Acta Crystallographica Section D
Biological Crystallography
ISSN 1399-0047

Multi-crystal native SAD analysis at 6 keV

Qun Liu,[a,b]* Youzhong Guo,[a,c] Yanqi Chang,[a] Zheng Cai,[d] Zahra Assur,[e] Filippo Mancia,[e] Mark I. Greene[d] and Wayne A. Hendrickson[a,b,c,e]*

[a]NYCOMPS, New York Structural Biology Center, New York, NY 10032, USA, [b]New York Structural Biology Center, NSLS X4, Brookhaven National Laboratory, Upton, NY 11973, USA, [c]Department of Biochemistry and Molecular Biophysics, Columbia University, New York, NY 10032, USA, [d]Department of Pathology and Laboratory Medicine, University of Pennsylvania, Philadelphia, PA 19104, USA, and [e]Department of Physiology and Cellular Biophysics, Columbia University, New York, NY 10032, USA

Correspondence e-mail: qunliu@bnl.gov, wayne@xtl.cumc.columbia.edu

Received 2 May 2014
Accepted 8 June 2014

PDB references: YetJ, 4tkq; ThiT, 4tkr; EGFR kinase domain, 4tks

Anomalous diffraction signals from typical native macromolecules are very weak, frustrating their use in *de novo* structure determination. Here, native SAD procedures are described to enhance signal to noise in anomalous diffraction by using multiple crystals in combination with synchrotron X-rays at 6 keV. Increased anomalous signals were obtained at 6 keV compared with 7 keV X-ray energy, which was used for previous native SAD analyses. A feasibility test of multi-crystal-based native SAD phasing was performed at 3.2 Å resolution for a known tyrosine protein kinase domain, and real-life applications were made to two novel membrane proteins at about 3.0 Å resolution. The three applications collectively serve to validate the robust feasibility of native SAD phasing at lower energy.

1. Introduction

Single-wavelength anomalous diffraction (SAD) and multi-wavelength anomalous diffraction (MAD) are two dominant methods for *de novo* evaluation of biological macromolecular structures (Hendrickson, 1991, 2014). In each case, one must first locate the substructure of anomalously scattering atoms and then evaluate the phases in order to produce an image of the entire structure. Both substructure determination and phase evaluation use anomalous signals that in general are only a few percent of the overall diffraction intensities. Therefore, the acquisition of accurate anomalous signals from the overall diffraction intensities is critical for a successful SAD/MAD experiment. With advances in synchrotron instrumentation and computational methods, SAD phasing has become highly effective for macromolecular structural analysis (Hendrickson, 1999; Adams *et al.*, 2013; Duke & Johnson, 2010; Dauter *et al.*, 2010, 2000; Schiltz *et al.*, 2003). For the most part, this success is based on resonances from heavier anomalous scatterers ($Z \geq 25$, Mn), including those intrinsic to metalloproteins, such as Fe or Zn, those added in conventional heavy-atom derivatizations ($Z \geq 73$, Ta) or those incorporated covalently, as for Se ($Z = 34$) in selenomethionyl proteins (Hendrickson *et al.*, 1990) or for Br ($Z = 35$) in brominated nucleic acids (Dauter *et al.*, 2000).

Lighter anomalous scatterers ($Z \leq 20$) are also prevalent in biological macromolecules, intrinsically sulfur ($Z = 16$) in proteins and phosphorus ($Z = 15$) in nucleic acids. These and other lighter elements may also bind specifically to crystallized macromolecules as ions [*e.g.* PO_4^{3-}, SO_4^{2-}, Na^+ ($Z = 11$), Mg^{2+}

(Z = 12), Cl^- (Z = 17), K^+ (Z = 19) and Ca^{2+} (Z = 20)] or as constituents of ligands (*e.g.* ATP). The resonant edges of these lighter elements may not be readily accessible; however, off-resonance anomalous scattering from these light elements increases as the X-ray energy is lowered and, although the signals may be weak, effective experiments are possible for *de novo* structure determination. The proof-of-principle for native SAD phasing was first demonstrated with the structure of crambin (Hendrickson & Teeter, 1981). The subsequent development of density-modification procedures (Wang, 1985; Chen *et al.*, 1991) led to more effective applications of native SAD to proteins (Dauter *et al.*, 1999; Liu *et al.*, 2000) and nucleic acids (Dauter & Adamiak, 2001). These and other important advances have been reviewed (Doutch *et al.*, 2012; Hendrickson, 2013; Dauter, 2006).

A further advance has come with procedures for enhancing the signal-to-noise ratios for anomalous diffraction signals by incorporating data from multiple crystals (Liu *et al.*, 2011). The more recent extension of such multi-crystal procedures to measurements at 7 keV, where light-atom anomalous diffraction signals are increased, has led to routine and robust *de novo* structure determination of native SAD structures (Liu *et al.*, 2012, 2013). Real-life applications have been made in determining various structures at resolutions between 2.3 and 2.8 Å with unique protein sizes of between 144 and 1200 amino-acid residues. Most recently, multi-crystal data at 7 keV have been used for the native SAD analysis of viral protein NS1 at a marginal resolution of 3.0 Å (Akey *et al.*, 2014), further demonstrating the utility of multiple crystals in *de novo* structure determination of native macromolecules.

In comparison to heavier atom SAD phasing, native SAD phasing requires measurements at energy lower than 9 keV to realise appreciable anomalous signals. Complications such as air scattering, absorption, sample size, radiation damage and detector geometry may arise when performing experiments at lower energy (Liu *et al.*, 2013; Weiss *et al.*, 2001, 2005). As a compromise to mitigate such effects, we have limited our previous native SAD phasing to 7.1 keV (the Fe K edge). Nevertheless, the advantage of increased anomalous signals at lower X-ray energy is clear; Bijvoet differences, on which SAD phasing depends, are proportional to the imaginary component of anomalous scattering, f'', and f'' increases as energy is lowered towards the K-edge resonances of low-Z elements. Thus, the f'' of sulfur increases by 36%, from 0.70 to 0.95 electrons, when the energy is reduced from 7.1 to 6.0 keV. The adverse effects of air scattering and absorption are largely eliminated by using a helium beam path, and our calculations show that the effects of sample absorption are readily addressed by using appropriately smaller crystals (Liu *et al.*, 2013). With such attention to experimental conditions, the previously suggested X-ray energy of 6 keV (λ = 2.1 Å; Mueller-Dieckmann *et al.*, 2005) may be routinely suitable for multi-crystal native SAD phasing. In this study, we compare the quality of diffraction data and phasing statistics at 7 and 6 keV (Cr K edge) using a previously solved test problem, and we perform three multi-crystal native SAD applications at 6 keV. One is a test problem at 3.2 Å resolution and two are real-life applications for solving novel membrane-protein structures at about 3.0 Å resolution. Collectively, we conclude that multi-crystal native SAD phasing can be more effectively achieved at 6 keV than at 7 keV if performed properly.

2. Materials and methods

2.1. Sample preparation

Protein production and crystallization of CysZ from *Idiomarina loihiensis* were carried out as described previously (Liu *et al.*, 2012).

Human epidermal growth-factor receptor kinase domain (EGFRK; residues 695–1022) was expressed in baculovirus-driven insect cells and was purified following existing procedures (Stamos *et al.*, 2002). Crystallization experiments were performed by the hanging-drop vapor-diffusion method at room temperature. Crystals of EGFRK were produced by equilibrating equal volumes of protein and well solution consisting of 1.0 *M* sodium/potassium tartrate, 100 m*M* 2-(*N*-morpholino)ethanesulfonic acid (MES) pH 7.0 followed by two cycles of macroseeding. Cubic crystals of ∼100 µm on a side were cryoprotected by supplementation with 20% glycerol before harvesting them into liquid nitrogen.

The details of the expression, purification and crystallization of ThiT will be reported separately (Guo *et al.*, 2014). Briefly, ThiT from *Listeria monocytogenes* was concentrated to 5 mg ml^{-1} and used for crystallization by mixing equal volumes of ThiT solution and well solution consisting of 0.5% *n*-octyl-β-D-glucoside (β-OG), 0.2 *M* ammonium acetate, 0.1 *M* trisodium citrate dihydrate pH 5.6, 30%(*w*/*v*) polyethylene glycol (PEG) 4000, 3%(*v*/*v*) glycerol. Crystallization experiments were performed by the sitting-drop vapor-diffusion method at room temperature. Crystals appeared after one week and took seven to eight weeks to mature. Crystals with dimensions of about 75 × 75 × 75 µm were cryocooled directly in liquid nitrogen.

The details of the expression, purification and crystallization of YetJ have been reported elsewhere (Chang *et al.*, 2014). Briefly, YetJ from *Bacillus subtilis* (strain 168) was overexpressed in *Escherichia coli* and purified by affinity column, ion-exchange and gel-filtration column chromatography. Concentrated protein at ∼13 mg ml^{-1} in 25 m*M* Tris–HCl pH 8.0, 400 m*M* NaCl, 0.08%(*v*/*v*) detergent C10E5 was used for crystallization by mixing an equal volume of protein solution with well solution consisting of 30% PEG 600, 100 m*M* $CaCl_2$, 100 m*M* Tris–HCl pH 8.0. Crystallization experiments were performed by the sitting-drop vapor-diffusion method at 4°C. Crystals are thin hexagonal plates of width ∼75 µm and thickness 10–20 µm. Crystals were harvested directly into liquid nitrogen without the addition of cryoprotectants.

2.2. Beamline setup and diffraction data acquisition

Native SAD data sets were collected on NSLS beamline X4A using a Quantum 4R detector at a cryogenic temperature of 100 K. The X-ray energies were calibrated by fluorescence

Table 1
Data-collection and reduction statistics for two CysZ crystals at 6 and 7 keV.

Values in parentheses are for the highest resolution shell.

Crystal	CysZ-1		CysZ-2			
Data	6 keV (first)	7 keV (second)	6 keV (second)	7 keV (first)	6 keV merged	7 keV merged
Unit-cell parameters						
a (Å)	128.92	128.97	128.83	128.79	128.87	128.88
b (Å)	81.49	81.48	81.34	81.30	81.42	81.39
c (Å)	100.31	100.38	100.20	100.14	100.25	100.26
β (°)	124.96	124.93	124.91	124.94	124.93	124.93
No. of frames	360	360	360	360	720	720
Rotation span (°)	360	360	360	360	720	720
Bragg spacings (Å)	40–3.20 (3.29–3.20)	40–3.20 (3.29–3.20)	40–3.20 (3.29–3.20)	40–3.20 (3.29–3.20)	40–3.20 (3.29–3.20)	40–3.20 (3.29–3.20)
Total reflections	102357	103490	102808	104401	205102	207799
Unique reflections	13722	13829	13725	13807	14019	14007
Multiplicity	7.5 (6.7)	7.5 (7.2)	7.5 (6.8)	7.6 (7.5)	14.6 (12.8)	14.8 (14.1)
Completeness (%)	98.2 (95.0)	98.8 (95.9)	98.1 (95.7)	98.9 (97.1)	99.8 (98.3)	99.8 (98.7)
$R_{\rm meas}$†	0.043 (0.205)	0.039 (0.158)	0.052 (0.247)	0.044 (0.155)	0.066 (0.268)	0.057 (0.204)
$R_{\rm p.i.m.}$‡	0.022 (0.108)	0.020 (0.082)	0.027 (0.129)	0.022 (0.079)	0.024 (0.103)	0.020 (0.075)
$\langle I/\sigma(I)\rangle$§	40.8 (10.1)	45.4 (14.4)	34.1 (7.9)	41.7 (13.8)	39.7 (10.9)	44.5 (16.2)
$\langle \Delta F/\sigma(\Delta F)\rangle$¶	1.10 (0.86)	0.98 (0.80)	1.04 (0.86)	0.99 (0.85)	1.01 (0.79)	0.84 (0.70)
Anomalous CC†† (%)	35.5	32.9	28.9	25.9	22.8	12.6
FOM	0.275	0.264	0.252	0.309	0.279	0.300
MapCC before DM (%)	27.9	23.4	23.1	22.2	31.3	30.0
MapCC after DM (%)	46.1	36.0	39.6	32.9	51.7	49.9

† $R_{\rm meas}$ is the redundancy-independent (multiplicity-weighted) $R_{\rm merge}$ as reported from *SCALA*. ‡ $R_{\rm p.i.m.}$ is the precision-indicating (multiplicity-weighted) $R_{\rm merge}$ as reported from *SCALA*. § $\langle I/\sigma(I)\rangle = \langle\langle I(hkl)\rangle/\langle\sigma[\langle I(hkl)\rangle]\rangle\rangle$, where $\langle I(hkl)\rangle$ is the weighted mean of all measurements for a 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). ¶ $\langle \Delta F/\sigma(\Delta F)\rangle$ is the average anomalous signal from data truncated to $d_{\rm min}$ = 3.5 Å. The values were derived using *CCP*4 programs and were computed by *SFTOOLS* as $\langle \Delta F/\sigma(\Delta F)\rangle$ where $\Delta F = |F(h)| - |F(-h)|$. †† Anomalous correlation coefficient evaluated from data truncated to $d_{\rm min}$ = 3.2 Å.

Table 2
Data-collection and reduction statistics for EGFRK at 6 keV.

Notes are the same as for Table 1 except that anomalous CC, RACC and $\Delta F/\sigma(\Delta F)$ were evaluated from data truncated to $d_{\rm min}$ = 4.0 Å. In this table and Tables 3 and 4, the use of the inverse-beam mode of data collection is denoted by $N \times 2$, where N is the number of frames.

Crystal/data	EGFRK-1	EGFRK-2	EGFRK-3	EGFRK-4	Merged 1 to 4	Merged (3)
Unit-cell parameter a (Å)	146.17	146.60	146.86	145.72	146.34	146.16
No. of frames	720 × 2	720 × 2	720 × 2	720 × 2	5760	4320
Rotation span (°)	360 × 2	360 × 2	360 × 2	360 × 2	2880	2160
Bragg spacings (Å)	40–3.20 (3.29–3.20)	40–3.20 (3.29–3.20)	40–3.20 (3.29–3.20)	40–3.20 (3.29–3.20)	40–3.20 (3.29–3.20)	40–3.20 (3.29–3.20)
Total reflections	742205	752241	751501	728119	2966150	2218779
Unique reflections	8743	8810	8867	8666	8749	8731
Multiplicity	84.9 (69.9)	85.4 (75.8)	84.8 (71.7)	84.0 (67.4)	339.0 (285.4)	254.1 (211.0)
Completeness (%)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)	100.0 (100.0)
$R_{\rm meas}$	0.122 (1.519)	0.199 (3.123)	0.236 (4.250)	0.118 (1.660)	0.319 (2.623)	0.229 (2.013)
$R_{\rm p.i.m.}$	0.018 (0.250)	0.030 (0.498)	0.035 (0.697)	0.018 (0.277)	0.024 (0.217)	0.020 (0.192)
$\langle I/\sigma(I)\rangle$	50.0 (4.6)	35.4 (2.7)	27.3 (1.8)	49.9 (4.0)	48.5 (5.6)	50.3 (6.1)
$\langle \Delta F/\sigma(\Delta F)\rangle$	1.52 (0.86)	1.24 (0.78)	1.10 (0.75)	1.69 (0.93)	1.34 (0.73)	1.39 (0.79)
RACC (%) [rank]	84.0 [1′]	80.1 [2′]	52.1 [4′]	73.2 [3′]	100.0	—
Anomalous CC (%)	73.6	60.3	56.2	78.3	51.1	70.7

scans from a Fe foil for a nominal 7 keV (actually 7.112 keV) and a Cr foil for a nominal 6 keV (actually 5.989 keV). For all data collections, a helium gas-purged path of 120 mm was inserted between the sample and detector to reduce air scattering and absorption. The sample-to-detector distance was thus fixed at ∼120 mm. The beamstop was placed near the detector at ∼120 mm from the crystal. The orientation of the crystals was random without special consideration of crystal alignment.

Two CysZ crystals (CysZ-1 and CysZ-2) of comparable size and diffraction capability were used for complete data collection at both 7 keV (Fe K edge) and 6 keV (Cr K edge). For crystal CysZ-1, data were first collected at 7 keV for 360° followed by a repeated data set at 6 keV for a further 360°. For crystal CysZ-2, data were first collected at 6 keV for 360° followed by a repeated data set at 7 keV for a further 360°. An oscillation angle of 1° was used for the CysZ data sets. For EGFRK, ThiT and YetJ crystals, data were only collected at 6 keV with an oscillation angle of either 1 or 0.5°. For native SAD data collection at 6 keV, a total of four EGFRK crystals, five ThiT crystals and 12 YetJ crystals were used, each to obtain a complete data set without appreciable radiation damage. We used standard Hampton Research loops for EGFRK crystals and MiTeGen dual-thickness micromounts for CysZ, ThiT and YetJ crystals. The beam size was adjusted to either 100 or 75 µm to best match the crystal size. An inverse-beam strategy was used for all data sets except ThiT-4 and CysZ tests. We note that the crystals used in each of these

Table 3
Data-collection and reduction statistics for ThiT at 6 keV.

Notes are the same as for Table 1 except that anomalous CC, RACC and $\Delta F/\sigma(\Delta F)$ were evaluated from data truncated to $d_{\min}$ = 4.0 Å.

Crystal/data	ThiT-1	ThiT-2	ThiT-3	ThiT-4	ThiT-5	Merged (5)
Unit-cell parameters						
a (Å)	95.64	95.38	95.70	95.49	95.878	95.65
c (Å)	125.21	124.66	124.83	124.81	125.28	125.03
No. of frames	360 × 2	220 × 2	180 × 2	360	360 × 2	2600
Rotation span (°)	360 × 2	220 × 2	180 × 2	360	360 × 2	2600
Bragg spacings (Å)	40–3.01 (3.09–3.01)	40–3.21 (3.29–3.21)	40–3.10 (3.18–3.10)	40–3.10 (3.18–3.10)	40–3.01 (3.09–3.01)	40–3.0 (3.08–3.00)
Total reflections	531220	286404	226488	255206	538576	1918529
Unique reflections	13578	11179	12392	12338	13674	13687
Multiplicity	39.1 (21.7)	25.6 (22.6)	18.3 (11.4)	20.7 (14.9)	39.4 (21.0)	140.2 (68.0)
Completeness (%)	99.7 (96.8)	99.8 (97.9)	99.6 (94.8)	99.6 (95.9)	99.9 (99.0)	99.8 (98.1)
$R_{\rm meas}$	0.195 (1.883)	0.182 (1.438)	0.163 (1.243)	0.254 (1.880)	0.206 (1.498)	0.271 (2.047)
$R_{\rm p.i.m.}$	0.043 (0.545)	0.049 (0.412)	0.052 (0.499)	0.076 (0.650)	0.045 (0.446)	0.031 (0.332)
$\langle I/\sigma(I)\rangle$	26.3 (1.7)	23.6 (3.0)	21.7 (1.9)	15.0 (1.5)	28.9 (2.3)	39.9 (3.0)
$\langle \Delta F/\sigma(\Delta F)\rangle$	1.10 (0.94)	0.95 (0.83)	0.86 (0.74)	0.86 (0.81)	1.11 (0.93)	1.31 (0.98)
RACC (%) [rank]	68.4 [2′]	53.3 [3′]	49.0 [4′]	40.0 [5′]	77.8 [1′]	100.0
Anomalous CC (%)	36.7	24.6	22.5	13.3	37.7	51.5

studies were sufficiently small that the transmitted anomalous signals exceed absorption (Liu *et al.*, 2013).

2.3. Data reduction and analysis

Native SAD data sets were individually indexed, integrated and corrected by *XDS* (Kabsch, 2010). Bijvoet mates were treated separately during data reduction by *XDS* and were kept separate after *XDS*. The *CCP*4 programs (Winn *et al.*, 2011) *POINTLESS* and *SCALA* (Evans & Murshudov, 2013; Evans, 2011) were used for further data combination, scaling and merging of single-crystal and multi-crystal data sets. For the two CysZ crystals collected at 7 and 6 keV, data from the same energy were merged together for further analyses (Table 1). For native SAD data sets, outlier data rejections were performed following our previously described procedures with unit-cell parameters and diffraction intensities extracted from single-crystal data sets (Liu *et al.*, 2013). Relative anomalous correlation coefficients (RACCs) of individual data sets to the merged data were calculated for further data validation and rejection. After outlier data rejection, the remaining data sets were scaled and merged for structural analyses. Data-collection and reduction statistics for single-crystal and multiple-crystal data sets are listed in Table 2 for EGFRK, Table 3 for ThiT and Table 4 for YetJ.

2.4. Structure determination

Substructure solutions were found by *SHELXD* (Sheldrick, 2010) and were further refined and completed by *Phaser* (Read & McCoy, 2011) for calculation of SAD phases at the data limit (Read & McCoy, 2011). Phases were generated for both enantiomers in each case, and then subjected to density modification as implemented in the *CCP*4 programs *DM* and *PHENIX* (Cowtan & Zhang, 1999; Adams *et al.*, 2011) to improve the phases and also to break the phase ambiguity.

PHENIX (Terwilliger *et al.*, 2008) was used for initial model building into the experimental electron-density maps at the data limit. The initial models were completed after iterative model building in *Coot* (Emsley *et al.*, 2010) and refinement using the *PHENIX* module *phenix.refine*. TLS parameters and isotropic *B* factors were refined without noncrystallographic symmetry (NCS) restraints. Friedel mates were treated as two reflections in all refinements. The stereochemistry of the refined structures was validated with *PROCHECK* (Laskowski *et al.*, 1993) and *MolProbity* (Chen *et al.*, 2010) for quality assurance. The refinement statistics for EGFRK, ThiT and YetJ are listed in Table 5.

3. Results

3.1. Comparison of anomalous diffraction at 7 and 6 keV

Based on values of f'', the effect of anomalous diffraction as measured in Bijvoet differences is expected to be increased by 36% at our nominal 6 keV (Cr *K* edge, 5.989 keV) over that at our nominal 7 keV (Fe *K* edge, 7.112 keV). We also expect less readily quantifiable changes such as absorption, background scattering and detector efficiency. To evaluate the effects of such factors experimentally, we combined data from two CysZ crystals (named CysZ-1 and CysZ-2), each contributing two data sets, one measured at 6 keV and one at 7 keV. In order to control for radiation damage, at least partially, we first collected a complete data set at 7 keV followed by a repeated data set at 6 keV for CysZ-1; we then reversed the order for data collection for CysZ-2. The $R_{\rm d}$ plot (Diederichs, 2006) was used to assure that radiation damage is not apparent after completion of data collection at two different energies. These four data sets were then scaled together and merged into separate two-crystal data sets: one for 6 keV and one for 7 keV (Table 1).

Idiomarina CysZ crystals diffract X-rays to Bragg spacings of about 2.3 Å; however, owing to detector limitations imposed by the placement of the helium path, data were collected and processed to 3.2 Å resolution. The two crystals are compatible with each other as judged by the very similar unit-cell parameters and correlated diffraction intensities. For the four single-crystal data sets (two from CysZ-1 and two from CysZ-2), the multiplicity and completeness are very

Table 4
Data-collection and reduction statistics for YetJ at 6 keV.

Notes are the same as for Table 1 except that anomalous CC, RACC and $\Delta F/\sigma(\Delta F)$ were evaluated from data truncated to $d_{\min}$ = 3.5 Å.

Crystal/data set	YetJ-1	YetJ-2	YetJ-3	YetJ-4	YetJ-5	YetJ-6	YetJ-7
Unit-cell parameters							
a (Å)	62.90	61.98	61.91	62.07	62.09	62.08	61.60
b (Å)	288.07	287.41	287.76	287.38	287.74	286.60	288.71
No. of frames	720 × 2	720 × 2	600 × 2	360 × 2	360 × 2	360 × 2	360 × 2
Rotation span (°)	360 × 2	360 × 2	300 × 2	180 × 2	360 × 2	360 × 2	360 × 2
Bragg spacings (Å)	40–3.19 (3.27–3.19)	40–3.18 (3.27–3.18)	40–3.21 (3.29–3.21)	40–3.19 (3.28–3.19)	40–3.20 (3.28–3.20)	40–3.20 (3.29–3.20)	40–2.80 (2.87–2.80)
Total reflections	465641	462985	384208	233004	457338	457168	530451
Unique reflections	6158	6174	6051	6138	6118	6077	8507
Multiplicity	75.6 (63.5)	75.0 (56.8)	63.5 (59.5)	38.0 (32.7)	74.8 (68.0)	75.2 (69.7)	62.4 (23.8)
Completeness (%)	99.7 (96.3)	99.7 (96.3)	99.8 (98.2)	99.8 (98.6)	99.8 (98.2)	99.9 (99.1)	96.9 (83.3)
$R_{\rm meas}$	0.201 (1.397)	0.385 (2.582)	0.236 (1.773)	0.347 (2.108)	0.279 (1.604)	0.281 (1.753)	0.129 (0.870)
$R_{\rm p.i.m.}$	0.030 (0.229)	0.057 (0.432)	0.039 (0.304)	0.073 (0.482)	0.042 (0.257)	0.042 (0.277)	0.021 (0.236)
$\langle I/\sigma(I)\rangle$	35.6 (4.3)	20.7 (2.2)	29.9 (3.2)	15.5 (2.0)	23.6 (3.7)	26.4 (3.6)	39.9 (4.3)
$\langle \Delta F/\sigma(\Delta F)\rangle$	1.03	0.88	0.93	0.84	0.92	0.96	1.39
RACC (%) [rank]	34.2 [5′]	25.5 [10′]	33.2 [6′]	21.2 [11′]	28.5 [9′]	29.6 [8′]	32.5 [7′]
Anomalous CC (%)	26.3	8.3	12.5	3.6	22.5	21.1	54.4

Crystal/data set	YetJ-8	YetJ-9	YetJ-10	YetJ-11	YetJ-12	YetJ1–12	Merged (10)
Unit-cell parameters							
a (Å)	61.55	61.79	61.81	61.60	61.62	61.87	61.87
b (Å)	287.13	290.13	290.44	289.75	289.94	288.30	288.40
No. of frames	360 × 2	360 × 2	360 × 2	360 × 2	360 × 2	10560	9120
Rotation span (°)	360 × 2	360 × 2	360 × 2	360 × 2	360 × 2	8160	6720
Bragg spacings (Å)	40–2.80 (2.87–2.80)	40–2.81 (2.88–2.81)	40–2.81 (2.88–2.81)	40–3.00 (3.07–3.00)	40–2.81 (2.88–2.81)	40–2.80 (2.87–2.80)	40–2.80 (2.87–2.80)
Total reflections	536689	552887	551875	513974	548275	5694070	4706387
Unique reflections	8590	8797	8656	7298	8542	8784	8784
Multiplicity	62.5 (21.8)	62.8 (20.7)	63.8 (23.4)	70.4 (39.3)	64.2 (24.4)	648.2 (96.2)	535.8 (77.7)
Completeness (%)	98.3 (90.2)	99.3 (94.5)	97.9 (86.9)	99.7 (97.4)	97.3 (85.8)	98.9 (88.7)	98.9 (88.7)
$R_{\rm meas}$	0.270 (2.073)	0.153 (1.579)	0.168 (1.576)	0.264 (1.636)	0.174 (1.487)	0.366 (1.720)	0.359 (1.896)
$R_{\rm p.i.m.}$	0.044 (0.592)	0.025 (0.459)	0.027 (0.430)	0.041 (0.346)	0.028 (0.397)	0.017 (0.227)	0.019 (0.205)
$\langle I/\sigma(I)\rangle$	19.0 (1.8)	35.2 (2.0)	34.0 (2.1)	26.3 (2.8)	33.1 (2.4)	57.7 (3.3)	56.8 (3.0)
$\langle \Delta F/\sigma(\Delta F)\rangle$	0.97	1.22	1.19	0.98	1.25	1.50	1.51
RACC (%) [rank]	19.2 [12′]	62.9 [1′]	57.7 [2′]	43.1 [4′]	55.9 [3′]	100.0	—
Anomalous CC (%)	31.2	39.6	37.4	17.9	48.8	56.5	55.7

similar. The data sets do differ systematically, however; the 7 keV data sets have better $R_{\rm meas}$ and average $I/\sigma(I)$ statistics, whereas the 6 keV data sets show better anomalous signals in terms of average $\Delta F/\sigma(\Delta F)$ and anomalous correlation coefficient (CC) (Table 1). As a consequence, the stronger anomalous signals benefit native SAD analysis, as shown by higher figure-of-merit (FOM) and map correlation coefficient (MapCC) values before as well as after density modification (Table 1).

To better characterize the differential effects of 6 keV *versus* 7 keV radiation on anomalous signals from multiple crystals, we merged the data at 6 and 7 keV, each from two crystals. As for the single-crystal data sets, the merged 7 keV data have lower $R_{\rm meas}$ and higher average $I/\sigma(I)$ relative to the merged 6 keV data. Nevertheless, the anomalous diffraction signals, such as anomalous CC and average $\Delta F/\sigma(\Delta F)$, are all higher for the 6 keV data (Table 1). Fig. 1 shows a detailed comparison of various statistics of the two merged data sets. The 6 keV data have higher anomalous CC, average $\Delta F/\sigma(\Delta F)$, average $\Delta F/F$, Bijvoet difference Fourier peaks and MapCC values before and after density modification. The only decreased indicator is the overall diffraction intensity as defined as average $I/\sigma(I)$, which is 10.7% lower than at 7 keV. This deterioration in overall diffraction quality at 6 keV may be owing to increased sample absorption and scattering as well as non-optimal detector efficiency.

3.2. Native SAD phasing at 6 keV

After seeing the benefits from using lower-energy X-rays at 6 keV, we next tested whether we could extend our multi-crystal native SAD phasing to 6 keV for robust and routine native SAD structure determination. We carried out structure determinations on three proteins, one for re-solving a known structure and two for real-life applications to novel membrane proteins.

3.2.1. EGFRK. Human epidermal growth-factor receptor kinase domain (EGFRK) is a drug target for pharmaceutical development for the treatment of various kinds of cancers (Kumar *et al.*, 2008). EGFRK contains 318 residues and 15 S atoms. The estimated Bijvoet diffraction ratio is 1.5% at 6 keV. Growing EGFRK crystals has not been straightforward and macroseeding was reportedly necessary to obtain crystals of about 100 µm or larger. The crystal structure of EGFRK has been reported at 2.6 Å resolution (Stamos *et al.*, 2002); however, the crystals that we obtained diffracted only weakly to ∼3.2 Å resolution as demonstrated by an outer-shell $R_{\rm meas}$ of over 1.0 for all single-crystal data sets. We could have put

Table 5
Structure-refinement statistics.

Data	EGFRK	ThiT	YetJ
Space group	$I23$	$P3_121$	$P6_522$
Resolution (Å)	3.2	3.0	2.8
No. of reflections	16644	25496	15009
R_{work}/R_{free}	0.189/0.209	0.198/0.234	0.221/0.260
No. of atoms			
Total	2459	3172	1636
Protein	2459	2830	1632
Ligand/ion/detergent	—	342	4
Average B (Å^2)			
Overall	94.9	56.5	49.6
Protein	94.9	54.2	49.6
Ligand/ion	—	77.6	66.4
R.m.s. deviations			
Bond lengths (Å)	0.002	0.002	0.002
Bond angles (°)	0.494	0.660	0604
PDB code	4tks	4tkr	4tkq

more effort into optimizing the crystallization conditions; but these crystals served as excellent testing examples for native SAD phasing at lower resolution.

Since these EGFRK crystals are in space group $I23$, a very high multiplicity of ∼85 was achieved even in single-crystal data sets (Table 2). A total of four native SAD data sets were collected, each from a different single crystal, and the data were processed to 3.2 Å Bragg spacings. Crystal EGFRK-3 was identified as an outlier because it deviates from the others by ∼8% in the diffraction dissimilarity analysis (Fig. 2*b*), although it was compatible based on unit-cell variation and RACC analyses (Figs. 2*a* and 2*c*). We thus merged the remaining three data sets for structure determination. Figs. 2(*d*), 2(*e*) and 2(*f*) show diffraction signals of single crystals as well as the merger of the three compatible crystals. Because of the extremely high redundancy reached in single-crystal data, there is no improvement in anomalous CC, average $\Delta F/\sigma(\Delta F)$ or even average $I/\sigma(I)$. This is consistent with our previous analysis showing an asymptotic effect on diffraction-quality indicators with respect to increasing multiplicity (Liu *et al.*, 2011, 2012).

Both merged and single-crystal data sets were tested for substructure determination. Although measures of anomalous signal had similar values for single-crystal data sets as for the merged data, none of them supported substructure determination. In contrast, successful solutions identifying the 15 expected sulfur sites were found from the merged data as demonstrated by a clean separation of correct *SHELXD* solutions (CC_{weak} = 18.5%; CC_{all} = 42.3%) from the random results (Fig. 3*a*). We refined and completed the *SHELXD* substructure using the *Phaser* module in *PHENIX* and calculated SAD phases with an overall figure of merit (FOM) of 0.31. After density modification using the *RESOLVE* module in *PHENIX*, a traceable electron-density map was obtained (Fig. 3*b*). A partial model of 193 residues (60.7%) was then built automatically by *PHENIX* and refined to R and R_{free} values of 0.355 and 0.418, respectively, which validated the correctness of the three-crystal SAD phases. Further model building was readily accomplished by alignment of PDB entry 1m17 (Stamos *et al.*, 2002) with the partial model, whereupon TLS refinement against our SAD amplitudes, treating them as separate reflections, completed the analysis. The refined model has excellent geometry and low R_{free}

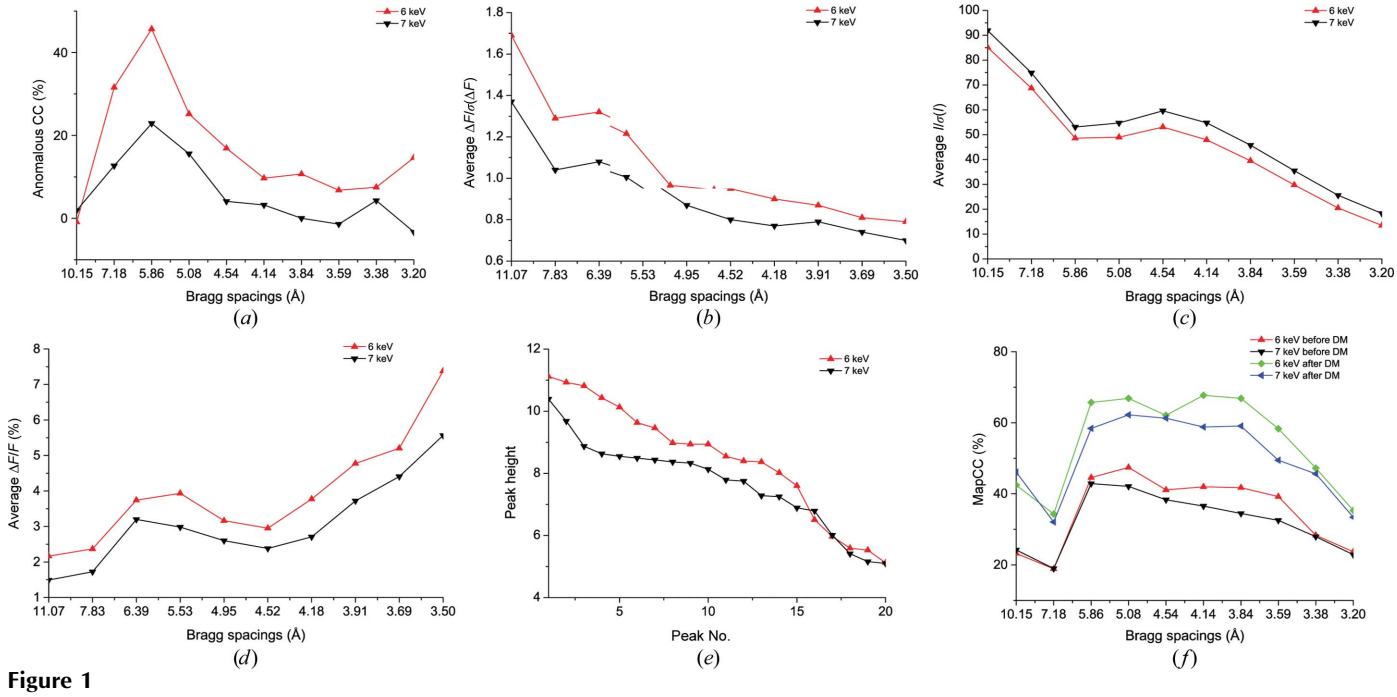


Figure 1
Evaluation of two-crystal CysZ data at 6 and 7 keV. (*a*) Anomalous correlation coefficients. (*b*) Average $\Delta F/\sigma(\Delta F)$. (*c*) Average $I/\sigma(I)$. (*d*) Average $\Delta F/F$. (*e*) Heights of Bijvoet difference Fourier peaks. Ordered peak-height profiles are shown for the map from the 6 keV data set (identified by the inset keys and shown in red) and for the map from the 7 keV data set (identified by the inset keys and shown in black). Peak heights are given in units of root-mean-squared deviation over the entire respective Fourier synthesis. (*f*) MapCC before and after density modification (DM). $|S| = 2\sin\theta/\lambda$, which is labeled as the Bragg spacing, $d = 1/|S|$.

(Table 5). Fig. 3(*c*) shows a ribbon diagram of the EGFRK structure re-solved by three-crystal native SAD phasing at a resolution of 3.2 Å using an X-ray energy of 6 keV. Although the structure is known, we did not need prior structural information to assure success in our native SAD phasing.

3.2.2. ThiT. ThiT is a novel membrane protein used in the uptake of thiamine and its derivatives into cells. ThiT from *L. monocytogenes* was identified from the NYCOMPS pipeline as a promising target for structural analysis. Crystallized in space group $P3_121$, ThiT forms a dimer in the asymmetric unit. Each subunit has 186 residues including eight S atoms, giving an estimated Bijvoet diffraction ratio of 1.5% at 6 keV. Diffraction data from five native ThiT crystals were collected and processed to marginal limits at a Bragg spacing of 3.0 Å (Table 3). Similar to the EGFRK data sets, the outer-shell $R_{\rm meas}$ values are all beyond 1.0 in the single-crystal ThiT data sets, indicating relatively poor diffraction. Outlier crystal rejection analyses indicated that all five crystals are compatible (Figs. 4*a*, 4*b* and 4*c*). The merged five-crystal data set had an accumulated multiplicity of 140. Compared with the best single-crystal data set ThiT-5, anomalous CC, average $\Delta F/\sigma(\Delta F)$ and average $I/\sigma(I)$ are all increased in the merged data set, by 36.6, 18.0 and 38.1%, respectively. The enhancements of both average $\Delta F/\sigma(\Delta F)$ and average $I/\sigma(I)$ are for all resolution shells (Figs. 4*d*, 4*e* and 4*f*), suggesting an extension of useful anomalous and normal signals from averaging.

For substructure determination by *SHELXD*, there is no special consideration of the number of molecules in the asymmetric unit. The solvent-content estimation suggested a solvent content of 69% for two molecules and 53% for three molecules per asymmetric unit. For *SHELXD* runs, we performed substructure searches seeking 10, 15, 20 and 30 sites and the best substructure solution came from the 15-site search with the five-crystal set, which gave highest $CC_{\rm weak}$ and $CC_{\rm all}$ values of 14.8 and 35.4%, respectively (Fig. 5*a*), giving the answer of two molecules in the asymmetric unit. Similar to the case for EGFRK, there were no correct *SHELXD* solutions from any of single-crystal data sets.

The *SHELXD* substructure was used for automatic structure determination using *PHENIX* as for EGFRK. The FOM from *PHENIX* is 0.267 and the twofold NCS was automatically identified from the substructure and was applied for density modification and model building. The density-modified electron-density map was of excellent quality and almost indistinguishable from the refined map (Fig. 5*b*). Owing to the high quality of the electron-density map, *PHENIX* was able to build 351 residues with refined R and $R_{\rm free}$ values of 0.33 and 0.38, respectively. Further cycles of model building and refinement were performed iteratively in *Coot* and *PHENIX*. Bijvoet mates were separated for final refinements and Bijvoet difference Fourier peaks helped in the identification and validation of both S and P atoms of the ligand thiamine diphosphate (Guo *et al.*, 2014). In fact, these atoms had already been found by *Phaser* during the substructure-completion procedure and contributed to the calculation of the SAD phases. The refined model is a quasi-dyad symmetric dimer (Fig. 5*c*). To test whether NCS is required to solve the structure, we re-ran the SAD phasing by turning off the NCS feature. Although fewer residues were built automatically (278 residues out of 372), the refined R and

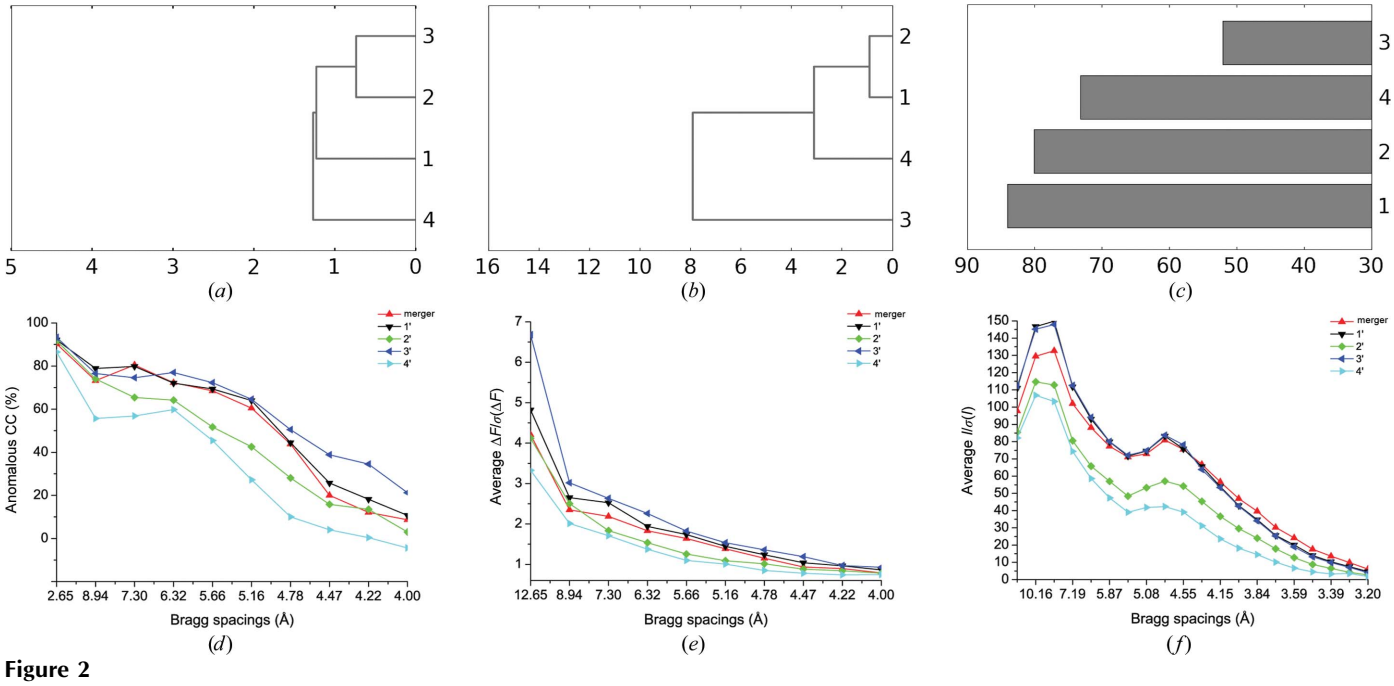
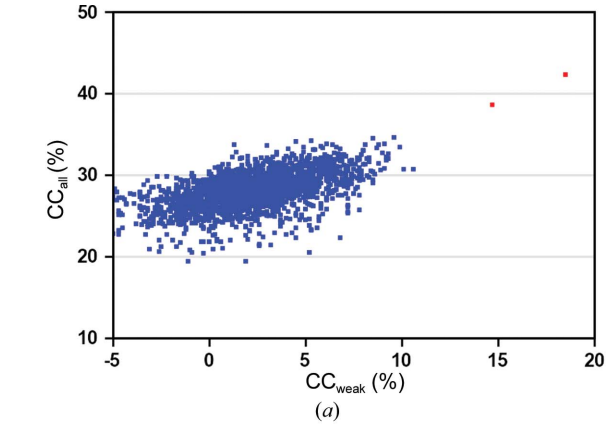
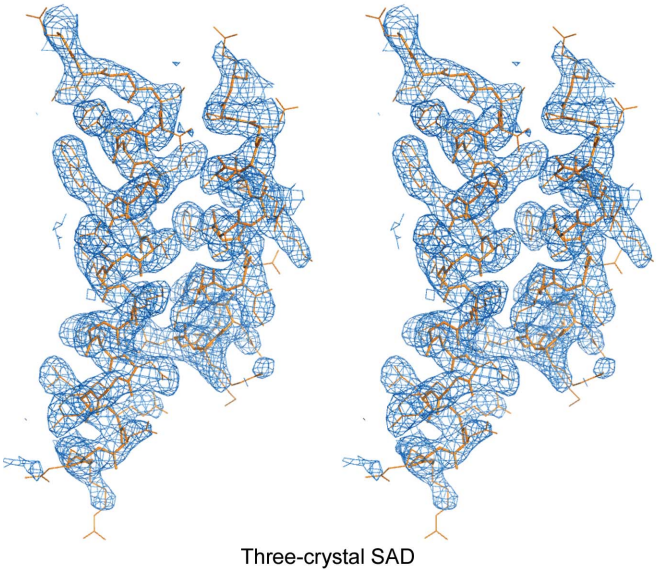
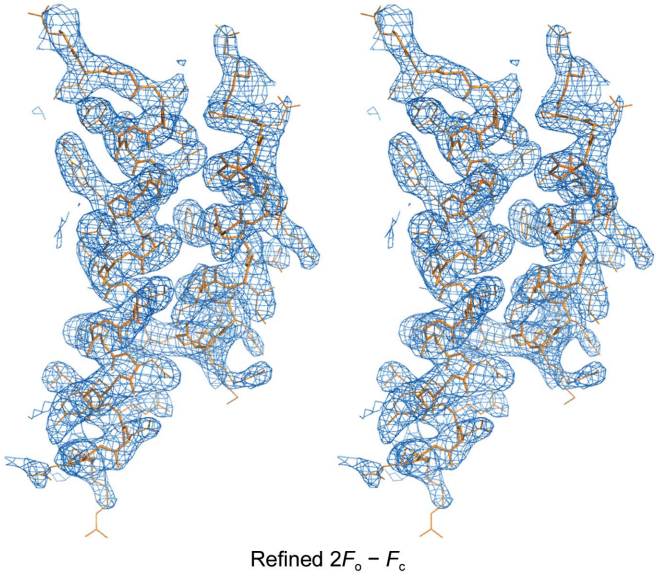


Figure 2
Diffraction data analyses of EGF receptor kinase domain (EGFRK). (*a*) Unit-cell variation. (*b*) Diffraction dissimilarity. (*c*) Relative correlation coefficient. (*d*, *e*, *f*) Diffraction signal strength in single-crystal and multi-crystal data. Anomalous CC (*d*), average $\Delta F/\sigma(\Delta F)$ (*e*) and average $I/\sigma(I)$ (*f*) for single-crystal data and merged data as a function of Bragg spacings. The overall diffraction dissimilarity between crystals i and j is defined as $D_{i,j} = 1.0 - C_{i,j}$, where $C_{i,j}$ is the correlation coefficient between all Bragg intensities in common between the two diffraction data sets. Data sets are identified by the inset keys.



(*a*)



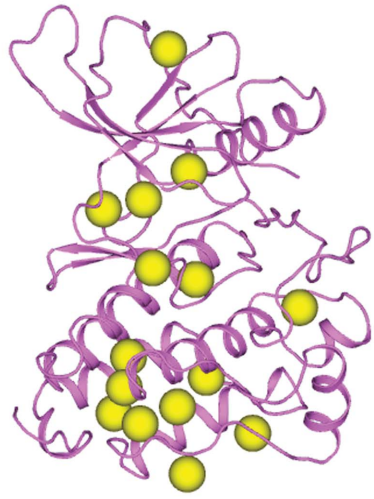


(*b*)

R_{free} for the partial structure were 0.43 and 0.50, respectively, indicating that the structure can also be solved with no need for the twofold NCS.

3.2.3. YetJ. YetJ is a previously uncharacterized membrane protein from the BI-1 family, which includes proteins important in the suppression of cell death and the regulation of calcium flux. Like ThiT from *L. monocytogenes*, YetJ from *B. subtilis* was identified to be suitable for structural analysis through high-throughput screening at NYCOMPS. YetJ contains 214 residues including eight S atoms, corresponding to an estimated Bijvoet diffraction ratio of 1.4% at 6 keV. Compared with the crystals used in the other 6 keV experiments reported here, the YetJ crystals are small. They are hexagonal plates typically 50–75 µm in width and 10–20 µm thick. Attempting to match the crystal size, we used a 75 µm beam for all YetJ data collections to reduce background diffraction. A total of 12 data sets were collected, each from a different single crystal with an oscillation angle of either 1.0 or 0.5°. These crystals diffracted variably to Bragg limits ranging from 2.8 to 3.2 Å (Table 4).

From our comparative analyses, we identified two data sets as outliers for rejection owing to incompatibility in diffraction dissimilarity, although they are comparable with the other crystals in unit-cell parameters and relative anomalous correlation coefficient (RACC) (Figs. 6*a*, 6*b* and 6*c*). After rejection of these two crystals (YetJ-7 and YetJ-8), the remaining data were merged and used for structure determination. Compared with the best single-crystal data set, the merged data show increased values for anomalous CC, average $\Delta F/\sigma(\Delta F)$ and average $I/\sigma(I)$ by 12.9, 7.9 and 42.4%, respectively. Enhancements were observed over almost the entire range (Figs. 6*d*, 6*e* and 6*f*), more significantly for the average $I/\sigma(I)$ distributions relative to those for anomalous CC and average $\Delta F/\sigma(\Delta F)$. In fact, the anomalous CC values



(*c*)

Figure 3
Phasing of EGFRK. (*a*) *SHELXD* substructure determination. Successful solutions are colored red and random solutions are colored blue. (*b*) Representative portion of the experimental electron-density map and its comparison with the refined map. Electron-density distributions calculated from phases after density modification at 3.2 Å are shown as sky-blue meshes contoured at 1.5σ. For reference, the model of the refined structure is shown as sticks and lines (orange). (*c*) Ribbon diagram of the model showing S atoms.

in single-crystal data sets were rather noisy, with an unrealistic increase at high Bragg angles. By comparison, the merged anomalous CC was more reasonable, being with much less high-angle noise. It is noted that such high-angle data fluctuations were not obvious in overall diffraction as defined by average $I/\sigma(I)$, perhaps underscoring that weak anomalous diffraction rather than inadequate data strength is responsible for the behavior of the anomalous CC and average $\Delta F/\sigma(\Delta F)$ values.

The YetJ protein was crystallized in space group $P6_522$ and the estimated solvent content is 63% with only one molecule in the asymmetric unit. Substructure determination by *SHELXD* was straightforward by searching for the expected number of sulfur sites. From 2000 tries at cutoffs of 3.8 Å for resolution and 1.3 for $E_{\rm min}$, successful substructures were found with $CC_{\rm weak}$ and $CC_{\rm all}$ of 21.1 and 44.2%, respectively (Fig. 7*a*). Again, there was no *SHELXD* solution from any of the YetJ single-crystal data sets. The *SHELXD* substructure was used for automated structure determination by *PHENIX* at 2.8 Å resolution, resulting in a FOM of 0.25 and a traceable electron-density map (Fig. 7*b*). Initial model building by *PHENIX* found 131 residues with a refined partial model of R and $R_{\rm free}$ of 0.46 and 0.52, respectively. Additional cycles of model building by *phenix.autobuild* produced a nearly complete model by building 200 residues out of 214 with refined R and $R_{\rm free}$ of 0.238 and 0.290, respectively. Further model completion and refinement were performed in *Coot* and *phenix.refine*. With SAD amplitudes being treated as separate reflections for refinement, we were able to identify three chloride ions and one loosely bound calcium ion on the surface of the protein by our previously defined method (Liu *et al.*, 2013). The statistics of the refined YetJ model are given in Table 5 and its ribbon diagram is shown in Fig. 7(*c*).

4. Discussion

4.1. Anomalous signal indicators

Demonstrable anomalous signals are key to successful substructure determination and SAD phasing (Dauter, 2006). For the three examples in this study, the estimated Bijvoet diffraction ratio is around 1.5% at 6 keV. In our previous analysis of anomalous signals in multiple crystal data, we have shown the utility of anomalous signal indicators such as anomalous CC and average $\Delta F/\sigma(\Delta F)$ (Liu *et al.*, 2011). The merged five-crystal ThiT data and ten-crystal YetJ data have significantly enhanced average $\Delta F/\sigma(\Delta F)$ compared with their single-crystal counterparts (Figs. 4*e* and 6*e*); however, we did not observe an enhancement in average $\Delta F/\sigma(\Delta F)$ values for the merged three-crystal EGFRK data (Fig. 2*e*). In fact, the merged three-crystal EGFRK data ranked only third based on either average $\Delta F/\sigma(\Delta F)$ or anomlaous CC values (Table 2, Figs. 2*d* and 2*e*). Nonetheless, only the merged data set supported substructure determination for EGFRK; none of the single-crystal data sets yielded a solution. Evidently, the merging really did produce superior anomalous signals.

To further explore the phasing effectiveness of single *versus* merged EGFRK data, we performed SAD phasing for individual data sets using the substructure that could only be found from the merged data set. Fig. 8 compares phasing statistics from the single-crystal data sets with the merged counterparts both before and after density modification. Here,

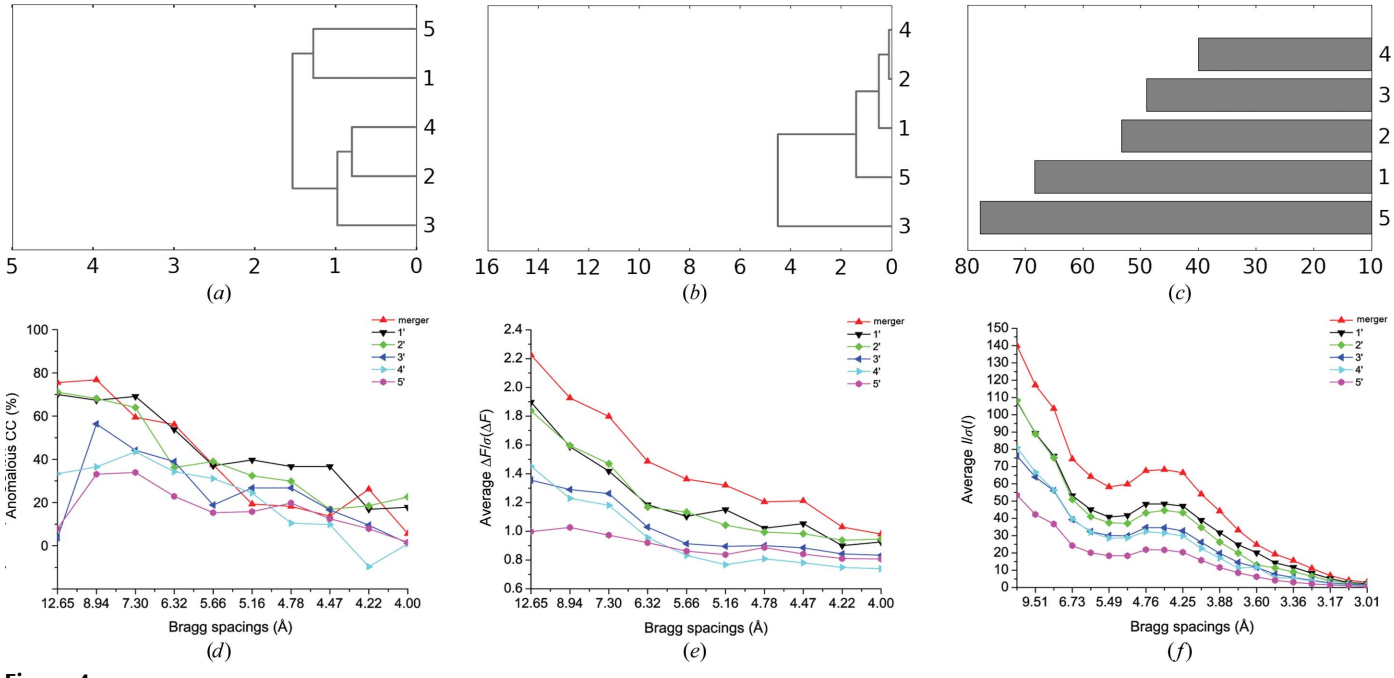
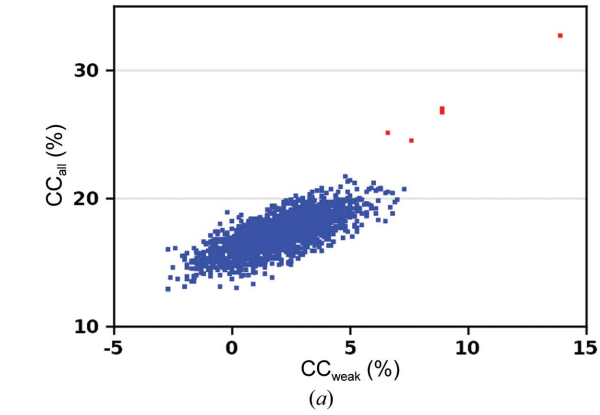
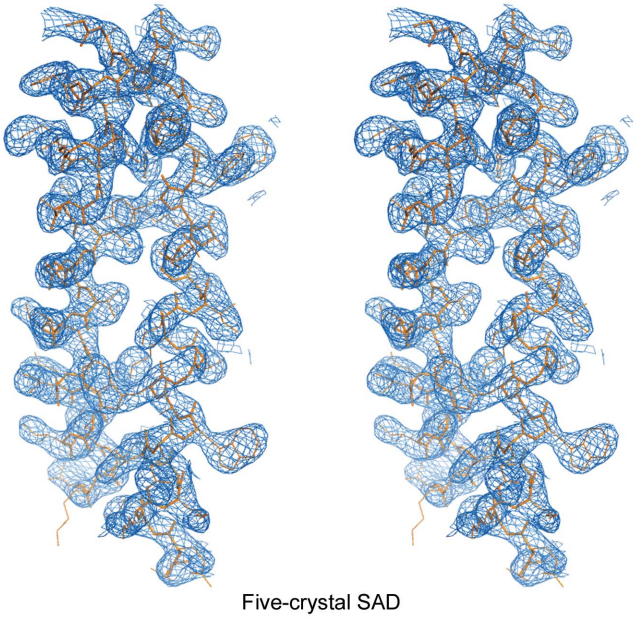


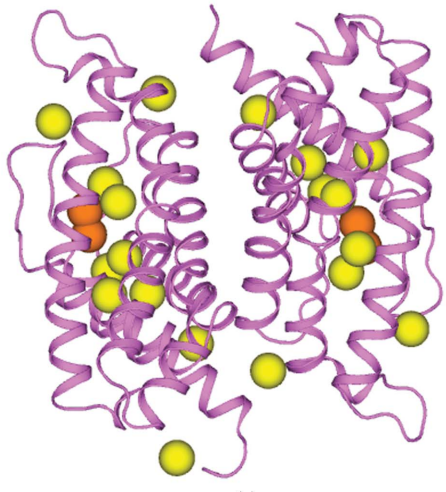
Figure 4
Diffraction data analyses of ThiT. (*a*) Unit-cell variation. (*b*) Diffraction dissimilarity. (*c*) Relative correlation coefficient. (*d*, *e*, *f*) Diffraction signal strength in single-crystal and multi-crystal data. Anomalous CC (*d*), average $\Delta F/\sigma(\Delta F)$ (*e*) and average $I/\sigma(I)$ (*f*) for single-crystal data and merged data as a function of Bragg spacings. Data sets are identified by the inset keys.



(*a*)



(*b*)



(*c*)

Figure 5
Phasing of ThiT. (*a*) *SHELXD* substructure determination. (*b*) Representative portion of the experimental electron-density map and its comparison with the refined map. (*c*) Ribbon diagram of the model showing S atoms. Electron-density maps were produced as described for Fig. 3.

in contrast to the statistics for anomalous signal indicators (Figs. 2*d* and 2*e*), the merged data do perform best in giving the highest MapCC values. The two MapCC plots also show that after including 420 frames of data, adding more frames into the single-crystal data sets did not improve the MapCC values either before or after density modification, suggesting that an asymptotic limit had been reached, as we had previously characterized (Liu *et al.*, 2012). The high symmetry of the *I*23 space group for EGFRK, giving an extremely high overall multiplicity of ∼85 in single-crystal data sets, makes this rapid approach to an asymptote unsurprisingly. What is remarkable, however, is the evident complementarity of the information from the individual data sets. The MapCC values for the merged data do continue to increase with as many as 660 frames being included, thus realising a special advantage in multiplicity from multiple crystals.

Why should the efficacy of merged data for substructure determination and phasing outperform the expectations from anomalous signal indicators? Behavior of this kind, as found for EGFRK, was also observed in the CysZ results reported here. Both at 6 and 7 keV, the anomalous CC values for the merged data sets were lower than for the corresponding single-crystal data sets, yet the MapCC phasing statistics improved with the mergers (Table 1). A clue for resolving the puzzle of phasing efficacy that defies anomalous signal statistics may be given in the frame-by-frame progressions of Fig. 8. As reported by MapCC values, asymptotic limits in phasing efficacy were reached for the individual data sets but these limits were exceeded for the merged data set. The very existence of an asymptotic limit with multiplicity suggests that a systematic error, such as uncorrected bias in Friedel-mate intensities, limits the expected improvements owing to statistical reduction of random errors. The further improvement that is observed upon merging from multiple crystals implies that a crystal-specific systematic error, not radiation damage, is the limiting factor. We suggest that slight variations among

the crystals with respect to physical characteristics, such as shape, may affect Bijvoet mate reflections systematically. Differing specific errors will result and the multiplicity coming from added crystals can address the problem effectively

4.2. Outlier data rejection

To merge data from multiple crystals, it is essential to assure that the crystals to be merged are all statistically compatible. Based on our experience with multi-crystal native SAD applications at 7 keV, we devised three criteria for rejection of crystals as outliers: unit-cell variation greater than 3σ, diffraction dissimilarity greater than 5% and relative anomalous correlation coefficient (RACC) less than 30%. Although we based these numbers on our experience with multi-crystal data at 7 keV, where native anomalous signals are weaker than at 6 keV, but absorption and background scattering are lower, these criteria remain useful for 6 keV applications. We have found variations, however.

One variation was found with EGFRK. Here, both the unit-cell variation and RACC tests suggested that all four EGFRK data sets are compatible with one another; however, EGFRK-3 differed by nearly 8% in the diffraction dissimilarity test. Including the EGFRK-3 data in fact deteriorated the overall data quality by decreasing the anomalous CC, average $\Delta F/\sigma(\Delta F)$ and average $I/\sigma(I)$ (Table 2). More seriously, the data merged from the four crystals did not support the substructure determination by *SHELXD*, further emphasizing the importance of outlier data rejection in multi-crystal-based native-SAD phasing.

Another variation on outlier data rejection came with YetJ. Based on diffraction dissimilarity analysis, YetJ-7 and YetJ-8 are two outliers. However, this is not obvious from the unit-cell variation and RACC tests. The subset consisting of YetJ-7 and YetJ-8 differs by as much as 14% in the diffraction dissimilarity test relative to the remaining subset. Here, we did find that the structure could still be solved even when including the two outlier data sets. On the other hand, in a comparison with the majority group consisting of ten crystals, the outlier subset makes a small contribution and is less likely to affect the overall phasing, as we have previously demonstrated for multi-crystal SeMet SAD phasing at low resolution (Liu *et al.*, 2011).

We present here a set of outlier rejections that permitted structure determinations. The question arises as to whether improvement might come from more stringent tests. For EGFRK and ThiT this was not possible; no substructure solutions could be obtained if even the poorest crystal were omitted. For YetJ, the substructure and phasing could be accomplished when omitting seven crystals (five besides the two diffraction dissimilarity outliers), and slightly better MapCC values were obtained when the two crystals with poorest RACC values were rejected as well as the two diffraction dissimilarity outliers. In retrospect, we might suggest RACC < 25% as an outlier test independent of other criteria.

4.3. Considerations in multi-crystal native SAD phasing

Multi-crystal data collection provides an effective way of strengthening anomalous signal-to-noise ratios. For an optimized native SAD phasing from multiple crystals, a few considerations may be noted. The first consideration is the number of frames or snapshots from a single crystal. To make

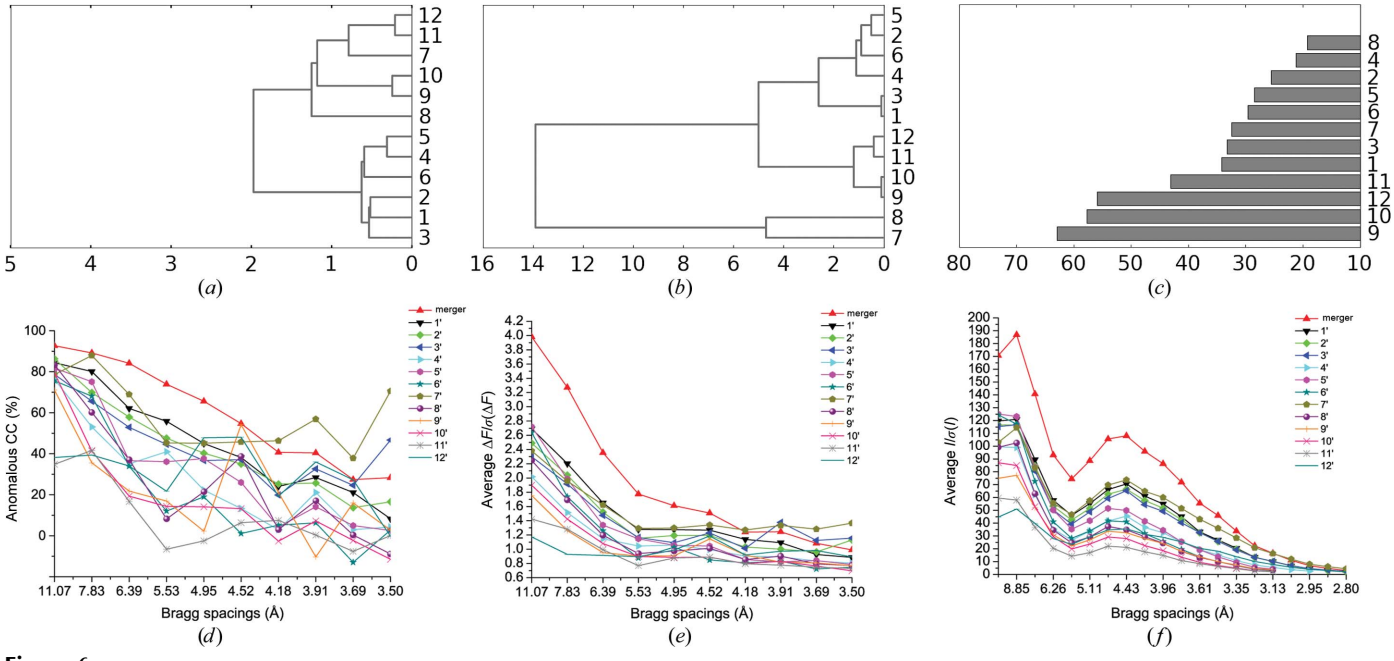


Figure 6
Diffraction data analyses of YetJ. (*a*) Unit-cell variation. (*b*) Diffraction dissimilarity. (*c*) Relative correlation coefficient. (*d*, *e*, *f*) Diffraction signal strength in single-crystal and multi-crystal data. Anomalous CC (*d*), average $\Delta F/\sigma(\Delta F)$ (*e*) and average $I/\sigma(I)$ (*f*) for single-crystal data and merged data as a function of Bragg spacings. Data sets are identified by the inset keys.

data scaling more reliable, it might be advantageous to use a low-dose approach, *e.g.* less than 5 MGy per crystal, to have complete single-crystal data sets, in general 360° of data. More ambitiously, multiple low-dose data sets might be accumulated with fine-slicing data collection, as suggested for use with pixel-array-based detectors (Mueller *et al.*, 2012). A second consideration is the cutoff limit on resolution. Anticipating that averaging from multiple crystals will tend to improve outer-shell statistics, the single-crystal data sets can be processed to an average outer-shell $I/\sigma(I)$ of under 1.0. A third consideration concerns outlier rejection. We propose that crystal rejection tests should be performed first based on unit-cell variations followed by tests on diffraction dissimilarity and RACC. This is because it is much easier to obtain unit-cell parameters from indexing prior to computation-intensive data integration and scaling. A fourth consideration is radiation damage. To best address radiation damage, we suggest merging multi-crystal data as accumulated wedges and making substructure-determination and phasing attempts in a succession of these accumulations. It should be noted that sometimes including damaged data can still enhance phasing owing to increasing multiplicity, as noted above for the EGFRK analysis.

The use of multiple crystals has proved essential for several challenging SAD analyses, including those here at 6 keV, and the addition of statistically equivalent data does improve accuracy. Nevertheless, there is interest in what is minimally needed. To find the minimal number of crystals required for successful structure determination, we merged data progressively crystal by crystal (from the best to the worst) based on the values of the relative anomalous correlation coefficient (RACC). For each merged data set, we performed both substructure determination by *SHELXD* and phasing by *Phaser* followed by density modification with *DM*. After checking for both *SHELXD* solutions and experimental electron-density maps, we identified that the minimal numbers of crystals required for structure determinations are three for EGFRK, five for ThiT and five for YetJ. Although we may have collected more data sets than needed for YetJ, we found that adding more data still benefited both substructure determination, with higher CC_{weak}/CC_{all} values, and also electron-density map quality, with higher MapCCs until the two lowest RACC data sets were added.

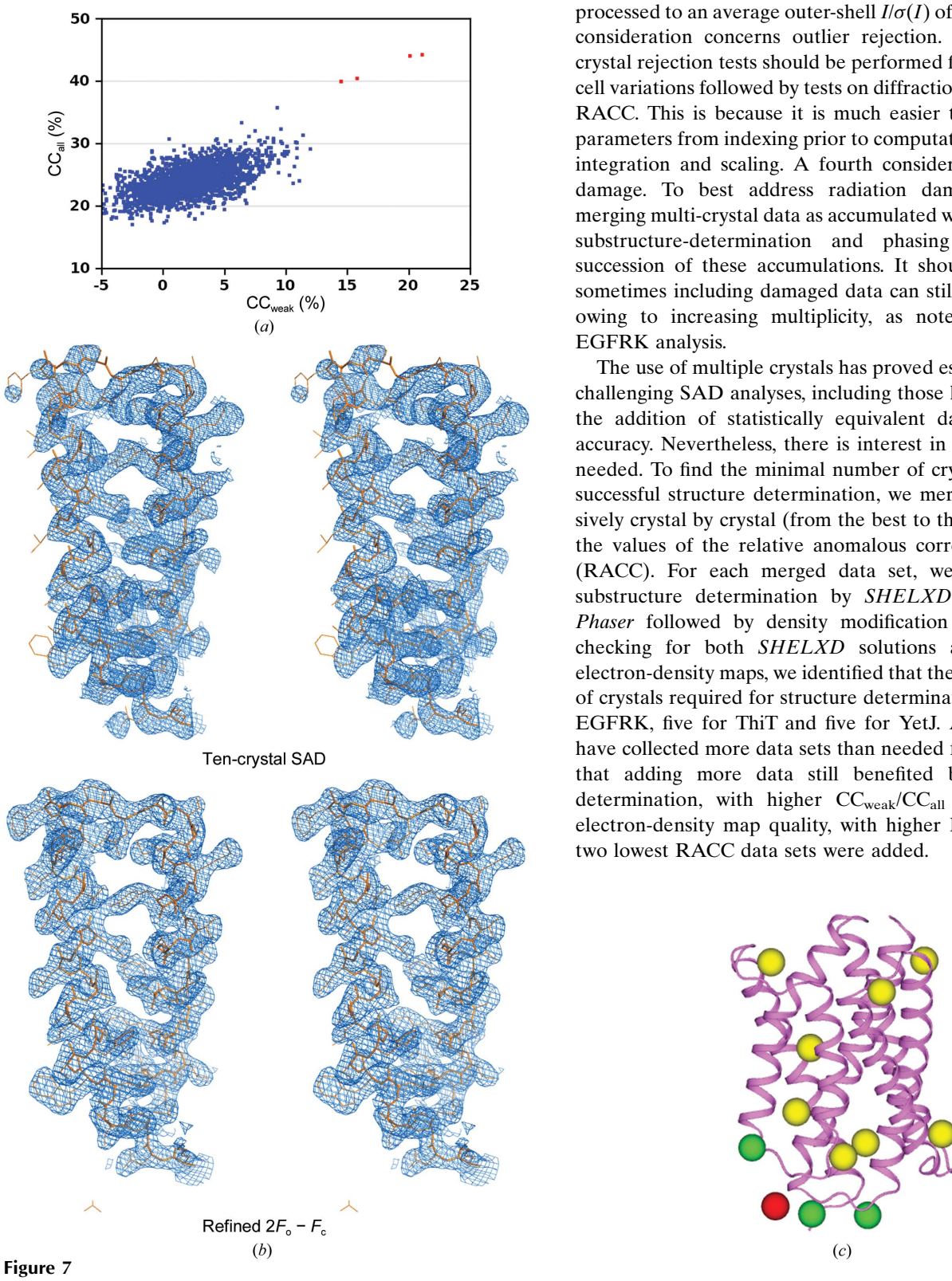


Figure 7
Phasing of YetJ. (*a*) Substructure determination. (*b*) Representative portion of the experimental electron-density map and its comparison with the refined map. (*c*) Ribbon diagram of the model showing S atoms. Electron-density maps were produced as described for Fig. 3.

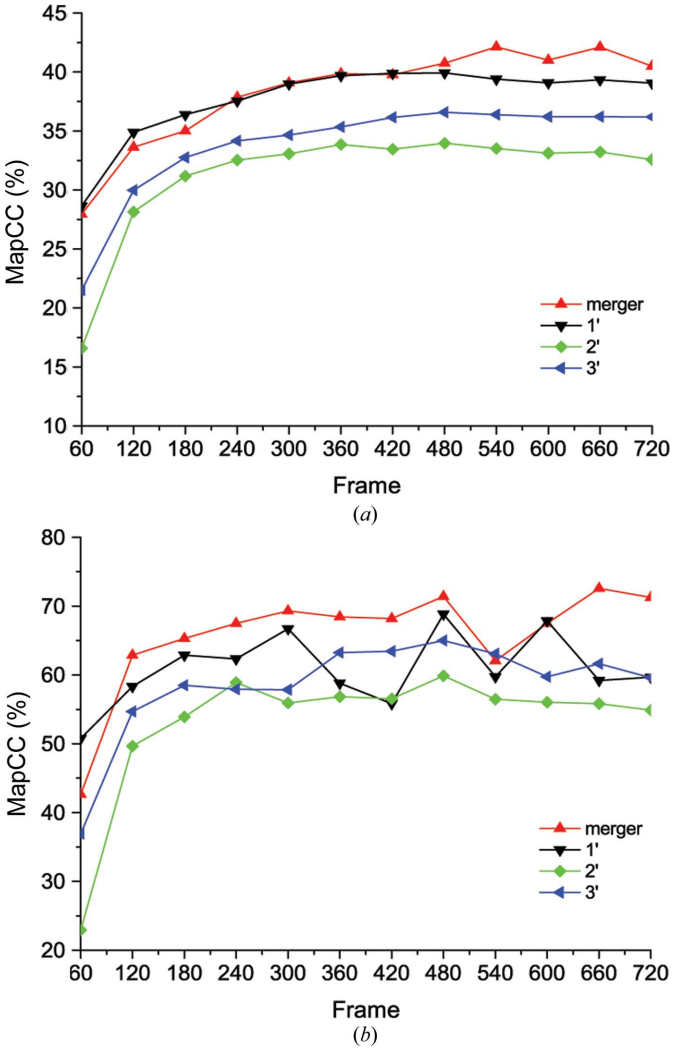


Figure 8
Plot of MapCC with respect to increasing number of frames. The known EGFRK model was first used for substructure determination by MR-SAD following by SAD phasing with the found substructure. The electron densities from SAD phasing before and after density modification were compared with the model-derived phases as shown by MapCC. (*a*) MapCC before density modification for the three compatible data sets and their merger. (*b*) MapCC after density modification for the three compatible EGFRK data sets and their merger.

Structure determination by multi-crystal native SAD phasing is distinguished from current practice in requiring a supply of equivalent crystals. Meeting this condition may be an obstacle in some cases; however, an abundance of crystals is quite common and such production would be likely to increase should new technology take hold. With anticipated advances in synchrotron instrumentation and automation, it is plausible to expect that multi-crystal native SAD phasing could come to predominate for *de novo* structure determination.

5. Concluding remarks

Native SAD phasing may be realised by using lower energy X-rays at 6 keV. Although some complications may arise at 6 keV, anomalous signals are enhanced intrinsically and signal-to-noise ratios are increased through the averaging of multiple data sets. The three examples presented here demonstrate the routine utility of multi-crystal native SAD phasing at low energy even for more challenging lower resolution structures.

We thank John Schwanof and Randy Abramowitz for help with synchrotron data collection and Zhiqiang Zhu for useful discussion. This work was supported in part by NIH grants GM095315, GM098617 and GM107462. 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

Adams, P. D. *et al.* (2011). *Methods*, **55**, 94–106.
Adams, P. D., Baker, D., Brunger, A. T., Das, R., DiMaio, F., Read, R. J., Richardson, D. C., Richardson, J. S. & Terwilliger, T. C. (2013). *Annu. Rev. Biophys.* **42**, 265–287.
Akey, D. L., Brown, W. C., Dutta, S., Konwerski, J., Jose, J., Jurkiw, T. J., DelProposto, J., Ogata, C. M., Skiniotis, G., Kuhn, R. J. & Smith, J. L. (2014). *Science*, **343**, 881–885.
Chang, Y., Bruni, R., Kloss, B., Assur, Z., Kloppmann, E., Rost, B., Hendrickson, W. A. & Liu, Q. (2014). *Science*, **344**, 1131–1135.
Chen, L., Rose, J. P., Breslow, E., Yang, D., Chang, W.-R., Furey, W. F. Jr, Sax, M. & Wang, B.-C. (1991). *Proc. Natl Acad. Sci. USA*, **88**, 4240–4244.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Cowtan, K. D. & Zhang, K. Y. J. (1999). *Prog. Biophys. Mol. Biol.* **72**, 245–270.
Dauter, Z. (2006). *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. & Rajashankar, K. R. (2000). *Acta Cryst.* D**56**, 232–237.
Dauter, Z., Jaskolski, M. & Wlodawer, A. (2010). *J. Synchrotron Rad.* **17**, 433–444.
Diederichs, K. (2006). *Acta Cryst.* D**62**, 96–101.
Doutch, J., Hough, M. A., Hasnain, S. S. & Strange, R. W. (2012). *J. Synchrotron Rad.* **19**, 19–29.
Duke, E. M. H. & Johnson, L. N. (2010). *Proc. R. Soc. A*, **466**, 3421–3452.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Evans, P. R. (2011). *Acta Cryst.* D**67**, 282–292.
Evans, P. R. & Murshudov, G. N. (2013). *Acta Cryst.* D**69**, 1204–1214.
Guo, Y., Liu, Q. & Hendrickson, W. A. (2014). In preparation.
Hendrickson, W. A. (1991). *Science*, **254**, 51–58.
Hendrickson, W. A. (1999). *J. Synchrotron Rad.* **6**, 845–851.
Hendrickson, W. A. (2013). *Acta Cryst.* A**69**, 51–59.
Hendrickson, W. A. (2014). *Q. Rev. Biophys.* **47**, 49–93.
Hendrickson, W. A., Horton, J. R. & LeMaster, D. M. (1990). *EMBO J.* **9**, 1665–1672.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Kumar, A., Petri, E. T., Halmos, B. & Boggon, T. J. (2008). *J. Clin. Oncol.* **26**, 1742–1751.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Liu, Q., Dahmane, T., Zhang, Z., Assur, Z., Brasch, J., Shapiro, L., Mancia, F. & Hendrickson, W. A. (2012). *Science*, **336**, 1033–1037.

Liu, Q., Liu, Q. & Hendrickson, W. A. (2013). *Acta Cryst.* D**69**, 1314–1332.
Liu, Q., Zhang, Z. & Hendrickson, W. A. (2011). *Acta Cryst.* D**67**, 45–59.
Liu, Z.-J., Vysotski, E. S., Chen, C.-J., Rose, J. P., Lee, J. & Wang, B.-C. (2000). *Protein Sci.* **9**, 2085–2093.
Mueller, M., Wang, M. & Schulze-Briese, C. (2012). *Acta Cryst.* D**68**, 42–56.
Mueller-Dieckmann, C., Panjikar, S., Tucker, P. A. & Weiss, M. S. (2005). *Acta Cryst.* D**61**, 1263–1272.
Read, R. J. & McCoy, A. J. (2011). *Acta Cryst.* D**67**, 338–344.
Schiltz, M., Fourme, R. & Prangé, T. (2003). *Methods Enzymol.* **374**, 83–119.
Sheldrick, G. M. (2010). *Acta Cryst.* D**66**, 479–485.
Stamos, J., Sliwkowski, M. X. & Eigenbrot, C. (2002). *J. Biol. Chem.* **277**, 46265–46272.
Terwilliger, T. C., Grosse-Kunstleve, R. W., Afonine, P. V., Moriarty, N. W., Zwart, P. H., Hung, L.-W., Read, R. J. & Adams, P. D. (2008). *Acta Cryst.* D**64**, 61–69.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
Weiss, M. S., Panjikar, S., Mueller-Dieckmann, C. & Tucker, P. A. (2005). *J. Synchrotron Rad.* **12**, 304–309.
Weiss, M. S., Sicker, T. & Hilgenfeld, R. (2001). *Structure*, **9**, 771–777.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.