Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Multi-crystal anomalous diffraction for low-resolution macromolecular phasing

Qun Liu,[a] Zhen Zhang[b] and Wayne A. Hendrickson[a,b,c]*

[a]New York Structural Biology Center, NSLS X4, Building 725, Brookhaven National Laboratory, Upton, NY 11973, USA, [b]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA, and [c]Department of Physiology and Cellular Biophysics, Columbia University, New York, NY 10032, USA

Correspondence e-mail: wayne@convex.hhmi.columbia.edu

Received 13 September 2010
Accepted 10 November 2010

Multiwavelength anomalous diffraction (MAD) and single-wavelength anomalous diffraction (SAD) are the two most commonly used methods for *de novo* determination of macromolecular structures. Both methods rely on the accurate extraction of anomalous signals; however, because of factors such as poor intrinsic order, radiation damage, inadequate anomalous scatterers, poor diffraction quality and other noise-causing factors, the anomalous signal from a single crystal is not always good enough for structure solution. In this study, procedures for extracting more accurate anomalous signals by merging data from multiple crystals are devised and tested. SAD phasing tests were made with a relatively large (1456 ordered residues) poorly diffracting ($d_{\min}$ = 3.5 Å) selenomethionyl protein (20 Se). It is quantified that the anomalous signal, success in substructure determination and accuracy of phases and electron-density maps all improve with an increase in the number of crystals used in merging. Structure solutions are possible when no single crystal can support structural analysis. It is proposed that such multi-crystal strategies may be broadly useful when only weak anomalous signals are available.

1. Introduction

Multiple-wavelength anomalous diffraction (MAD) and single-wavelength anomalous diffraction (SAD) experiments use anomalous scattering effects to solve the 'phase problem' for the determination of macromolecular structures. Both MAD and SAD phasing can be treated as a two-step process (Hendrickson, 1991; Dauter *et al.*, 2002). The first step is to find the substructure of anomalous scatterer components from within the whole structure by interpretation of anomalous Patterson maps or by direct methods (Hendrickson & Ogata, 1997; Weeks & Miller, 1999; Burla *et al.*, 2003; Grosse-Kunstleve & Adams, 2003; Sheldrick, 2008). The substructure is then refined against the anomalous differences and is used for the calculation of initial phases for the whole structure. By collecting anomalous diffraction data at two or more wavelengths, ideally, the definitive phase angle can be determined algebraically by MAD (Karle, 1980; Hendrickson *et al.*, 1988; Hendrickson, 1991). In comparison, SAD uses anomalous data collected at one wavelength, resulting in an intrinsic phase-ambiguity problem that has to be resolved by other means, such as resolved anomalous phasing (Hendrickson & Teeter, 1981), density modification (Wang, 1985; Cowtan & Zhang, 1999), direct methods (Fan *et al.*, 1990; Hao, 2000) or averaging when noncrystallographic symmetry (NCS) is

doi:10.1107/S0907444910046573

available (Kleywegt & Read, 1997). With advances in techniques at synchrotron sources, CCD detectors, cryogenic preservation and methods of data reduction and analysis, MAD and SAD phasing have evolved substantially over the past twenty years (Dauter *et al.*, 1999; Hendrickson, 1999; Ealick, 2000; Dauter, 2006*a*). The routine incorporation of anomalous scatterers by the expression of selenomethionine proteins (Hendrickson *et al.*, 1990) has dramatically boosted the application of anomalous diffraction in macromolecular phasing to an unprecedented level.

MAD and SAD phasing techniques greatly rely on the accurate measurement of anomalous diffraction differences, which are typically very weak when compared with normal diffraction from the whole structure. The anomalous signals from macromolecular crystal structures are affected by many factors. The most important factors are the ratio of the Bijvoet differences to normal diffraction (Bijvoet difference ratio) and the quality of diffraction (Hendrickson & Teeter, 1981). It is generally considered that higher Bijvoet difference ratios and higher resolution data significantly benefit a successful MAD/SAD phasing experiment. In an ideal case where anomalous signals can be measured very accurately, an anomalous signal with a Bijvoet difference ratio of as low as 0.53% can be utilized for successful structure determination (Ramagopal *et al.*, 2003).

For anomalous phasing, MAD is, in principle, superior to SAD as it allows the direct calculation of phase angles and has no phase-ambiguity issue (Hendrickson, 1991). This is only true if all of the data can be measured accurately, however. Radiation-induced irreversible damage, even with cryogenic preservation, can severely restrict the collection of usable data at multiple wavelengths for structural studies (Rice *et al.*, 2000). Although this complication is sometimes attributed to the brilliance of modern synchrotrons and may play a role in specific radiation damage (Leiros *et al.*, 2006), it appears that radiation damage is independent of dose rate even for the most powerful of undulator beams (Sliz *et al.*, 2003). Rather, failed MAD experiments often result from the extension of measurements beyond the radiation limit of the samples (Rice *et al.*, 2000; Nave & Garman, 2005). For large macromolecular complexes and membrane proteins, which are often more delicate and X-ray sensitive, the damage will be even more serious (Carugo & Carugo, 2005). SAD can thus be an option to mitigate the effects of radiation damage by relying on fewer measurements and counting on phasing programs to resolve the phase-ambiguity problem. Alternatively, recent developments in microbeam crystallography may help to diminish radiation damage by collecting wedged data using a microbeam focused on different portions of a crystal (Clemons *et al.*, 2001; Li *et al.*, 2004; Moukhametzianov *et al.*, 2008; Sanishvili *et al.*, 2008).

MAD and SAD phasing techniques work most favorably for crystals that diffract X-rays well to fine Bragg spacings. Unlike normal diffraction, which decreases with scattering angle, anomalous diffraction arises from the interaction between X-ray photons and inner shell electrons and is not in itself dependent on scattering angle. However, both normal diffraction and anomalous diffraction are modulated by dynamic and static displacements of atoms in macromolecular structures, as reflected by the *B* factors, which systematically reduce the diffracted intensity as the scattering angle increases (Shen *et al.*, 2003). Such modulation adds to background noise and degrades the anomalous signal, especially at higher angles. For this reason, the limit of usable anomalous data must be truncated appropriately in order to obtain reliable anomalous signals. To measure weak anomalous signals accurately, in particular for crystals that only diffract to low angle, increasing the data redundancy through multiplicity of measurement has been proposed as an effective approach (Dauter & Adamiak, 2001; Debreczeni *et al.*, 2003; Wagner *et al.*, 2006). On the other hand, radiation damage can spoil this advantage by ‘killing’ crystals before high redundancy is reached.

To overcome radiation damage, as well as to increase data redundancy, the acquisition of anomalous diffraction data from multiple crystals might be a promising route (Clemons *et al.*, 2001). This is not a new idea; multiple crystals were used routinely in solving radiation-damage problems before the advent of cryogenic data collection. More recently, data from multiple naturally formed crystals have been used in solving a baculovirus polyhedra structure by the SIRAS technique (Ji *et al.*, 2010). In terms of MAD/SAD phasing, however, single-crystal data acquisition has been typical and may be advisable in order to avoid adverse effects on small signals from variability in frozen crystal lattices. This might be true for high-angle data; however, for cases in which diffraction is limited, *e.g.* minimal Bragg spacings of 3.5 Å, a certain amount of non-isomorphism among crystals might be tolerated. We hypothesize that anomalous signals would be more accurately measured as highly redundant multi-crystal data sets, very effectively at low scattering angles and perhaps also at higher angles if crystal variability is small.

In this study, we evaluated the possibility of using multi-crystal anomalous diffraction data for phasing at 3.5 Å resolution. We collected anomalous diffraction data from eight selenomethionine-substituted crystals of the 748-residue sensor domain of a histidine kinase receptor for which the structure has been determined and analyzed independently (Z. Zhang, Q. Liu & W. A. Hendrickson, unpublished work). The selenomethionyl protein crystals diffracted X-rays weakly to a Bragg spacing limited to approximately 3.5 Å. By merging multiple single-crystal data together, we show that multi-crystal anomalous diffraction data can dramatically increase the anomalous signal. This facilitated both Se-substructure determination and subsequent phasing. The multi-crystal strategy could be used effectively to solve some challenging macromolecular structures that would prove intractable using a single-crystal route.

2. Materials and methods

2.1. Crystallization and data collection

The genome of the human gut symbiont *Bacteroides thetaiotaomicron* encodes a large family of histidine kinase

Table 1
Data-collection and reduction statistics for single-crystal data.

Values in parentheses are for the highest resolution shell. The inverse-beam mode of data collection was used to collect all data sets except for data 2 and is distinguished by reporting the number of frames as $N + N$.

Crystal/data set	1	2	3	4	5	6	7	8
Unit-cell parameters								
a (Å)	88.24	88.11	88.18	87.94	88.26	88.16	88.04	88.21
c (Å)	434.08	433.22	433.60	432.85	433.19	432.99	432.24	433.44
No. of frames	160 + 160	180	120 + 120	120 + 120	130 + 130	120 + 120	120 + 120	180 + 90
Bragg spacing (Å)	40–3.5 (3.69–3.5)	40–3.5 (3.69–3.5)	40–3.8 (4.01–3.8)	40–4.0 (4.22–4.0)	40–3.7 (3.9–3.7)	40–3.2 (3.37–3.2)	40–3.5 (3.69–3.5)	40–3.9 (4.11–3.9)
Measurements	283745	161130	164197	142143	197564	222176	192753	173981
Multiplicity	12.4 (12.0)	7.7 (7.8)	9.6 (9.7)	10.1 (10.0)	11.0 (11.1)	8.1 (4.9)	10.0 (6.9)	11.6 (11.8)
Completeness (%)	99.9 (99.9)	93.4 (94.5)	96.1 (97.2)	94.3 (96.8)	93.6 (94.7)	92.9 (69.9)	87.4 (86.8)	91.8 (93.0)
$R_{\rm merge}$†	0.147 (0.400)	0.190 (0.394)	0.207 (0.510)	0.177 (0.286)	0.227 (0.432)	0.129 (0.338)	0.220 (0.454)	0.236 (0.438)
$I/\sigma(I)$‡	17.7 (6.5)	11.2 (5.5)	11.5 (4.8)	11.1 (4.8)	11.7 (7.0)	14.9 (4.2)	11.1 (5.1)	11.5 (6.8)
$\Delta F/\sigma(\Delta F)$§	1.60 (0.70)	1.22 (0.77)	1.17 (0.68)	1.27 (0.76)	1.07 (0.78)	0.97 (0.68)	1.02 (0.70)	1.11 (0.76)
Anomalous CC¶ (%)	65.3 (34.8)	42.2 (22.3)	34.1 (8.9)	44.2 (13.8)	29.2 (10.0)	24.7 (12.1)	32.8 (8.0)	34.7 (16.8)
Bragg limit for anomalous CC of 30%†† (Å)	5.5	7.2	7.6	7.1	7.4	8.2	7.7	7.4

† $R_{\rm merge} = \sum_{hkl}\sum_i |I_i(hkl) - \langle I(hkl)\rangle| / \sum_{hkl}\sum_i I_i(hkl)$, where $I_i(hkl)$ is the intensity of reflection hkl and $\langle I(hkl)\rangle$ is the average over i measurements of reflection hkl. ‡ $I/\sigma(I) = \langle\langle I(hkl)\rangle/\sigma[\langle I(hkl)\rangle]\rangle$, where $\langle I(hkl)\rangle$ is the weighted mean of all measurements for reflection hkl and $\sigma[\langle I(hkl)\rangle]$ is the standard deviation of the weighted mean. The values are as reported from *SCALA* as Mn(*I*/sd). § $\Delta F/\sigma(\Delta F)$ is the average anomalous signal from data truncated to $d_{\rm min}$ = 6 Å. The values were derived using *CCP*4 programs and are computed as $\langle|\Delta F|/\sigma(\Delta F)\rangle$, where $\Delta F = |F(h)| - |F(-h)|$. ¶ Anomalous correlation coefficient evaluated from data truncated to $d_{\rm min}$ = 6 Å. †† Estimated Bragg limit for anomalous CC greater than 30%.

receptors, one of which is BT4673. The sensor domains of these and related receptors, which are distinctive and exceptionally large, are classified as family 3 histidine kinase sensors. We identify the sensor domain (residues 29–776) of the HK3 receptor BT4673 as $\rm Bt_{4673}HK3_S$. The expression and purification of recombinant selenomethionyl (SeMet) $\rm Bt_{4673}HK3_S$ are the same as for native $\rm Bt_{4673}HK3_S$ (Z. Zhang, Q. Liu & W. A. Hendrickson, unpublished work). For the crystallization of SeMet $\rm Bt_{4673}HK3_S$, 1 µl of 10 mg ml^{-1} protein buffered in 10 m*M* Tris pH 8.0 was mixed with an equal volume of well solution consisting of 0.1 *M* sodium cacodylate pH 6.5, 5% PEG MME 2000 and 0.1 *M* Tacsimate (Hampton Research). At 291 K, crystals appeared in one week and grew to full size in 4–6 weeks. After optimization by microseeding, bipyramidal crystals were obtained with dimensions of about 30 × 30 × 60 µm. The crystals of the SeMet protein are much smaller than those obtained of the native protein, which is at least part of the reason for their inferior diffraction.

For cryocrystallography, crystals were first soaked in cryoprotectant composed of well solution supplemented with 25% glycerol and then flash-frozen in liquid nitrogen for data collection at 100 K. X-ray diffraction data were measured from the eight best diffracting crystals on National Synchrotron Light Source (NSLS) beamline X4A using an ADSC Q4R CCD detector. The wavelength used for data collection was at the peak of Se f'' (0.97912 Å). The inverse-beam mode of data collection was used (the crystal was rotated 180° every 5–10 frames to measure Friedel mates), except for data 2. To better resolve the reflections corresponding to the long axis of 433 Å, the crystals were aligned in the loop with the long axis roughly parallel to the rotational spindle axis. In addition, a long sample-to-detector distance (350 mm) and an oscillation of 0.5° were used to reduce overlap. The strategy function in *HKL*-2000 (Otwinowski & Minor, 1997) was used to reduce overlap as well as to maximize data completeness. Diffraction data sets were collected from a total of 20 crystals and eight of these were selected for analysis here.

2.2. Multi-crystal data reduction

Each single-crystal data set was indexed and integrated by *XDS* (Kabsch, 1993, 2010*a*,*b*) and corrected for absorption, decay and modulation as implemented in *XDS*. The *CCP*4 program *SCALA* (Collaborative Computational Project, Number 4, 1994; Evans, 2006) was used for data scaling and merging with secondary beam correction and rotational restraints for scale and *B* factors. The 'anomalous' option in *SCALA* was turned on to allow the separation of Friedel mates in the merged data. For scaling, Friedel mates were not treated separately. The diffraction data statistics for individual single-crystal data sets are listed in Table 1. To generate multi-crystal data, *SCALA* was run multiple times with different combinations of single-crystal data. We generated two sets of multi-crystal data: one including data 1 (data 1 to data 8) and the other excluding data 1 (data 2 to data 8). This was performed to test the impact of data 1, which proved to have appreciably stronger anomalous signal than the others. Diffraction data statistics for the multi-crystal mergings are presented in Table 2 (for data 1 to data 8) and Table 3 (for data 2 to data 8).

2.3. Model used for phase comparison

The crystals of SeMet $\rm Bt_{4673}HK3_S$ belonged to space group $P4_12_12$, with unit-cell parameters a = 88.16, c = 433.26 Å, and are essentially isomorphous to the $\rm Ta_6Br_{12}$ derivative of native $\rm Bt_{4673}HK3_S$ (a = 88.17, c = 432.96 Å), the structure of which was refined at 2.2 Å resolution based on SAD phasing from the tantalum-cluster derivative. This structure (PDB entry 3ott) provided us with an independent model at high resolution for use in phase and map comparisons with the

Table 2
Data-processing statistics for the merging of data sets data 1 to data 8.

Data 1to1 is the same as single-crystal data 1; the other data sets are merges of single-crystal data. For example, data 1to2 means the merging of single-crystal data 1 and data 2 and data 1to8 means the merging of all eight single-crystal data sets. Definitions are as given in the footnotes to Table 1.

Data set	1to1	1to2	1to3	1to4	1to5	1to6	1to7	1to8
Unit-cell parameters								
a (Å)	88.24	88.20	88.19	88.13	88.18	88.17	88.16	88.16
c (Å)	434.08	433.77	433.72	433.51	433.47	433.40	433.24	433.26
Resolution (Å)	40–3.5 (3.69–3.5)	40–3.5 (3.69–3.5)	40–3.5 (3.69–3.5)	40–3.5 (3.69–3.5)	40–3.5 (3.69–3.5)	40–3.5 (3.69–3.5)	40–3.5 (3.69–3.5)	40–3.5 (3.69–3.5)
Measurements	283745	444755	651332	848840	1046696	1243129	1435807	1675067
Multiplicity	12.4 (12.0)	19.5 (19.2)	28.5 (27.5)	37.2 (33.5)	45.9 (33.3)	54.5 (39.2)	63.0 (45.0)	73.5 (55.5)
Completeness (%)	99.9 (99.9)	99.9 (100.0)	99.9 (100.0)	99.9 (100.0)	99.9 (100.0)	99.9 (100.0)	99.9 (100.0)	99.9 (100.0)
$R_{\rm merge}$	0.147 (0.400)	0.246 (0.497)	0.275 (0.598)	0.300 (0.605)	0.293 (0.604)	0.292 (0.580)	0.290 (0.579)	0.306 (0.619)
$I/\sigma(I)$	17.7 (6.5)	17.9 (7.4)	17.3 (7.8)	17.3 (8.0)	18.1 (7.7)	19.5 (8.6)	22.2 (9.8)	23.9 (10.8)
$\Delta F/\sigma(\Delta F)$	1.60 (0.70)	1.60 (0.68)	1.49 (0.68)	1.47 (0.66)	1.48 (0.65)	1.48 (0.65)	1.65 (0.68)	1.78 (0.71)
Anomalous CC (%)	65.3 (34.8)	63.5 (34.7)	61.3 (37.8)	59.5 (39.2)	64.4 (34.0)	61.6 (33.6)	68.0 (43.3)	70.0 (39.7)

Table 3
Data-processing statistics for the merging of data sets data 2 to data 8.

Data 2to2 is the same as single-crystal data 2; definitions are as given in the footnotes to Table 1.

Data set	2to2	2to3	2to4	2to5	2to6	2to7	2to8
Unit-cell parameters							
a (Å)	88.11	88.15	88.07	88.16	88.16	88.14	88.15
c (Å)	433.22	433.44	433.22	433.27	433.21	433.04	433.11
Resolution (Å)	40–3.5 (3.69–3.5)	40–3.5 (3.69–3.5)	40–3.5 (3.69–3.5)	40–3.5 (3.69–3.5)	40–3.5 (3.69–3.5)	40–3.5 (3.69–3.5)	40–3.5 (3.69–3.5)
Measurements	161130	367779	565421	763307	959663	1152426	1391559
Multiplicity	7.7 (7.8)	16.1 (15.6)	24.8 (21.5)	33.5 (21.3)	42.1 (27.2)	50.6 (33.0)	61.1 (53.5)
Completeness (%)	93.4 (94.5)	99.9 (100.0)	99.9 (100.0)	99.9 (100.0)	99.9 (100.0)	99.9 (100.0)	99.9 (100.0)
$R_{\rm merge}$	0.190 (0.394)	0.292 (0.637)	0.303 (0.622)	0.295 (0.620)	0.282 (0.557)	0.278 (0.556)	0.296 (0.611)
$I/\sigma(I)$	11.2 (5.5)	10.9 (5.6)	14.0 (6.8)	16.0 (6.5)	18.5 (8.5)	20.8 (9.5)	22.6 (10.4)
$\Delta F/\sigma(\Delta F)$	1.22 (0.77)	1.02 (0.73)	1.32 (0.74)	1.42 (0.71)	1.40 (0.72)	1.52 (0.71)	1.68 (0.73)
Anomalous CC (%)	42.2 (18.7)	35.3 (16.6)	54.9 (32.2)	59.9 (28.1)	62.5 (34.5)	66.8 (36.0)	69.4 (44.4)

multi-crystal SeMet results at 3.5 Å resolution. We used it for the calculation of model phases for comparison with experimental phases.

2.4. Substructure determination and phasing

Selenium-substructure determinations were performed with the *SHELXD* program through *HKL2MAP* (Pape & Schneider, 2004; Sheldrick, 2008). A resolution cutoff at 6 Å and an $E_{\rm min}$ cutoff at 1.4 were used to find Se substructures with *SHELXD*. Trials were made for each data set and for various merged data sets. For each case, 1000 attempts were made to find the expected 20 Se sites.

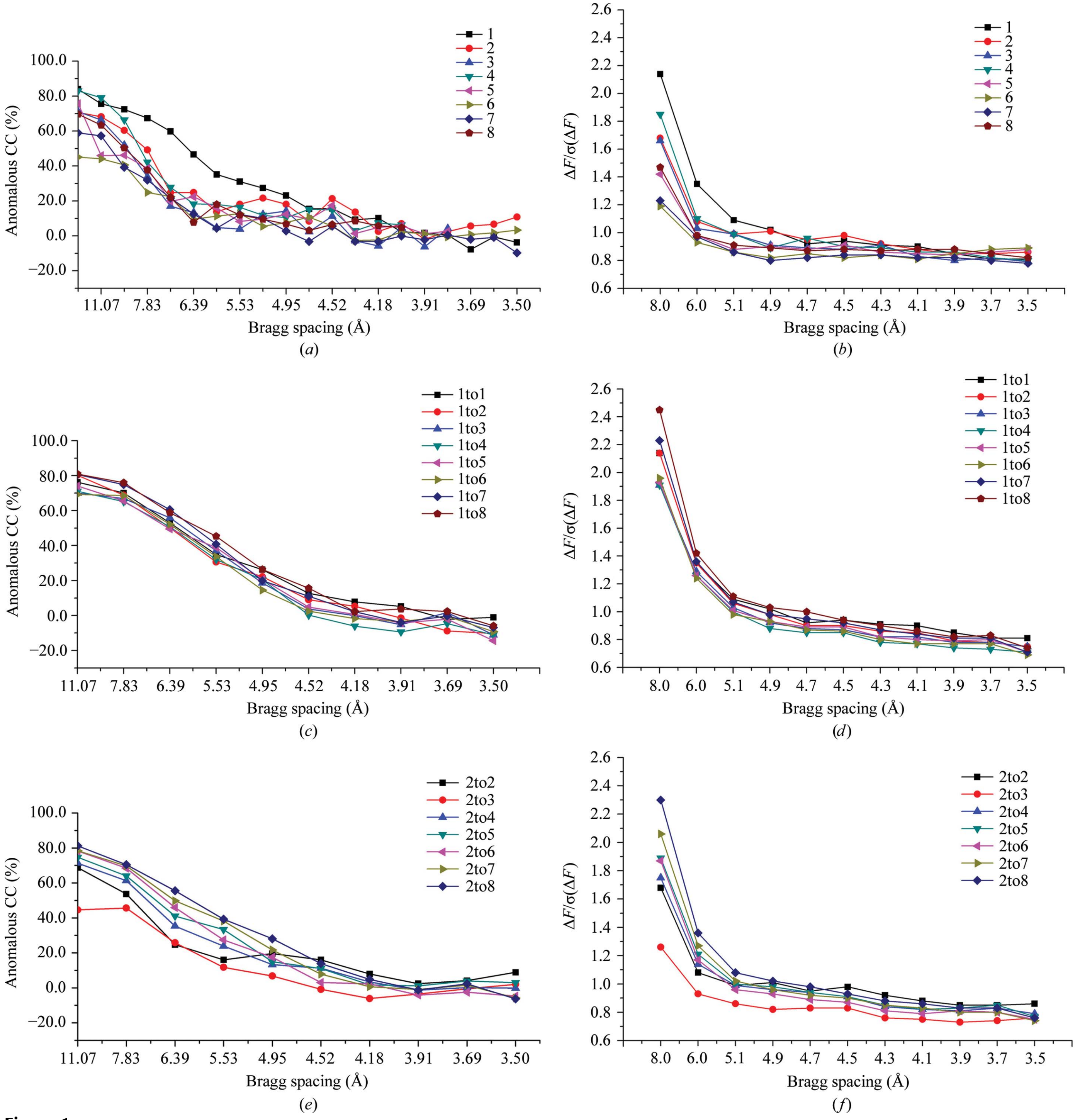
For those single-crystal and multi-crystal data sets that did not yield successful Se-substructure determinations using *SHELXD*, Se substructures were obtained by running *Phaser* (McCoy *et al.*, 2007) in its MR-SAD mode with phases from the model. The model was only used for Se-substructure determination and was excluded from the subsequent SAD phasing.

For all cases, initial SAD phases at the data limit (3.5 Å for multi-crystal data) were calculated by *Phaser*. These initial phases were subjected to automatic density modification with solvent flattening and histogram matching as implemented in the *CCP*4 program *DM* (Cowtan & Zhang, 1999). An estimated solvent content of 52% was used for the density-modification procedure involved in this study. Map correlation coefficients (map CCs) and mean phase errors were calculated to compare the resulting experimental phases with model phases derived from PDB coordinates.

3. Results

3.1. Anomalous signals

Observing reliable anomalous signals is a prerequisite for any anomalous phasing. Several anomalous indicators have been proposed for the evaluation of anomalous signals (Hendrickson *et al.*, 1988; Zwart, 2005; Dauter, 2006*b*). Of these, in this study we used the anomalous correlation coefficient (anomalous CC as reported by *SCALA*) and the anomalous difference signal [$\Delta F/\sigma(\Delta F)$ as reported by *HKL2MAP*]. The anomalous CC evaluates the correlation between Bijvoet differences in two randomly divided subsets from a data set (Schneider & Sheldrick, 2002). Typically, the anomalous CC and $\Delta F/\sigma(\Delta F)$ decrease with increase in scattering angle as the signal-to-noise ratio becomes worse. Figs. 1(*a*) and 1(*b*) show this fall-off in anomalous CC and $\Delta F/\sigma(\Delta F)$, respectively, for the eight single-crystal data sets. Data 1 has the strongest anomalous CC, with a gradual drop to ~30% at 5.5 Å Bragg spacing. The anomalous CCs for all other data are much weaker, with sharp drops at 7–9 Å Bragg spacing to only 18% or lower at 6 Å. At higher diffraction angles, increased noise makes the anomalous CC very unreliable, as shown by zero or negative anomalous CC values. Therefore, we limited the reported anomalous CC to 6 Å resolution data (Table 1). The data set with highest anomalous CC at 6 Å was identified as data 1; the other data sets are named at random. Consistent with the anomalous CC analysis, data 1 also has the highest $\Delta F/\sigma(\Delta F)$ up to 4.9 Å Bragg spacing. Taken together, Figs. 1(*a*) and 1(*b*) demonstrate the existence of anomalous signals in single-crystal data; however, the anomalous signals have substantial variations, with data 1 being superior. The Bragg limits for anomalous CC greater than 30% are reduced to 7.1–8.2 Å for the seven other data sets (Table 1).



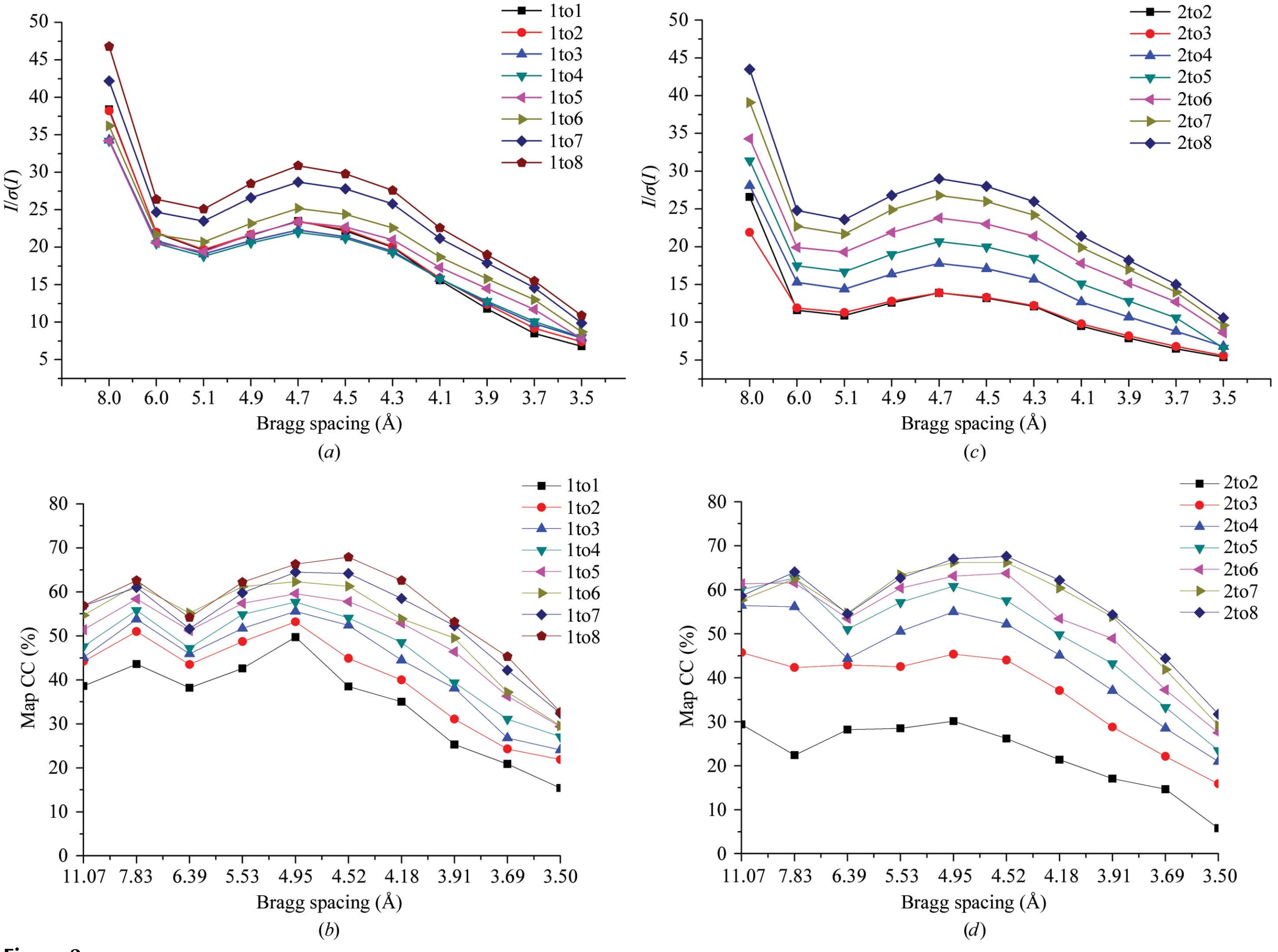
(*a*) (*b*) (*c*) (*d*) (*e*) (*f*)

Figure 1
Anomalous signals in single-crystal and multi-crystal data. (*a*, *c*, *e*) Anomalous correlation coefficients (anomalous CC) as a function of scattering-vector length, $|S| = 2\sin\theta/\lambda$, which is labeled as the Bragg spacing, $d = 1/|S|$. (*b*, *d*, *f*) Anomalous difference signals $\Delta F/\sigma(\Delta F)$ as a function of scattering-vector length. Data sets are identified by inset keys.

Anomalous signals for the multi-crystal data sets in Table 2 are also displayed as anomalous CC (Fig. 1*c*) and $\Delta F/\sigma(\Delta F)$ (Fig. 1*d*) with respect to Bragg spacing. Data 1to2 means the merging of data 1 and data 2 together and so on for the other data. Data 1 is renamed as data 1to1 and is included in Figs. 1(*c*) and 1(*d*) for comparison. In contrast to the single-crystal data sets, anomalous signals [anomalous CC and $\Delta F/\sigma(\Delta F)$] are apparent in multi-crystal data sets, with anomalous CCs of 30% or higher at 5.5 Å Bragg spacing (Fig. 1*c*). Although adding a few poor data (*e.g.* data 1to2 to data 1to6) to data 1 yields fluctuating anomalous signals, with more poor data sets added the anomalous CC and $\Delta F/\sigma(\Delta F)$ are highest out to 5.3 Å spacing.

There is a concern that data 1 might outweigh the multi-crystal data by dominating the anomalous contribution. This is plausible since the plots of anomalous CC for multi-crystal data 1to2 to 1to8 have curves that are more similar to that of the single-crystal data 1 than to any other single-crystal data (Figs. 1*a* and 1*c*). To examine this possibility, we purposely excluded data 1 and constructed a new set of multi-crystal data from data 2 to data 8 (Table 3). Surprisingly, the anomalous CC from multi-crystal data 2to8 is almost as high as that from multi-crystal data 1to8 (Tables 2 and 3), suggesting that once the overall quality becomes sufficiently high no single data set overly influences the anomalous signal. The course of improvement is shown in Figs. 1(*e*) and 1(*f*) in terms of anomalous CC and $\Delta F/\sigma(\Delta F)$, respectively. Although the inclusion of single-crystal data 3 with data 2 actually causes a general decrease in anomalous CC and $\Delta F/\sigma(\Delta F)$, the gradual increase in anomalous CC and $\Delta F/\sigma(\Delta F)$ when more single-crystal data are included is rather dramatic. By merging seven single-crystal data sets together, the anomalous CC increases to 69.4% (multi-crystal data 2to8) from 24.7% at worst (data 6) or 44.2% at best (data 4). Our multi-crystal merging process appears to be robust in enhancing anomalous signal.

For multi-crystal data 1to1 to 1to8, as more data are included the multiplicity increases from 12.4 (data 1to1) to 73.5 (data 1to8). $R_{\rm merge}$, a commonly adopted data-quality indicator, becomes worse as more single-crystal data are included. For example, $R_{\rm merge}$ increases from 0.147 for data 1to1 to 0.306 for data 1to8. It has been recognized that



(*a*) (*c*)

(*b*) (*d*)

Figure 2
Plots of $I/\sigma(I)$ and map CC distribution with respect to Bragg spacing for multi-crystal data sets. (*a*, *c*) $I/\sigma(I)$, (*b*, *d*) map CC. The definition of Bragg spacing is as in Fig. 1. Data sets are identified by inset keys.

Table 4
Se-substructure determinations and phase evaluations for single-crystal data.

All map comparison values are truncated at 4.0 Å.

Crystal/data set	1	2	3	4	5	6	7	8
Substructure success rate†	16	0	0	0	0	0	0	0
FOM‡ before DM	0.341	0.266	0.243	0.276	0.245	0.221	0.259	0.246
FOM after DM	0.603	0.405	0.528	0.683	0.652	0.662	0.666	0.311
Map CC§ before DM (%)	30.9	26.8	22.1	29.8	20.3	20.9	24.8	20.7
Map CC after DM (%)	48.2	33.4	22.7	38.5	19.1	18.8	31.1	21.9
Mean phase error (°)	68.1	76.4	81.4	72.1	81.4	81.5	78.3	82.0

† *SHELXD* solutions per 1000 attempts. ‡ Figure of merit. § Correlation coefficient between experimental and model-phased maps.

Table 5
Se-substructure determinations and phase evaluations for multi-crystal data sets data 1 through data 8.

Definitions are as given in Table 4, except that all data to 3.5 Å resolution were used.

Data set	1to1	1to2	1to3	1to4	1to5	1to6	1to7	1to8
Substructure success rate	16	5	24	52	41	112	60	45
Successful $CC_{\rm all}$† (%)	43.78	46.99	47.23	49.84	47.71	51.43	53.47	55.20
Successful $CC_{\rm weak}$‡ (%)	13.91	17.93	16.10	20.01	17.39	20.92	23.90	27.67
FOM before DM	0.300	0.300	0.309	0.310	0.315	0.312	0.322	0.331
FOM after DM	0.610	0.641	0.665	0.605	0.757	0.622	0.629	0.630
Map CC before DM (%)	27.6	28.7	30.6	31.7	33.7	34.8	36.4	37.0
Map CC after DM (%)	40.5	44.6	48.9	51.0	54.3	56.2	58.5	60.5
Mean phase error (°)	74.0	71.3	68.9	67.8	64.9	64.3	63.1	62.0

† Average $CC_{\rm all}$ for successful substructure determinations. ‡ Average $CC_{\rm weak}$ for successful substructure determinations.

Table 6
Se-substructure determinations and phase evaluations for multi-crystal data sets data 2 through data 8.

Definitions are as given in Table 4, except that all data to 3.5 Å resolution were used.

Crystal/data set	2to2†	2to3†	2to4	2to5	2to6	2to7	2to8
Substructure success rate	0	0	6	26	61	43	25
Successful $CC_{\rm all}$ (%)	—	—	45.39	47.40	48.75	54.91	53.94
Successful $CC_{\rm weak}$ (%)	—	—	15.31	17.95	18.40	24.70	26.02
FOM before DM	0.232	0.255	0.282	0.307	0.301	0.312	0.318
FOM after DM	0.363	0.555	0.573	0.652	0.598	0.712	0.617
Map CC before DM (%)	24.6	28.2	31.5	32.6	34.2	35.7	36.2
Map CC after DM (%)	30.2	41.8	48.7	50.6	53.7	57.2	58.3
Mean phase error (°)	79.0	72.9	68.9	67.0	65.8	62.8	63.0

† FOM, map CC and mean phase error values were calculated based on Se substructures from running *Phaser* in MR-SAD mode. The native structure (PDB entry 3ott) was used for MR-SAD.

$R_{\rm merge}$ and redundancy are inversely correlated (Weiss & Hilgenfeld, 1997). In contrast, the average $I/\sigma(I)$ tends to improve, albeit not continuously, as data redundancy increases. For example, $I/\sigma(I)$ increases from 17.7 for data 1to1 to 23.9 for data 1to8. For multi-crystal data 2to2 to 2to8, a similar but more dramatic increase in $I/\sigma(I)$ is obtained from merging of more single-crystal data. For example, $I/\sigma(I)$ is almost doubled from 11.2 for data 2to2 to 22.6 for data 2to8. The course of improvement from multi-crystal merging is shown graphically in Fig. 2(*a*) for the series of data 1to1 to 1to8 and in Fig. 2(*c*) for data 2to2 to 2to8. With highly redundant data merged from multiple crystals, *e.g.* 1to8 and 2to8, $I/\sigma(I)$ values are significantly higher than those from any single-crystal data set. $I/\sigma(I)$ might be effective for assessing the overall quality of multi-crystal data, as it is used for the single-crystal case (Evans, 2006).

3.2. Substructure determinations

For each of the eight single-crystal data sets, *SHELXD* was used to find Se substructures. *SHELXD* derives E values from Bijvoet-difference amplitudes, $||F(\mathbf{h})| - |F(-\mathbf{h})|| \rightarrow E_{\rm obs}$, but it excludes the weak $E_{\rm obs}$ data (here, the lowest 30%) from the substructure calculations because they are least reliable. The program then uses correlation coefficients (CC) between $E_{\rm obs}$ and $E_{\rm calc}$ as criteria to evaluate the validity of substructure solutions (Sheldrick, 2008). $CC_{\rm all}$ is the correlation coefficient based on all data and $CC_{\rm weak}$ is that from the weakest 30% of the data, which, akin to $R_{\rm free}$, provides a sensitive measure of validity (Schneider & Sheldrick, 2002). By exploring different parameters of *SHELXD*, 16 correct solutions can be obtained from 1000 attempts for data 1 at 6 Å. There is no solution if higher or lower angle data are included for substructure determination. The preponderance of non-solutions have low $CC_{\rm all}$ and $CC_{\rm weak}$ values, here with a mean of 29.3% ± 1.8% for $CC_{\rm all}$ and 2.4% ± 2.5% for $CC_{\rm weak}$ for data 1. Successful solutions are differentiated by having higher values for both coefficients, here with a mean of 43.8% ± 0.7% for $CC_{\rm all}$ and 13.9% ± 1.1% for $CC_{\rm weak}$ for data 1. However, using *SHELXD* with the same set of parameters, we could find no solutions from any other single data set (Table 4). This is consistent with the anomalous CC shown in Fig. 1(*a*) and Table 1: data 1 has the strongest anomalous CC and allows Se-substructure determination, whereas the seven other data sets have lower anomalous CCs and do not allow substructure determination. The crystal for data 1 was not apparently larger than the others; instead, variations in crystal perfection could be a reason for the differential data quality.

The multi-crystal data sets in Table 2 were also used to find Se substructures by using the program *SHELXD* with the same parameters as used for the single-crystal data. We found that each multi-crystal data set in Table 2 allows the determination of the Se substructure but with different rates of success (Table 5). For data 1to1 through 1to8 success rates fluctuate but the $CC_{\rm all}$ and $CC_{\rm weak}$ averages for these successes increase steadily. $CC_{\rm weak}$ is nearly doubled from single-crystal data 1to1 to multi-crystal data 1to8. Similarly, CC_{all} increases from 43.8% for data 1to1 to 55.2% for data 1to8.

From the progression of anomalous signal in the successive mergings from data 2 through data 8 (Table 3), we expected

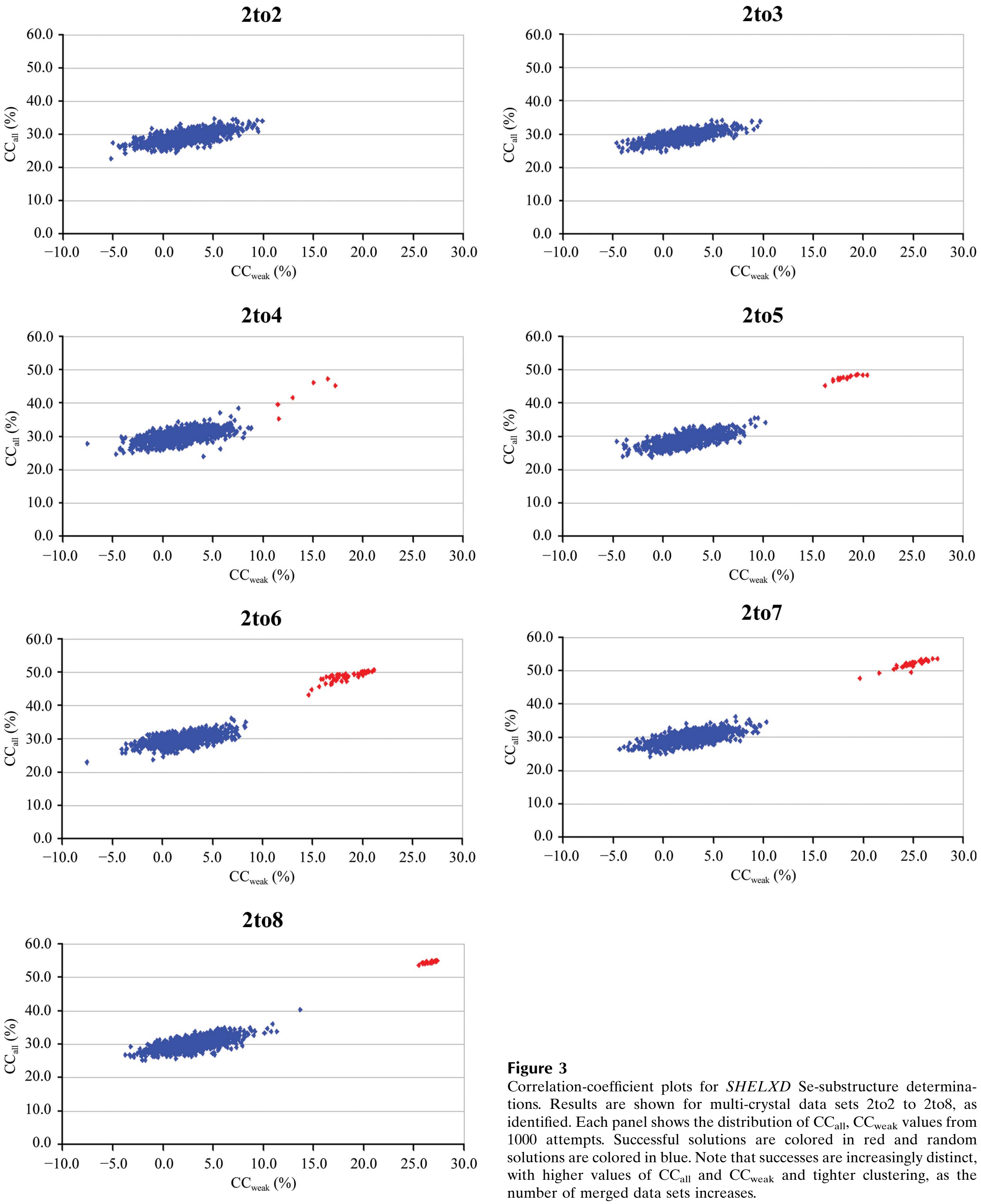


Figure 3

Correlation-coefficient plots for *SHELXD* Se-substructure determinations. Results are shown for multi-crystal data sets 2to2 to 2to8, as identified. Each panel shows the distribution of CC_{all}, CC_{weak} values from 1000 attempts. Successful solutions are colored in red and random solutions are colored in blue. Note that successes are increasingly distinct, with higher values of CC_{all} and CC_{weak} and tighter clustering, as the number of merged data sets increases.

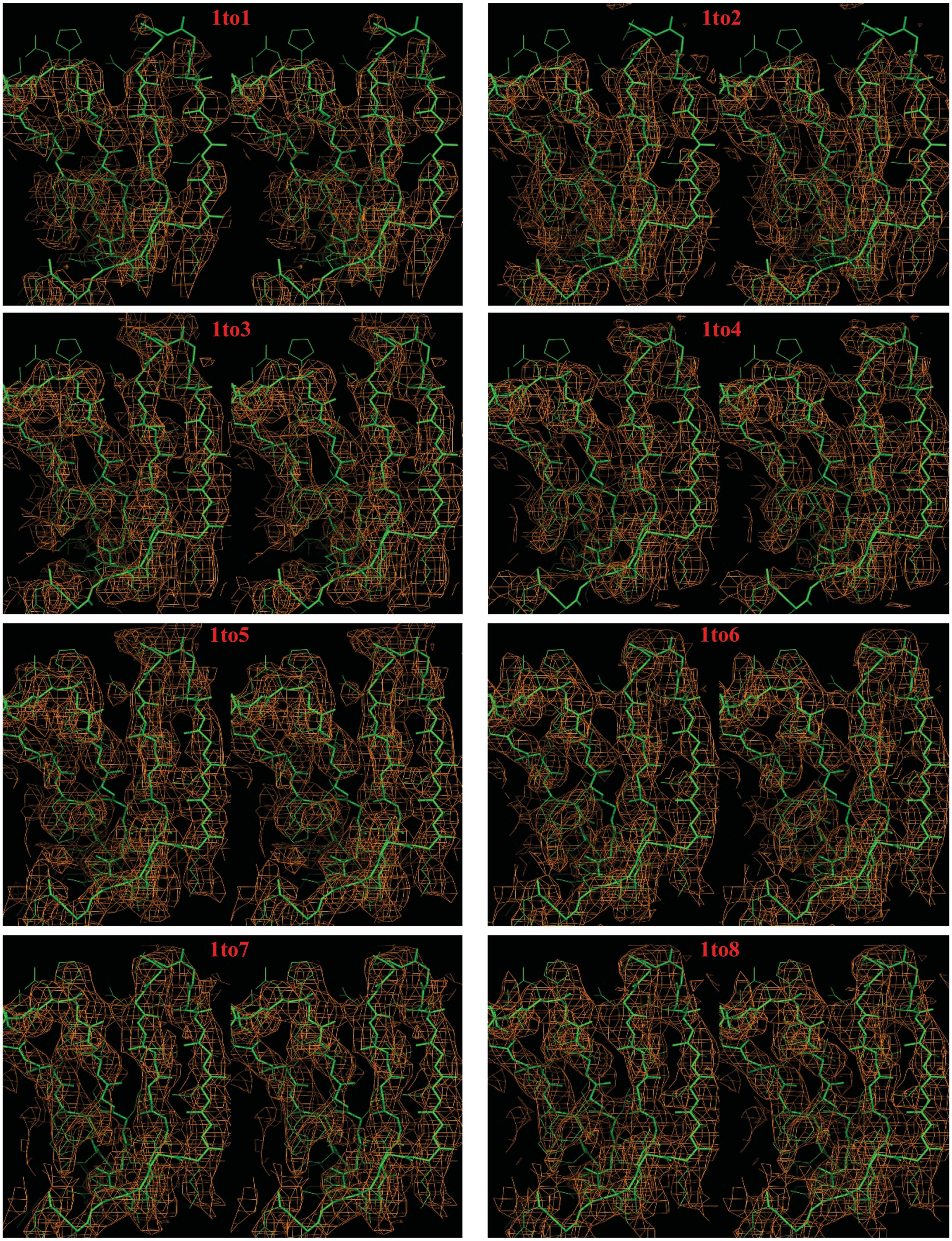


Figure 4
Stereo representations of electron densities for successive mergings of multi-crystal data sets. Electron-density distributions calculated from phases after density modification at 3.5 Å resolution are shown as orange meshes contoured at 1.0σ. For reference, the model from the native structure (residues 256–294) is shown as sticks for bonds between main-chain atoms and selected side chains (green). Note the improved continuity and coverage in successive maps. This figure was prepared using *PyMOL* (http://www.pymol.org).

that some of these multi-crystal data sets without data 1 would support substructure determination. Indeed, *SHELXD* runs in this series did produce Se substructures. Success rates for each data set are listed in Table 6 and the visual evolution of CC_{all} and CC_{weak} with respect to the number of merged single-crystal data is shown in Fig. 3. There are no credible substructure solutions from data 2to2 and 2to3. Gratifyingly, from data 2to4 to 2to8 successful solutions emerged from nonsolutions with increased CC_{all} and CC_{weak} values. With all seven sub-threshold single-crystal data sets merged, data 2to8 gives a very clear contrast between successful solutions and nonsolutions. As for the progression from multi-crystal data 1to1 to 1to8 (Table 5), the average CC_{weak} and CC_{all} values for successful solutions also increase from data 2to4 to 2to8 (Fig. 3, Table 6). The appearance of signature clusters of *SHELXD* solutions in the *Shake-and-Bake* approach (Miller *et al.*, 1993) validates the efficacy of multi-crystal merging of anomalous diffraction data.

3.3. Phasing

Although the Se substructure could be obtained from single-crystal data 1, this did not support the production of an interpretable electron-density map by SAD phasing. To further evaluate the single-crystal data, the Se substructure found by running *Phaser* in the MR-SAD mode was used for SAD phasing in each case. The resulting figure of merit (FOM), map CC and mean phase error are tabulated in Table 4 as computed at 4.0 Å. In accordance with its highest anomalous CC, data 1 gives the highest FOM before DM, highest map CC and lowest mean phase error of all the single-crystal data sets. The FOM after DM, which is mostly quite variable, might be biased from an automatically terminated density-modification process. The phasing statistics are quite poor from these single-crystal data sets; for example, the mean phase errors range up to 82.0°, approaching random. Even the best values (for data 1 and data 4) are marginal and hence it is not surprising that none of these single-crystal data sets produced an interpretable electron-density map.

Having established the effectiveness of multi-crystal data in substructure determinations, we then moved forward to determine whether multi-crystal data can also be used effectively for phase evaluation to determine complete structures. To make the results comparable to the case of single-crystal phasing (Table 4), we employed the same experimental phasing procedure again for the data sets listed in Table 2 (data 1to1 to 1to8). Instead of using data truncated to 4 Å as for the single-crystal case, however, here we included all data to 3.5 Å Bragg spacing in the analyses. The phasing statistics for the data sets in Table 2 clearly show that as increasing numbers of data are merged, the map CC values increase and the mean phase errors decrease (Table 5). (If not specified, hereafter map CC means the map CC after DM.) Thus, the map CC starts at 40.5% for single-crystal data 1to1 and steadily climbs to 60.5% for multi-crystal data 1to8; in parallel, the mean phase error gradually decreases from 74.0° for single-crystal data 1to1 to 62.0° for multi-crystal data 1to8. These statistics indicate that multi-crystal data can help to obtain improved phases.

We also calculated phases for multi-crystal data assembled from single-crystal data sets 2 to 8; the resultant phasing statistics are given in Table 6. Since the anomalous signals and substructure solutions improve with these multi-crystal mergings, it is not surprising to observe that the phases also improve. Thus, the mean phase error for multi-crystal data 2to8 at 4.0 Å decreased to 57.6° from single-crystal values ranging from 72.1° (data 4) in the best case to 82.0° (data 8) in the worst case (Table 4). The map CC and mean phase error values for multi-crystal data 2to8 (58.3% and 63.0°, respectively) are nearly as good as those for data 1to8 (60.5% and 62.0, respectively). Figs. 2(*b*) and 2(*d*) plot the gradual increase in map CC for the mergings of data 1 through 8 and data 2 through 8, respectively. As for use in substructure determination, the best single-crystal data are helpful, but not indispensable, for excellent phase evaluation.

To see visually what differences the use of multi-crystal data can have in structure determination, we examined the electron density derived from alternative situations. As an illustration, we compare stereo drawings of electron densities for a representative protein segment (residues 256–294) based on eight progressively improved data sets: 1to1 to 1to8 (Fig. 4). The map from the best single-crystal data 1to1 (map CC = 40.5%) could not be interpreted: there were gaps in β-strands and no continuous electron density for side chains and surface turns. In comparison, electron densities gradually improved when more single-crystal data are added. The electron density from multi-crystal data 1to8 (map CC = 60.5%) reveals much of the structural detail including traceable β-strands, large side chains and structured turns.

4. Discussion

4.1. Difficulties in single-crystal phasing

In this study, we evaluated the utility of multi-crystal anomalous diffraction in phasing the exceptionally large histidine kinase sensor domain, $Bt_{4673}HK3_S$, at 3.5 Å resolution using otherwise standard SeMet SAD procedures. This structure proved challenging to solve from single-crystal Se-SAD data; indeed, we failed with attempts from many single-crystal data sets. The first challenge is that the structure is large, containing 1456 ordered amino-acid residues in the asymmetric unit, and entirely organized into all-β domains. The absence of helical structures makes it difficult to resolve secondary-structural features from a low-resolution Se-SAD experiment. The second challenge comes from weak diffraction. Of all of the single-crystal data sets, the best R_{merge} value is 0.129 (Table 1). Thus, although the estimated Bijvoet ratio from 20 Se sites (one per 73 residues) is respectable at 3%, the anomalous signal tends to be lost in the noise. Indeed, based on the anomalous CC measure, there is virtually no detectable signal beyond 4 Å Bragg spacing for any of the crystals (Fig. 1*a*). The Se substructure could only be deduced in one case (at 6 Å resolution); however, even then, or when the

Table 7
Phase evaluations for multi-crystal data but excluding one single-crystal data set each time.

Data set	Exclude data 1	Exclude data 2	Exclude data 3	Exclude data 4	Exclude data 5	Exclude data 6	Exclude data 7	Exclude data 8
Map CC (%)	58.3	55.6	58.4	57.9	56.8	58.4	54.1	58.5
Mean phase error (°)	63.0	63.7	62.9	62.9	62.6	63.4	64.0	63.1

correct substructure was assumed for the other cases, phase evaluations led to uninterpretable maps with unacceptably high mean phase errors and low map CCs (Table 4). Altogether, these problems prevented our determination of the structure from conventional single-crystal Se-SAD phasing. The effective application of a multi-crystal strategy with this challenging structure may establish its applicability for phasing other difficult structures.

4.2. Multiplicity

For single-crystal anomalous phasing, high redundancy has been found to be more important than low $R_{\rm merge}$ (Dauter & Adamiak, 2001). This is also likely to be true for multi-crystal anomalous phasing. In our multi-crystal mergings of data, we found that as multiplicity increased with the addition of data sets, various measures of structure-determination effectiveness also improved (Tables 5 and 6). For example, multi-crystal data 1to8 has the highest redundancy of 73.5 and was also the best data set for substructure determination (highest $\rm CC_{all}$ and $\rm CC_{weak}$) and subsequent phasing (highest map CC). Fig. 5 shows the strongly positive progression of map CC for both multi-crystal data sets with respect to multiplicity. Multiplicity seems to be a positive indicator of data quality in a multi-crystal strategy.

The multi-crystal anomalous phasing strategy presented here is quite robust and is not likely to be biased by any single data set. For example, the exclusion of data 1, which is by far the best, from the multi-crystal merging does not obviously affect substructure determination and phasing (Table 6). The mean phase error increases by only 1.0° when excluding data set 1. To further test the impact of individual data sets, we systematically excluded one data set at a time (Table 7). We found only slight changes in map CC (decreases ranging from 2.0 to 6.4%) and mean phase error (increases ranging from 0.6 to 2.0°). The fact that the effect was detrimental in each case indicates that all data sets are compatible with one another and the dispensability of the best data set shows that multiplicity can overcome intrinsic weakness.

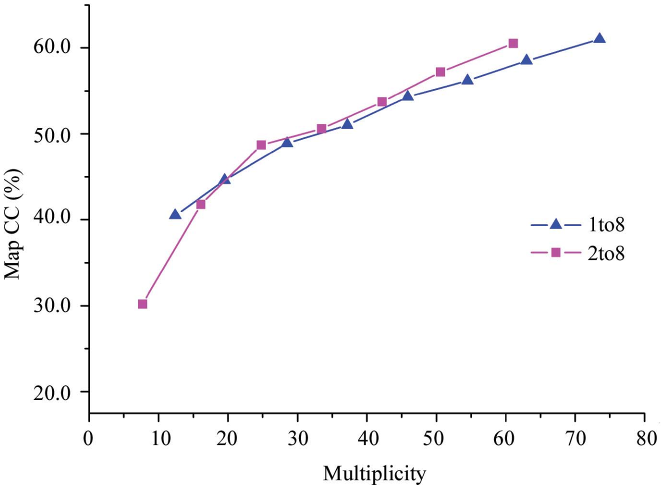


Figure 5
Progression of map CC values for multi-crystal data sets with respect to multiplicity. The improvement in map CC can be seen clearly with the accumulation of multiplicity from merging of data 1 through data 8 (blue) and data 2 through data 8 (magenta).

The benefits from data-merging additions do gradually approach a limit. Based on this study, a multiplicity of above 40 provides most of the phase improvement. For example, multi-crystal data 1to5 has a multiplicity of about 46 (Table 2), corresponding to a mean phase error of 64.9° (Table 5). The addition of three more single-crystal data sets only reduces the mean phase error by 2.9°. Similarly, multi-crystal data 2to6 has a multiplicity of about 42 (Table 3), corresponding to a mean phase error of 65.8° (Table 6). The addition of two more single-crystal data sets only reduces the mean phase error by 2.8°. Therefore, we may postulate that for a defined experimental situation there might be a limitation of accuracy from multi-crystal merging. The multi-crystal strategy may only help to enhance the anomalous signals to approach that limit but cannot be better.

4.3. Data-quality assessment

In single-crystal data analyses, $R_{\rm merge}$ and $I/\sigma(I)$ are the two most common criteria for the control of data quality. For high-redundancy single-crystal data, other measures such as $R_{\rm meas}$ and $R_{\rm p.i.m.}$ may be better criteria for the control of data quality (Diederichs & Karplus, 1997; Weiss, 2001; Evans, 2006). As the purpose of the multi-crystal strategy implemented here is to enhance the detection of anomalous diffraction signals, it would be natural to use an anomalous indicator to quantitatively control data quality. From our experience, the anomalous CC could be useful for this purpose as it is calculated from measured intensities and is independent of the scaling-error model. Moreover, the anomalous CC is well suited for the evaluation of multi-crystal data. Most significantly, the anomalous CC correlates well with measures of success in structure determination; the higher the anomalous CC (Table 3), the more reliable substructure determination is (Fig. 3) and the lower the mean phase error (Table 6). In fact, for the case of single-crystal data, in which the phase evaluations are inadequate to support structure determination, the anomalous CC nevertheless correlates almost perfectly with the map CC (Table 8).

Taking advantage of the high correlation between anomalous CC and map CC for single-crystal data, we may use the anomalous CC to evaluate and to truncate single-crystal data

Table 8
Correlation coefficients relating estimates of data accuracy to phase evaluations.

Only the correlations of map CC with estimators of data accuracy are recorded, since map CC is almost perfectly anticorrelated with mean phase error and is also highly correlated with the substructure parameters CC_{all} and CC_{weak}.

	Single crystals	Multi-crystals† 1 to 8	Multi-crystals‡ 2 to 8	All data sets§
Map CC *versus* anomalous CC (%)	92.9	66.1	97.2	94.1
Map CC *versus* $\Delta F/\sigma(\Delta F)$ (%)	89.0	47.9	96.4	85.7
Map CC *versus* multiplicity (%)	31.4	99.1	96.9	84.0
Map CC *versus* $I/\sigma(I)$ (%)	41.9	83.2	98.7	81.5
Map CC *versus* R_{merge} (%)	−37.6	82.0	−37.5	69.5
Map CC *versus* mean phase error (%)	−97.8	−99.6	−99.5	−99.2
Map CC *versus* successful CC_{all} (%)	—	88.2	96.8	88.1
Map CC *versus* successful CC_{weak} (%)	—	80.7	96.7	84.5

† Single-crystal data 1to1 is excluded. ‡ Single-crystal data 2to2 is excluded. § The two data sets that are duplicated between single crystals and the two multi-crystal mergings are counted only once as single-crystal data sets.

Table 9
Phase evaluations from multi-crystal data merged in separated wedges.

Each of the eight single-crystal data sets was split into wedges of frames 1–30, 31–60, 61–90 and 91–120. Data collected as inverse-beam pairs were split and both were included in the merging. Frames from data 2, which did not use the inverse-beam approach, were combined as frames 1–60 in wedge 1, frames 61–120 in wedge 2 and frames 121–180 in wedge 3. Corresponding wedges from all single-crystal data were then merged together. Map CC and mean phase error were calculated based on Se substructures from running *Phaser* in MR-SAD mode. The native structure (PDB entry 3ott) was used for MR-SAD.

Wedge (frames)	1 (1–30)	2 (31–60)	3 (61–90)	4 (91–120)
Multiplicity	16.6 (12.4)	16.8 (12.5)	16.6 (12.3)	17.8 (13.3)
Completeness (%)	99.7 (99.7)	99.6 (99.5)	99.9 (100.0)	99.9 (99.9)
Map CC (%)	43.5	39.6	40.0	28.9
Mean phase error (°)	70.8	73.9	73.7	77.4

sets in order to avoid radiation damage. To examine the possible utility of the anomalous CC in the assessment of radiation damage, we followed the course of this measure over the data range for each of the eight single-crystal data sets (Fig. 6*a*). The anomalous CC tends to increase as more data frames are included. Without radiation damage, we would expect to see a steady increase in the anomalous CC with increased multiplicity; when radiation damage is present, however, deterioration in the agreement of Bijvoet differences is expected as the range of data frames increases. Increased multiplicity then ceases to be beneficial. Single-crystal data 7 provides such an example; its anomalous CC is highest (37.1%) at frame 140 and then decreases sharply to 16.5% at frame 200, indicative of radiation damage (Fig. 6*a*). Thus, in this particular case it seemed prudent to truncate the data, which we did at frame 120. The radiation damage may also be observed in a 'decay *R* factor' (R_d) plot (Diederichs, 2006). Consistent with the anomalous CC plot, data 7 shows increased decay at larger frame-number differences (Fig. 6*b*).

4.4. Radiation damage

Although radiation damage did not seem to be crippling for the samples used in this study, damage does exist. When we separated the single-crystal data into different wedges and merged the multi-crystal data wedge by wedge, we found that the map CC and mean phase error degrade through successive wedges (Table 9). For example, for wedge 1 of merged data from eight single-crystal data sets (30 × 2 images for inverse-beam data) the mean phase error is 70.8°, whereas for wedge 4 (frames 91–120) the mean phase error is 77.4°, an increase of 6.6°. Since all other factors are equal as the frames accumulate, surely so in the average, we attribute this deterioration in phasing efficacy with frame number to a time-dependent effect, namely radiation damage. Indeed, whereas wedge 1 produces phasing statistics (map CC and mean phase error) that are nearly as good as those from data 1, which is by far the best single-crystal data set, wedge 4 produces phasing statistics that are only comparable to or worse than those from each of the other single-crystal data sets, despite having a multiplicity that is 44–131% greater than in those data sets (Tables 4 and 9).

Using the program *RADDOSE* (Murray *et al.*, 2004; Paithankar *et al.*, 2009), we estimated the absorbed doses in these experiments. For the longest-exposure data set 7, the absorbed dose for 200 × 2 images with inverse-beam mode data collection is 23 MGy, which is slightly above the Henderson limit (Henderson, 1990) but within the Garman limit (Owen *et al.*, 2006); however, the absorbed dose for the 120 × 2 images actually used in our analyses (Table 1) is about 14 MGy, which is well below the Henderson limit. The absorbed doses for all other data sets are also within this limit. Based on the plot of anomalous CC *versus* number of frames (Fig. 6*a*), data set 7 could be truncated at the first 140 images (140 × 2), which is equivalent to an acceptable absorbed dose of about 16 MGy. We conclude that the anomalous CC can be useful for quality control to avert radiation damage in single-crystal data sets.

Exposure of macromolecular crystals at third-generation synchrotrons inevitably causes radiation damage to crystals, with symptoms such as loss of high-angle diffraction spots, increased scaling *B* factors and deteriorated R_{merge} values, even under cryogenic protection at 100 K (Sliz *et al.*, 2003; Garman & Nave, 2009). The multi-crystal strategy used in this study may provide an effective way to deal with the radiation-damage problem.

4.5. Trade-off between multiplicity and radiation damage

For the multi-crystal data sets used in this study, there is a high overall correlation of multiplicity with the phasing statistics (map CC and mean phase error). The coefficient of correlation between multiplicity (Tables 2 and 3) and map CC (Tables 5 and 6) is 97.3% over multi-crystal data sets. Multiplicity is not the only determinant of phasing accuracy, of course. Thus, this correlation coefficient is only 31.4% among the single-crystal data sets, where the differences in multi-

Table 10
Phase evaluations from multi-crystal data merged in accumulated wedges.

Wedges of data were separated as defined for Table 9, but then accumulated together in successive mergings before phase evaluations were performed as described in Table 9.

Wedge (frames)	1 (1–30)	1–2 (1–60)	1–3 (1–90)	1–4 (1–120)	All
Multiplicity	16.6 (12.4)	33.1 (24.7)	49.7 (36.9)	65.0 (47.8)	73.5 (55.5)
Completeness (%)	99.7 (99.7)	99.9 (100.0)	99.9 (100.0)	99.9 (100.0)	99.9 (100.0)
Map CC (%)	43.5	54.6	58.1	59.6	61.3
Mean phase error (°)	70.8	64.2	62.5	62.2	60.5

plicity are very slight, whereas it rises to 99.1 and 96.9% for the 1 to 8 and 2 to 8 series, respectively, where phase accuracy gradually increases with multiplicity (Fig. 5).

Other studies (Dauter & Adamiak, 2001) as well as the results here point to the benefits of high redundancy for phasing accuracy. Yet, as shown in the previous section, there is also a cost in phasing accuracy from redundant accumulations made at the expense of increased exposure. One might expect that the merging of later more damaged data with earlier more pristine data would be deleterious to the overall statistical results. Remarkably, and contrary to our *a priori* expectation, phasing efficacy improves with the accumulation of data into the merging despite the evident radiation damage (Table 10). Upon combining wedges 1 and 2 from eight single-crystal data sets, the phasing statistics improve appreciably (the map CC increases from 43.5% to 54.6% and the mean phase error decreases from 70.8° to 64.2°). This trend continues through wedge 3 and the merging of all data ends up with the highest map CC and the lowest mean phase error.

We anticipate that when crystal supply is not limiting, it should be possible to use a large number of crystals with each restricted to a limited data range (such as in wedge 1 here or less) where radiation damage is minimized. Thereby, the advantages of high multiplicity might be realised without adverse effects from radiation damage. In the event of variability in crystal lattices, perhaps because of variations accompanying flash-cryoprotection, diffraction patterns might be classified into distinct subsets to be phased and analyzed separately. Collecting Bijvoet pairs separately for individual crystals should also be helpful in minimizing the effects of lack of isomorphism (Clemons *et al.*, 2001).

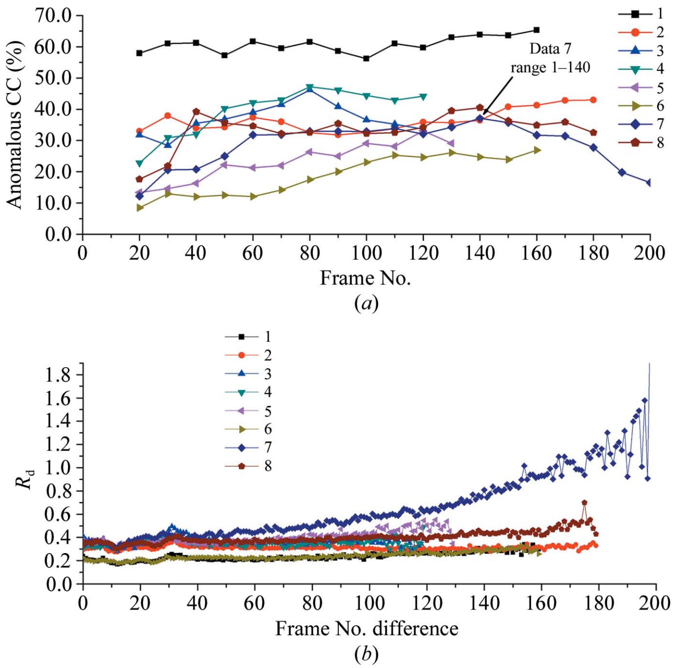


Figure 6
Data-quality assessments for single-crystal data sets. (*a*) Progression of anomalous CC with increased number of frames included in merging. Anomalous CC tends to increase at the beginning of each single-crystal data range. The anomalous CC may drop dramatically after the collection of a certain amount of data, perhaps owing to radiation damage. For example, obvious radiation damage could be seen in data 7 after 140 data frames. (*b*) R_d plot to show radiation-induced decay with respect to frame-number difference. The dramatically increased R_d for data 7 with a frame-number difference beyond 140 indicates radiation damage consistent with the anomalous CC plot in (*a*). Data sets are identified by inset keys.

4.6. Resolution

Using only the most reliable data for phasing is important in any phasing experiment. A straightforward way to achieve reliability is to limit data to correspond to a certain resolution. In this study, owing to variations among crystals, the effective resolution for eight single-crystal data sets ranges from 3.2 Å for data 6 to 4.0 Å for data 4. All multi-crystal data were processed at 3.5 Å resolution, although in retrospect perhaps limitation to lower resolution might have given better results.

Two different nominal resolution limits were used in this study. The first is the limit for Se-substructure determination, which was set at 6 Å. For the single-crystal data sets, the anomalous signals are at the noise level at Bragg spacings beyond 4.5 Å in the best case (Figs. 1*a* and 1*b*). Indeed, only single-crystal data 1 permitted determination of the substructure and even then the resolution limit had to be cut to 6 Å for success. In multi-crystal cases, very reliable substructure solutions could also be obtained readily from multi-crystal data sets 1to2 to 1to8 and from 2to4 to 2to8 when cut at 6 Å (Tables 5 and 6). Since the anomalous signal in data 1to8 is much more accurate than in any single-crystal data set, substructures can actually be found even at 3.5 Å, with a success rate of 25 from 1000 attempts. This implies that usable anomalous signals have been extended to higher Bragg spacings than before. There is a concern that Se atoms might not be resolved at 6 Å resolution since a pair of SeMet residues could be in van der Waals contact at 4 Å or even closer in the case of disorder into alternate rotamers. This is not a serious concern, however, since the preponderance of distances in a real structure will be much longer and unresolved sites would emerge on refinement. In the event, the average of closest realised Se—Se distances here is 13.6 Å, with the shortest of all being 7.2 Å.

The second resolution cutoff is the limit used for phase calculation and subsequent electron-density modification.

Based on the objective comparisons afforded by map CC (and comparably by mean phase error), we conclude that multi-crystal data can support effective phase evaluation to a conventional data limit. Across the entire range of scattering angles, multi-crystal data 1to8 and 2to8 give the highest map CCs, including the outermost shell, in which the anomalous signals are prone to be least reliable (Figs. 2*b* and 2*d*). Moreover, consistent with the increase in overall map CC values in multi-crystal data, before density modification as well as after (Tables 5 and 6), there is a steady increase of shell-by-shell map CC values as single-crystal data sets are added.

5. Concluding remarks

We have described the application of multi-crystal anomalous diffraction to phasing of a large all-β structure at 3.5 Å resolution. To assess the effectiveness of the multi-crystal strategy in anomalous phasing, we systematically analyzed a series of tests using a standard SAD phasing protocol. We conclude that a multi-crystal strategy can significantly enhance anomalous signal extraction. Occasionally, one can encounter an exceptional crystal that could possibly yield a structure; however, the combined data from other crystals examined in the search for this exceptional crystal may yield results superior to those extracted from the diffraction of this best crystal. The improved multi-crystal anomalous data can be used in standard SAD phasing with obvious advantages: (i) multi-crystal anomalous data permit substructure determinations that cannot be found from the single-crystal data sets and (ii) multi-crystal anomalous data yield improved phases superior to those obtained from the single-crystal data sets. Based on this study, we propose that the multi-crystal strategy may be a useful tool for solving difficult structures such as large macromolecular complexes and membrane proteins that diffract poorly.

We thank John Schwanof and Randy Abramowitz for help with synchrotron data collection. This work was supported in part by NIH grant GM34102. Beamline X4A of the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory, a DOE facility, is supported by the New York Structural Biology Center.

References

Burla, M. C., Carrozzini, B., Cascarano, G. L., Giacovazzo, C. & Polidori, G. (2003). *Acta Cryst.* D**59**, 662–669.

Carugo, O. & Carugo, K. D. (2005). *Trends Biochem. Sci.* **30**, 213–219.

Clemons, W. M., Brodersen, D. E., McCutcheon, J. P., May, J. L. C., Carter, A. P., Morgan-Warren, R. J., Wimberly, B. T. & Ramakrishnan, V. (2001). *J. Mol. Biol.* **310**, 827–843.

Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.

Cowtan, K. D. & Zhang, K. Y. J. (1999). *Prog. Biophys. Mol. Biol.* **72**, 245–270.

Dauter, Z. (2006*a*). *Acta Cryst.* D**62**, 1–11.

Dauter, Z. (2006*b*). *Acta Cryst.* D**62**, 867–876.

Dauter, Z. & Adamiak, D. A. (2001). *Acta Cryst.* D**57**, 990–995.

Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.

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

Debreczeni, J. É., Bunkóczi, G., Girmann, B. & Sheldrick, G. M. (2003). *Acta Cryst.* D**59**, 393–395.

Diederichs, K. (2006). *Acta Cryst.* D**62**, 96–101.

Diederichs, K. & Karplus, P. A. (1997). *Nature Struct. Biol.* **4**, 269–275.

Ealick, S. E. (2000). *Curr. Opin. Chem. Biol.* **4**, 495–499.

Evans, P. (2006). *Acta Cryst.* D**62**, 72–82.

Fan, H.-F., Hao, Q., Gu, Y.-X., Qian, J.-Z., Zheng, C.-D. & Ke, H.-M. (1990). *Acta Cryst.* A**46**, 935–939.

Garman, E. F. & Nave, C. (2009). *J. Synchrotron Rad.* **16**, 129–132.

Grosse-Kunstleve, R. W. & Adams, P. D. (2003). *Acta Cryst.* D**59**, 1966–1973.

Hao, Q. (2000). *J. Synchrotron Rad.* **7**, 148–151.

Henderson, R. (1990). *Proc. R. Soc. Lond. Ser. B*, **241**, 6–8.

Hendrickson, W. A. (1991). *Science*, **254**, 51–58.

Hendrickson, W. A. (1999). *J. Synchrotron Rad.* **6**, 845–851.

Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.

Hendrickson, W. A. & Ogata, C. M. (1997). *Methods Enzymol.* **276**, 494–523.

Hendrickson, W. A., Smith, J. L., Phizackerley, R. P. & Merritt, E. A. (1988). *Proteins*, **4**, 77–88.

Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.

Ji, X., Sutton, G., Evans, G., Axford, D., Owen, R. & Stuart, D. I. (2010). *EMBO J.* **29**, 505–514.

Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.

Kabsch, W. (2010*a*). *Acta Cryst.* D**66**, 125–132.

Kabsch, W. (2010*b*). *Acta Cryst.* D**66**, 133–144.

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

Kleywegt, G. J. & Read, R. J. (1997). *Structure*, **5**, 1557–1569.

Leiros, H.-K. S., Timmins, J., Ravelli, R. B. G. & McSweeney, S. M. (2006). *Acta Cryst.* D**62**, 125–132.

Li, J., Edwards, P. C., Burghammer, M., Villa, C. & Schertler, G. F. X. (2004). *J. Mol. Biol.* **343**, 1409–1438.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst.* **40**, 658–674.

Miller, R., DeTitta, G. T., Jones, R., Langs, D. A., Weeks, C. M. & Hauptman, H. A. (1993). *Science*, **259**, 1430–1433.

Moukhametzianov, R., Burghammer, M., Edwards, P. C., Petitdemange, S., Popov, D., Fransen, M., McMullan, G., Schertler, G. F. X. & Riekel, C. (2008). *Acta Cryst.* D**64**, 158–166.

Murray, J. W., Garman, E. F. & Ravelli, R. B. G. (2004). *J. Appl. Cryst.* **37**, 513–522.

Nave, C. & Garman, E. F. (2005). *J. Synchrotron Rad.* **12**, 257–260.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Owen, R. L., Rudiño-Piñera, E. & Garman, E. F. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 4912–4917.

Paithankar, K. S., Owen, R. L. & Garman, E. F. (2009). *J. Synchrotron Rad.* **16**, 152–162.

Pape, T. & Schneider, T. R. (2004). *J. Appl. Cryst.* **37**, 843–844.

Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003). *Acta Cryst.* D**59**, 1020–1027.

Rice, L. M., Earnest, T. N. & Brunger, A. T. (2000). *Acta Cryst.* D**56**, 1413–1420.

Sanishvili, R., Nagarajan, V., Yoder, D., Becker, M., Xu, S., Corcoran, S., Akey, D. L., Smith, J. L. & Fischetti, R. F. (2008). *Acta Cryst.* D**64**, 425–435.

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

Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.

Shen, Q., Wang, J. & Ealick, S. E. (2003). *Acta Cryst.* A**59**, 371–373.

Sliz, P., Harrison, S. C. & Rosenbaum, G. (2003). *Structure*, **11**, 13–19.

Wagner, A., Pieren, M., Schulze-Briese, C., Ballmer-Hofer, K. & Prota, A. E. (2006). *Acta Cryst.* D**62**, 1430–1434.

Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
Weeks, C. M. & Miller, R. (1999). *J. Appl. Cryst.* **32**, 120–124.
Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.
Weiss, M. S. & Hilgenfeld, R. (1997). *J. Appl. Cryst.* **30**, 203–205.
Zwart, P. H. (2005). *Acta Cryst.* D**61**, 1437–1448.