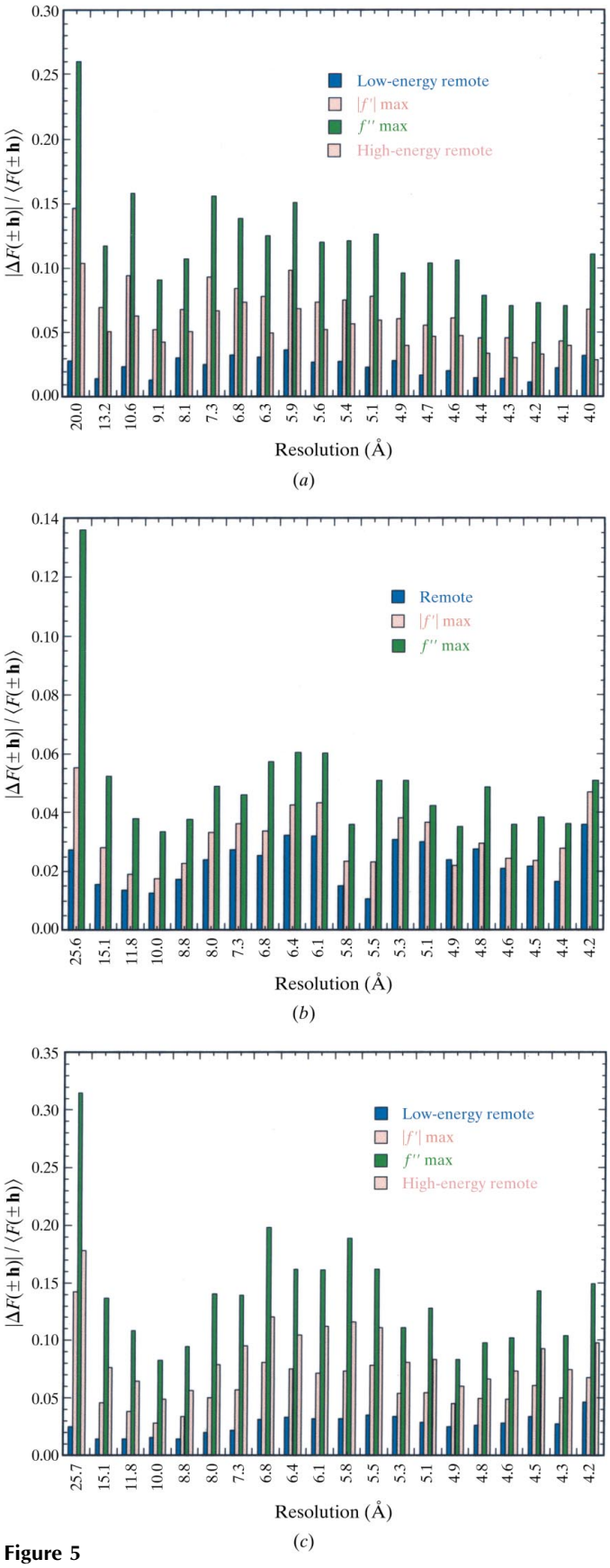


Figure 5
Anomalous R factors as a function of resolution from MASC data. (*a*) HEWL in 0.5 M $YbCl_3$, (*b*) P64k in 3.5 M $(NH_4)_2SeO_4$ and (*c*) XI in 2.0 M $(NH_2)_2SeO_4$ (Ramin *et al.*, 1999).

anomalous signal extends well beyond 10 Å resolution, indicating that Yb^{3+} ions have bound to the protein in the crystal lattice. The metal sites have been identified from phased anomalous difference Fourier maps (Fig. 6*a*) and their number and occupancies vary for each crystal. In one case using *SHARP* (de La Fortelle & Bricogne, 1997), as many as six metal sites could be refined (Ramin *et al.*, 1999). It is important to note that the data in the lowest resolution shells are rather incomplete as the crystals were aligned along the crystallographic fourfold axis. This is a consequence of the combination of the large blind zone formed by the curvature of the Ewald sphere at the wavelength of the Yb L_{III} edge (1.3862 Å) and the relatively small unit-cell parameters of HEWL.

5.2. Outer membrane meningitidis protein (P64k) soaked in $(NH_4)_2SeO_4$

One of the proteins under study in our laboratory is a 54 kDa domain of the outer membrane protein (P64k) from *Neisseria meningitidis* (Li de la Sierra *et al.*, 1994, 1997) that crystallizes from ammonium sulfate solutions and for which the mother liquor can be substituted with multimolar concentrations of ammonium selenate *via* simple soaking techniques. The selenate anion is isostructural with sulfate and has a sharp white-line feature at the Se K-edge (~0.9796 Å). Crystals of P64k withstand 3.5 M $(NH_4)_2SeO_4$ solutions which bring the solvent electronic density equal to the average protein electronic density, *i.e.* the contrast matching point. MASC data were collected at the Se K-edge on the TROIKA station at the ESRF. Severe X-ray fluorescence effects were observed because of the sharp white-line resonance of the selenate anion coupled with its high concentration and the fine X-ray bandpass from the monochromator (Fig. 4). The X-ray data in the lowest resolution shells are virtually complete because the curve of the Ewald sphere is relatively flat at ~1 Å wavelength, so much so that Friedel pairs often appeared on the same image. Despite the difficulties arising from fluorescence, the anomalous signals are large in the low-resolution shells but do not completely disappear at higher resolution (Fig. 5*b*); phased anomalous difference Fourier maps (Fig. 6*b*) confirmed the existence of at least one selenate site bound in a pocket of the macromolecule. Anomalous difference Patterson maps recovered this site which could be refined using *SHARP*, but no other sites could be located (Ramin *et al.*, 1999).

5.3. Xylose isomerase soaked in $(NH_4)_2SeO_4$

Xylose isomerase (XI) also crystallizes from ammonium sulfate solutions but as a large tetramer of 173.2 kDa in the asymmetric unit (Rey *et al.*, 1988). This represents a fairly large macromolecular structure on the scale of those typically solved by the MAD method. Crystals were soaked in 2 M $(NH_4)_2SeO_4$ solutions and MASC data were collected at the Se K-edge on the TROIKA station in the same fashion as for the P64k crystals. Evidence for several selenate anion sites have been found in the anomalous signal and in the phased anomalous difference Fourier maps of XI (Figs. 5*c* and 6*c*). All

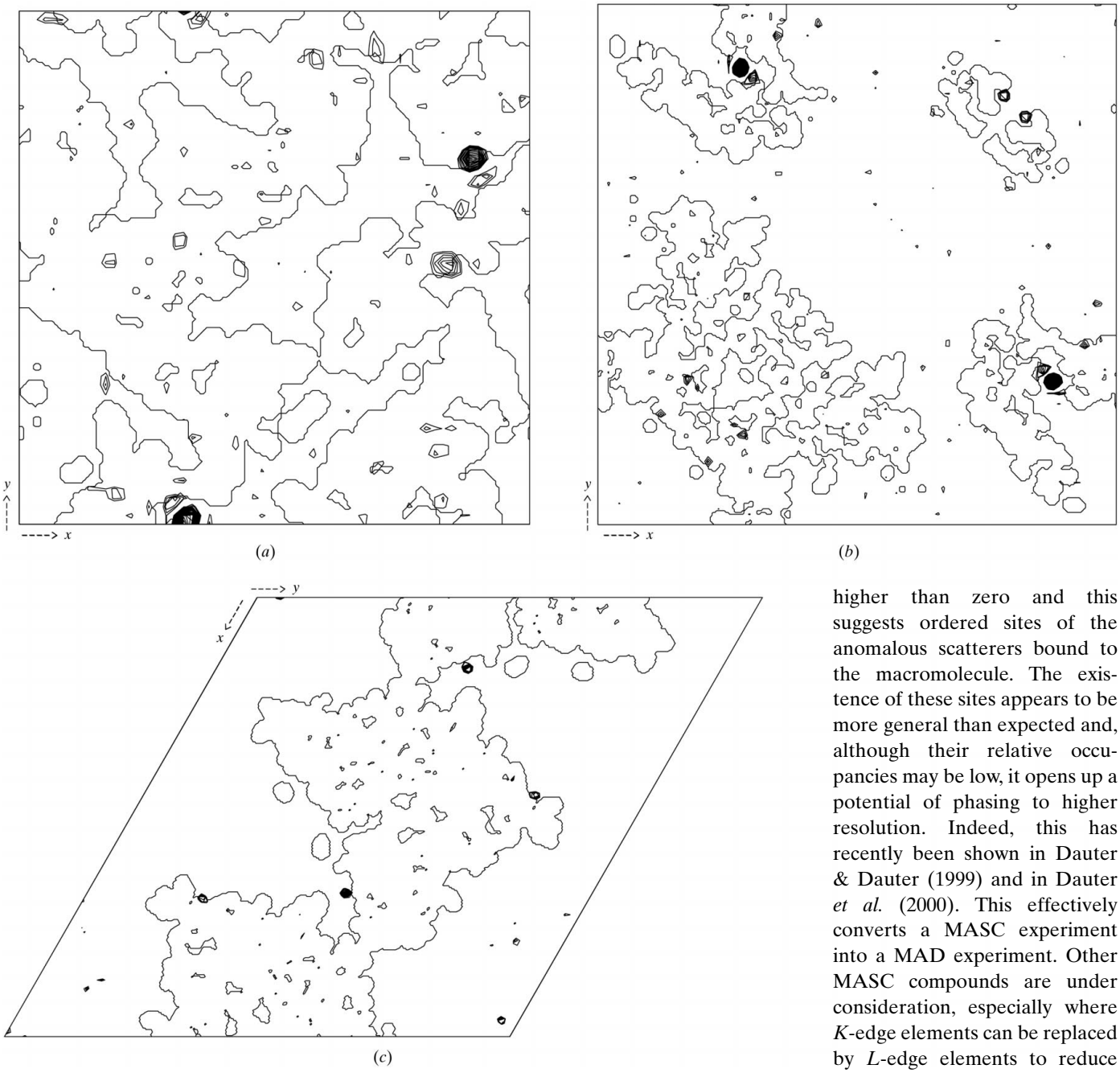


Figure 6
Ordered sites of anomalous scatterers found in (*a*) HEWL, (*b*) P64k and (*c*) XI. In each case, a phased anomalous Fourier map is superimposed onto a map of the protein envelope. Dark spots on each map show anomalous scattering atom positions in crevices and near the surface of the protein (Ramin *et al.*, 1999).

of the sites are at or near the macromolecular boundary; however, attempts to refine any of these have not proved fruitful (Ramin *et al.*, 1999).

5.4. General comments

The anomalous signal for all test cases follows the expected trend, being very large at lowest resolution and decreasing rapidly with increasing resolution. At higher resolution, the Bijvoet ratios for all protein crystals remains significantly higher than zero and this suggests ordered sites of the anomalous scatterers bound to the macromolecule. The existence of these sites appears to be more general than expected and, although their relative occupancies may be low, it opens up a potential of phasing to higher resolution. Indeed, this has recently been shown in Dauter & Dauter (1999) and in Dauter *et al.* (2000). This effectively converts a MASC experiment into a MAD experiment. Other MASC compounds are under consideration, especially where *K*-edge elements can be replaced by *L*-edge elements to reduce the background owing to fluorescence as well as to prevent binding to the macromolecule. From this last point of view, non-detergent sulfobetains (NDSB), which are neutral zwitterions commonly used as solubilizing agents to help the crystallization of macromolecules, are of special interest. Girard *et al.* (1999) have tested tetragonal crystals of HEWL co-crystallized with a monobrominated sulfobetain at a concentration of 0.72 *M*. As verified by a Fourier difference map, this anomalous contrast agent does not bind to the macromolecule and as predicted the MASC signal was observed at low resolution whereas no anomalous signal could be detected at higher resolution.

6. Extracting envelope structure-factor moduli, $|G_U(\mathbf{h})|$, from MASC data

Extraction of the envelope structure-factor moduli, $|G_U(\mathbf{h})|$, from MASC data can be performed in two ways. The first is to treat the MASC data as a MAD data by determining the anomalous partial structure, $|^{\circ}F_A(\mathbf{h})|$, with the program *MADLSQ* (Hendrickson, 1985). The second method is to consider the MASC data as a chemical contrast-variation series, $|^{i}F_{\rm obs}(\mathbf{h})|$, using a modified version of the program *GFROMF* (Carter & Bricogne, 1987). Both methods give satisfactory results for data to at least 20 Å resolution.

6.1. *MADLSQ*

The program *MADLSQ* was originally designed for MAD data and solves the set of algebraic MADLSQ equations by non-linear least-squares for $|^{\circ}F_T|$, $|^{\circ}F_A|$ and the phase difference $\Delta\varphi_{T-A}$.[2] These equations can be applied to MASC data by substituting $|\Gamma(\mathbf{h})| = -\rho_{sA}|G_U(\mathbf{h})|$ for $|^{\circ}F_A(\mathbf{h})|$ and the phase difference for $\Delta\varphi_{T-G}$. The program has the ability to refine or fix the values of $^{\lambda}f'$ and $^{\lambda}f''$ of the different wavelengths. Comparisons of the $|G_U(\mathbf{h})|$ from *MADLSQ* with the $|G_U(\mathbf{h})|$ calculated from the three-dimensional coordinates in the PDB are shown in Figs. 7(*a*)–7(*c*) for the three different protein crystals (*R* factors of ~16–33%; see Table 1). Note the sharp asymptotic decrease in $|G_U(\mathbf{h})|$ with increasing resolution. The agreement between model and experiment deteriorates beyond 10–20 Å resolution for several reasons: (i) the relative magnitudes of $|G_U(\mathbf{h})|$ are small, (ii) the absorption effects are more pronounced at higher diffracting angles and (iii) the possibility of ordered anomalous scattering sites contributing to the partial structure extracted from the MADLSQ equations [*i.e.* $|\Gamma(\mathbf{h})|$ is more precisely defined as $|\Gamma(\mathbf{h}) + {}^{\circ}F_A(\mathbf{h})|$].

6.2. *GFROMF*

In chemical contrast-variation studies the diffraction data are a contrast series of $|^{i}F_{\rm obs}(\mathbf{h})|$ for $i = 1, \ldots, N$, where i corresponds to a different solvent-density level $^{i}\rho_s$. To extend this to a multiple-wavelength case, we simply substitute in for the contrast series $|^{\lambda_i}F_{\rm obs}(\mathbf{h})|$, where $\lambda_i = \lambda_1, \ldots, \lambda_N$ and the solvent density becomes $^{\lambda_i}\rho_s$. The same formalism is used to describe the overall structure factor in terms of the Fourier transforms of the envelope $[G_U(\mathbf{h})]$ and the internal density fluctuations $[\Delta(\mathbf{h})]$. If $X(\mathbf{h})$ and $Y(\mathbf{h})$ are the real and imaginary components of $\Delta(\mathbf{h})$ relative to $G_U(\mathbf{h})$, one has

$$|^{i}F_{\rm calc}(\mathbf{h})| = K_i\{[X(\mathbf{h}) + (\langle\rho_p\rangle - {}^{i}\rho_s)|G_U(\mathbf{h})|]^2 + Y(\mathbf{h})^2\}^{1/2}.$$

The *GFROMF* (Carter & Bricogne, 1987) program carries out the non-linear least-squares refinement of $|G_U(\mathbf{h})|$, $X(\mathbf{h})$ and $Y(\mathbf{h})$ from scaled data summed over all contrasts series and minimizes the function

$$\sum_i \sum_{hkl} \sigma_{\rm obs}(\mathbf{h})^{-2}[|^{i}F_{\rm obs}(\mathbf{h})| - |^{i}F_{\rm calc}(\mathbf{h})|]^2,$$

where $\sigma_{\rm obs}(\mathbf{h})$ is the standard deviation of $|^{i}F_{\rm obs}(\mathbf{h})|$. $X(\mathbf{h})$ and $Y(\mathbf{h})$ effectively represent the magnitude and the phase difference between $G_U(\mathbf{h})$ and $\Delta(\mathbf{h})$. In practice, a scale factor between the different data sets, K_i, should be refined for all but one contrast or wavelength.

The original program was modified to incorporate anomalous scattering contributions such that,

$$\begin{aligned}|^{\lambda_i}F_{\rm calc}(\pm\mathbf{h})| = {}^{\lambda_i}K\big(\{&X(\mathbf{h}) \\ &+ [\langle\rho_p\rangle - \rho_s - ({}^{\lambda_i}f'/f^{\circ})^{\circ}\rho_{sA}]|G_U(\mathbf{h})|\}^2 \\ &+ \{y(\mathbf{h}) \pm [-({}^{\lambda_i}f'/f^{\circ})^{\circ}\rho_{sA}]|G_U(\mathbf{h})|\}^2\big)^{1/2}.\end{aligned}$$

Trials on simulated MASC data of kallikrein (52 kDa) at three different contrasts of selenate and three wavelengths per contrast returned exact values of $|G_U(\mathbf{h})|$, $X(\mathbf{h})$ and $Y(\mathbf{h})$ of the simulated observed data (Ramin, 1999). With experimental data, the results gave *R* factors of 27–45% for the test-case crystals (see Table 1). This level of agreement is satisfactory considering that many of the parameters are left unrefined. In particular, the values of $^{\lambda_i}f'$ and $^{\lambda_i}f''$ utilized were derived from previous runs of *MADLSQ* and theoretical values of the contrast were used rather than allowing them to refine. The scale factors between different wavelengths ($^{\lambda_i}K$) were set to unity since the data were already set on a common scale. In principle, all of these parameters should be refined in the *GFROMF* scheme, even though the number of observations in the lower resolution shells is not large. What is certain is that prior precise knowledge of the values of the contrasts and the anomalous scattering factors $^{\lambda_i}f'$ and $^{\lambda_i}f''$ is important to extract $|G_U(\mathbf{h})|$ values of satisfactory quality.

6.3. Separation of anomalous signals from punctual sites and the disordered solvent

Separation of the anomalous signal from punctual sites (MAD) and from the disordered region of the solvent (MASC) can at least be partially accomplished by delimiting the X-ray diffraction data into the very lowest resolution shell, say below 20 Å resolution, where the MASC effects are strongest, or into the higher resolution shells, say above 5 Å resolution, where the MAD effects are strongest. This however is not a deconvolution, since even in the lowest resolution shells, which are dominated by the enormous MASC signals, the MAD signals are present albeit weak. The same is also true for MASC signals in the higher resolution shells. At moderate resolution, roughly between 20 and 5 Å resolution, MAD and MASC signals are mixed in non-negligible proportions. In order to deconvolute punctual sites and the disordered solvent region, additional information is necessary either in the form of another contrast series or as structural information on the punctual sites or of the solvent region. As such, a model for the punctual sites may be determined from the high-resolution anomalous data and these sites could then be used to extract out a model for the solvent at lower resolution. Both models could then in principle be refined against the entire data set of $|\Gamma(\mathbf{h}) + {}^{\circ}F_A(\mathbf{h})|$.

[2] $|^{\circ}F_T|$ is the moduli of normal scattering for the total structure, $|^{\circ}F_A|$ is the moduli of normal scattering for the anomalous partial structure and $\Delta\varphi_{T-A}$ is the phase difference between $^{\circ}F_T$ and $^{\circ}F_A$.

7. Phasing the moduli of envelope structure factors

7.1. The phase problem at low resolution

The problem of directly phasing the structure-factor moduli of macromolecular envelopes is rather unusual in crystallography since it does not adhere to the characteristics normally found in small-molecule or protein crystallography. For example, envelopes do not have atomic or point scatterer character and thus the principles of 'atomicity' do not apply. Furthermore, by definition the density of an envelope of a macromolecule is anything but randomly distributed throughout the unit cell; quite the contrary, its distribution is binary and can be considered – in a first approximation – to be confined to a 'compact' domain (*i.e.* a continuous single volume described by a closed single surface). For an assembly of macromolecules this approximation still holds true, but whereas in solution these compact domains are well separated, in a crystal they are adjoining owing to the crystal packing contacts made between macromolecules in the crystal lattice.

An important problem is that there is no prior information on the scale factor between the observed diffraction data (structure-factor moduli of macromolecular envelope) and the calculated data (from the envelope model). In small-molecule or macromolecule crystallography, the scale factors used are based upon Wilson statistics (Wilson, 1942), which assumes a random distribution of point scatterers in the unit cell. However, Wilson statistics can only be successfully applied to diffraction data at moderate to high resolution; at resolutions below 3.5 Å Wilson statistics breaks down because the diffraction process is then dominated by the diffraction of secondary structures. The determination of the scale factor is further complicated by the large 'dynamic range' of values for the structure-factor moduli of macromolecular envelopes, which can vary over at least two orders of magnitude in the

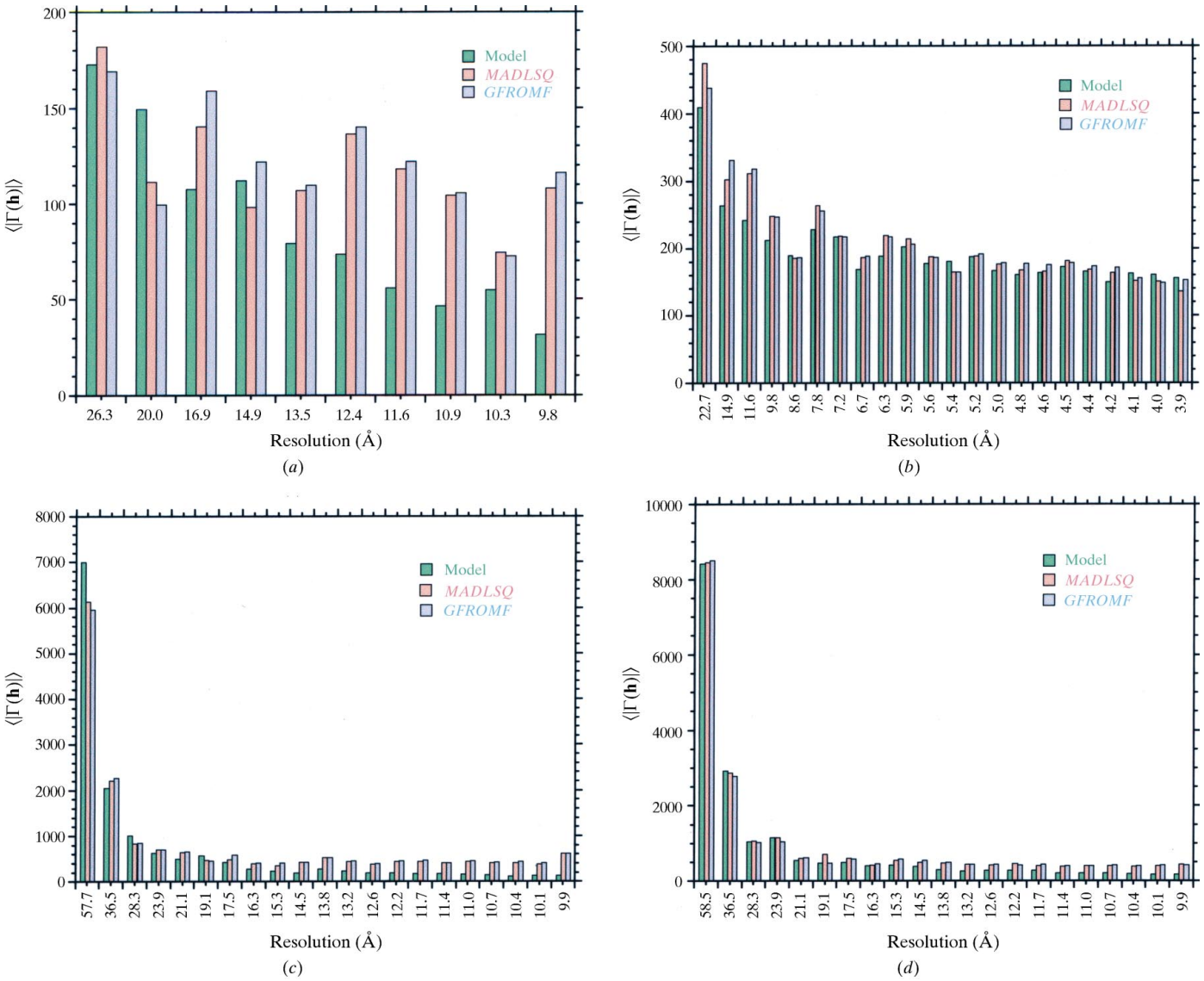


Figure 7
Experimental $|\Gamma(\mathbf{h})|$ values extracted by *MADLSQ* and *GFROMF* compared with the values expected from their models (*a*) HEWL, (*b*) HEWL with six Yb sites, (*c*) P64k and (*d*) XI (Ramin *et al.*, 1999).

resolution range extending from 100 to 10 Å and are very large in the lowest resolution shells, diminishing sharply with increasing resolution.

Another aspect which should be kept in mind is that at low resolution there is a relatively small number of observations; in fact, the number of observations is a small fraction of the number of non-H atoms in the asymmetric unit. For example, lysozyme, a 14 kDa protein, consists of approximately 1500 non-H atoms, but its tetragonal crystal form has only 202 independent observations to 7.5 Å resolution.

It should now begin to be clear to the reader that 'traditional' direct methods used in solving the phase problem for small-molecule structures are unsuitable in the determination of macromolecular envelopes. Even the application of Sayre's equation[3] (Sayre, 1952) is complicated by the limited number of observations at low resolution and their large dynamic range, since this equation is a self-convolution process requiring the products of $F(\mathbf{k})F(\mathbf{h}-\mathbf{k})$ be summed over all $\mathbf{k}$, which should cover as much of reciprocal space as possible.

Unless constraints to the model of the envelope are applied, knowledge of only the moduli of the envelope structure factors would yield an infinite number of solutions to the phase problem. Our approach is to constrain the solution in real space by imposing a constant electron density inside the volume delimited by the envelope of a single molecule that is reproduced by the crystal symmetry.

7.2. Method and theory

7.2.1. Parameterization of the macromolecular envelope[4]. Since there are a small number of observations at low resolution, the envelope of the macromolecule must be described by a small number of parameters. We start with the hypothesis that the envelope delimits a mathematically compact domain U_0 in which the indicator function is unity inside the domain and zero outside the domain (Fig. 8). The boundary of the domain is unknown. The proportion of the cell volume occupied by this domain is presumed to be known. The parameters describing the frontier of the domain U_0 are refined in order to improve the agreement between calculated and observed structure factors, $G_{\rm cal}(\mathbf{h})$ and $G_{\rm obs}(\mathbf{h})$, respectively. For reasons of convenience, spherical coordinates (r, θ, φ) are used to describe the macromolecular envelope (Fig. 8*a*). The origin C of this coordinate system is set arbitrarily inside the molecule. The surface is described by the value $R(\theta, \varphi)$, taken as the distance from the origin to the boundary of the domain for given angular variables (θ, φ)

$$\mathbf{R}(\theta, \varphi) = R(\theta, \varphi)\hat{\mathbf{u}}(\theta, \varphi).$$

It should be noted that $R(\theta, \varphi)$ is a function limited to a single value for each (θ, φ) and as a consequence certain concave features of a surface cannot be described by this method (see Fig. 8*b*). $R(\theta, \varphi)$ is expressed as a linear combination of a set of functions $f_k(\theta, \varphi)$

$$R(\theta, \varphi) = \sum_k a_k f_k(\theta, \varphi)$$

and the refined parameters thus become the coefficients a_k. In principle, any set of linearly independent functions may be used to describe the envelope. Here, we develop the use of spherical harmonics $Y_{l,m}(\theta,\varphi)$ as the set of linear functions which describe the envelope, because this set of functions is orthonormal and allows one to control the 'order of detail' which increases with the spherical harmonic order. Mathematical methods employing spherical harmonics are well developed. Its use in small-angle scattering has been introduced by Stuhrmann (1970), for which extensive use of the orthonormal properties of spherical harmonics can be made because the scattering pattern results from an average over all possible orientations of the envelope. In our application, where the orientation of the macromolecule is not time averaged, no specific simplification occurs from the orthonormal properties of the spherical harmonics during the computation of structure factors.

7.2.2. Expression for $G_U(\mathbf{h})$ and its derivatives for a single molecule. An expression for $G_U(\mathbf{h})$ and its derivatives for a single macromolecule can be derived in spherical coordinates. The Fourier integrals are estimated numerically taking into account the peculiar structure of the indicator function $\chi_U(\mathbf{r})$. The volume integral $G_U(\mathbf{h})$ of a single macromolecule is transformed into a surface integral to minimize the number of integration points and to improve precision,

$$\begin{aligned} G_U(\mathbf{h}) &= \int_{R^3} \chi_U(\mathbf{r}) \exp(i2\pi\mathbf{h}\cdot\mathbf{r})\,\mathrm{d}^3\mathbf{r} = \int_V \exp(i2\pi\mathbf{h}\cdot\mathbf{r})\,\mathrm{d}^3\mathbf{r} \\ &= \frac{1}{i2\pi\mathbf{h}^2}\int_S \exp(i2\pi\mathbf{h}\cdot\mathbf{r})\mathbf{h}\cdot\hat{\mathbf{n}}\,\mathrm{d}S, \end{aligned}$$

where $\hat{\mathbf{n}}$ is the unit vector normal to the surface S and pointing toward the exterior of the volume enclosed by S. From $R(\theta, \varphi)$ one can easily deduce $\hat{\mathbf{n}}\,\mathrm{d}S$ in terms of polar coordinates (see Fig. 8*c* for definitions of $\hat{\mathbf{u}}$, $\hat{\mathbf{u}}_\theta$ and $\hat{\mathbf{u}}_\varphi$). Since

$$\mathrm{d}\mathbf{R} = \left(\frac{\partial R}{\partial\theta}\hat{\mathbf{u}} + R\hat{\mathbf{u}}_\theta\right)\mathrm{d}\theta + \left(\frac{\partial R}{\partial\varphi}\hat{\mathbf{u}} + R\sin\theta\hat{\mathbf{u}}_\varphi\right)\mathrm{d}\varphi,$$

one obtains

$$\begin{aligned} \hat{\mathbf{n}}\,\mathrm{d}S &= \left(\frac{\partial R}{\partial\theta}\hat{\mathbf{u}} + R\hat{\mathbf{u}}_\theta\right)\mathrm{d}\theta \times \left(\frac{\partial R}{\partial\varphi}\hat{\mathbf{u}} + R\sin\theta\hat{\mathbf{u}}_\varphi\right)\mathrm{d}\varphi \\ \hat{\mathbf{n}}\,\mathrm{d}S &= \left(\hat{\mathbf{u}} - \frac{1}{R}\frac{\partial R}{\partial\theta}\hat{\mathbf{u}}_\theta - \frac{1}{R\sin\theta}\frac{\partial R}{\partial\varphi}\hat{\mathbf{u}}_\varphi\right)R^2\sin\theta\,\mathrm{d}\theta\,\mathrm{d}\varphi. \end{aligned}$$

In order to refine the model, the derivatives of $G_U(\mathbf{h})$ should be computed with respect to the parameters a_k. From the expression of $G_U(\mathbf{h})$ in polar coordinates,

$$G_U(\mathbf{h}) = \int_0^\pi \sin\theta\,\mathrm{d}\theta \int_{-\pi}^{\pi} \mathrm{d}\varphi \int_0^{R(\theta,\varphi)} \exp(i2\pi\mathbf{h}\cdot\mathbf{r})r^2\,\mathrm{d}r,$$

one can deduce derivatives of $G_U(\mathbf{h})$ for the parameters a_k,

[3] In principle, Sayre's equation is valid for envelope structure factors because the square of the indicator function, equal to zero or unity, is virtually identical to itself.

[4] A complete mathematical treatment will be submitted for publication in the near future (Kahn, in preparation).

$$\frac{\partial G_U(\mathbf{h})}{\partial a_k} = \int_0^{\pi} \sin\theta \,\mathrm{d}\theta \int_{-\pi}^{\pi} \mathrm{d}\varphi \frac{\partial R}{\partial a_k} \frac{\partial}{\partial R} \int_0^{R(\theta,\varphi)} \exp(i2\pi\mathbf{h}\cdot\mathbf{r}) r^2 \,\mathrm{d}r$$

$$\frac{\partial G_U(\mathbf{h})}{\partial a_k} = \int_0^{\pi} \sin\theta \,\mathrm{d}\theta \int_{-\pi}^{\pi} f_k(\theta,\varphi) R^2(\theta,\varphi) \exp[i2\pi\mathbf{h}\cdot\mathbf{R}(\theta,\varphi)] \,\mathrm{d}\varphi.$$

$G_U(\mathbf{h})$ and its derivatives $[\partial G_U(\mathbf{h})/\partial a_k]$ are calculated by numerical integration using the same sample points (θ, φ) in angular space.

7.2.3. Application of crystallographic symmetry. One can compute $G_U(\mathbf{h})$ for the entire unit cell by defining $G_0(\mathbf{h})$ as the inverse Fourier transform of the indicator function $\chi_{U_0}(\mathbf{r})$ for a single molecule occupying the domain U_0 centred at C. The unit-cell origin O is taken as the origin for the calculation and we define $\overrightarrow{OC} = \mathbf{r}_C$, $\mathbf{u} = \overrightarrow{CM}$, $\overrightarrow{OM} = \mathbf{r}$ $(\mathbf{r} = \mathbf{r}_C + \mathbf{u})$. Then, $G_0(\mathbf{h})$ for a single molecule becomes

$$\begin{aligned} G_0(\mathbf{h}) &= \int_{U_0} \chi_U(\mathbf{r}_C + \mathbf{u}) \exp(i2\pi\mathbf{h}\cdot\mathbf{u}) \,\mathrm{d}^3\mathbf{u} \\ &= \int_{U_0} \exp(i2\pi\mathbf{h}\cdot\mathbf{u}) \,\mathrm{d}^3\mathbf{u}. \end{aligned}$$

The indicator function can be rotated and translated for any symmetry operation, g, of the space group such that $\chi_U(\mathbf{t}_g + \underline{\underline{R}}_g\mathbf{r}) = \chi_U(\mathbf{r})$ and then $G_U(\mathbf{h})$ for the entire unit cell develops to

$$\begin{aligned} G_U(\mathbf{h}) &= \int_V \chi_U(\mathbf{r}) \exp(i2\pi\mathbf{h}\cdot\mathbf{r}) \,\mathrm{d}^3\mathbf{r} \\ &= \sum_g \int_{V_0} \chi_U(\mathbf{r}) \exp[i2\pi\mathbf{h}\cdot(\mathbf{t}_g + \underline{\underline{R}}_g\mathbf{r})] \,\mathrm{d}^3\mathbf{r} \\ &= \sum_g \int_{U_0} \chi_U(\mathbf{r}_C + \mathbf{u}) \exp\{i2\pi\mathbf{h}\cdot[\mathbf{t}_g + \underline{\underline{R}}_g(\mathbf{r}_C + \mathbf{u})]\} \,\mathrm{d}^3\mathbf{u} \\ &= \sum_g \exp[i2\pi\mathbf{h}\cdot(\mathbf{t}_g + \underline{\underline{R}}_g\mathbf{r}_C)] \\ &\quad \times \int_{U_0} \chi_U(\mathbf{r}_C + \mathbf{u}) \exp(i2\pi\mathbf{h}\cdot\underline{\underline{R}}_g\mathbf{u}) \,\mathrm{d}^3\mathbf{u} \\ &= \sum_g \exp[i2\pi\mathbf{h}\cdot(\mathbf{t}_g + \underline{\underline{R}}_g\mathbf{r}_C)] \\ &\quad \times \int_{U_0} \chi_U(\mathbf{r}_C + \mathbf{u}) \exp(i2\pi\mathbf{u}\cdot\underline{\underline{R}}_g^T\mathbf{h}) \,\mathrm{d}^3\mathbf{u} \end{aligned}$$

and finally,

$$G_U(\mathbf{h}) = \sum_g \exp[i2\pi\mathbf{h}\cdot(\mathbf{t}_g + \underline{\underline{R}}_g\mathbf{r}_C)] G_0(\underline{\underline{R}}_g^T\mathbf{h}).$$

More simply put, the overall $G_U(\mathbf{h})$ for the entire unit cell is defined as the sum of the $G_0(\mathbf{h})$ for each single molecule in the unit cell after applying the corresponding rotations and translations of the crystallographic symmetries in reciprocal space. Non-crystallographic symmetry can also be applied in an analogous way.

7.3. The program *NVLOP*: refinement procedure and criteria

The program *NVLOP* (Kahn, in preparation) determines a parameterized envelope using spherical harmonics given a set of $|G_{\rm obs}(\mathbf{h})|$. A sphere is used as a starting model and the volume occupied by this sphere is considered to be known. The centre of the sphere is determined from a few random trials followed by a positional refinement minimizing the residual $R_{\rm ref}$,

$$R_{\rm ref} = \left\{ \frac{\sum_{\mathbf{h}} m(\mathbf{h}) [|G_{\rm obs}(\mathbf{h})| - K|G_{\rm cal}(\mathbf{h})|]^2}{\sum_{\mathbf{h}} m(\mathbf{h}) |G_{\rm obs}(\mathbf{h})|^2} \right\}^{1/2},$$

where $m(\mathbf{h})$ is the number of reflections equivalent to $\mathbf{h}$ by symmetry and K is a scale factor. After the initial sphere positioning, the clusters of these solutions are identified and the refinement using the linear combination of spherical harmonics is started. The refined parameters are the coefficients a_k, the coordinates of the centre C of one molecule and the scale factor K. The only constraint is the constant volume imposed onto the molecular envelope. Thus, during the refinement process, the component along the gradient of the volume of the vector describing the variations of the set of parameters is forced to vanish. Attempts to introduce a global thermal factor as a refined parameter were unsuccessful and led to abnormally high values for the thermal parameter coupled with very high values for the coefficients a_k of highest order.

The program requires the maximum order of the spherical harmonics describing the envelope to be set for the refinement. This order can be incremented in steps and the refinement can be restrained to the even-order spherical harmonics which are centrosymmetric. This is of particular importance, because the odd-order (non-centrosymmetric) spherical harmonics are strongly coupled to shifts in the centre of mass. Subsequently, eigenvalue filters have been incorporated and are very important when refining the odd-order spherical harmonics. Another attribute of *NVLOP* includes the possibility of introducing a starting model from a predefined set of parameters (the centre C, the scale factor K and the coefficients a_k). The refinement procedure is performed using a least-squares normal matrix. This procedure is also rather insensitive to missing data.

After the last cycle of refinement, the parameterized envelope in the unit cell might overlap with its symmetry-related counterparts. The overlapping grid points are set to unity and play the same role as non-overlapping pixels in the calculation of the final map (*i.e.* overlapping and non-overlapping pixels are equally weighted). To make the comparison of pre-existing maps easier, the program can output all of the maps using different possible origins.

7.3.1. Simulated data. A simulated set, $|G_{\rm sim}(\mathbf{h})|$, of $|G_{\rm obs}(\mathbf{h})|$ for hen egg-white tetragonal lysozyme has been generated to a resolution of 7.5 Å using the following procedure.

(i) A mask for a single molecule was generated from the PDB model 6lyt using the program *MAMA* (Kleywegt & Jones, 1999). Structural water molecules were included, but no attempt was made to rebuild the missing H atoms. The mask was built onto a fine grid (0.5 Å pitch) using an atom radius of 2 Å. A complete cycle of expansion, filling, contraction and removal of islands was applied.

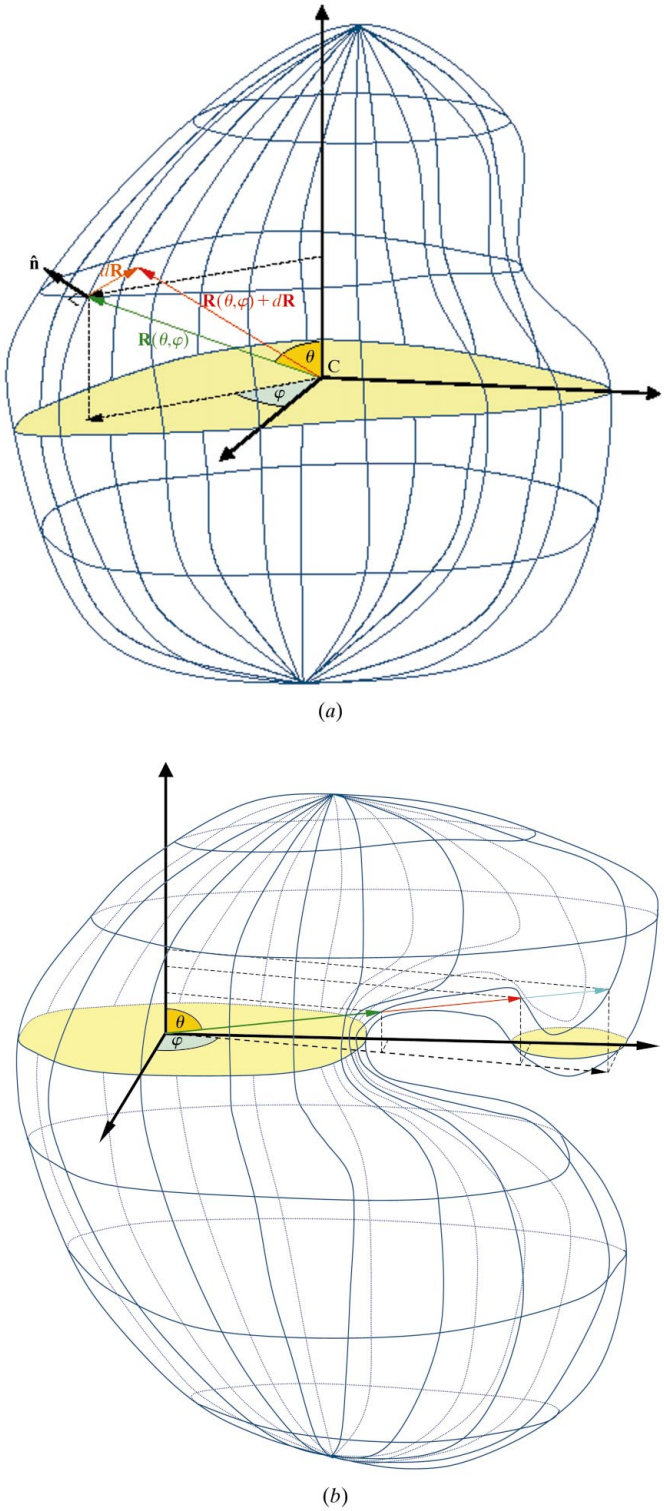
(ii) The mask for the whole unit cell was generated by setting the grid points inside the protein to unity. Overlapping

grid points were reset to unity (0.5% of total). The total fraction of grid points inside the mask is 67.3%.

(iii) The set of 202 unique $G_U(\mathbf{h})$ extending to a resolution of 7.5 Å was generated by Fourier transform of the indicator function at rest, $\chi_U^0(\mathbf{r})$, and an exponential correction using a Debye–Waller factor, B, of 50 Å^2 was applied to the set of $G_U(\mathbf{h})$ giving the final set of $G_{\rm sim}(\mathbf{h})$.

7.3.2. Initial sphere positioning and envelope refinement. The best results of the envelope-determination process were obtained using a volume significantly higher than the actual value (75% instead of 67.3%). 200 random trials of initial sphere positioning were undertaken. (It should be remembered that the initial sphere positioning fixes the hand and the origin of the final solution.) These trials converged into 11 different clusters of solutions defined as those solutions which are separated by no more than 0.005 in their fractional coordinates or by ~0.4 Å. This was found to be the coarsest spacing possible before neighbouring spherical solutions would refine into non-identical spherical harmonic solutions. The resulting refined coordinates X_c, Y_c, Z_c of the sphere centre and the corresponding value for $R_{\rm ref}$ as well as the number $N_{\rm freq}$ of trials corresponding to equivalent solutions are reported in Table 2. Each of these 11 cluster solutions have been used as a starting model for subsequent envelope refinement. Spherical harmonics up to the order l = 4 (29 refined parameters) and up to l = 5 (40 refined parameters) were successively tested. These refinements converged to 11 different solutions. Results for these solutions are summarized in Tables 3 and 4. Sections z = 0 of the indicator functions for these 11 solutions and the corresponding section for the model are shown in Fig. 9.

7.3.3. General comments on the phasing method. Several solutions with a high proportion of grid points correctly set have been found. Nevertheless, these results address the problem of selecting the correct solutions from a few indica-



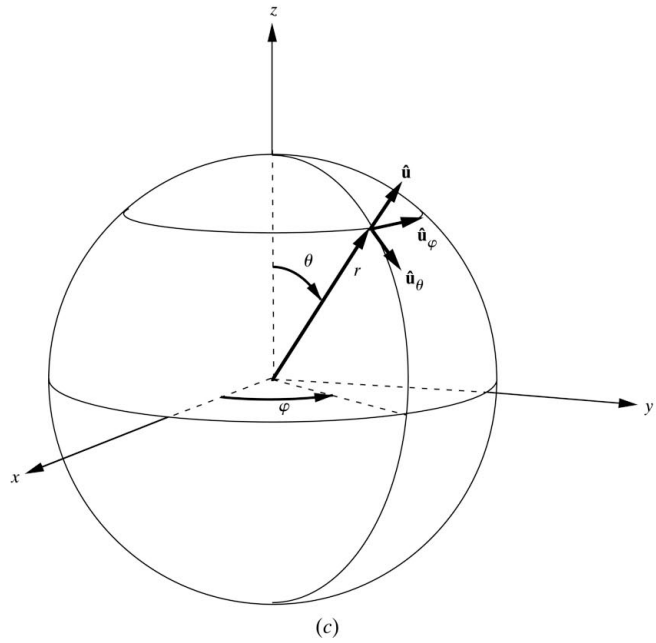


Figure 8
(*a*) Description of the surface of the domain U_0 using spherical coordinates (θ, φ). (*b*) An example of a domain for which no origin can be found to describe the surface by the function $R(\theta, \varphi)$ which is limited to a single value of r for every (θ, φ). In this case, three values of r are necessary to describe the surface at the (θ, φ) illustrated. Note that the origin may be located anywhere inside the domain. As a consequence, certain concave features, depending upon their severity, cannot be represented using spherical coordinates, *e.g.* those features approaching the shape of a 'donut hole'. (*c*) Graphical definitions of $\hat{\mathbf{u}}$, $\hat{\mathbf{u}}_\theta$ and $\hat{\mathbf{u}}_\varphi$.

Table 2
Results of random trials of initial sphere positioning.

Solution number	X_c	Y_c	Z_c	R_{ref}	N_{freq}
1	0.7650	0.0376	0.0572	0.6021	71
2	0.7348	0.0461	0.0555	0.6101	34
3	0.7215	0.0228	0.0685	0.6360	24
4	0.2819	0.0179	0.0840	0.6597	12
5	0.2700	0.0295	0.0768	0.6652	9
6	0.2196	0.0153	0.0841	0.6688	18
7	0.2208	0.0124	0.0781	0.6693	6
8	0.2303	0.0251	0.0836	0.6794	5
9	0.6938	0.1336	0.0446	0.7246	12
10	0.6871	0.1317	0.0370	0.7256	7
11	0.8972	0.1807	0.1166	0.8689	2

tors. First of all, it should be noted that the correct solution corresponds to the lowest residual R_{ref} before overlap removal using spherical harmonics up to the order $l = 4$ and its residual is ranked second lowest after removal of the overlaps. In both refinements using spherical harmonics of orders $l = 4$ and $l = 5$, solutions 9–11 can be eliminated because their number of overlapping grid points are too high and consequently their residuals after removal are high. Similarly, solution 6 can be eliminated according to its high residual after overlap removal. When compared with the known model, the remaining seven solutions refined using spherical harmonics of orders $l = 5$ have more than 70% of grid points correctly set.

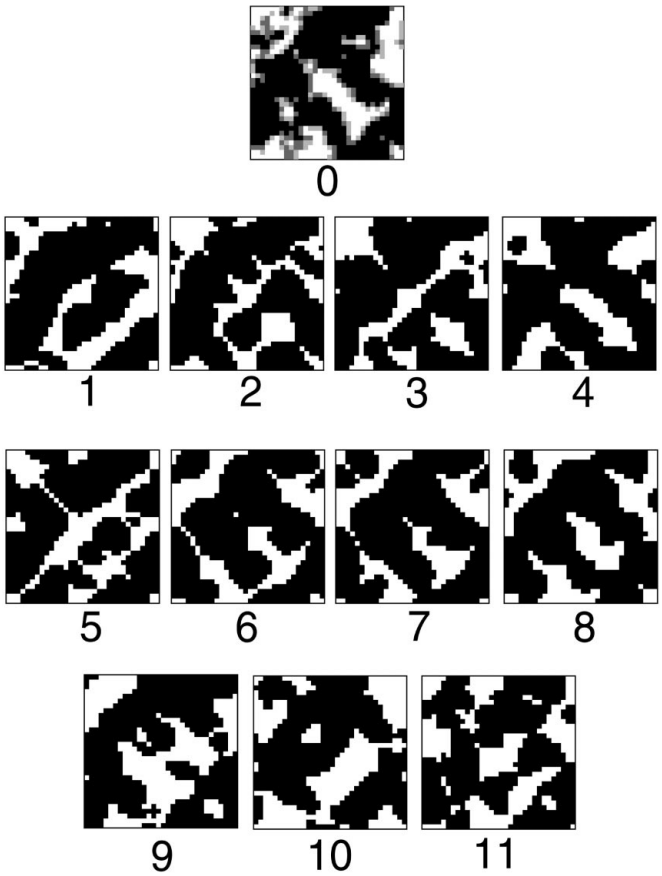


Figure 9
Section $z = 0$ ($0 \leq x < 1$; $0 \leq y < 1$) of the 11 different maps found using spherical harmonics up to the order $l = 5$ compared with the equivalent section (noted 0) of the model. For each map, the allowed origin of the space group that gave the highest ratio of grid points correctly set has been chosen as the origin. Solutions 9–11, as well as solution 5, can be rejected from their high residuals after overlap removal (see text).

When higher orders of spherical harmonics are employed (*i.e.* above $l = 6$, the equivalent of 53 refinable parameters), the refinement diverges. This gives an idea of the practical upper limit of the number of spherical harmonic orders which can be effectively refined, at least for this crystal form of lysozyme. In theory, the upper limit of the number of spherical harmonic orders would be restricted by the number of independent observations. For the above case of lysozyme with 202 reflections at 7.5 Å resolution, this leads to an upper limit of $l = 14$ for a 1:1 ratio of parameters to observations.

The preliminary results of this phasing method are very encouraging. However, these results need to be confirmed on simulated data from other macromolecules (globular proteins as well as DNA, RNA *etc.*) and on experimentally obtained envelope structure-factor moduli. This latter point raises the concern of the effect of ordered sites of the anomalous scatterer, since such sites begin to 'pollute' the experimental moduli of envelope structure factors in the higher resolution shells and where these values are generally weak.

The program gives a fair estimation of the absolute scale factor, without any prior knowledge except for the solvent content of the crystal. In addition, non-crystallographic symmetry can be easily implemented for cases containing several molecules per asymmetric unit, but the method is not adapted to cases involving mixtures or complexes of more than one macromolecule (*e.g.* heterodimers), unless the entire complex was included into a single envelope.

Improvements can be made in both the method and algorithm, most notably in the treatment of the overlapping regions between symmetry-related envelopes. A particular cause for concern is the loss in discrimination between refined solutions after the overlapping regions have been forced to unity. Work on a more robust minimization algorithm will be needed, especially when one considers that the effect of ordered sites of the refinement is as yet uncertain and may prove detrimental to resolving the envelope of the macromolecule. The difficulties associated with possible ordered sites of the anomalous scatterer could in principle be resolved by implementing a maximum-entropy refinement procedure using a low-order spherical harmonic envelope as a starting point.

8. Conclusions and perspectives

There are three principal advantages to generating contrast variation in the MASC method: (i) the contrast is applied externally to a single-crystal ensuring isomorphism, (ii) the f'' scattering factor supplies extra phase information owing to the breakdown of Friedel's law and (iii) it uses the intense synchrotron-radiation X-ray sources which are much brighter than neutron sources. However, the disadvantages are the effects arising from absorption and fluorescence, and the weaker contrast between data sets compared with H/D isotope

Table 3
Statistics for the refinement of the HEWL molecular envelope in tetragonal crystals.

Spherical harmonics up to the order $l = 4$ have been used. The starting points for the refinement are the refined centres derived from the initial sphere positioning. The fraction of grid points correctly set according to the model is calculated using the origin shift allowed by the space group that gives the highest correlation between both maps.

Solution number	R_{ref} before overlap removal	Absolute scale factor before overlap removal	Fractional volume of overlapping grid points	Fractional volume after overlap removal	R_{ref} after overlap removal	Absolute scale factor after overlap removal	Fraction of grid points correctly set
1	0.304	0.719	0.038	0.712	0.381	0.803	0.731
2	0.324	0.709	0.032	0.718	0.399	0.786	0.735
3	0.317	0.698	0.040	0.710	0.400	0.778	0.707
4	0.289	0.665	0.057	0.693	0.383	0.768	0.755
5	0.337	0.719	0.042	0.708	0.418	0.795	0.696
6	0.318	0.720	0.032	0.718	0.395	0.783	0.742
7	0.315	0.705	0.031	0.719	0.392	0.783	0.736
8	0.319	0.689	0.051	0.699	0.394	0.770	0.756
9	0.378	0.547	0.093	0.657	0.443	0.710	0.662
10	0.358	0.561	0.090	0.660	0.457	0.712	0.646
11	0.322	0.597	0.079	0.671	0.415	0.751	0.687

Table 4
Statistics for the refinement of the HEWL molecular envelop in tetragonal crystals.

Spherical harmonics up to the order $l = 5$ have been used. The starting points for the refinement are taken from the refinement using spherical harmonics up to the order $l = 4$.

Solution number	R_{ref} before overlap removal	Absolute scale factor before overlap removal	Fractional volume of overlapping grid points	Fractional volume after overlap removal	R_{ref} after overlap removal	Absolute scale factor after overlap removal	Fraction of grid points correctly set
1	0.275	0.723	0.041	0.709	0.357	0.820	0.703
2	0.282	0.718	0.030	0.720	0.374	0.799	0.744
3	0.280	0.732	0.033	0.717	0.345	0.808	0.704
4	0.269	0.666	0.054	0.696	0.388	0.767	0.758
5	0.307	0.739	0.043	0.707	0.409	0.801	0.685
6	0.291	0.736	0.031	0.719	0.367	0.802	0.736
7	0.289	0.712	0.031	0.719	0.389	0.787	0.744
8	0.261	0.723	0.057	0.693	0.379	0.813	0.735
9	0.284	0.573	0.092	0.658	0.426	0.713	0.630
10	0.285	0.567	0.105	0.645	0.429	0.717	0.620
11	0.294	0.607	0.089	0.661	0.413	0.752	0.693

exchange. The risk of anomalous scatterers binding to ordered sites will contribute to the anomalous signal and contaminate the extraction of the structure-factor moduli of the envelope in the intermediate and higher resolution shells. This is where the disordered solvent effects are mixed in with the ordered solvent effects and thus MASC signals are combined with MAD signals.

In principle, ordered anomalous sites could help to phase the protein structure to higher resolution, as has been shown recently by Dauter *et al.* (2000). The attempts to directly phase with the ordered anomalous sites found in the anomalous difference Patterson maps of HEWL and P64k test cases were hindered by the moderate resolution of the data (3.9–4.1 Å) because of the long crystal-to-detector distances required. Ideally, the multiple-wavelength data should be collected in two runs, one to collect the low-resolution data (MASC signal) and another to collect high-resolution data (MAD signal). Such a strategy is complicated by absorption and fluorescence effects washing out the typically weak reflections at high resolution and the risk of generating many low-occupancy sites which would render the determination of their location difficult either from Patterson maps or *via* direct methods by using programs such as *Shake-and-Bake* (Miller & Weeks, 1998) and *SHELXD* (Sheldrick, 1998). However, if a few well occupied sites could be generated and found then other sites may be localized using difference-map techniques.

The presence of ordered anomalous scatterers suggests that the biphasic model of the unit cell in a macromolecular crystal needs to be revised. In fact, the macromolecule and solvent domains are not identical to the ordered and disordered domains in the unit cell. Since all macromolecular crystals contain solvent shells of ordered molecules and crystals often contain disordered segments of the macromolecule (*e.g.* termini), an intermediate zone exists between the macromolecule and the bulk solvent where the electron density shows some order but is still in exchange with the bulk solvent. In a MASC experiment, the anomalous partial structure may not necessarily be restricted to the disordered zone of the bulk solvent, although it can permeate into this intermediate zone. Indeed, the anomalous scatterers may not be evenly distributed throughout the bulk-solvent zone. One would expect ions to be absent near a macromolecular surface which has a like charge. How an anomalous scatterer is distributed about the crystal will depend on its physical chemistry character (charge, size *etc.*), the bulk-solvent medium and the surface of the macromolecule crystallized.

We have demonstrated that the envelope moduli can be phased by treating the envelope as a compact domain of known volume and then expressing it as a linear combination of a set of functions, such as spherical harmonics. This allows the macromolecular envelope to be expressed as a small number of parameters, given that there are relatively few independent observations. The preliminary results presented here are encouraging, but further research is necessary to

optimize this phasing step and other methods may also prove to be successful (*e.g.* Andersson & Hovmöller, 1996; Badger, 1996; Harris, 1995; Lunin *et al.*, 1995; Urzhumtsev *et al.*, 1996, Subbiah, 1991, 1993). As such, it turns the MASC method into a complete technique of envelope determination.

References

Andersson, K. M. & Hovmöller, S. (1996). *Acta Cryst.* D**52**, 1174–1180.

Badger, J. (1996). *Basic Life Sci.* **64**, 333–343.

Bentley, G. A., Lewit-Bentley, A., Finch, J. T., Podjarny, A. D. & Roth, M. (1984). *J. Mol. Biol.* **176**, 55–75.

Bragg, W. L. & Perutz, M. F. (1952). *Acta Cryst.* **5**, 277–289.

Bricogne, G. (1993). *Acta Cryst.* D**49**, 37–60.

Carter, C. W. Jr & Bricogne, G. (1987). *GFROMF: A Computer Program for Scaling and Estimating Envelope Structure Factors from Contrast Variation Data.* Department of Biochemistry, CB#7260, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599–7260, USA.

Carter, C. W., Crumley, K. V., Coleman, D. E., Hage, F. & Bricogne, G. (1990). *Acta Cryst.* A**46**, 57–68.

Crumley, K. V. (1989). MSc thesis. University of North Carolina, Chapel Hill, USA.

Dauter, Z. & Dauter, M. (1999). *J. Mol. Biol.* **289**, 93–101.

Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Acta Cryst.* D**56**, 232–237.

Dumas, C. (1988). PhD thesis. Université de Paris-Sud, Orsay, France.

Evans, G., Roversi, P. & Bricogne, G. (2000). *Acta Cryst.* D**56**, 1304–1311.

Fourme, R., Shepard, W., Kahn, R., L'Hermite, G. & Li de La Sierra, I. (1995). *J. Synchrotron Rad.* **2**, 36–48.

Girard, E., Kahn, R. & Vicat, J. (1999). *IUCr XVIIth Congress and General Assembly Abstracts*, p. 254. Abstract P12.02.019.

Harris, G. W. (1995). *Acta Cryst.* D**51**, 695–702.

Harrison, S. C. (1969). *J. Mol. Biol.* **42**, 457–483.

Hendrickson, W. A. (1985). *Trans. Am. Crystallogr. Assoc.* **21**, 11–21.

Hütsch, M. (1993). PhD thesis. University of Hamburg, Germany.

Ibel, K. & Stuhrmann, H. B. (1975). *J. Mol. Biol.* **93**, 255–265.

Jack, A., Harrison, S. C. & Crowther, R. A. (1975). *J. Mol. Biol.* **97**, 163–172.

Junemann, R., Burkhardt, N., Wadzack, J., Schmitt, M., Willumeit, R., Stuhrmann, H. B. & Nierhaus, K. H. (1998). *Biol. Chem.* **379**, 807–818.

Kahn, R., Fourme, R., Bosshard, R. & Saintagne, V. (1986). *Nucl. Instrum. Methods A*, **246**, 596.

Karle, J. (1980). *Int. J. Quant. Chem. Quant. Biol. Symp.* **7**, 357–367.

Kleywegt, G. J. & Jones, T. A. (1999). *Acta Cryst.* D**55**, 941–944.

Kortright, J. B. (1986). *Center for X-ray Optics X-ray Data Booklet*, edited by D. Vaughan, pp. 2.19–2.20. Lawrence Berkeley Laboratory, University of California, Berkeley, California.

La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.

Li de la Sierra, I., Pernot, L., Prangé, T., Saludjian, P., Schiltz, M., Fourme, R. & Padron, G. (1997). *J. Mol. Biol.* **269**, 129–141.

Li de la Sierra, I., Prangé, T., Fourme, R., Padron, G., Fuentes, P., Musacchio, A. & Madrazo, J. (1994). *J. Mol. Biol.* **235**, 1154–1155.

Lunin, V. Y., Lunina, N. L., Petrova, T. E., Vernoslova, E. A., Urzhumtsev, A. G. & Podjarny, A. D. (1995). *Acta Cryst.* D**51**, 896–903.

Miller, R. & Weeks, C. M. (1998). *Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier, pp. 389–400. Dordrecht: Kluwer Academic Publishers.

Moras, D., Lorber, B., Romby, P., Ebel, J.-P., Gig, R., Lewit-Bentley, A. & Roth, M. (1983). *J. Biol. Struct. Dyn.* **1**, 209–223.

Pebay-Peyroula, E., Garavito, R. M., Rosenbusch, J. P., Zulauf, M. & Timmins, P. A. (1995). *Structure*, **3**, 1051–1059.

Penel, S., Pebay-Peyroula, E., Rosenbusch, J., Rummel, G., Schirmer, T. & Timmins, P. A. (1998). *Biochimie*, **80**, 543–551.

Pignol, D., Hermoso, J., Kerfelec, B., Crenon, I., Chapus, C. & Fontecilla-Camps, J. C. (1998). *Chem. Phys. Lipids*, **93**, 123–129.

Podjarny, A., Bhat, T. N. & Zwick, M. (1987). *Annu. Rev. Biophys. Biophys. Chem.* **16**, 351–374.

Porod, G. (1951). *Kolloid. Z.* **124**, 83–114.

Ramin, M. (1999). PhD thesis. Université de Paris-Sud, Orsay, France.

Ramin, M., Shepard, W., Fourme, R. & Kahn, R. (1999). *Acta Cryst.* D**55**, 157–167.

Rey, F., Jenkins, J., Janin, J., Lasters, I., Alard, P., Matthyssens, G. & Wodak, S. (1988). *Proteins Struct. Funct. Genet.* **4**, 165–172.

Roth, M. (1991). *Crystallographic Computing 5: From Chemistry to Biology*, edited by D. Moras, A. D. Podjarny & J. C. Thierry, pp. 229–248. Oxford University Press.

Roth, M., Arnoux, B., Ducruix, A. & Reiss-Husson, F. (1991). *Biochemistry*, **30**, 9403–9413.

Roth, M., Lewit-Bentley, A. & Bentley, G. A. (1984). *J. Appl. Cryst.* **17**, 77–84.

Sayre, D. (1952). *Acta Cryst.* **5**, 60–65.

Sheldrick, G. M. (1998). *Direct Methods for Solving Macromolecular Structures*, edited by S. Fortier, pp. 401–411. Dordrecht: Kluwer Academic Publishers.

Stuhrmann, H. B. (1970). *Z. Phys. Chem.* **72**, 177–198.

Stuhrmann, H. B. (1980). *Acta Cryst.* A**36**, 996–1001.

Stuhrmann, H. B. & Kirste, R. (1965). *Z. Phys. Chem.* **46**, 247–250.

Stuhrmann, H. B. & Nierhaus, K. H. (1996). *Basic Life Sci.* **64**, 397–413.

Subbiah, S. (1991). *Science*, **252**, 128–133.

Subbiah, S. (1993). *Acta Cryst.* D**49**, 108–119.

Timmins, P. A. & Pebay-Peyroula, E. (1996). *Basic Life Sci.* **64**, 267–272.

Timmins, P. A., Poliks, B. & Banaszak, L. (1992). *Science*, **257**, 652–655.

Urzhumtsev, A. G., Vernoslova, E. A & Podjarny, A. D. (1996). *Acta Cryst.* D**52**, 1092–1097.

Williams, C. E., May, R. P. & Guinier, A. (1994). *Materials Science and Technology: A Comprehensive Treatment – Characterization of Materials Part II*, Vol. 2B, edited by E. Lifshin, pp. 611–656. Weinheim: VCH.

Wilson, A. J. C. (1942). *Nature (London)*, **150**, 151–152.

Zhao, J., Wang, J., Chen, D. J., Peterson, S. R. & Trewhella, J. (1999). *Biochemistry*, **38**, 2152–2159.