



ISSN: 1399-0047

journals.iucr.org/d

S-SAD phasing of monoclinic histidine kinase from *Brucella abortus* combining data from multiple crystals and orientations: an example of data-collection strategy and a *posteriori* analysis of different data combinations

Sebastián Klinke, Nicolas Foos, Jimena J. Rinaldi, Gastón Paris, Fernando A. Goldbaum, Pierre Legrand, Beatriz G. Guimarães and Andrew Thompson

Acta Cryst. (2015). D71, 1433–1443



Copyright © International Union of Crystallography

Author(s) of this paper may load this reprint on their own web site or institutional repository provided that this cover page is retained. Republication of this article or its storage in electronic databases other than as specified above is not permitted without prior permission in writing from the IUCr.

For further information see <http://journals.iucr.org/services/authorrights.html>

S-SAD phasing of monoclinic histidine kinase from *Brucella abortus* combining data from multiple crystals and orientations: an example of data-collection strategy and a *posteriori* analysis of different data combinations

Sebastián Klinke,^{a*} Nicolas Foos,^b Jimena J. Rinaldi,^a Gastón Paris,^{a‡} Fernando A. Goldbaum,^a Pierre Legrand,^b Beatriz G. Guimarães^{b*§} and Andrew Thompson^b

Received 5 February 2015

Accepted 17 April 2015

Edited by Z. Dauter, Argonne National Laboratory, USA

^aFundación Instituto Leloir, IIBBA–CONICET, Avenida Patricias Argentinas 435, C1405BWE Buenos Aires, Argentina, and ^bExperimental Division, Synchrotron SOLEIL, BP 48, L'Orme des Merisiers, 91192 Gif-sur-Yvette, France.

*Correspondence e-mail: sklinke@leloir.org.ar, beatriz.guimaraes@fiocruz.br

‡ Present address: Centro de Investigaciones en Bionanociencias, CIBION-CONICET, Godoy Cruz 2390, C1425FQD Buenos Aires, Argentina.

§ Present address: Instituto Carlos Chagas, FIOCRUZ-PARANÁ, Rua Prof. Algacyr Munhoz Mader 3775, PR 81350-010, Brazil.

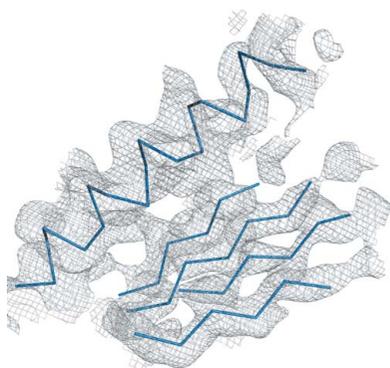
Keywords: experimental phasing; anomalous signal; two-component system; HWE family.

Supporting information: this article has supporting information at journals.iucr.org/d

The histidine kinase (HK) domain belonging to the light–oxygen–voltage histidine kinase (LOV-HK) from *Brucella abortus* is a member of the HWE family, for which no structural information is available, and has low sequence identity (20%) to the closest HK present in the PDB. The ‘off-edge’ S-SAD method in macromolecular X-ray crystallography was used to solve the structure of the HK domain from LOV-HK at low resolution from crystals in a low-symmetry space group ($P2_1$) and with four copies in the asymmetric unit (~108 kDa). Data were collected both from multiple crystals (diffraction limit varying from 2.90 to 3.25 Å) and from multiple orientations of the same crystal, using the κ -geometry goniostat on SOLEIL beamline PROXIMA 1, to obtain ‘true redundancy’. Data from three different crystals were combined for structure determination. An optimized HK construct bearing a shorter cloning artifact yielded crystals that diffracted X-rays to 2.51 Å resolution and that were used for final refinement of the model. Moreover, a thorough *a posteriori* analysis using several different combinations of data sets allowed us to investigate the impact of the data-collection strategy on the success of the structure determination.

1. Introduction

The possibility of phasing protein structures using the anomalous signal from sulfur, close to or away from its absorption edge, has long been an objective of macromolecular crystallographers (Wang, 1985; Dauter *et al.*, 1999), and many reports can be found in the literature of successful structure determinations either using synchrotron radiation or in-house sources (Liu *et al.*, 2000; Debreczeni, Bunkoczi, Girmann *et al.*, 2003; Ramagopal *et al.*, 2003; Yang *et al.*, 2003; Sarma & Karplus, 2006; Vasur *et al.*, 2006; Wagner *et al.*, 2006; Cianci *et al.*, 2008; Goulet *et al.*, 2010; Liu *et al.*, 2011; Zhu *et al.*, 2012). Indeed, one of the first protein structures to be solved with the help of anomalous signal was that of the 4.8 kDa protein crambin (Hendrickson & Teeter, 1981). Notwithstanding, the anomalous signal from sulfur is very weak, often with Bijvoet ratios at the 1–1.5% level, and consequently structure solution is strongly limited by the requirement to collect data of sufficient quality prior to structural changes resulting from radiation damage (Sarma & Karplus, 2006). In addition, very few structures have been reported in the absence of high-resolution data in the 2.0–2.5 Å range (Micossi *et al.*, 2002;



Debreczeni, Bunkoczi, Girmann *et al.*, 2003; Doan & Dokland, 2003; Goulet *et al.*, 2010; Liu *et al.*, 2012, 2014; Ru *et al.*, 2012; Akey *et al.*, 2014; El Omari *et al.*, 2014; Nawrotek *et al.*, 2014; Weinert *et al.*, 2015). For these reasons, many laboratories in need of *ab initio* phasing tend to launch the production of selenomethionated protein (if at all practicable) as soon as native crystallization conditions have been established. The consequence of this is that the number of reported structures over the last ten years using S-SAD phasing has been modest although constant (Liu *et al.*, 2013).

As with all anomalous structure determination, the starting point of phasing *via* sulfur is to find the heavy-atom substructure (Usón *et al.*, 2003), which is often a major difficulty for many sulfur sites but can be simplified if disulfide bridges are present (which is unfortunately not always the case). In 2003, Sheldrick and coworkers (Debreczeni, Bunkoczi, Girmann *et al.*, 2003; Debreczeni, Bunkoczi, Ma *et al.*, 2003) reported the routine use of in-house sources for sulfur phasing experiments, using a CCD detector and multiple sample orientations to collect highly redundant data. Whereas commercial laboratory X-ray diffraction equipment regularly includes multi-axis goniometry and the software required for its exploitation, synchrotron sources have rarely extracted its full benefit. Indeed, beamlines have been equipped with multi-axis goniometers used for the accurate measurement of anomalous signals since the early days of synchrotron-radiation macromolecular crystallography (see, for example, Phillips *et al.*, 1979; Kahn *et al.*, 1980). Since these early efforts, multi-circle goniometry has been rediscovered at synchrotron sources from time to time. The mini- κ goniometer head is used at most of the macromolecular crystallography beamlines at the ESRF and elsewhere (Brockhauser *et al.*, 2013), and a high-precision multi-axis goniometer, PRIGo, has recently been developed by the Macromolecular Crystallography group at the Swiss Light Source (unpublished work). Although interesting, a full survey of hardware and software advances in multi-circle goniometry at synchrotron sources is beyond the scope of this article. The use of different sample orientations implies the collection of the same (or symmetry-related) reflections in a different geometry and hence subjected to different systematic errors (radiation damage, sample or cryo-buffer absorption *etc.*), therefore addressing concerns about the accuracy of the anomalous signal. The application of this data-collection strategy for S-SAD phasing, as suggested by Sheldrick, has only become frequent practice at synchrotron beamlines recently, probably as a result of the length of the experiments with relatively slow-readout detectors and concerns about radiation damage. Since the inception of the PROXIMA 1 beamline, the multiple crystal orientation approach has been regularly used in low-energy SAD phasing experiments, taking advantage of the availability of a full κ -geometry goniostat. For example, Goulet and coworkers applied this method to the structure determination of a small protein from the archaeal two-tailed *Acidanus virus* (Goulet *et al.*, 2010) and Nawrotek and coworkers determined the structure of a tubulin-interacting protein (Nawrotek *et al.*, 2014).

In 2012, Hendrickson and coworkers reported on the combination of data from multiple crystals of the same sample based on several criteria, including unit-cell dimension differences, correlation of anomalous signal per data set to the average anomalous signal of multiple data sets and diffraction dissimilarity of measured intensities from different crystal samples (Liu *et al.*, 2012). This approach combines data from multiple crystals in different orientations, distributing the radiation damage endured over several crystals in order to obtain data with a sufficient signal-to-noise ratio to measure the weak S anomalous signal. Finally, the recent generation of photon-counting (and hence noise-free) pixel-array detectors (Broennimann *et al.*, 2006; Hülsen *et al.*, 2006), when combined with background elimination in the diffraction experiment, have allowed good signal-to-noise measurement of extremely weak diffraction spots by measuring highly redundant data (Ben-Shem *et al.*, 2011).

The collection of highly redundant data is a major prerequisite for S-SAD phasing. For this reason, it is not surprising that most of the structures determined using this method belong to high-symmetry space groups. Based on the statistics of a recent study (Weinert *et al.*, 2015), there are no S-SAD structures in the triclinic *P1* space group and only 12 in the monoclinic *P2*₁ or *C2* space groups (Hendrickson & Teeter, 1981; Gentry *et al.*, 2005; Kitamura *et al.*, 2008; Lakomek *et al.*, 2009; Nan *et al.*, 2009; Liu *et al.*, 2012; Weinert *et al.*, 2015). Additionally, S-SAD structures bearing a high number of sulfur sites (over 50) and/or large asymmetric units (over 100 kDa) deserve special interest. Amongst them, it is worth mentioning the *Sarcocystis muris* lectin SML-2 (PDB entry 2yil; Müller *et al.*, 2011), which was the first reported structure with more than 100 S sites, and two major complexes described by Weinert and coworkers, namely the *Escherichia coli* DNA polymerase IV–DNA complex (PDB entry 4r8u) and the *Bos taurus* ($\alpha\beta$ -tubulin)₂–RB3–tubulin–tyrosine ligase complex (PDB entry 4wbn) (Weinert *et al.*, 2015). The latter structure contains the largest asymmetric unit for a S-SAD structure of 266 kDa with 118 S sites.

Here, we report the structure determination of a 242-residue protein solved by S-SAD from crystals belonging to a low-symmetry space group with four copies in the asymmetric unit and diffracting to 2.9 Å resolution. The studied protein corresponds to the histidine kinase domain (HK) of the light–oxygen–voltage histidine kinase (LOV-HK) from the pathogen *Brucella abortus* (Swartz *et al.*, 2007). Typical HK domains are composed of a globular catalytic and ATP-binding (CA) subdomain connected N-terminally to a helical hairpin linker called the dimerization and histidine phosphotransfer (DHP) subdomain. HKs usually dimerize through the latter subdomain, forming a central four-helix bundle surrounded by two CA subdomains.

The particular HK domain studied in this work belongs to the HWE family (Karniol & Vierstra, 2004), for which no structural information is available, and shows low sequence identity (20%) to the closest HK present in the PDB, namely the blue light-activated HK from *Erythrobacter litoralis* HTCC2594 (PDB entry 4r3a; Rivera-Cancel *et al.*, 2014). The

facts that (i) there is an increased plasticity in the relative location of the CA subdomains in HKs (Albanesi *et al.*, 2009) and (ii) there are no three-dimensional structures available for the HWE family made the application of molecular replacement as a phasing method unsuccessful for this particular case.

It is important to clarify that the complete structural description of *B. abortus* HK, together with its biological implications, will be the key subject of a future publication; this work focuses mainly on the applied methodology that led to the successful phasing of the reflections in this difficult yet interesting example.

2. Methods and results

2.1. Gene cloning, protein expression and purification

The gene fragment corresponding to the HK domain from *B. abortus* LOV-HK (residues 266–489, gi:221272017) was amplified by PCR using *B. abortus* strain 2308 genomic DNA as template. The PCR product was inserted into the pET-24d plasmid (Novagen, Billerica, Massachusetts, USA) between the NdeI and XhoI restriction sites. Two slightly different constructs were used which included different C-terminal cloning artifacts. The construct used for the solution of the structure comprised a total of 242 residues: residues 266–489 from the HK domain, a three-residue N-terminal cloning artifact (MAS) and a 15-residue C-terminal cloning artifact that includes a 6×His tag (VDKLAALAEHHHHHH). A shorter, optimized construct with a smaller C-terminal cloning artifact (HHHHHH, 233 residues) was used for refinement of the structure. The final constructs were checked by DNA sequencing.

E. coli BL21(DE3) competent cells (Stratagene, La Jolla, California, USA) were transformed with the expression plasmids mentioned above. Precultures were grown overnight in

5 ml LB medium with 25 µg ml⁻¹ kanamycin at 37°C with agitation (200 rev min⁻¹) and were then diluted to 500 ml and grown to an absorbance (at 600 nm) of 0.6. At this point, isopropyl β-D-1-thiogalactopyranoside was added to a final concentration of 0.5 mM and the cultures were further incubated overnight at 20°C with agitation (200 rev min⁻¹). The bacteria were centrifuged at 5000g for 10 min at 4°C. Pellets were resuspended and sonicated in a solution consisting of 50 mM sodium phosphate, 0.5 M sodium chloride, 20 mM imidazole, 1 mM DTT, 1 mM PMSF pH 7.5 (buffer A) and then centrifuged at 160 000g in an L7-65 ultracentrifuge (Beckman Coulter, Brea, California, USA) for 60 min at 4°C. The supernatant was filtered through a 0.45 µm membrane and loaded onto a HisTrap HP column (all columns were from GE Healthcare, Little Chalfont, England). Elution was performed with a linear gradient of buffer B consisting of 50 mM sodium phosphate, 0.5 M sodium chloride, 0.5 M imidazole, 1 mM DTT, 1 mM PMSF pH 7.5. A major peak was observed. The appropriate fractions were pooled and dialyzed overnight at 4°C against buffer C (50 mM MES, 0.25 M sodium chloride, 1 mM DTT, 1 mM PMSF pH 6.5). A size-exclusion purification step was then performed on a Superdex S75 column in the same buffer. A single peak was observed. Fractions were concentrated to 30 mg ml⁻¹ by centrifugation in Amicon Ultra-4 devices (Millipore, Billerica, Massachusetts, USA) and simultaneously exchanged into crystallization buffer (10 mM MES, 50 mM sodium chloride pH 7.5). The protein was aliquoted and stored at -70°C. The quality of the final preparation was checked by SDS-PAGE (15% gel) and UV spectrophotometry.

2.2. Protein crystallization

Both HK constructs were crystallized at room temperature using the hanging-drop vapour-diffusion method. The construct used for phasing crystallized in 10% (w/v) PEG 8000, 10% (w/v) ethylene glycol, 0.1 M HEPES pH 7.5, while the crystallization condition of the construct used for refinement consisted of 7.5% (w/v) PEG 8000, 0.1 M HEPES pH 8.3. A total of 1.5 µl protein solution (7 mg ml⁻¹) with 5 mM AMP-PCP (a nonhydrolyzable ATP analogue) and 5 mM magnesium chloride was mixed with 1.5 µl crystallization solution. Long bars appeared within one week of equilibration and reached approximate dimensions of 0.40 × 0.10 × 0.05 mm (Fig. 1). Samples were then cryoprotected in their respective mother liquors with 25% (w/v) PEG 400, mounted in Hampton Research loops (Aliso Viejo, California, USA) and cryocooled in liquid nitrogen for data collection.

2.3. X-ray diffraction data collection and processing

2.3.1. Data collection. The PROXIMA 1 beamline at the SOLEIL synchrotron is equipped with a PILATUS 6M 25 Hz large surface-area detector (Dectris, Baden, Switzerland) and a three-circle κ-geometry goniostat (Crystal Logic, Los Angeles, California, USA), which allows data collection with rotation about the φ or ω axes. The beamline was configured to provide a focused X-ray beam (125 × 65 µm) at a wavelength

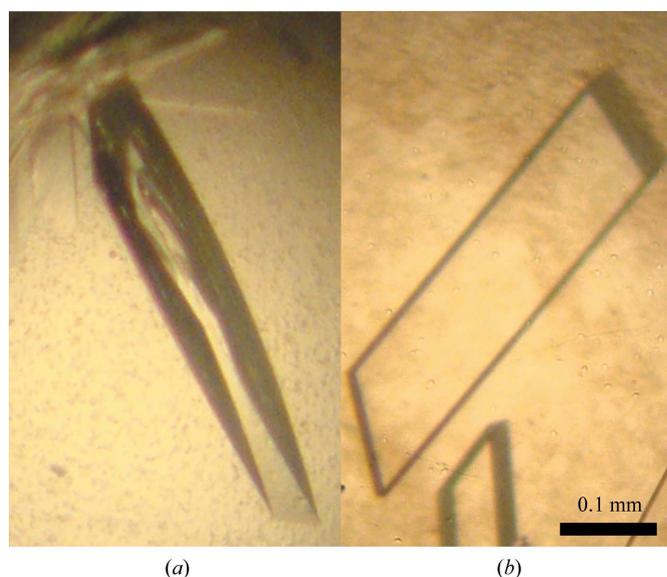


Figure 1
HK crystals. (a) Sample from the longer construct, used for phasing. (b) Sample from the shorter construct, used for refinement.

Table 1

Data-collection parameters and statistics from data sets that were combined for HK structure determination by S-SAD (longer construct).

In the data-set name, a, b and c refer to translations of the crystal in the X-ray beam, whereas k refers to the different crystal orientations used (κ offset). Ten data sets from three crystals were used: X113 (two positions, three orientations in each), X112 (single position, three orientations) and X111 (only one data set used for structure determination). $\Delta F/\sigma(\Delta F)$ is the average anomalous signal from data truncated to 4.2 Å resolution. The anomalous correlation coefficients are also calculated from data truncated to 4.2 Å resolution. Values in parentheses are for the last shell.

Crystal	X113	X113	X113	X113	X113	X113	X112	X112	X112	X111
Data-set name	a	ak10	ak-10	b	bk10	bk-10	a	ak10	ak-10	c
κ angle (°)	0	10	−10	0	10	−10	0	10	−10	0
Total range (°)	400	400	400	400	400	400	400	400	400	400
Unit-cell parameters										
<i>a</i> (Å)	72.2	72.3	72.3	72.0	72.1	72.1	72.1	72.1	72.2	72.1
<i>b</i> (Å)	109.2	109.1	109.1	109.3	109.2	109.4	109.1	109.1	109.2	108.9
<i>c</i> (Å)	73.3	73.4	73.5	73.3	73.3	73.3	73.4	73.4	73.4	73.3
β (°)	103.3	103.3	103.3	103.1	103.1	103.1	103.1	103.1	103.1	102.9
Resolution (Å)	50.0–3.10	50.0–3.10	50.0–3.10	50.0–2.90	50.0–2.90	50.0–2.90	50.0–2.90	50.0–2.90	50.0–2.90	50.0–2.85
	(3.28–3.10)	(3.28–3.10)	(3.28–3.10)	(3.07–2.90)	(3.07–2.90)	(3.07–2.90)	(3.07–2.90)	(3.07–2.90)	(3.07–2.90)	(3.01–2.85)
Unique reflections	39390 (6169)	39485 (6167)	39518 (6125)	47967 (7657)	48044 (7667)	48108 (7669)	47087 (7380)	47187 (7403)	47251 (7413)	49744 (7652)
Average multiplicity	3.7 (3.6)	3.7 (3.6)	3.7 (3.6)	3.8 (3.7)	3.8 (3.7)	3.8 (3.7)	3.9 (3.8)	3.9 (3.8)	3.9 (3.8)	3.8 (3.7)
Completeness (%)	98.8 (95.3)	98.8 (95.4)	98.6 (94.5)	99.5 (98.8)	99.5 (98.7)	99.5 (98.6)	97.7 (95.2)	97.7 (95.2)	97.7 (95.2)	97.8 (93.2)
R_{meas} (%)	5.6 (91.0)	5.7 (104.6)	6.0 (124.7)	4.3 (99.4)	4.4 (112.7)	4.3 (109.0)	4.7 (79.4)	4.7 (79.5)	4.7 (93.6)	6.4 (76.5)
$CC_{1/2}$ (%)	100.0 (88.6)	100.0 (86.2)	100.0 (81.7)	100.0 (83.6)	100.0 (81.5)	100.0 (81.5)	99.9 (90.1)	99.9 (90.0)	99.9 (87.8)	99.8 (91.7)
$\langle I/\sigma(I) \rangle$	17.3 (1.5)	16.8 (1.3)	16.2 (1.1)	19.8 (1.4)	19.4 (1.3)	19.5 (1.3)	17.5 (1.7)	17.2 (1.7)	17.3 (1.5)	12.8 (1.6)
Anomalous CC (%)	33	38	38	33	35	35	32	31	33	27
$\Delta F/\sigma(\Delta F)$	1.15	1.20	1.21	1.22	1.25	1.27	1.16	1.19	1.22	1.02

of 1.80 Å. At this energy and with the storage ring operating in ‘top-up’ mode at 400 mA and 2.75 GeV, the measured photon flux at the sample was 1.3×10^{12} photons s^{-1} for the whole focal spot. In order to limit the beam intensity on the sample, the beam size and divergence were adjusted using two sets of X-ray slits, one in front of the monochromator at 12 m from the source position, which exerts a large effect on the beam divergence, and a second set in front of the sample position, which exerts a large effect on the beam size at the sample. The cumulative effect of these measures reduced the beam size at the sample position to $90 \times 60 \mu\text{m}$ and reduced the overall flux at the sample by a factor of ten. Under these circumstances, the focal beam was smaller than the crystal size in all directions.

Firstly, diffraction data were collected using the *MxCuBE* interface (Gabadinho *et al.*, 2010) from crystals of HK cooled to 100 K in an attempt to obtain the highest resolution possible using a wavelength of 0.98 Å. Further data sets were then collected at 1.80 Å wavelength in order to judge the size of the S anomalous signal. These data showed the beginnings of well measured anomalous signal at low resolution; consequently, we attempted S-SAD prior to a search for standard heavy-atom derivatives or the production of selenomethionated protein.

The morphology of the HK crystals was such that their long dimension could be mounted along the goniometer rotation axis such that the largest crystals allowed translation to multiple positions during data collection (see Fig. 1). Data-collection statistics from six crystals are presented in Table 1 and Supplementary Table S1. The following protocols of data collection were used. (i) Rotation around the φ axis with the crystal in a random setting, sometimes with and sometimes without inverse-beam data collection. (ii) Angular offset of the crystal using the κ rotation of the goniostat. κ offsets of 10°

were used (the authors have not completely explored the variation of different offsets on the subsequent scaling, but presume that the optimal value will be dependent on the space group and/or the crystal shape; a value of $10\text{--}15^\circ$ has proved to be suitable for other S-SAD data collections on PROXIMA 1). Note that in this case the rotation axis φ is no longer in the plane of the storage ring, requiring the calculation of the rotation-axis orientation with respect to the laboratory system. (iii) Following the first round of data collection, a ‘fresh’ part of the crystal was translated into the X-ray beam, a different orientation was selected and further data sets were collected as explained above.

The crystals belonged to space group $P2_1$, with unit-cell parameters $a = 72.0$, $b = 109.2$, $c = 73.2$ Å, $\beta = 103.1^\circ$. Based on the size of the protein and the unit cell, solvent-content analysis indicated a most probable value of 52% and consequently the presence of four molecules per asymmetric unit; each copy is expected to contain two Cys and nine Met residues, or potentially 44 S sites and 968 amino acids in the asymmetric unit, with a predicted Bijvoet ratio (Smith, 1997) of 1.5%. A total of 34 900 images were collected from six crystals, the best of which diffracted to 2.9 Å resolution but which typically diffracted to between 3.0 and 3.1 Å. The detector distance used ranged from 262.3 to 276.3 mm. These data were collected from large crystals with, in some cases, translation of the sample between successive data collections. At each orientation of the crystal 400° of data were typically collected, resulting in an absorbed X-ray dose per orientation of 0.4 MGy as calculated by *RADDOSE-3D* (Zeldin *et al.*, 2013). Data were reduced using the *XDS* program (Kabsch, 2010) via the *xdsme* command-line interface (<https://code.google.com/p/xdsme/>) for semi-automatic data processing. After verification of the consistent indexing, data sets were combined using *XSCALE* (Kabsch, 2010). Solution of the S

Table 2

Data-collection and phasing statistics.

Values in parentheses are for the highest resolution shell.

Data-set name	S-SAD†	Native 1	Native 2
Data collection			
No. of crystals	3	1	1
No. of data sets	10	3	1
No. of frames	20000	2900	1000
Oscillation step (°)	0.2	0.2	0.2
Wavelength (Å)	1.80000	0.98011	0.98011
Exposure per frame (s)	0.2	0.2	0.2
Indexing and scaling			
Unit-cell parameters			
<i>a</i> (Å)	72.01	71.33	70.39
<i>b</i> (Å)	109.16	105.60	100.84
<i>c</i> (Å)	73.23	73.09	71.41
β (°)	103.1	102.7	102.8
Space group	$P2_1$	$P2_1$	$P2_1$
Resolution limit (Å)	2.90 (2.98–2.90)	2.70 (2.77–2.70)	2.51 (2.66–2.51)
Total No. of reflections	1713187	317026	125068
No. of unique reflections	48120	29297	32760
Average multiplicity	35.6 (25.7)	10.9 (10.7)	3.8 (3.8)
$\langle I/\sigma(I) \rangle$	39.6 (3.2)	22.3 (1.4)	14.3 (2.0)
R_{meas} (%)	6.4 (111.8)	5.7 (163.1)	6.0 (59.1)
$CC_{1/2}$ (%)	100.0 (96.8)	100.0 (84.3)	99.9 (86.7)
Completeness (%)	99.9 (100.0)	99.8 (99.4)	97.7 (95.1)
Chains per asymmetric unit		4	4
Solvent content (%)		48	49
<i>B</i> factor, Wilson plot (Å ²)		98	72
Phasing			
SigAno (3.9 Å/overall)	1.029/1.655		
Best $CC_{\text{all}}/CC_{\text{weak}}$	41.74/15.84		
Best PATFOM	4.80		
No. of sulfur sites	27 [44 expected]		

† The S-SAD data set results from the merging of the data sets presented in Table 1.

partial structure was attempted for different combinations of data sets.

The data set used for structure determination was based on three crystals (X113, X112 and X111; Table 1) and a total of 20 000 images each of 0.2°, corresponding to 4000° of data to give a total multiplicity of approximately 35 (Bijvoet pairs were counted as separate reflections). It should be noted that this particular combination of data was chosen by inspection of the indicators of data quality and anomalous signal after merging for phasing ‘on the beamline’ and without running any clustering programs. An initial analysis of the data combinations was performed mainly by looking at the values of R_{meas} , anomalous correlation and SigAno calculated by *XSCALE* for the lower resolution shells, roughly up to 6.0 Å, and the resolution where SigAno fell below 1. The chosen combination presented R_{meas} values ranging from 3 to 4% and SigAno starting from above 3 in the inner shells. Combinations with SigAno starting from around 2.5 in the inner shells were considered poorer for phasing at this point. Crystal X113 was large enough to be translated once, and at each position data were collected at three different κ angles (–10, 0 and 10°). Crystal X112 permitted the collection of data from a single position at three different κ angles (–10, 0 and 10°). 2000 images were collected in any given orientation from a given position of the crystals. Crystal X111 allowed two translations, with two inverse-beam data collections from the

Table 3

B factors (Å²) of the different S and P sites in the refined structure to 2.70 Å resolution (Native 1) and in the refined structure of the shorter construct to 2.51 Å resolution (Native 2).

Chains IDs and residue numbers are indicated.

Chain	Native 1				Native 2			
	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>	<i>A</i>	<i>B</i>	<i>C</i>	<i>D</i>
Met294	95	102	—	—	66	65	—	—
Met296	79	92	—	—	58	61	—	—
Met331	105	109	110	109	66	80	86	91
Cys346	102	105	93	100	63	69	77	81
Met363	108	105	112	137	75	71	84	99
Met409	142	129	130	153	94	93	106	101
Met419	115	108	95	132	70	72	74	89
Met430	126	156	151	—	79	100	107	—
Cys454	102	112	101	135	68	74	70	93
Met479	—	—	158	—	—	—	103	—
AMP-PCP ‘PA’	84	101	143	171	54	69	105	138
AMP-PCP ‘PB’	82	109	162	171	53	67	125	149
AMP-PCP ‘PG’	81	110	168	171	48	63	133	157

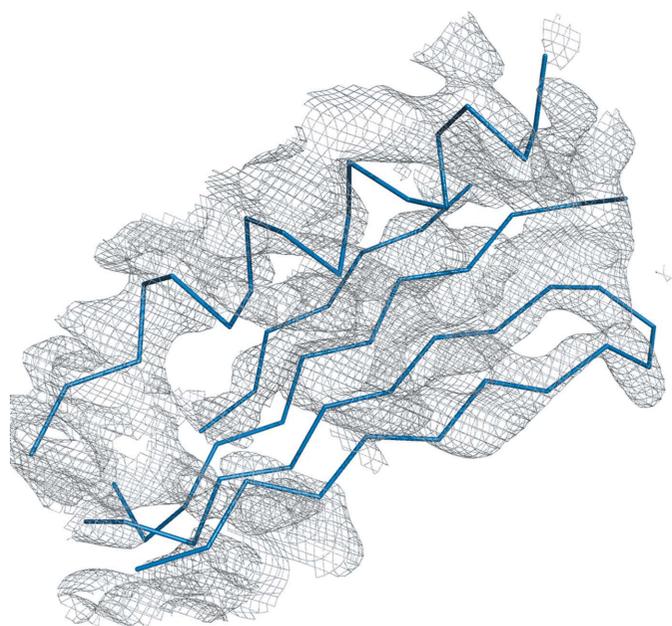
first two positions and a single rotation from the third. Only the data from this third position (X111c, 2000 images) were used for initial phasing. Statistics of the merged data set used for structure solution are presented in Table 2. The overall anomalous signal (estimated from the merged data and by the crossover where the SigAno calculated by *XSCALE* falls below 1) extended to 3.9 Å resolution and the overall resolution of the merged data set was 2.9 Å. Separate statistics from the different data collections are presented in Table 1 and Supplementary Table S1.

2.3.2. Heavy-atom location and structure determination.

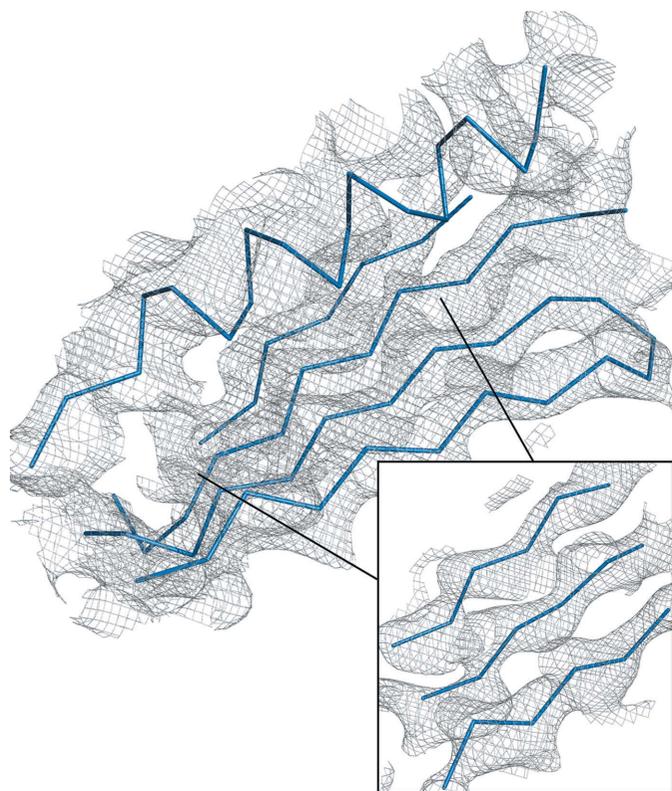
Substructure determination was performed using *SHELXC* and *SHELXD* (Sheldrick, 2008), making use of the *HKL2MAP* interface (Pape & Schneider, 2004). A total of five *SHELXD* runs (5000 trials each) were performed with different resolution-cutoff values. When the results were compared using *SITCOM* (Dall’Antonia & Schneider, 2006), a total of 27 sites occurring more than three times in these five *SHELXD* runs were found. In fact, after refinement we observed that only 32 sulfur-containing residues (24 Met and eight Cys) were ordered in the final model, with most of the sites presenting very high *B* factors (see Table 3). Also, three of the 27 consensus sites determined by *SITCOM* were shown to correspond to phosphorus ‘supersites’ from the ligand AMP-PCP, since the three P atoms from the ligand could not be distinguished at the substructure-determination stage. For comparison, the average *B*-factor values for S and P atoms in the refined structure are very similar (116 and 129 Å², respectively). Taking into consideration the observed sulfur positions and the solvent content after refinement, the Bijvoet ratio based on the observed model becomes 1.25%. The concatenation of a large multimer, disordered sites, low resolution, low crystal symmetry and a not particularly favourable Bijvoet ratio demonstrates that the structure solution was a difficult case for sulfur phasing.

The 27 initial sites were used in *Phaser* (McCoy *et al.*, 2007) to calculate phases for the two hands of the heavy-atom partial structure. One hand showed poor but nonetheless

interpretable electron density for a section of an α -helix and a β -sheet after solvent flattening with *Parrot* (Winn *et al.*, 2011; Fig. 2*a*). An automatic search for secondary-structure



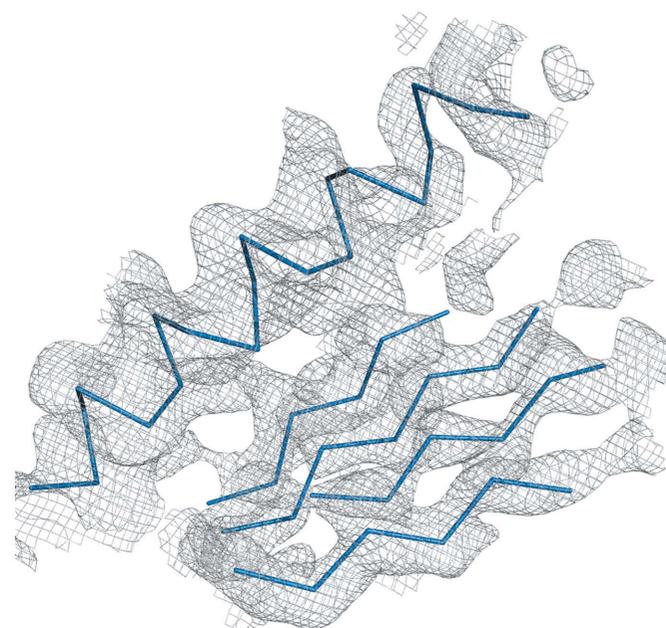
(a)



(b)

elements using *Coot* (Emsley *et al.*, 2010) succeeded in placing a helix and three β -strands in this map section. At this point, we took advantage of the availability of homologous structures of the CA subdomain in the PDB to position a truncated model in the electron density by means of the MR-SAD method, using the procedure of a six-dimensional search using phases implemented in *MOLREP* (Vagin & Teplyakov, 2010). The selected CA model for this step was based on PDB entry 3zxo (Cho *et al.*, 2013), which was one of the first hits in a homology-detection search performed with the *HHpred* server (<http://toolkit.tuebingen.mpg.de/hhpred>; Söding *et al.*, 2005). Further inspection with *Coot* allowed us to generate a partial model which was used, along with the completed heavy-atom sites from the previous *Phaser* run, to recalculate phases. The 'model' was also used without *a priori* heavy-atom information, allowing *Phaser* to determine heavy-atom sites and compare these with those from *SHELXD*, testing the validity of both the initially modelled residues and the correctness of the heavy-atom sites.

After solvent flattening with *Parrot*, the electron density showed significant improvement, allowing the positioning of additional secondary-structure elements in the map. A few cycles of model rebuilding and phase recalculation using a new partial model were performed, resulting in successive improvement of the electron density (Fig. 2*b*). Completion of the model was then performed using *Buccaneer* (Cowtan, 2006). The model based on S-SAD phases was later refined to 2.70 Å resolution against a native data set collected from newly grown crystals using 12.65 keV photons ($\lambda = 0.98$ Å;



(c)

Figure 2

(a) Solvent-flattened experimental electron-density map after phase calculation with *Phaser* using the 27 initial sites found by *SHELXD*. The relative and absolute contour levels for the map are 1.2σ and 0.176 e \AA^{-3} , respectively. A section of an α -helix and four β -strands (backbone trace in blue), which were used as a partial model for the second round of *Phaser*, are superposed on the map. (b) The same region of the electron-density map after a few cycles of model rebuilding, phase recalculation and solvent flattening. (c) Improved experimental electron-density map obtained after *a posteriori* analysis of different data-set combinations. The solvent-flattened electron-density map presented was obtained from data-set combination 2 (Table 4) using 30 initial sites found by *SHELXD* (Patterson search disabled) for phasing. The relative and absolute contour levels for the map are 1.2σ and 0.234 e \AA^{-3} , respectively. The backbone trace shown in blue corresponds to a model automatically built by *Coot*.

Native 1 in Table 2 and Supplementary Table S2). The model was refined using *BUSTER* (Bricogne *et al.*, 2011) with manual inspection, correction and completion as required and with 5% of reflections marked for the calculation of R_{free} using *XDSCONV* (Kabsch, 2010). The refined model to 2.70 Å resolution contained *A* chain residues 270–475, *B* chain residues 274–477, *C* chain residues 311–478 and *D* chain residues 310–475, with 32 ordered S sites. In addition, a molecule of AMP-PCP could be observed associated with each of the four monomers, clearly appearing in monomers *A* and *B* at the 2.8σ level (absolute contour level $0.294 e \text{ \AA}^{-3}$) in the $2mF_o - DF_c$ map based on the refined model. In a subsequent stage, a new HK construct was produced with a shorter cloning artifact at its C-terminus. Crystals from this new construct diffracted to 2.51 Å resolution and were used for the final refinement of the structure as shown in Table 2 and Supplementary Table S2 (Native 2).

Table 3 shows the sulfur locations from Cys and Met residues in the primary sequence from each chain, together with their *B* factors taken from the model refined against the ‘high-resolution native’ data at 2.70 Å resolution and from the final refined model to 2.51 Å resolution based on the shortened construct. Interestingly, only six sulfur sites appear in all four chains, and almost all sites show high *B* factors ($>90 \text{ \AA}^2$) in the longer construct. The large numbers of disordered sites hampered the determination of the NCS operators during the initial density-modification cycle.

2.4. *A posteriori* data analyses

Advances in ‘off-edge’ S-SAD phasing have been reported by Hendrickson and coworkers (Liu *et al.*, 2012, 2013) by combining data from many crystals and paying careful attention to the internal coherence of the data using a clustering approach. A slightly different approach was used in the present work, in which data sets from multiple orientations of the same crystal are combined as well as data from multiple crystals. The relative scale constants and correlation between data sets, as well as the unit-cell dimension differences, were used to cluster data sets together in a procedure similar to that described by Liu *et al.* (2012) (Supplementary Fig. S1). Multiple data sets from the same crystal were always very close together in the cluster plot, and globally the overall diffraction dissimilarity is quite low among all data sets. The ability to solve the structure from different data sets combinations was analysed *a posteriori* in an attempt to answer two questions in this specific case: (i) how many data were actually needed for a successful structure determination and (ii) whether the data-collection strategy had an impact on this minimum quantity of data.

In order to assess the ‘quality’ of the experimental phases obtained from the several data-set combinations, the highest resolution 2.51 Å model was refined against the original S-SAD data (Table 2) to produce the structure factors that were used as reference in our correlation analysis.

Data sets were combined in different ways. Initially, we tested five data-set combinations, for which we kept a

comparable multiplicity (Table 4), in order to investigate the following data-collection scenarios: (1) a single crystal, beam exposure at two positions of the crystal, multiple orientations ($\kappa = 0, 10$ and -10°); (2) three crystals, two orientations of each ($\kappa = 0$ and 10°); (3) the same as (2) but using a different combination of crystals; (4) five crystals, all data collections with $\kappa = 0^\circ$; (5) five crystals, the first data set collected from each (‘fresh crystals’), all data collections with $\kappa = 0^\circ$. For each data combination, nine runs of *SHELXD* were carried out combining different resolution cutoffs (4.0, 4.2 and 4.4 Å) and different numbers of heavy atoms for the search (44, 36 and 28). The heavy-atom consensus sites from the nine runs of *SHELXD* were selected by including the high- and medium-reliability groups as determined by *SITCOM* (Dall’Antonia & Schneider, 2006). Moreover, the same combination of parameters was tested with the Patterson search enabled and

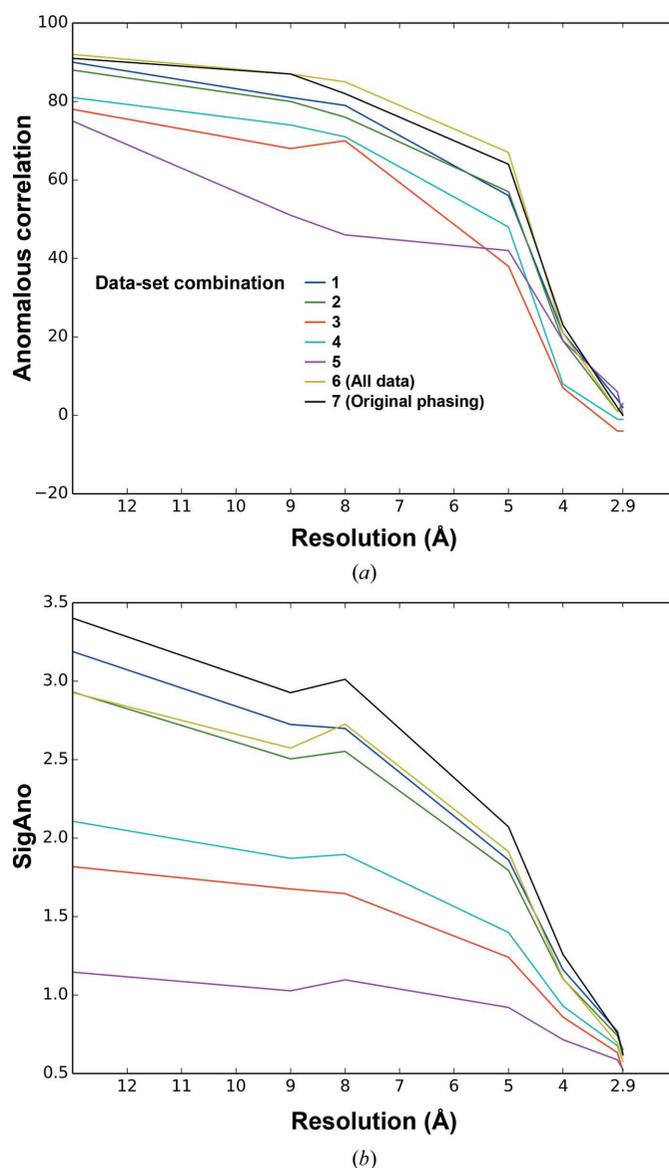


Figure 3
Anomalous correlation (a) and SigAno (b) calculated by *XSCALE* as a function of resolution. Plots correspond to the combinations presented in Table 4 [IDs 1–5, All data (6) and Original phasing (7)].

Table 4
Analysis of the ability to solve the structure using different data-set combinations.

The overall multiplicity was calculated from data merging to 2.9 Å resolution. Overall $\Delta F/\sigma(\Delta F)$ and anomalous correlation coefficients were calculated from data truncated to 4.2 Å resolution. Initial sites: consensus sites determined with *SITCOM* based on nine cycles of *SHELXD* using 4.0, 4.2 and 4.4 Å resolution cutoffs and looking for 44, 36 and 28 sites; the results obtained from enabled and disabled Patterson-based searches (PATS on and PATS off) were analyzed separately. Completed sites: number of sites refined/completed by *Phaser*. Correct sites: number of consensus sites determined by *phenix.emma* (tolerance = 4 Å) between the completed sites and the refined S and P sites from the reference model. CC without NCS: overall map correlation coefficient calculated with *SFTOOLS* to 3.5 Å resolution using the experimental phases after *Parrot* (no NCS automatically found) against the 'reference'. CC with NCS from heavy atoms: overall map correlation coefficient calculated with *SFTOOLS* to 3.5 Å resolution using the experimental phases after *Parrot* (ncs-mask-filter-radius set to 22 Å) against the 'reference'.

ID	1	2	3	4	5	2b	4b	6	7
Data sets	X113-all	X113a X113ak10 X112 x112k10 X104 X104k10	X112 X112k10 X104 X104k10 X102 X102k10	X113a X113b X112a X111c X104a X93a	X113a X112a X104a X102a X93a	X113a X113ak10 X112 X112k10 X104	X113a X113b X112a X104a X93a	All data	Original phasing
Data multiplicity	20.5	20.6	21.6	21.5	17.9	17.2	17.8	60.1	35.6
$\Delta F/\sigma(\Delta F)$	1.66	1.47	0.84	1.22	0.77	1.38	1.15	1.58	1.80
Anomalous CC (%)	52	45	37	40	30	42	41	58	57
PATS on									
Initial sites	37	40	39	28	19	42	7	59	
Completed sites	33	32	38	26	18	41	6	56	
Correct sites (r.m.s.)	12 (2.28)	7 (2.85)	6 (2.78)	6 (2.91)	5 (3.39)	8 (1.47)	1 (0.51)	23 (1.10)	
CC without NCS	0.10	0.01	0.03	-0.01	-0.02	0.08	0.04	0.37	
CC with NCS†	0.11	0.026 (NF)	0.007 (NF)	0.009 (NF)	-0.03 (NF)	0.07	0.03 (NF)	0.58	
PATS off									
Initial sites	26	30	29	21	5	5	15	48	27
Completed sites	28	31	30	22	5	6	15	50	29
Correct sites (r.m.s.)	22 (1.32)	25 (1.05)	20 (0.98)	13 (1.26)	3 (2.31)	1 (1.21)	2 (1.13)	32 (0.99)	23
CC without NCS	0.31	0.42	0.33	0.20	0.04	0.001	0.01	0.52	0.39
CC with NCS†	0.39	0.50	0.52	0.30	0.022 (NF)	0.01	-0.03 (NF)	0.65	(0.60)

† NCS from heavy atoms. NF, NCS not found.

disabled, and the resulting sites were treated separately (Table 4). The consensus sites were used in *Phaser* (McCoy *et al.*, 2007) to calculate phases, followed by density modification using *Parrot*. Two runs of *Parrot* were assayed, the first using the default parameters set in *CCP4* and the second setting the 'ncs-mask-filter-radius' parameter to 22 Å instead of 6 Å (the default value), which allows a larger mask filter radius for the NCS research. Map correlation coefficients were then calculated in reciprocal space with *SFTOOLS* (B. Hazes, unpublished) between the *Parrot* output and the reference (see above) for each data-set combination. We also compared the heavy-atom sites from *Phaser* output with the sulfur and phosphate sites from the refined model using *phenix.emma* (Adams *et al.*, 2002) in order to determine the number of 'correct sites' that were refined/completed by *Phaser* (Table 4).

Several observations arise from these first combinations of data (Table 4; IDs 1–5). (i) The highest correlation coefficients came from combinations 2 and 3 (multi-crystals, multi-orientation), followed by combination 1 (single crystal, multi-orientation), whereas combination 4 (multi-crystals, $\kappa = 0^\circ$) gives a borderline phasing solution. Thus, data collection at this wavelength using multiple orientations seems to play an important role in the success of structure determination. (ii) The potential contribution of the data sets to initial phasing could not always be reliably estimated from the anomalous correlation coefficients, as shown by the observation that the highest anomalous CCs do not necessarily correspond to the highest map correlation coefficients (Table 4 and Fig. 3). (iii)

No NCS could be automatically determined from the heavy-atom sites when *Parrot* (or *RESOLVE*) was run using the default parameters. As expected, a significant improvement in the map correlation coefficient was obtained after the application of NCS averaging. (iv) Phasing was unsuccessful for all data-set combinations when the heavy-atom search in *SHELXD* was performed using the Patterson search. A large number of consensus sites among the runs was found, but the number corresponding to 'correct sites' was significantly lower when compared with the sites obtained from the random-seeded search. At first sight, the number of consensus sites can be misleading in terms of the real success of the substructure determination, indicating that the Patterson seeding method could favour false consensus solutions. (v) Combination 5 (multi-crystals, $\kappa = 0^\circ$, data collection from a 'fresh crystal') was unsuccessful for phase determination. Considering that the overall multiplicity is slightly lower for this data combination than for the others (limited by the availability of data sets collected under these conditions), we considered the question of whether we have exceeded the minimal limit of multiplicity necessary for the consistent estimation of the anomalous differences.

In order to assay the 'minimal' multiplicity needed to solve the structure, we analysed two further data-set combinations, 2b and 4b (Table 4), for which one data set was omitted from combination 2 (a combination giving one of the best results in terms of phasing) and from a borderline case (combination 4). Interestingly, phase determination was unsuccessful in both

cases after applying the same procedure as described previously. This result, together with the result obtained with combination 5, suggests that an overall multiplicity of approximately 20 is necessary to solve the structure in our case.

For comparative purposes, Table 4 also includes a combination of all of the data sets collected (6) as well as the data combination originally used for phasing (7) (Table 2, S-SAD data set). It is important to remark that in our original phasing procedure model building was started using the experimental map generated by a run of *Parrot* for which no NCS had been automatically determined from the heavy-atom sites. Our *a posteriori* analysis shows that the correlation coefficient calculated using the reference structure factors is reasonable but, as expected, it would have been largely improved by the use of NCS averaging (CC of 0.60 instead of 0.39), which would have resulted in an experimental map of better quality and easier model building. In our case, the use of a partial model as input to recalculate phases allowed the initial difficulty of automatic and/or manual model building to be overcome. The NCS operators could then be determined based on the model, largely contributing to the improvement of the subsequent solvent-flattened maps. However, when no homology models are available it is worth trying different density-modification programs and/or different protocols in order to determine the NCS operators from the heavy-atom sites alone. Fig. 2(c) shows the solvent-flattened experimental electron-density map obtained from data-set combination 2 (Table 4) with the use of NCS averaging. The quality of the map allowed the automatic building of secondary-structure segments using *Coot*.

Finally, we analyzed the impact of the data-resolution cutoff on the phasing results. The cutoff value originally applied for the S-SAD data sets was rather conservative considering the $CC_{1/2}$ values presented in Tables 1 and 2. Taking the data-set combination 1 as an example (Table 4, X113-all), reprocessing the individual data sets keeping the data to 2.5 Å resolution and applying a new cutoff of 2.75 Å for the merged data (instead of the 2.9 Å previously used) resulted in $CC_{1/2} = 45\%$ and $\langle I/\sigma(I) \rangle = 0.5$ for the outer shell (data not shown). We then used the same protocol as previously described for the heavy-atom search and phasing. Interestingly, we observed that the new cutoff impaired the substructure determination (a smaller number of correct sites was found) and consequently the phasing step (data not shown). However, when the initial sites previously determined using the conservative cutoff were used for phasing with the new data set, the extended resolution was revealed to be largely beneficial to the density-modification procedure. In particular, additional NCS operators were found and consequently the quality of the solvent-flattened experimental electron-density map was improved.

3. Conclusion

It is interesting to contrast the example presented in this work (a relatively large 108 kDa asymmetric unit, low-symmetry $P2_1$ space group and moderate 2.9 Å resolution) with other

S-SAD structures of similar complexity solved previously. A literature survey failed to find a structure solved under a completely unfavourable scenario for all three parameters. Two main observations can be derived from the comparisons between HK and other S-SAD structures: (i) the crystals used to solve structures in low-symmetry space groups ($P2_1$ and $C2$) with comparable asymmetric units in size, as for example in Kitamura *et al.* (2008), Lakomek *et al.* (2009) and Weinert *et al.* (2015), always diffracted to better resolutions than HK and (ii) structures at resolutions equal or worse than 3 Å (Debrezzeni, Bunkoczi, Girmann, *et al.*, 2003; Akey *et al.*, 2014; El Omari *et al.*, 2014) always showed higher symmetry space groups than that of HK. A special case corresponds to the human CIB1 protein (PDB entry 1xo5; Gentry *et al.*, 2005), which bears fairly comparable parameters to those of HK (41 kDa asymmetric unit, space group $P2_1$ and 2.6 Å resolution). However, this structure presents the important advantage of having eight Ca^{2+} ions in the asymmetric unit that contribute a significant amount to the anomalous scattering of the crystal.

Owing to the fact that the very first goal of our experiment was ‘determining the structure’, the applied data-collection strategy was not optimized to allow a complete methodological analysis. However, based on this experiment and on our *a posteriori* analyses we believe we can come to some practical observations. (i) A well known success factor in anomalous phasing is data multiplicity. This is clearly exemplified in the present case, where strong data multiplicity (Table 4, All data) allowed us to move away from borderline solutions. (ii) In addition to the use of multiple crystals, our results reinforce the advantage of the use of multiple sample orientations to increase ‘true redundancy’. In brief, multiple crystals plus multiple orientations worked better for phasing than a single crystal in multiple orientations, and that was better than multiple crystals in single orientations. The simple method of data collection (several exposures at different κ offsets) provides a straightforward protocol to improve the quality of anomalous difference measurements. This approach is well illustrated in the very recent article of Weinert *et al.* (2015). (iii) The total X-ray dose for a 400° data set used in the present work (estimated at 0.4 MGy) proved to be suitable and might be used as a hint for a data-collection strategy in cases of low-symmetry space groups. (iv) When a large number of heavy-atom sites is expected, the use of a *SHELXD* Patterson-based search can mislead the interpretation of the substructure-determination results and lead to failure of the phasing process. The simple change to a random-seeded search may favour the correct determination of the heavy-atom sites in these cases. (v) When more than one equivalent monomer is expected in the asymmetric unit, it is worth trying to obtain the most that we can from NCS averaging, even in the presence of incomplete partial structures, in order to improve the quality of the initial unbiased experimental map. Also, the NCS operators can help to complete the substructure by creating a NCS consensus model. (vi) The use of all available information about the protein under study can be essential to the structure determination; even a very incomplete partial model provides valuable input into phase calculation.

Although none of the observations above is new to crystallographers' debates, the present case illustrates that taking them together is particularly important when treating adverse cases of S-SAD phasing (large asymmetric unit, low-symmetry space group, medium-to-low data resolution). Hopefully, these observations will be useful to less experienced macromolecular crystallography beamline users. Also, we hope that this example will encourage the solution of other *a priori* unfavourable cases, widening the use of light elements that are naturally present in protein crystals (S, P, K, Ca, Cl *etc.*) in off-edge low-energy SAD phasing experiments.

4. Related literature

The following references are cited in the Supporting Information for this article: Chen *et al.* (2010) and Engh & Huber (1991).

Acknowledgements

This work was supported by the Argentinian Research Council (CONICET) and the Argentinian Agency for Scientific and Technological Development (ANPCyT). We acknowledge access to the SOLEIL synchrotron, France. We are grateful to the Argentinian Ministry of Science (MINCyT) and the Center for Structural Biology from the Mercosur (CEBEM) for travel support and to Dr Alejandro Buschiazzo for initial diffraction tests at the Institut Pasteur Montevideo, Uruguay.

References

- Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Cryst.* **D58**, 1948–1954.
- Akey, D. L., Brown, W. C., Konwerski, J. R., Ogata, C. M. & Smith, J. L. (2014). *Acta Cryst.* **D70**, 2719–2729.
- Albanesi, D., Martín, M., Trajtenberg, F., Mansilla, M. C., Haouz, A., Alzari, P. M., de Mendoza, D. & Buschiazzo, A. (2009). *Proc. Natl Acad. Sci. USA*, **106**, 16185–16190.
- Ben-Shem, A., Garreau de Loubresse, N., Melnikov, S., Jenner, L., Yusupova, G. & Yusupov, M. (2011). *Science*, **334**, 1524–1529.
- Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., Smart, O. S., Vornrhein, C. & Womack, T. O. (2011). *BUSTER v. 2.10.0*. Cambridge: Global Phasing Ltd.
- Brockhauser, S., Ravelli, R. B. G. & McCarthy, A. A. (2013). *Acta Cryst.* **D69**, 1241–1251.
- Broennimann, C., Eikenberry, E. F., Henrich, B., Horisberger, R., Huelsen, G., Pohl, E., Schmitt, B., Schulze-Briese, C., Suzuki, M., Tomizaki, T., Toyokawa, H. & Wagner, A. (2006). *J. Synchrotron Rad.* **13**, 120–130.
- 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.* **D66**, 12–21.
- Cho, H. Y., Lee, Y.-H., Bae, Y.-S., Kim, E. & Kang, B. S. (2013). *J. Biol. Chem.* **288**, 12437–12447.
- Cianci, M., Helliwell, J. R. & Suzuki, A. (2008). *Acta Cryst.* **D64**, 1196–1209.
- Cowtan, K. (2006). *Acta Cryst.* **D62**, 1002–1011.
- Dall'Antonia, F. & Schneider, T. R. (2006). *J. Appl. Cryst.* **39**, 618–619.
- Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.
- Debreczeni, J. É., Bunkóczi, G., Girmann, B. & Sheldrick, G. M. (2003). *Acta Cryst.* **D59**, 393–395.
- Debreczeni, J. É., Bunkóczi, G., Ma, Q., Blaser, H. & Sheldrick, G. M. (2003). *Acta Cryst.* **D59**, 688–696.
- Doan, D. N. & Dokland, T. (2003). *Structure*, **11**, 1445–1451.
- El Omari, K., Iourin, O., Kadlec, J., Fearn, R., Hall, D. R., Harlos, K., Grimes, J. M. & Stuart, D. I. (2014). *Acta Cryst.* **D70**, 2197–2203.
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* **D66**, 486–501.
- Engh, R. A. & Huber, R. (1991). *Acta Cryst.* **A47**, 392–400.
- Gabadiño, J. *et al.* (2010). *J. Synchrotron Rad.* **17**, 700–707.
- Gentry, H. R., Singer, A. U., Betts, L., Yang, C., Ferrara, J. D., Sondek, J. & Parise, L. V. (2005). *J. Biol. Chem.* **280**, 8407–8415.
- Goulet, A., Vestergaard, G., Felisberto-Rodrigues, C., Campanacci, V., Garrett, R. A., Cambillau, C. & Ortiz-Lombardía, M. (2010). *Acta Cryst.* **D66**, 304–308.
- Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
- Hülsen, G., Broennimann, C., Eikenberry, E. F. & Wagner, A. (2006). *J. Appl. Cryst.* **39**, 550–557.
- Kabsch, W. (2010). *Acta Cryst.* **D66**, 125–132.
- Kahn, R., Fourme, R., Caudron, B., Bosshard, R., Benoit, R., Bouclier, R., Charpak, G., Santiard, J. C. & Sauli, F. (1980). *Nucl. Instrum. Methods*, **172**, 337–344.
- Karniol, B. & Vierstra, R. D. (2004). *J. Bacteriol.* **186**, 445–453.
- Kitamura, M., Okuyama, M., Tanzawa, F., Mori, H., Kitago, Y., Watanabe, N., Kimura, A., Tanaka, I. & Yao, M. (2008). *J. Biol. Chem.* **283**, 36328–36337.
- Lakomek, K., Dickmanns, A., Mueller, U., Kollmann, K., Deuschl, F., Berndt, A., Lübke, T. & Ficner, R. (2009). *Acta Cryst.* **D65**, 220–228.
- Liu, Q., Dahmane, T., Zhang, Z., Assur, Z., Brasch, J., Shapiro, L., Mancina, F. & Hendrickson, W. A. (2012). *Science*, **336**, 1033–1037.
- Liu, Q., Guo, Y., Chang, Y., Cai, Z., Assur, Z., Mancina, F., Greene, M. I. & Hendrickson, W. A. (2014). *Acta Cryst.* **D70**, 2544–2557.
- Liu, Q., Liu, Q. & Hendrickson, W. A. (2013). *Acta Cryst.* **D69**, 1314–1332.
- Liu, Q., Zhang, Z. & Hendrickson, W. A. (2011). *Acta Cryst.* **D67**, 45–59.
- Liu, Z.-J., Vysotski, E. S., Vysotski, E. S., Chen, C.-J., Rose, J. P., Lee, J. & Wang, B.-C. (2000). *Protein Sci.* **9**, 2085–2093.
- 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.
- Micossi, E., Hunter, W. N. & Leonard, G. A. (2002). *Acta Cryst.* **D58**, 21–28.
- Müller, J. J., Weiss, M. S. & Heinemann, U. (2011). *Acta Cryst.* **D67**, 936–944.
- Nan, J., Zhou, Y., Yang, C., Brostromer, E., Kristensen, O. & Su, X.-D. (2009). *Acta Cryst.* **D65**, 440–448.
- Nawrotek, A., Guimarães, B. G., Velours, C., Subtil, A., Knossow, M. & Gigant, B. (2014). *J. Biol. Chem.* **289**, 25199–25210.
- Pape, T. & Schneider, T. R. (2004). *J. Appl. Cryst.* **37**, 843–844.
- Phillips, J. C., Cerino, J. A. & Hodgson, K. O. (1979). *J. Appl. Cryst.* **12**, 592–600.
- Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003). *Acta Cryst.* **D59**, 868–875.
- Rivera-Cancel, G., Ko, W. H., Tomchick, D. R., Correa, F. & Gardner, K. H. (2014). *Proc. Natl Acad. Sci. USA*, **111**, 17839–17844.
- Ru, H., Zhao, L., Ding, W., Jiao, L., Shaw, N., Liang, W., Zhang, L., Hung, L.-W., Matsugaki, N., Wakatsuki, S. & Liu, Z.-J. (2012). *Acta Cryst.* **D68**, 521–530.
- Sarma, G. N. & Karplus, P. A. (2006). *Acta Cryst.* **D62**, 707–716.
- Sheldrick, G. M. (2008). *Acta Cryst.* **A64**, 112–122.
- Smith, J. L. (1997). In *Proceedings of the CCP4 Study Weekend. Recent Advances in Phasing*, edited by K. S. Wilson, G. Davies, A. W. Ashton & S. Bailey. Warrington: Daresbury Laboratory.
- Söding, J., Biegert, A. & Lupas, A. N. (2005). *Nucleic Acids Res.* **33**, W244–W248.

- Swartz, T. E., Tseng, T.-S., Frederickson, M. A., Paris, G., Comerci, D. J., Rajashekara, G., Kim, J.-G., Mudgett, M. B., Splitter, G. A., Ugalde, R. A., Goldbaum, F. A., Briggs, W. R. & Bogomolni, R. A. (2007). *Science*, **317**, 1090–1093.
- Usón, I., Schmidt, B., von Bülow, R., Grimme, S., von Figura, K., Dauter, M., Rajashankar, K. R., Dauter, Z. & Sheldrick, G. M. (2003). *Acta Cryst. D* **59**, 57–66.
- Vagin, A. & Teplyakov, A. (2010). *Acta Cryst. D* **66**, 22–25.
- Vasur, J., Kawai, R., Larsson, A. M., Igarashi, K., Sandgren, M., Samejima, M. & Ståhlberg, J. (2006). *Acta Cryst. D* **62**, 1422–1429.
- Wagner, A., Pieren, M., Schulze-Briese, C., Ballmer-Hofer, K. & Protz, A. E. (2006). *Acta Cryst. D* **62**, 1430–1434.
- Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
- Weinert, T. *et al.* (2015). *Nature Methods*, **12**, 131–133.
- Winn, M. D. *et al.* (2011). *Acta Cryst. D* **67**, 235–242.
- Yang, C., Pflugrath, J. W., Courville, D. A., Stence, C. N. & Ferrara, J. D. (2003). *Acta Cryst. D* **59**, 1943–1957.
- Zeldin, O. B., Gerstel, M. & Garman, E. F. (2013). *J. Appl. Cryst.* **46**, 1225–1230.
- Zhu, J.-Y., Fu, Z.-Q., Chen, L., Xu, H., Chrzas, J., Rose, J. & Wang, B.-C. (2012). *Acta Cryst. D* **68**, 1242–1252.