Experimental methods for measuring accurate high-amplitude phases and their importance in isomorphous replacement experiments

Alexei S. Soares[a]* and Yanina Vekhter[b]

[a]Biology Department, Brookhaven National Laboratory, Upton, NY 11973, USA, and [b]Laboratory for the Structure of Matter, Naval Research Laboratory, Washington DC 20375-5000, USA

Correspondence e-mail: soares@bnl.gov

Conventional experimental phasing methods are most accurate for moderate-resolution reflections, with progressively greater ambiguity in the phases of reflections away from this optimal point. Frequently, very strong (usually low-resolution) reflections are either poorly phased or altogether unrecorded. While the spatial frequency of these reflections is predominantly too low to dramatically affect the calculated electron density at an atomic level, they have a dominant impact on the determination of the large-scale distribution of matter in the unit cell. Consequently, while these few strong reflections play only a peripheral role in the latter stages of a structure-determination project, they are crucial to the success of initial phasing and model-building efforts. Here, the pivotal importance of a limited number of strong/low-resolution reflection phases is shown and a procedure to derive these phases is described. The improvement in map correlation coefficients after density modification of a marginal 'starting' MAD data set (obtained from two Zn atoms at special positions in rhombohedral insulin crystals) was compared with the improvement in map correlation coefficients observed after density modification of an 'expanded' data set obtained by combining a limited number of highly accurate phases measured using three-beam diffraction with the 'starting' MAD data. It is concluded that a small number of high-amplitude/low-resolution reflections contribute disproportionately to generating an initial structure and it is suggested that a small number of triplet phases could be measured quickly and combined with experimental isomorphous replacement phases in order to move stubborn structures for novel proteins down the structure-solution pathway.

1. Introduction

Conventional MAD and MIR phasing rely on an explicit atomic model of a few strongly scattering atoms. The contribution of these atoms to low-resolution reflections is dwarfed by the dominant contribution of disordered protein and solvent (Badger & Caspar, 1991; Podjarny & Urzhumtsev, 1997). Consequently, experimental phases are often optimal at moderate resolution and little or no phase information is available for reflections at the high- and low-resolution limits of diffraction. The impact of improved high-resolution phase information on both initial model building (Schneider & Sheldrick, 2002; Cowtan & Main, 1993) and subsequent refinement (Burling *et al.*, 1996) is well documented and numerous methods have been proposed to enhance the quality of experimental high-angle phasing power (Banumathi *et al.*, 2003; Brodersen *et al.*, 2000). In this work, we examine the importance of accurately phased high-amplitude/low-resolution reflections and propose a novel method for

obtaining experimental high-amplitude phases with high accuracy.

Strong low-resolution reflections are pivotal in establishing the large-scale distribution of matter in a crystal and play a disproportionate role in solvent flattening and other density-modification procedures (Vekhter, 2005). The rhombohedral insulin MAD phases used here had a starting mean phase error of 63.7° and most of the phases with amplitude over 1000 e$^-$ were either absent or had very low figures of merit (36% had FOM < 0.1, ~30% were unmeasured). To accurately measure the missing high-amplitude/low-resolution phases, three-beam interference methods were used to record triplet phases with an experimental uncertainty under 12° (Soares et al., 2003)[1]. These triplet phases were then used to derive highly accurate individual phases for selected high-amplitude reflections (described in §3.2). Rhombohedral insulin was selected for this study because the anomalous signal from two Zn atoms located on a threefold crystallographic axis is very small (0.66 Zn atoms in 102 residues, which is one heavy atom per 153 residues) and the resulting MAD data set was marginal, meaning that if the structure were unknown at the outset, a successful structure determination would be in doubt. It should be noted that this weak anomalous zinc signal has been successfully used to determine a SAD structure for insulin in a recent demonstration paper focusing on problem structures (Dauter et al., 2002), but considerable efforts involving novel methods were needed to supplement the poor data.

2. Three-beam phasing

Experimental triplet phases for a protein structure were first published in 1991 from crystals of myoglobin (Hümmer et al., 1991) and hemoglobin (Chang et al., 1991) using techniques for measuring triplets one at a time that were first developed with small molecules (Hümmer et al., 1990; Weckert & Hümmer, 1990). A theoretical treatment of the three-beam diffraction method is necessarily lengthy (Weckert & Hümmer, 1997) and only a brief description is given here.

Three-beam interference profiles are measured by maintaining one reciprocal-space node in reflecting position and bringing a second node into and then out of reflecting position. The node that is maintained on the sphere of reflection is called the primary reflection **h** and the node that is moved through the sphere of reflection is called the secondary reflection **g**. Maximum interference occurs when the incident beam optimally excites both **h** and **g**. This is called a three-beam case (see Fig. 1). In a three-beam experiment, during a Ψ-scan around the strong reflection **h**, a second strong reflection **g** is excited simultaneously by rotating it through the

sphere of reflection. The primary reflection **h** with phase $\varphi(\mathbf{h})$ interferes with the wave that is diffracted twice, by the reciprocal-lattice plane **g** and by the reciprocal lattice plane **h** − **g**, which has phase $\varphi(\mathbf{h}) + \varphi(\mathbf{h} - \mathbf{g})$. This double-diffracted wave is called the umweg or detour wave.

As illustrated in Fig. 1, the geometry of diffraction directs the umweg wave to propagate in the same direction as the strong **h** reflection. That is, the secondary reflection **g** is diffracted into the direction of the primary reflection **h** by (**h** − **g**). Consequently, **h** and **g** interfere with each other via ±(**h** − **g**). The interference between the strong **h** wave and the weaker double diffracted umweg wave results in a change in the measured intensity of the **h** wave while the **g** lattice node traverses the sphere of reflection. The interference between the two waves depends on the difference between their phases $\Phi = \varphi(\mathbf{g}) + \varphi(\mathbf{h} - \mathbf{g}) - \varphi(\mathbf{h})$. The primary wave **h** with the
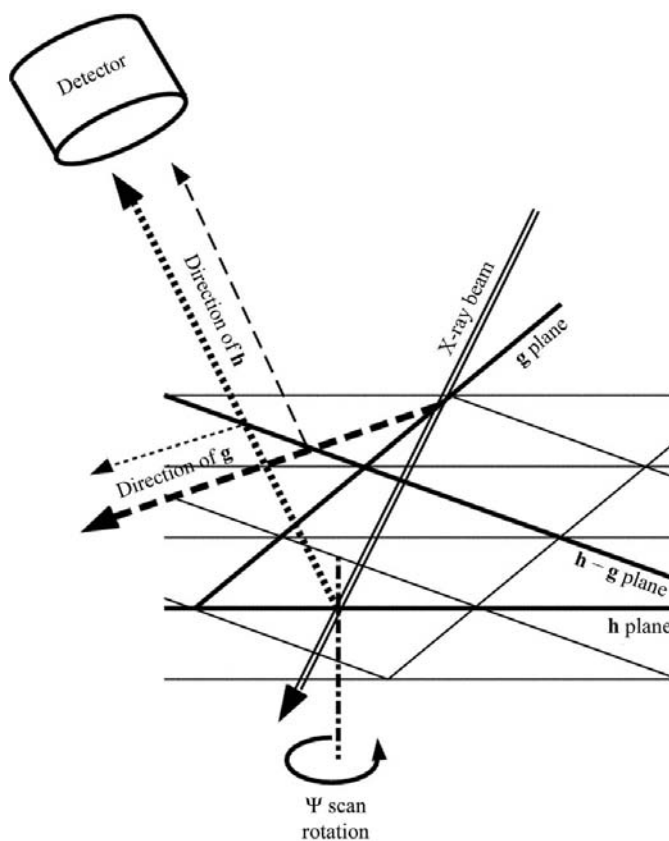

Figure 1
Graphical representation of a three-beam case. The lattice planes of **g**, of **h** and of **h** − **g** are shown as bold lines and the incident beam is shown as double lines. The primary wave **h** is shown as a dotted line, as is the weaker umweg wave, which is the second reflection of the **h** wave from the (**h** − **g**) planes in the **g** diffracted-beam direction. The secondary wave **g** is shown as a dashed line, as is the weaker umweg wave in the **h** direction. Strong waves are shown in bold. The experimentally measured quantity is the intensity in the direction of **h** as **g** is brought in and out of reflecting condition by a sample rotation called a Ψ scan. The rotation axis of the Ψ scan is illustrated with a dot–dashed line. The umweg wave in the direction of the **h** diffracted beam will interfere with the primary wave **h** (the fine dashed line interferes with the bold dotted line). The observed intensity change will depend on the relative amplitudes and the difference between the phase of the umweg wave and the phase of **h**: $\varphi(\mathbf{g}) + \varphi(\mathbf{h} - \mathbf{g}) - \varphi(\mathbf{h})$.

---

[1] The phasing uncertainty of 12° was measured using triplet phase data from cubic insulin. When **h**, **g** and **h** − **g** were all semi-invariant with only two possible phase values, the triplet phase also had only two possible phase values and consequently could be known exactly. The interference data from these known triplets was input to computer phasing software that did not enforce the centric phase restrictions and the mean error of the measurement was then readily identified.

Table 1
Number of individual phases obtained using different methods.

Individual phases can be obtained from the triplet data set plus the MAD data (§3.2, method 1, labeled 'MAD/triplet' in the table) or from triplet data and a starting phase set (§3.2, method 2, labeled 'Triplet only' in the table). Each method was used three times. Each time the MAD/triplet phases (method 1) expanded the size of the starting phase set, which was then used to generate additional phases by elementary algebra (method 2). The table shows the number of phases obtained using each method in each phasing cycle (column 2), the average phase residual against model calculated phases (column 3), the cumulative total number of individual phases from all methods (column 4) and the cumulative phase residual (column 5). The figure of merit (FOM) assigned to each phase before density modification is also shown (column 6). At the bottom of the table, the number of MAD phases and their phase residual against the model are shown. The average residual between measured and calculated phases for the third cycle of phasing was higher than for the MAD data and we conclude that additional cycles would be unfruitful.

	No. of phases in this set	Phase residual for this set	Cumulative No. of phases	Cumulative phase residual	FOM assigned to this set
MAD/triplet	59	17.61	59	17.61	1.00
Triplet only	331	22.78	390	22.00	1.00
MAD/triplet	220	36.77	610	27.33	0.70
Triplet only	162	29.31	772	27.74	0.70
MAD/triplet	18	74.90	790	28.82	0.40
Triplet only	38	62.04	828	30.34	0.40
MAD FOM > 0.20	5152	62.25			
MAD data	5383	63.67			

phase $\varphi(\mathbf{h})$ interferes with the *umweg* wave with phase $\varphi(\mathbf{h}) + \varphi(\mathbf{h} - \mathbf{g})$. The intensity of reflection $\mathbf{h}$ changes owing to the interference between the two waves. There is a unique correspondence between every possible phase difference and the shape of the interference profile. Consequently, the phase difference can be obtained from the interference profile

$$\Phi = \varphi(\mathbf{g}) + \varphi(\mathbf{h} - \mathbf{g}) - \varphi(\mathbf{h}). \qquad (1)$$
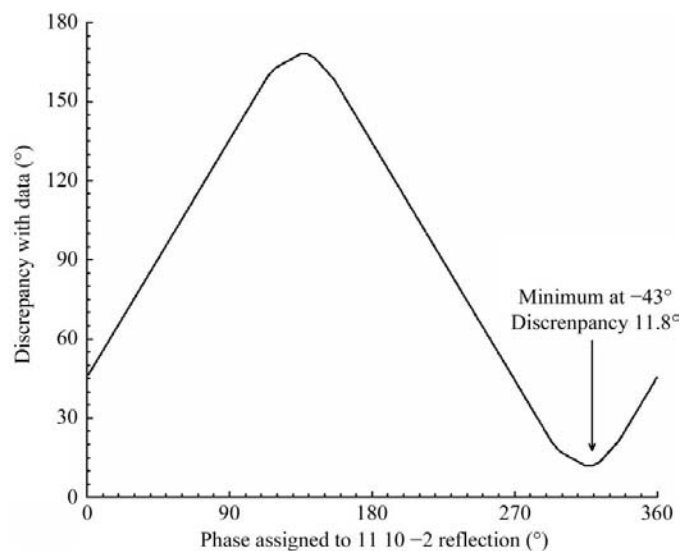
Figure 2
Combining triplet and MAD data to build a starting phase set. The average discrepancy between measured triplet phases and MAD data-calculated triplet phases is shown as a function of the phase assigned to the 11 10 −2 reflection. Six triplets were measured that contained both the 11 10 −2 and two additional reflections of known phase (phases for $\mathbf{h}$ and $\mathbf{h} - \mathbf{g}$ were known from the MAD data). An estimated phase of −43° was assigned to the 11 10 −2 reflection, since this value agreed best with the six triplets which contained this reflection (indicated by the arrow in the figure). The discrepancy between this estimated phase and the −26.5° phase calculated from the model is 16.5°, very close to the 17.6° average phase residual for the 'starting set' of 59 individual phases that were obtained using this procedure (see Table 1).

3. Methods

MAD phase information was obtained from rhombohedral insulin crystals using routine methods as detailed in §3.1. This MAD data set was later improved using conventional density modification. To explore the benefits of assigning a small number of accurate high-amplitude/low-resolution phases in the initial MAD data set, 865 triplet phases were measured from similar crystals. Individual phases can be derived from triplet phase data using simple algebra if a starting set of known phases is available. To obtain such a starting set of phases, the MAD data were used in combination with the triplet phase set in a novel way described as method 1 in §3.2. Method 2 in §3.2 describes how this starting phase set was used to bootstrap a phasing tree from the remaining triplet phases. In all, 828 individual phases were obtained from the 865 measured triplet phases (297 with the help of MAD phases). Although the effort involved in measuring many hundreds of triplet phases is nontrivial, we found that the majority of the refinement benefits resulted from just 100 strong reflections. Under optimal conditions, such a limited data set could be obtained in 2 d with conventional three-beam methods. Using new reference-beam diffraction methods (Shen & Wang, 2003), such data should be routinely measurable in a matter of hours.

3.1. Rhombohedral insulin crystals and MAD data

Triplet phases and MAD data were recorded from insulin crystals in space group $R3$, with unit-cell parameters $a = b = 82.5$, $c = 34.1$ Å, $\alpha = \beta = 90$, $\gamma = 120°$. The crystals were grown by dissolving 0.025 g pig insulin in crystallizing solution at 313 K (200 ml 0.02 M HCl, 100 ml 0.20 M sodium citrate, 60 ml acetone, 20 ml H_2O and 20 ml 0.12 M $ZnSO_4$) and allowing the solution to cool slowly to 293 K. There are two insulin monomers per asymmetric unit (chains A and B form the first monomer and chains C and D form the second quasi-symmetric monomer). Room-temperature triplet phase data were collected at NSLS beamline X26C from rhombohedral insulin crystals with mosaic spreads of ∼0.006°, volumes of

research papers

Table 2
A limited number of accurate high-amplitude phases strongly improves map correlation coefficients for starting and density-modified maps.

Map correlation coefficients are shown before ('Initial map CC') and after ('Final map CC') density modification against conventional MAD data only ('MAD data only') and against MAD data combined with up to 700 individual phases obtained from triplet phase data ('MAD +' data sets). The number of triplet-derived phases added to the MAD data is shown in each case. The improvement in the final maps over MAD data only observed for maps with triplet phase data is shown in column four ('Gain over MAD only'). The incremental benefit from the addition of each 100 added phases is in column five ('Incremental gain'). The total gain from all 700 added phases is 0.186, more than half of which arises from the first 100 added phases. Approximately 2 d of beam time are needed to measure 100 triplet phases.

Data set	Initial map CC	Final map CC	Gain over MAD only	Incremental gain
MAD only	0.297	0.374	—	—
MAD + 100	0.394	0.468	0.094	0.094
MAD + 200	0.442	0.498	0.124	0.030
MAD + 300	0.470	0.519	0.145	0.021
MAD + 400	0.491	0.527	0.153	0.008
MAD + 500	0.507	0.540	0.166	0.013
MAD + 600	0.517	0.548	0.174	0.008
MAD + 700	0.522	0.560	0.186	0.012

$\sim$0.5 mm^3 and $\sim$24 h survival in the unfocused beam. Cryogenic temperatures could not be used to increase survival because freezing drastically increased the mosaic spread. MAD data were collected from cryocooled but otherwise identical crystals using a standard protein crystallography configuration at NSLS beamline X12B.

3.2. Triplet phasing of high-amplitude/low-resolution reflections

Although experimental triplet phases accurately relate three reflections, further processing is needed to distill individual phases from a triplet-phase pool. Numerous methods have been developed to do this, all of which involve identifying the phases of two of the three reflections that make up the triplet (*e.g.* Shen & Wang, 2003). The phase of the third reflection can then be identified with elementary algebra. In this work, there were two independent sources of phase information for the two needed unknown phases.

Method 1. Triplet phases are most easily measured when the secondary reflection **g** is stronger than the primary reflection **h**, thus maximizing the signal of **g** *via* **h** − **g** that interferes with **h**. In practice, this means that the majority of triplet phases have **g** drawn from a limited pool of very strong/low-resolution reflections. Consequently, these strong reflections are measured with high redundancy in triplets where the two remaining elements are weaker reflections of moderate resolution that are often accurately phased and are assigned high 'figures of merit' in the MAD experiment. The best value for the phase of the strong **g** reflections optimized the aggregate agreement with the triplet phase and MAD data (see Fig. 2).

Method 2. Once a small ensemble of starting phases is known with high accuracy, additional individual phases can be obtained from cases where two of the three reflections in a measured triplet are in the known set. Each new identified phase then expands the pool of starting phases in an iterative

cycle. The number of individual phases obtained from each method in successive cycles is shown in Table 1, along with the mean phase errors for each group of identified phases. The first phase we identified in this way was for the 3 29 4 reflection, from the measured triplet equal to $0°$, where the phase of **g** was known to be $-40°$ and the phase of **h** − **g** was known to be $-12°$,

$$\varphi(\mathbf{g}|6\ 9\ 0| = -40°) + \varphi(\mathbf{h} - \mathbf{g}| - 3\ 20\ 4| = -12°)$$
$$- \varphi(\mathbf{h}|3\ 29\ 4|) = 0°$$
$$\Rightarrow \varphi(\mathbf{h}) = -(0° + 40° + 12°) = -52°.$$

The phase of **h** is consequently $-52°$, a discrepancy of $19°$ relative to the model phase of $-71°$.
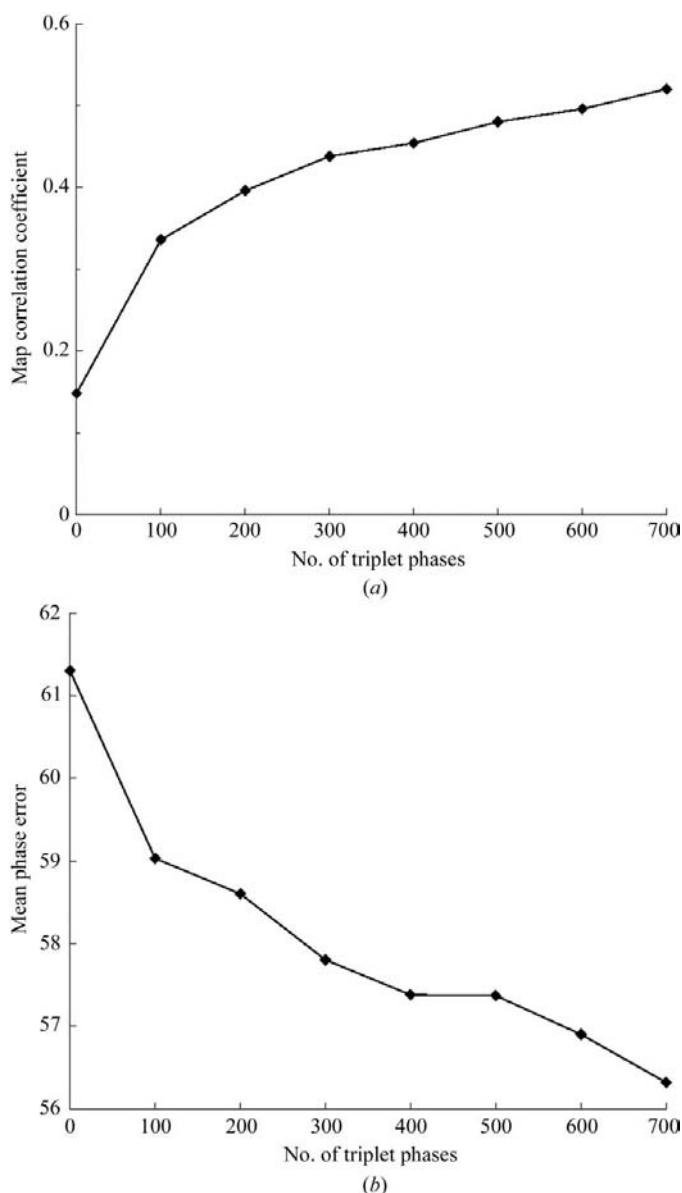

Figure 3
Improvement in map correlation coefficients and in the mean phase errors after density modification with 0, 100, 200 . . . 700 accurate phases assigned to the strongest reflections.
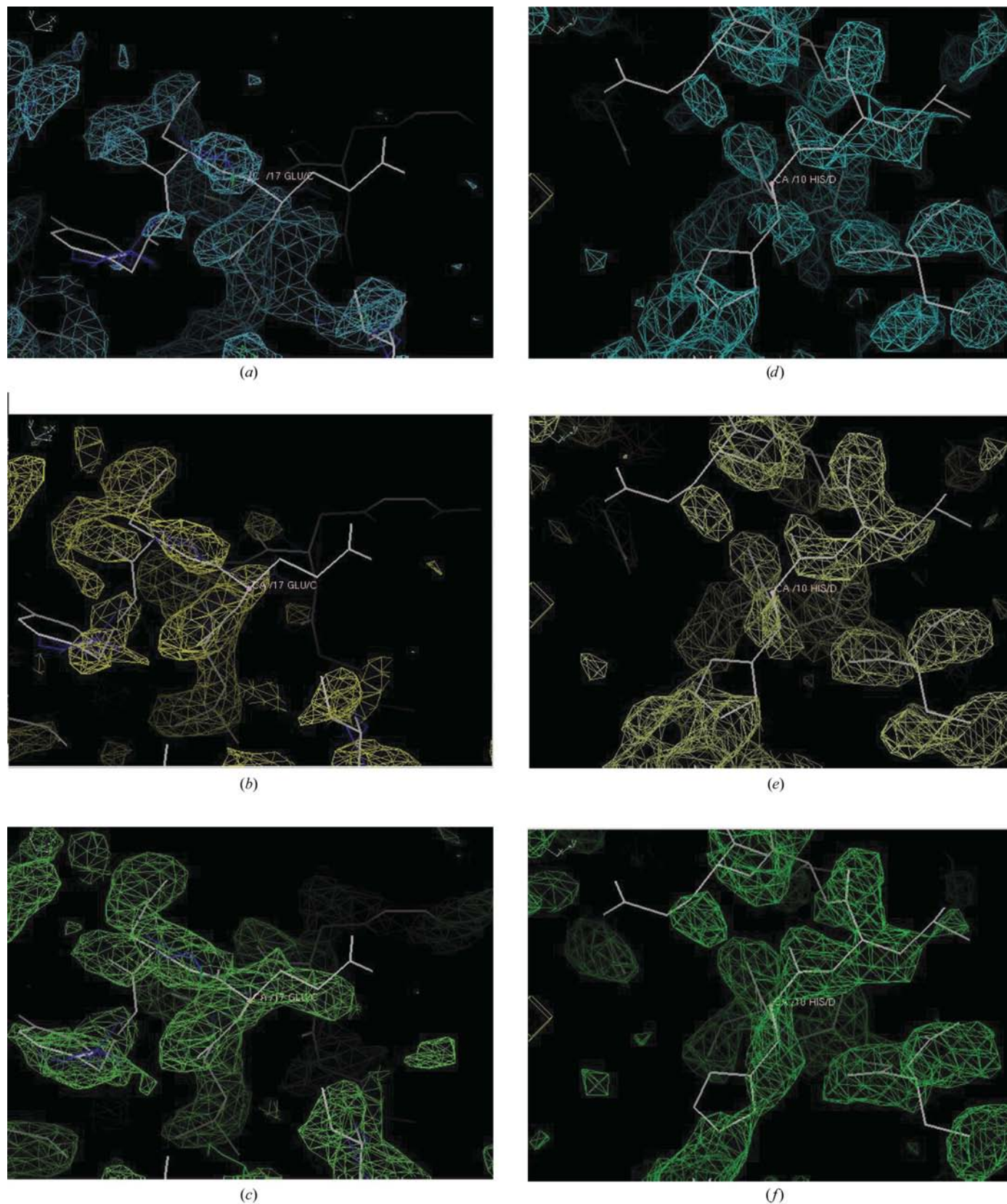
Figure 4
Strong reflections disproportionately impact electron density. Two regions of experimental electron density are shown, one from Glu17-Asn18-Tyr19 of chain *A* (*a*, *b* and *c*) and one from Ser9-His10 of chain *B* (*d*, *e* and *f*). The electron density is superposed on the known atomic model, subsequent to density modification. In each region, density is shown for the MAD only data set (*a* and *d*), for the MAD data set combined with 100 triplet derived phases (*b* and *e*) and for the MAD data set combined with 700 triplet derived phases (*c* and *f*). Although the number of added phases is small compared with the total number of reflections in the data set, the large amplitudes have a disproportionate impact on the overall electron density.

In our first phasing effort, 59 reflections were phased using method 1 and 331 reflections were then phased by method 2 using these 59 as a starting set. In total, 297 reflections were phased with method 1 and 531 with method 2, although the phase ambiguity grew quickly since efforts subsequent to the first involved triplets that are more difficult to measure accurately (see Table 1). Phasing was only attempted for reflections that either had a very low FOM (less than 0.2) or were altogether missing from the MAD set. Of the 754 possible reflections with a resolution of less than 4 Å, 530 (70%) could not be phased using the MAD data. Of these missing reflections, 488 (65%) were phased using three-beam methods, so that only 5% of the total possible reflections were unavailable from any source.

Although the specific methods described here have not been previously reported, numerous alternative means exist to distill individual phase information from triplet phase data (Weckert *et al.*, 1999) and these have been used to examine the properties of electron-density maps (Hölzer *et al.*, 2000).

3.3. Density modification and refinement

Map correlation coefficients and mean phase errors were evaluated for the structures obtained before and after density modification against MAD data only and against MAD data with both a small (100 phases) and a large (700 phases) set of accurate individual phases from the three-beam data (Table 2) combined with the main set of MAD phases. The MAD phases were obtained using conventional methods with *SHELX* and *SOLVE* phasing software. Density modification was performed with the package *DM* (*CCP*4 program suite). The small 100-phase data set was selected from the full set of available triplet-derived phases by applying an amplitude cutoff. An equivalent SAD solution was not possible because *SHELX* was unable to locate the heavy atoms using only one wavelength, which is an unsurprising result given the limited quality of the data. The signal-to-noise ratio was very low and the correlation between anomalous differences at different wavelengths was never higher than 0.60. Although the anomalous differences were weak, they were persistent and some signal was evident up to 2.0 Å. The density-modification strategy for initial map improvement outlined above is typical of steps taken by many crystallographers in the early stages of a structure-determination effort.

we compared the impact of a small data set (sufficient to derive 100 phases) to the impact of larger data sets (Table 2). We found that a substantial benefit could be obtained from a data set small enough to be measured in 2 d of synchrotron access.

The map correlation coefficient subsequent to density modification against the MAD data combined with all 700 triplet-derived phases was 0.56, compared with 0.37 for the MAD data only (see Fig. 3). The overall 18.6% improvement corresponds to a 2.7% improvement per 100 measured phases. However, measuring enough triplet phases to extract 700 individual phases might be prohibitive for many applications. However, the 100 strongest phases are responsible for the majority (51%) of the observed improvement. The improvement in map quality was also evident from visual inspection of the electron density (see Fig. 4). Maps with better initial contrast can be obtained from exceptional insulin crystals, but we found that the data from an average-quality crystal better illustrated the potential impact that a limited number of strong reflections can have on a marginal data set.

Our first efforts with three-beam phasing yielded only 0.25 triplet phases per hour of beam access. After resolving a number of experimental problems, this rate increased to two triplet phases per hour, which we believe to be the maximum sustainable rate using the methods outlined here. However, this rate of data acquisition is not constant. The first few hundred triplets can be measured much more rapidly, while subsequent triplets increasingly require extended exposure times to build acceptable signal to noise. However, phasing trees grow more efficient and easier to construct as the pool of available triplets increases. On balance, the rate at which individual phases can be obtained from the triplet data is fairly constant. We estimate that 2 d of beam time would be needed to record 100 individual phases from suitable crystals of a novel macromolecule. Such an investment is not unreasonable in cases where the extra data are sufficient for structure determination. 2 d also compares favorably with the time required to obtain better crystals or to identify new heavy-atom derivatives. Recent developments with area detector-based three-beam phasing methods also promise to greatly accelerate data acquisition ('reference-beam diffraction'; see Shen & Wang, 2003).

4. Results and discussion

Three-beam diffraction places stringent demands on both the specimen and the experimental setup. Although crystals need not diffract to very high resolution (our best data came from 2.0 Å samples), they must be tolerant to X-ray damage at room temperature and they must have low mosaic spreads (if the mosaicity drawn from conventional data-reduction software is close to 0.2°, three-beam phasing has a good chance of working). Experimentally, the endstation must be capable of unfocused operation so that X-rays incident on the sample are highly parallel. With these experimental difficulties in mind,

References

Badger, J. & Caspar, D. L. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 622–626.

Banumathi, S., Dauter, M. & Dauter, Z. (2003). *Acta Cryst.* D**59**, 492–498.

Brodersen, D. E., de La Fortelle, E., Vonrhein, C., Bricogne, G., Nyborg, J. & Kjeldgaard, M. (2000). *Acta Cryst.* D**56**, 431–441.

Burling, F. T., Weis, W. I., Flaherty, K. M. & Brünger, A. T. (1996). *Science*, **271**, 72–77.

Chang, S. L., King, H. E. Jr, Huang, M. T. & Gao, Y. (1991). *Phys. Rev. Lett.* **67**, 3113–3116.

Cowtan, K. D. & Main, P. (1993). *Acta Cryst.* D**49**, 148–157.

Dauter, Z., Dauter, M. & Dodson, E. (2002). *Acta Cryst.* D**58**, 494–506.

Hölzer, K., Weckert, E. & Schroer, K. (2000). *Acta Cryst.* D**56**, 322–327.

Hümmer, K., Schwegle, W. & Weckert, E. (1991). *Acta Cryst.* A**47**, 60–62.

Hümmer, K., Weckert, E. & Bondza, H. (1990). *Acta Cryst.* A**46**, 393–402.

Podjarny, A. D. & Urzhumtsev, A. G. (1997). *Methods Enzymol.* **276**, 641–658.

Schneider, T. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.

Shen, Q. & Wang, J. (2003). *Acta Cryst.* D**59**, 809–814.

Soares, A., Caspar, D. L. D., Weckert, E., Héroux, A., Hölzer, K., Schroer, K., Zellner, J., Schneider, D., Nolan, W. & Sweet, R. (2003). *Acta Cryst.* D**59**, 1716–1724.

Vekhter, Y. (2005). *Acta Cryst.* D**61**, 899–902.

Weckert, E., Hölzer, K., Schroer, K., Zellner, J. & Hümmer, K. (1999). *Acta Cryst.* D**55**, 1320–1328.

Weckert, E. & Hümmer, K. (1990). *Acta Cryst.* A**46**, 387–393.

Weckert, E. & Hümmer, K. (1997). *Acta Cryst.* A**53**, 108–143.