De novo crystal structure determination of double stranded RNA binding domain using only the sulfur anomalous diffraction in SAD phasing

Beatriz Gomes Guimarães a,1, Béatrice Golinelli-Pimpaneau b,∗

a Synchrotron Soleil, L’Orme des Merisiers, Saint-Aubin 91990 Gif-sur-Yvette, France
b Laboratoire de Chimie des Processus Biologiques, UMR 8229 CNRS, Collège de France, Sorbonne Université, 11 Place Marcellin Berthelot, 75005 Paris, France

1. Introduction

The general method for de novo solving the crystal structure of a protein is to prepare the selenomethionylated protein and use the anomalous signal of Se atoms to determine experimental phases using the multiple-wavelength anomalous dispersion (MAD) (Hendrickson, 1991) or single-wavelength anomalous dispersion (SAD) methods (Hendrickson, 2014). Compared to MAD, only one data set at or around the absorption edge of the anomalous atom (“anomalous” data set) is needed for SAD, which minimizes the effect of radiation damage. In addition, a so-called “native” data set may be used, in some cases, for final structure refinement. Although successfully used in a large number of cases, Se-SAD phasing has some drawbacks. First, it requires Se-labeling of the protein. Second, it is never known in advance if expression and solubility of the selenomethionylated protein will be sufficient and if crystallization will occur in the same conditions as the native protein.

In contrast, the native SAD method can be used to derive experimental phases directly from native protein crystals without the need of additional derivatives. Indeed, proteins contain different types of atoms that can be used as anomalous scatterers, although much weaker than selenium. In particular, sulfur is the heaviest atom in most native proteins, present within the cysteine and methionine amino acids that compose the polypeptide chain. In addition, other atoms such as Ca, Cl, Mg, K, Na, are occasionally - and sometimes fortuitously (Hegde et al., 2017) - found within the protein crystal, as determined by the X-ray fluorescence emission spectrum. Their anomalous signal can be exploited together with the sulfur signal for successful native SAD phasing (Dauter et al., 1999; Goulet et al., 2010; Liu et al., 2012; Doutch et al., 2012; Liu and Hendrickson, 2015; Rose et al., 2015; Weinert et al., 2015; Oleric et al., 2016; Olczak and Cianci, 2018).

Crambin was the first protein whose structure was solved by sulfur-SAD (Hendrickson and Teet, 1981). Favorable factors included a large (15%) sulfur content (i.e. percentage of sulfur-containing residues in the polypeptide chain), the presence of disulfide bonds, which account as “super-sulfur” sites (Kraatz et al., 2018), and diffraction to high resolution (1.5 Å). Later developments in diffraction hardware and crystallographic programs, together with extended and judicious use of data statistics (Foos et al., 2019; Assmann et al., 2020) made native-SAD phasing accessible to numerous other proteins (Liu et al., 2012; Liu and Hendrickson, 2015; Rose et al., 2015). The absorption edge of sulfur...
is $\lambda = 5.02$ Å. At such a high wavelength, where anomalous scattering is optimal, radiation damage is enhanced and X-ray absorption becomes severe. Hence, as a compromise between higher anomalous signal at low energy and lower noise at high energy, sulfur-SAD phasing is usually carried out “off-edge” at shorter wavelengths (1.5–3 Å). Home-source with Cu ($\lambda = 1.54$ Å) (Sarma and Karplus, 2006) or Cr ($\lambda = 2.29$ Å) radiations (Watanabe, 2006) have been used to solve the structure of large and well-diffracting crystals, whereas synchrotron beamlines have been developed to optimize beam stability and minimize X-ray absorption effects for native SAD data collection in more difficult cases (Doutch et al., 2012; Weinert et al., 2015; Wang et al., 2006; Basu et al., 2019a; De Sanctis et al., 2016; Cianci et al., 2017). Moreover, the development of pixel array detectors that operate in single photon counting mode such as the Pilatus (Broennimann et al., 2006) has been a breakthrough in macromolecular crystallography by enabling noise-free detection and novel data-acquisition modes.

To accurately measure the very weak sulfur anomalous signal at standard synchrotron beamlines and optimize the signal-to-noise ratio, highly redundant diffraction data are needed (Liu et al., 2012, 2014; Doutch et al., 2012; Ramagopal et al., 2005; Cianci et al., 2008), which may, however, increase radiation damage. Among the various synchrotron data-collection strategies that have been proposed to optimize native-SAD (Olieric et al., 2016) - including sulfur-SAD - phasing, the use of a long-wavelength (2.7 Å) over 1.9 Å (Cianci et al., 2016; Banerjee et al., 2016; Basu et al., 2019b) was shown to be advantageous for 100 μm or smaller crystals (Basu et al., 2019b; Liesbchner et al., 2016). Moreover, successful native SAD phasing was achieved in several difficult cases using low X-ray dose, high redundancy and multi-orientations data collection of multiple crystals (Liu et al., 2012, 2014; Assmann et al., 2020; Banerjee et al., 2016; Akey et al., 2014; El Omari et al., 2014; Klinke et al., 2015), or even a single crystal (Weinert et al., 2015; Basu et al., 2019a,b), which obviously removes the problem of non-isomorphism between various crystals. The latest technical advances include solution-free mounting systems to minimize X-ray absorption at long-wavelengths (Yu et al., 2020) and assembly of numerous partial single-crystal data sets from microcrystals with sizes less than 10 μm (Cianci et al., 2019; Guo et al., 2019; Nass et al., 2020). Despite these improvements, native SAD phasing is still not yet a routine method for successful structure solution (Rose et al., 2015; Basu et al., 2019a) because success is not only dependent on the content in anomalous scattering atoms but also on the crystal features, with a high diffracting power and high symmetry being advantageous.

In several cases, native SAD phasing was achieved using atoms other than S as scatterers, such as Ca, Cl, K, Na, Mg, which were constitutive of the protein or originated fortuitously from the crystallization solution (Dauter et al., 1999; Weinert et al., 2015; Olieric et al., 2016; Basu et al., 2019a, 2019b; Mueller-Dieckmann et al., 2007). Because the anomalous scattering factors $f$ of the atoms that are heavier than S (Ca, Cl, K) are higher than that of sulfur in the 1–2 Å wavelength range used at synchrotrons (see Fig. 1 in Mueller-Dieckmann et al., 2007), these atoms had often an important contribution to the phasing success (Goulet et al., 2010; Rose et al., 2015; Weinert et al., 2015; Olieric et al., 2016). Here, we review the recent cases, described in the literature, where native SAD phasing was used successfully using no atoms heavier than sulfur as anomalous scatterers (Table 1). Difficult cases include crystals belonging to low symmetry space groups (Basu et al., 2019a; Banerjee et al., 2016; Klinke et al., 2015), diffracting at poor resolution (Liu et al., 2014; Cianci et al., 2016; Basu et al., 2019b; Akey et al., 2014; El Omari et al., 2014; Klinke et al., 2015) with an extremely low sulfur content (Basu et al., 2019a; Gorgel et al., 2015) or large substructure (Basu et al., 2019b; Akey et al., 2014; Klinke et al., 2015). We also report our experience, where several data sets collected in various goniometer orientations on a single

![Fig. 1. Analysis of substructure determination using the D1, D5, D6 and D7 data sets. A Profiles of the anomalous signals $<\Delta F>/\sigma F$ as a function of resolution, as determined by SHELXC. B Plot of site occupancy versus peak heights for the top 25 peaks observed in the Bijvoet difference Fourier maps for the search of 15 S sites with SHELXD. C Histogram of the $CC_{n}$ correlation coefficients between observed and calculated Bijvoet differences out of 500 tries of SHELXD. D Contrast analysis of original and inverted hands versus cycle during phasing with SHELXE.](image-url)
Table 1
Protein structures solved using synchrotron radiation and only the sulfur anomalous signal in SAD phasing. Particular features (such as low symmetry, poor diffraction, low solvent content, large asymmetric unit content) that make structure determination by S-SAD more challenging are indicated in bold. Examples of other structures solved by native SAD using sulfur and another scattering atom can be found in Table 1 of Olczak and Cianci, 2018 and Supplementary Tables 1 of Weinert et al. (2015) and Gorgel et al. (2015).

| Protein | PDB code | S-SAD λ (Å) | Diffraction limit (Å) for native data set | Space group | Sulfur content | Number of Cys + Met in asymmetric unit | Number of residues in asymmetric unit (kDa) | Solvent % | Collection strategy | Total rotation range (˚)/redundancy | Predicted Bijvoet ratio ΔF/F (%) | substructure determination: initial number of sites and resolution cutoff (Å) | N° of crystals | Crystal size (μm³) | Beamline Reference |
|---------|----------|-------------|------------------------------------------|-------------|----------------|----------------------------------------|---------------------------------------------|---------|-------------------|---------------------------------|---------------------------------|---------------------------------------------|---------------|-----------------|-----------------|
| Lam16A 2c12 | 1.775 (7.0) | 1.34 | P2₁₂₂₁ | 35 | 13 (including 4 SS bridges) | 298 (36) | Single orientation of a unique crystal | 500/19.0 | 1.1 | 13 (2.5) | 1 | 500 x 500 x 500 | 1D14-4, ESRF Grenoble, France | Vasur et al. (2006) | Zhu et al. (2012) |
| ORF 1382 from Archaeoglobus fulgidus 3o3k | 1.9 (6.25) | 2.3 | P4₂ | 4 | 95 (11.1) | 54.4 | Multiple orientations of a single crystal | 720/25 | 1.4 | 4 (?) | 1 | | | |
| flavivirus NS1 4tpl | 1.746 (7.1) | 2.0 | P3₂₁ | 17 x 2 (including 12 SS bridges) | 377 x 2 (85) | 72.5 | Multiple crystals Inversed beam | 28 x 180/184.1 | 1.3 | 31 (5.2) | 28 | 50 x 50 x 150 | | Akey et al. (2014) |
| Thiamine transporter ThiT 4tkr | 2.066 (6.0) | 3.0 | P₃₂₁ | 8 x 2 | 184 x 2 (40.3) | 69.0 | Multiple crystals Inversed beam | 2600/140.2 | 1.5 | 15 (?) | 5 | 75 x 75 x 75 | | Liu et al. (2014) |
| EGFR kinase domain 4tkr | 2.066 (6.0) | 3.2 | I23 | 15 | 318 (36.8) | 64.3 | Multiple crystals Inversed beam | 2160/254.1 | 1.5 | 15 (?) | 4 | 100 x 100 x 100 | | Liu et al. (2014) |
| Nter domain of the ectodomain of HCV E1 4us7 | 1.77 (7.0) | 3.5 | P₄₁₂₂ | 4 x 6 (including 10 SS bridges) | 79 x 6 (57.6) | 52 | Multiple crystals Inversed beam | 90 x 32/121.5 | 1.1 | 12 (7) | 32 | 110 x 30 x 10 | | Liu et al. (2014) |
| D29 construct of CopN microtubule-destabilizing protein 4p3z | 1.9 (6.525) | 1.77 | P₂₁₂₂₁ | 11 (including four sulfate ions) | 424 (46.4) | 32 | Multiple orientations of a single crystal | ?/10.6 | 1.4 | 7 (?) | 1 | | | |
| Brucella abortus Histidine kinase domain 4p3x | 1.8 (6.888) | 2.0 | P₂₁ | 10 x 4 (108) | 52 | Multiple orientations of multiple crystals | 4000/35.6 | 1.2 | 27/32 (4) | 3 | 400 x 100 x 50 | | PXI SOLEIL France | Klinke et al. (2015) |
| Pilin PilBac1 4u7 | 1.8 (6.888) | 1.96 | I222 | 3 x 2 (including 2 SS bridges and 2 sulfate ions) | 89 x 2 (19.8) | 62.4 | Multiple orientations of a single crystal | ?/5.5 | 0.9 | 5 (?) | 1 | 70 x 70 x 700 | | BL-14.1, BESSY II, Germany | Gorgel et al. (2015) |
| hypothetical protein pf1771 4wbx | 2.066 (6.0) | 2.5 | F222 | 13 | 395 (44.0) | 53 | Low-dose & Multiple | 360 x 5/62.7 | 1.3 | 13 (3.7) | 1 | 150 x 100 x 100 | | Weinert et al. (2015) |

(continued on next page)
| Protein | PDB code | S-SAD \( \lambda \) (Å) | Diffraction limit (Å) for native data set | Space group | Sulfur content | Number of residues in asymmetric unit | Solvent % | Collection strategy | Total rotation range \( (\%) / \text{redundancy} \) | Predicted Bijvoet ratio \( \frac{\Delta F}{F}(\%)^a \) | Substructure determination: initial number of sites and resolution cutoff (Å) | N° of crystals | Crystal size (\( \mu m^3 \)) | Beamline | Reference |
|---------|----------|-------------------------|------------------------------------------|-------------|----------------|-------------------------------------|----------|-------------------|-----------------------------------------|------------------------------------------|------------------------------------------|-------------|---------------------|----------|---------------------|
| Human centromere protein M CENP-M 4awu | 2.066 (6.0) | 2.2 | P3 | 4.5 | 7 \times 2 | 155 \times 2 (34.1) | 55.3 | orientations of a single crystal Low-dose & Multiple orientations of a single crystal | \( 720 \times 5 + 2880/140.4 \) | 1.5 | 14 (2.7) | 1 | 100 x 100 \( < 150 \) | X06DA Swiss Light Source, Switzerland | Weinert et al. (2015) |
| Cdc-23 Nter Sfp | 2.7 (4.6) | P4_3 | 3.9 | 11 \times 2 | 282 \times 2 (65.4) | 53 | Long wavelength 2 data sets from a single crystal | \( 720/27.8 \) | 2.2 | 17 (4) | 1 | \( \sim 50 \times 50 \times 100 \) | P13 DESY Germany | Cianci et al. (2016) |
| Lili-Mip 4nyr | 2.7 (4.6) | P1 | 3.05 | 5 (including 2 SS bridges) | 164 (18.7) | 42.5 | Long wavelength Multiple crystals | \( 7 \times 720/23 \) | 2.0 | ? | 7 | 10 x 10 x 30 | BL-1A, Photon Factory, Japan | Banerjee et al. (2016) |
| Helicase Sen1 6i59 | 2.7 (4.6) | P2_1_2_2_1 | 4.4 | 31 | 749 (85.1) | 61.6 | Long wavelength Multiple orientations of a single crystal | 1440/49.1 | 2.3 | 22 (3.3) | 1 | 220 x 100 x 50 | BL-1A KEK Photon Factory, Japan | Basu et al. (2019b) |
| Streptavidin-biotin (SavB) 6m9b | 2.07 (6.0) | P2_1_2 | 0.8 | 1 \times 4 (4 S from biotin) | 127 \times 4 (53.1) | 46.5 | Low-dose Multiple orientations of a single crystal | 14 \times 360/39.0 | 0.6 | 4 (7) | 1 | 50 x 150 x 200 | X06DA Swiss Light Source, Switzerland | Basu et al. (2019a) |
| Thaumatin 6o8a | 2.48 (5.0) | P4_1_2_2_1 | 8.2 | 17 (including 8 SS bridges) | 207 (22.2) | 56.3 | Polymide well mounts & iterative frame rejection | \( 1381 \times 20/322.7 \) | 2.8 | ? | 1381 | Micro-crystals \(< 10 \mu m\) | FMX (NSLS-II), Brookhaven National Lab, NY, USA | Guo et al. (2019) |
| Bacteriophage T4 Spackle 6x6o | 1.653 (7.0) | I222 | 7.2 | 6 \times 2 (including 2 SS bridges) | 83 \times 2 (19.5) | 28.2 | Single orientation of a single crystal | \( 720/21.9 \) | 1.5 | 12 (2.2) | 1 | ? | 24-ID-C NE-CAT Advanced Photon Source, Argonne National | Shi et al. (2020) |
| Eisenia hydrolysis protein (EHEP) Not deposited at the PDB | 2.1 (5.9) | P2_1_2_2_1 | 9.5 | 20 | 209 (22.6) | 32.7 | Solution-less crystal mounting Multiple orientations of a single crystal | \( 8 \times 720/105 \) | 2.3 | 20 (7) | 1 | 270 \times 50 x 50 | Laboratory USA BL17A, Photon factory, Japan | Yu et al. (2020) |
| dsRBD of DUS 4WFT | 1.8 (6.888) | P4_3 | 4.1 | 5 \times 3 | 119 \times 3 (41.1) | 41.8 | Low-dose & Multiple orientations of a single crystal | See text | \( \text{1.08} \) | See text | 1 | 150 x 150 x 100 | PXI SOLEIL France | This paper |

\( \text{? not indicated in the original paper.} \\
^a \text{Except N-terminal methionine.} \\
^b \text{The expected Bijvoet ratio of the mean anomalous difference to the mean amplitude expected in the data at a given wavelength can be estimated by the formula} \( \text{\( \left( \frac{2N_A \lambda^2}{(346 N_B \lambda^2)} \right)^{1/2} f, \)} \text{where} \( N_A \) \text{and} \( N_B \) \text{are the number of anomalous atoms and number of protein residues in the asymmetric unit, respectively, and} \( f \) \text{is the atomic scattering factor of the anomalous atom (here sulfur), which depends on the wavelength (Hendrickson and Teet, 1981). If anomalous scatterers form a group (e.g. a disulfide or a metal cluster), this changes the dependence of the anomalous signal on resolution (Banumathi et al., 2003). The theoretical Bijvoet ratios for 5-SAD structures from the PDB range from 0.6 to 4.1%, with an average of 1.4% (Rose et al., 2015). Bijvoet ratios below the average number have been highlighted.} \)
| Crystal | D1 | D1 | D1 | D5 | D5 | D5 | D5 | D6 | D6 | D6 | D6 | D6 | D6 | D6 | D7 | D7 | D7 | D7 | D7 | D7 | D7 |
|---------|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|----|
| Data set name | native | sad1 | sad2 | native | sad1 | sad2 | sad3 | native | sad1 | sad2 | sad3 | native | sad1 | sad2 | sad3 | native | sad1 | sad2 | sad3 | native | sad1 | sad2 |
| Rotation axis | ϕ | ϕ | ω | ϕ | ω | ω | ϕ | ω | ω | φ | φ | φ | φ | φ | φ | φ | φ | ω | ω |
| κ angle | 001 | 5 | 001 | 5 | 001 | 5 | 001 | 5 | 001 | 5 | 001 | 5 | 001 | 5 | 001 | 5 | 001 | 5 | 001 | 5 | 001 | 5 |
| Unit cell parameters (Å) | | | | | | | | | | | | | | | | | | | | | |
| a, c | 55.5 115.3 | 55.6 115.4 | 55.7 115.7 | 55.5 115.3 | 55.8 115.7 | 55.5 115.3 | 55.5 115.3 | 55.5 115.3 | 55.5 115.3 | 55.5 115.3 | 55.6 115.3 | 55.8 115.7 | 55.8 115.9 | 55.5 115.3 | 55.5 115.3 | 55.5 115.3 | 55.5 115.3 | 55.5 115.3 | 55.5 115.3 |
| Transmission (%) | 50 | 50 | 50 | 40 | 40 | 40 | 40 | 40 | 30 | 20 | 20 | 20 | 20 | 5 | 5 | 5 | 5 | 5 |
| Number of frames | 600 | 800 | 900 | 500 | 800 | 900 | 900 | 900 | 2000 | 2000 | 2000 | 2000 | 2000 | 2500 | 2500 | 2500 | 900 | 900 |
| Resolution (Å) | 50 – 1.87 | 50 – 2.2 | 50 – 2.2 | 50 – 2.2 | 50 – 2.2 | 50 – 2.2 | 50 – 2.2 | 50 – 2.2 | 50 – 2.2 | 50 – 2.2 | 50 – 2.2 | 50 – 2.2 | 50 – 2.2 | (2.33 – 2.20) | (2.33 – 2.20) | (2.33 – 2.20) | (2.33 – 2.20) | (2.33 – 2.20) | (2.33 – 2.20) | (2.33 – 2.20) |
| Measurements | 128676 (20494) | 98820 (13852) | 111618 (15430) | 11988 (5489) | 13864 (5347) | 111618 (15430) | 146988 (23297) | 99638 (34465) | 113399 (15862) | 114873 (15953) | 255226 (33014) | 34083 (5286) | 34934 (5455) | 13864 (5347) | 35031 (5530) | 35236 (5555) | 21618 (35031) | 35105 (3550) | 35126 (3550) | 34949 (35031) | 35105 (3550) |
| Unique reflections | | | | | | | | | | | | | | | | | | | | | |
| Multiplicity | 6.3 (6.2) | 2.9 (2.6) | 3.2 (2.8) | 3.8 (3.7) | 2.9 (2.6) | 3.2 (2.9) | 3.3 (2.9) | 6.9 (8.8) | 7.3 (6.3) | 7.3 (6.5) | 6.8 (6.0) | 8.9 (6.45) | 35031 (5530) | 34934 (5455) | 21618 (35031) | 35236 (5555) | 34934 (5455) | 35031 (5530) | 34934 (5455) | 35236 (5555) |
| Completeness (%) | 99.6 (99.2) | 97.0 (92.8) | 98.3 (94.4) | 93.3 (99.0) | 98.1 (94.1) | 98.6 (95.5) | 98.8 (95.5) | 99.9 (99.7) | 99.9 (99.5) | 100 (99.9) | 99.8 (99.1) | 99.9 (99.6) | 100 (99.9) | 100 (99.9) | 99.4 (98.8) | 99.4 (98.9) |
| Rmeas (%) | 3.0 (10.0) | 4.3 (25.1) | 4.0 (21.0) | 6.1 (41.1) | 3.9 (24.6) | 4.3 (22.9) | 4.3 (23.0) | 4.4 (22.8) | 4.9 (13.7) | 5.6 (13.9) | 4.4 (13.4) | 4.4 (13.4) | 4.9 (13.5) | 3.9 (14.8) | 6.0 (33.3) | 3.5 (42.1) | 3.0 (10.0) | 6.1 (41.1) | 3.9 (24.6) | 4.3 (22.9) |
| ∆F/σ(ΔF) at 2.9 Å | 1.165 (0.860) | 1.185 (0.867) | 1.162 (0.867) | 1.185 (0.867) | 1.162 (0.867) | 1.185 (0.867) | 1.162 (0.867) | 1.185 (0.867) | 1.162 (0.867) | 1.185 (0.867) | 1.162 (0.867) | 1.185 (0.867) | 1.162 (0.867) | 1.185 (0.867) | 1.162 (0.867) | 1.185 (0.867) | 1.162 (0.867) | 1.185 (0.867) | 1.162 (0.867) | 1.185 (0.867) |

* ∆F/σ(ΔF) is the average anomalous signal from data truncated to dmin = 2.9 Å.
well-diffracting crystal of the double stranded RNA binding domain (dsRBD) of human dihydrouridine synthase h dus2 were used for successful sulfur-SAD phasing. This average case (tetragonal space group, diffraction to 1.8 Å, average number of sulfur atoms), together with the previous published literature, should encourage non-expert crystallographers to attempt sulfur-SAD phasing for de novo structure determination of proteins containing only light atoms, using only the sulfur anomalous signal.

2. Method: crystallization and data collection

The 13.7 kDa C-terminal domain (residues 345 to 457; 113 amino acids) of h dus2 dsRBD was expressed with a C-terminal 6-histidines tag. Crystals (~150 x 150 x 100 μm3) were obtained at 18 °C in hanging drops by vapor diffusion by mixing 1 μl of a solution containing 34 mg ml-1 protein in 25 mM Tris-HCl pH 8.0, 0.15 M NaCl with 1 μl of a 1 ml reservoir solution (28% PEG MMG, 2000, 0.1 sodium acetate, 0.1 M HEPES pH 7.5), as reported (Bou-Nader et al., 2015). Crystals were soaked for a few minutes in the same solution containing 25% glycerol and ~12 keV, respectively. The beamline was constructed to provide a focused X-ray beam of 125 x 65 mm2 and a photon flux of ~5.8 x 1014 and 7 x 1011 photons/s at ~7 keV and ~12 keV, respectively.

A so-called “native” data set, called D1, was collected at λ = 0.98 Å (12.7 keV) and 100 K. Then, using the same crystal, various consecutive “SAD” data sets were collected at a wavelength of 1.8 Å (6.89 keV) by rotating the crystal around the φ or ω axis at different fixed κ angles. Given the crystals dimensions and beam size, the crystal was not translated between the different data sets. Data were processed with XDS and merged with XSscale (Kabsch, 2010) using the native data set as a reference, the hkl and h-k-l reflections in an anomalous pair being treated as separate reflections during scaling and merging. The graphical user interface HKL2MAP (Pape and Schneider, 2004) was used for phasing with the SHELX programs (Usón and Sheldon, 2018). The heavy-atom substructure was solved with SHELXD (Schneider and Sheldon, 2002) using 500 trials and testing different resolution cut-offs. SHELXE (Sheldrick, 2002) was used to select the correct hand and attest for successful phasing. The substructures were refined and completed with Phaser (Adams et al., 2010) using all data. Phases were generated and density modification was performed using 5 cycles of PARROT (Cowtan, 2010) to improve phases. Buccaneer (Cowtan, 2006) (5 cycles) was used for automated model building. The final data and refinement statistics for the native and anomalous data sets were previously reported in Table S1 in Bou-Nader et al. (2015) and deposited in the PDB (code 4WFT).

3. Results and discussion

Dihydrouridine syntheses (DUS) catalyze the reduction of uridine to dihydrouridine mainly in the D-loop of tRNAs. The crystal structures of several DUS proteins have been determined: those of Thermus thermophilus DUS in complex with tRNA or a D-loop-containing small RNA substrate (Yu et al., 2011), E. coli DusC, alone (Chen et al., 2013) or in complex with tRNA (Byrne et al., 2015), E. coli DusB alone (Bou-Nader et al., 2018) and protein TM0096 from Thermotoga maritima (Park et al., 2004), which was not biochemically characterized. Compared to the bacterial enzymes, which are composed of a catalytic domain and a tRNA recognition domain, the human enzyme h dus2 possesses an additional C-terminal dsRBD. The structure of a construct of h dus2 comprising the first two N-terminal domains has been solved both using Se-SAD phasing (Whelan et al., 2015) and by molecular replacement (Bou-Nader et al., 2015; Whelan et al., 2015), the fold of this region being sufficiently conserved between species despite a low sequence identity (18%) (Griffiths et al., 2012). The dsRBD is a 65–70 amino acids domain present in proteins involved in various biological processes such as antiviral response, mRNA editing, RNA processing, RNA transport or gene silencing through RNA interference (Fierro-Monti and Mathews, 2000; Saunders and Barber, 2003; Tian et al., 2004). The dsRBDs fold is conserved despite large sequence variations (Gleghorn and Maquat, 2014); however, no crystal structure in the PDB presented sufficient sequence similarity to be used for molecular replacement. We have previously reported the structure of the dsRBD domain of h dus2 together with functional studies of the individual domains and the full-length protein (Bou-Nader et al., 2015). We detail here the data collection strategies and sulfur-SAD phasing methods that were used to solve the dsRBD structure.

The h dus2 dsRBD construct used in this work contains 3 cysteines and 2 methionines (apart from the possible N-terminal methionine) i.e. a sulfur content of 4.2%, which is the average value for eukaryotic proteins (Jauregui et al., 2000). The dsRBD crystals had tetragonal symmetry, 3 molecules in the asymmetric unit and a modest solvent content of 41.8%. Thus, the contribution of solvent flattening to resolve the phase ambiguity was not expected to be very important. The crystals diffracted well, up to 1.77 Å resolution at λ = 0.98 Å. At a wavelength of 1.8 Å (fλ = 0.7412), the anomalous signal estimated by the expected Bijvoet ratio (ΔF+/ΔF–) (Hendrickson and Teet, 1981) was 1.08%. This value is lower than the average ratio of 1.4% estimated for proteins for which the structure was solved by native SAD (Rose et al., 2015), but well above the current limit of 0.46% (Wang et al., 2006), which indicated that sulfur-SAD phasing should be feasible.

Because of the high diffracting power of the crystals, several highly redundant data sets were collected on a single crystal at the PROXIMA1 beam line at the SOLEIL synchrotron (Saint Aubin, France) using a Pilatus detector (Broennimann et al., 2006) and an oscillation angle of 0.2° per frame with an exposure time of 0.2 s. The beamline was configured to provide a focused X-ray beam of 125 x 65 mm2 and a photon flux of ~5.8 x 1014 and 7 x 1011 photons/s at ~7 keV and ~12 keV, respectively.

A so-called “native” data set, called D1, was collected at λ = 0.98 Å (12.7 keV) and 100 K. Then, using the same crystal, various consecutive “SAD” data sets were collected at a wavelength of 1.8 Å (6.89 keV) by rotating the crystal around the φ or ω axis at different fixed κ angles. Given the crystals dimensions and beam size, the crystal was not translated between the different data sets. Data were processed with XDS and merged with XSscale (Kabsch, 2010) using the native data set as a reference, the hkl and h-k-l reflections in an anomalous pair being treated as separate reflections during scaling and merging. The graphical user interface HKL2MAP (Pape and Schneider, 2004) was used for phasing with the SHELX programs (Usón and Sheldon, 2018). The heavy-atom substructure was solved with SHELXD (Schneider and Sheldon, 2002) using 500 trials and testing different resolution cut-offs. SHELXE (Sheldrick, 2002) was used to select the correct hand and attest for successful phasing. The substructures were refined and completed with Phaser (Adams et al., 2010) using all data. Phases were generated and density modification was performed using 5 cycles of PARROT (Cowtan, 2010) to improve phases. Buccaneer (Cowtan, 2006) (5 cycles) was used for automated model building. The final data and refinement statistics for the native and anomalous data sets were previously reported in Table S1 in Bou-Nader et al. (2015) and deposited in the PDB (code 4WFT).

| Crystal | D1 | D5 | D6 | D7 |
|---------|----|----|----|----|
| Resolution (Å) | 50–2.2 | 50–2.2 | 50–2.2 | 50–2.2 |
| Multiplicity | 9.5 (7.7) | 9.3 (7.6) | 21.3 (18.5) | 34.0 (29.3) |
| Completeness (%) | 100 (100) | 99.9 (99.5) | 100 (99.1) | 99.9 (99.0) |
| Rmerge | 5.2 (23.9) | 6.5 (21.5) | 4.3 (23.8) | 5.1 (14.3) |
| I/σ(I) | 32.5 (10.8) | 25.9 (9.9) | 61.4 (12.9) | 62.7 (24.2) |
| ΔF+/ΔF– | 1.51 (1.05) | 2.35 (1.43) | 1.40 (0.87) | 1.50 (1.01) |
| ΔF+/ΔF– 50–2.9 Å (2.98–2.9 Å) | 2.04 (1.7) | 2.85 (2.85) | 1.95 (1.24) | 1.92 (1.44) |

| Anomalous Corr (%) | 43 (27) | 73 (48) | 28 (10) | 44 (18) |
| Anomalous Corr (%) | 66 (54) | 85 (80) | 57 (33) | 71 (52) |
| Best CC1/2, CC1/2 soak (shelxd) | 22.2/11.7 | 15.7/6.7 | 35.9/20.1 | 42.2/24.0 |

| pseudoflfe CC original/inverted (%) (shelxd) | 49.2/48.8 | 43.2/43.8 | 58.7/70.5 | 59.7/69.7 |

* Data were truncated to 2.9 Å (D1, D5) or 2.2 Å (D6, D7).
**SHELXD** was used to solve the substructure using data truncated to 2.9 Å (crystals D1 and D5) or 2.2 Å (crystals D6 and D7), with the resolution cut-off corresponding to an anomalous signal $<\Delta F>/<\sigma(\Delta F)>$ higher than 1.0 (Fig. 1A). 14 S atoms out of 15 were successfully located with occupation factors greater than 0.5, for the D6 and D7 crystals. Success of substructure determination was demonstrated by the histograms showing 230 and 360 correct solutions with $CC_{all} = 36\%$ or 42\% ($CC_{all}$ is the correlation coefficient between all normalized structure factor differences in the measured data and those calculated from a given substructure solution), for D6 and D7, respectively, out of 500 trials (Fig. 1C). A clean separation of correct solutions from non-correct solutions was observed. Moreover, for both crystals D6 and D7, the correct sulfur positions were found with data truncated to any range between 2.2 and 3.5 Å resolution (data not shown). The D6 and D7 data sets led also to successful phasing, as shown by the contrast between the two enantiomers during **SHELXE** density modification (Fig. 1D), which indicates that the inverted heavy-atom substructure is the correct one. The contrast between the two possible solutions was similar for the D7 and D6 data (Fig. 1D). Both data sets allowed straightforward automated model building at 2.2 Å (Fig. 2A). For the D7 data set, 75\% of the residues were built by **SHELXE**, with the polypeptide chain being almost completely correctly traced.

In contrast, for the D1 and D5 data, only 4 or 8 sulfur sites, respectively, had an occupancy higher than 0.5 (Fig. 1B) and the histograms show that no successful solution with increased $CC_{all}$ value emerges from non-solutions (Fig. 1C), indicating random substructure positioning. Unsuccessful structure solving was confirmed by the absence of contrast between the two possible enantiomers when phases were calculated in **SHELXE** (Fig. 1D).

Interestingly, although the overall anomalous signals $<\Delta F>/<\sigma(\Delta F)>$...
<σ(ΔF)> calculated from D6 and D7 data were lower compared to those from D1 and D5 data sets (Table 3), the latter were not suitable for phasing. The higher signal-to-ratio values of D6 and D7 data sets is also noticeable, indicating the better accuracy of the estimated intensities and anomalous signal values. The substructure could be determined using the data up to 2.2 Å resolution, where ΔF/σ(ΔF) in the last resolution shell was 0.9–1.0 (with ΔF/σ(ΔF) of 1.4–1.5 for all data), indicating that the “rule” ΔF/σ(ΔF) >1.5 in all resolution ranges was not a necessary requirement for sulfur SAD substructure location, in contrast to other cases (Cianci et al., 2008).

The substructure determined from D6 and D7 data sets was refined and completed (one more sulfur site found) with PHASER, which also produced electron density maps that were subsequently modified by solvent flattening with PARROT. An easily interpretable electron-density map was obtained (Fig. 2B). The structure was successfully traced by BUCCANEER with, for D7, correct building during the first cycle of 257 residues (91% of the final structure) in 4 chains, the longest having 91 residues. After 5 cycles, the structure was 99% complete and refined to Rwork and Rfree values of 24.6% and 27.8%, respectively. Further building was done manually with COOT (Emsley et al., 2010) and refinement with REFMACS (Murshudov et al., 2011). The statistics for the refined crystal structure and structural analysis of the dsRBD have been previously been reported (Bou-Nader et al., 2015).

4. Conclusion

Proteins contain cysteine and methionine residues, whose sulfur atoms can be used for native phasing in the absence of exogenous anomalous scatterers. Yet, radiation damage in protein crystals is primarily located at the sulfur sites, and both specific and non-specific radiation damage degrade the anomalous signal. To get the best accurate measurements of the weak anomalous signal of sulfur in the hDus2 dsRBD crystals, variations in the incident beam intensity were studied. SAD-data sets were collected up to 2.2 Å resolution on a single well-diffracting crystal, which allowed us to get rid of crystal variations (in unit cells, overall diffraction, anomalous correlation coefficient) that occur when scaling multiple crystals. Highly redundant data sets (multiplicity of 3–9 per orientation) were obtained by rotating the crystal around the ϕ or ω axis with fixed k angle. An appropriate combination of high multiplicity and weak radiation damage was reached by using a highly attenuated beam (20% or 5% transmission compared to the data set collected at higher energy), ensuring a high signal-to-noise ratio and an accurate estimation of the anomalous signal. We consider that there is no unambiguous indicator, which is capable of predicting the A0(15ω)/2 in all resolution ranges was not a necessary requirement for sulfur SAD phasing and refinement of flavivirus NS1. Acta Crystallogr D Biol Crystallogr 70, 2719–2729.

Acknowledgements

We acknowledge the use of beamline PROXIMA 1 at the SOLEIL synchrotron, Saint Aubin, France and thank Pierre Legrand and Andrew Thomson from PROXIMA 1 for their help in data analysis and structure determination.

References

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapriel, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R.J., Richardson, D.C., Richardson, J.S., Terwilliger, T.C., Zwart, P.H., 2010. PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr D Biol Crystallogr 66, 213–221.

Akey, D.L., Brown, W.C., Konorswski, J.R., Ogata, C.M., Smith, J.L., 2014. Use of massively multiple moved data for low-resolution S-SAD phasing and refinement of flavivirus NS1. Acta Crystallogr D Biol Crystallogr 70, 2719–2729.

Axmann, G.M., Wang, M., Diederichs, K., 2020. Making a difference in multi-data-set crystallography: simple and deterministic data-scaling/selection methods. Acta Crystallogr D Biol Crystallogr 76, 636–652.

Banerjee, S., Cossens, N.P., Gallat, F., Sathyanarayanan, N., Skantham, J., Yagi, K.J., Gray, J.S., Tobe, S.S., Stay, B., Chavas, L.M., Ramazovany, S., 2016. Structure of a heterogeneous, glycosylated, lipid-bound, in vivo-grown protein crystal at atomic resolution from the viviparous cockroach Dipsopera puncuta. IUCrD 3, 283–293.

Banumathi, S., Dauter, M., Dauter, Z., 2003. Phasing at high resolution using Ta6Br12 cluster. Acta Crystallogr D Biol Crystallogr 59, 492–498.

Basi, S., Finke, A., Vera, L., Wang, M., Olieric, V., 2010a. Making routine native SAD a reality: lessons from beamline X06DA at the Swiss Light Source. Acta Crystallogr D Biol Crystallogr 75, 262–271.

Basi, S., Olieric, V., Leonardi, F., Matsukus, N., Kawano, Y., Takashi, T., Huang, C.-Y., Yamada, Y., Vera, L., Olieric, N., Basquin, J., Wojdyla, J.A., Bunk, O., Diederichs, K., Yamamotoc, M., Wang, M., 2019b. Long-wavelength native-SAD phasing: opportunities and challenges. IUCrD 6, 373–386.

Bou-Nader, C., Peccoeur, I., Breggeon, D., Kamah, A., Guerinine, V., Golinnelli-Pimpaneau, B., Guimaraes, B.G., Fontecave, M., Hamadane, D., 2015. An extended dsRBD is required for post-transcriptional modification in human tRNAs. Nucleic Acids Res. 43, 9446–9456.

Bou-Nader, C., Montemont, H., Guerinine, V., Jean-Jean, O., Breggeon, D., Hamadane, D., 2018. Unveiling structural and functional divergences of bacterial RNA dihydrodiximidine synthases: perspectives on the evolution scenario. Nucleic Acids Res. 46, 1386–1394.

Brouninnikov, C., Ikkenberry, E.F., Heinrich, B., Horisberger, R., Huebel, G., Pohl, E., Schmitt, B., Schulze-Briese, C., Suzuki, M., Tomizaki, T., Toyokawa, H., Wagner, A., 2006. The PILATUS 1M detector. J. Synchrotron Radiat. 13, 120–130.

Bunemathi, S., Dauter, M., Dauter, Z., 2003. Phasing at high resolution using Ta6Br12 cluster. Acta Crystallogr D Biol Crystallogr 59, 492–498.

Cianci, M., Nanao, M., Schneider, T.R., 2019. Long-wavelength Mesh phasing. Acta Crystallogr D Biol Crystallogr 75, 192–203.

Cianci, M., Groves, M.R., Barford, D., Schneider, T.R., 2016. Data collection with a tailored X-ray beam size at 2.69 Å wavelength (4.6 keV): sulfur SAD phasing of macromolecular structure. Acta Crystallogr D Biol Crystallogr 72, 403–412.

Cianci, M., Bourenkov, G., Fomilpil, G., Karpic, I., Kallio, J., Bento, I., Rosle, M., Ciprani, F., Fiedler, S., Schneider, T.R., 2017. P13, the EMBL macromolecular crystallography beamline at the low-emittance PETRA III ring for high- and low-energy phasing with variable beam focusing. J. Synchrotron Radiat. 24, 332–332.

Cianci, M., Nanao, M., Schneider, T.R., 2019. Long-wavelength Mesh&Collect native SAD phasing from microcrystals. Acta Crystallogr D Biol Crystallogr 75, 192–199.

Cowan, K., 2006. The Buccaneer software for automated model building. 1. Tracing protein chains. Acta Crystallogr D Biol Crystallogr 62, 1002–1011.

Cowan, K., 2010. Recent developments in classical density modification. Acta Crystallogr D Biol Crystallogr 66, 470–478.

Dauter, Z., Dauter, M., de la Fortelle, E., Bricegue, G., Sheldrick, G.M., 1999. Can anomalous signal of sulfur become a tool for solving protein crystal structures? J. Mol. Biol. 289, 83–92.

De Sanctis, D., Oscarsson, M., Popov, A., Svensson, O., Leonard, G., 2016. Facilitating best practices in collecting anomalous scattering data for de novo structure solution at the ESRF Structural Biology Beamlines. Acta Crystallogr D Biol Crystallogr 72, 413–420.

Douch, J., Hough, M.A., Hannain, S.S., Strange, R.W., 2012. Challenges of sulfur SAD phasing as a routine method in macromolecular crystallography. J. Synchrotron Radiat. 19, 19–29.

El Omari, K., Iourin, O., Kadlec, J., Fearn, R., Hall, D.R., Harlos, K., Grimes, J.M., Stuart, D.I., 2014. Pushing the limits of sulfur SAD phasing: de novo structure solution of the N-terminal domain of the ectodomain of HCV E1. Acta Crystallogr D Biol Crystallogr 70, 2197–2203.

Emsley, P., Lohkamp, B., Scott, W.G., Cowtan, K., 2010. Features and development of coot. Acta Crystallogr D Biol Crystallogr 66, 486–500.

Beatriz Gomes Guimarães: Conceptualization, Methodology, Software, Formal analysis, Data curation, Writing – review & editing.

Beatrice Golinnelli-Pimpaneau: Investigation, Writing – original draft, Writing – review & editing, Visualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.
Fierro-Monti, I., Mathews, M.B., 2000. Proteins binding to duplexed RNA: one motif, multiple functions. Trends Biochem. Sci. 25, 241–246.

Foos, N., Cianci, M., Nanao, M.H., 2019. Choosing your (frizzled) mates wisely: grouping data sets to improve anomalous signal. Acta Crystallogr D Struct Biol 75, 200–210.

Gleghorn, M.L., Maquat, L.E., 2014. Black sheep that don’t leave the double-stranded RNA-binding domain fold. Trends Biochem. Sci. 39, 328–340.

Gorzel, M., Boggild, A., Ustragr, J.J., Weiss, M.S., Müller, U., Nissen, P., Boesen, T., 2015. Against the odds? De novo structure determination of a pilin with two cysteine residues by sulfur SAD. Acta Crystallogr D Biol Crystallogr 71, 1095–1101.

Goulet, A., Vestergaard, G., Felisberto-Rodrigues, C., Campanacci, V., Garrett, R.A., Legrand, P., Guimaraes, B.G., Thelandersson, K.I., 2010. Getting the best out of low-wavelength X-rays: de novo chlorine/sulfur SAD phasing of a structural protein from ATV. Acta Crystallogr D Biol Crystallogr 66, 304–308.

Griffiths, S., Byrne, R.T., Antson, A.A., Whelan, F., 2012. Crystalization and preliminary X-ray crystallographic analysis of the catalytic domain of human dihydrouridine synthase. Acta Crystallogr Sect F Struct Biol Cryst Commun 68, 332–337.

Guo, G., Zhu, P., Fuchs, M.R., Shi, W., Andi, B., Gao, Y., Hendrickson, W.A., McSweeney, S., Liu, Q. 2019. Synchronisation of native-SAD phasing at a low energy. IUCrJ 7, 532–542.

Hegde, R.P., Fedorov, A.A., Sauer, J.M., Burley, S.K., Almob, S.C., Ramagopal, U.A., Klinke, S., Foos, N., Rinaldi, J.J., Paris, G., Goldbaum, F.A., Legrand, P., Guimaraes, B.G., Mancia, F., Corti, G., Mancia, F., 2010. The hidden treasure in your data: phasing with unexpected weak anomalous scatterers from routine data sets. Acta Crystallogr Sect F Struct Biol Cryst Commun 73, 184–195.

Hendrickson, W.A., 1991. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. Science 254, 51–58.

Hendrickson, W.A., 2014. Anomalous diffraction in crystallographic phase evaluation. Q. Rev. Biophys. 47, 49–93.

Hendrickson, W.A., Teet, M.M., 1981. Structure of the hydrophobic protein crambin determined directly from the anomalous scattering of sulfur. Nature 290, 107–113.

Jaureggi, R., Bolivar, F., Merino, E., 2000. Relationship between whole proteome aminocid composition and static DNA curvature. Microb. Comp. Genom. 5, 7–15.

Kabsch, W., 2010. Xds. Acta Crystallogr D Biol Crystallogr 66, 125–132.

Klinke, S., Foos, N., Rinaldi, J.J., Paris, G., Goldbaum, F.A., Legrand, P., Guimaraes, B.G., Thompson, A., 2015. S-SAD phasing of monoclinic histidine kinase from Brucella abortus combining data from multiple crystals and orientations: an example of datacollection strategy and a posteriori analysis of different data combinations. Acta Crystallogr D Biol Crystallogr 71, 1433–1445.

Knaus, S.H.W., Bianchi, S., Steinmetz, M.O., 2014. Combinatorial use of disulfide bridges and native-sulfur-SAD phasing for rapid structure determination of coiled-coils. Biocrys. Rep. 38, BSR20181073.

Liebscher, D., Yamada, Y., Matsuoki, N., Senda, M., Senda, T., 2016. On the influence of crystal size and wavelength on native SAD phasing. Acta Crystallogr D Struct Biol 72, 728–741.

Liu, Q., Hendrickson, W.A., 2015. Crystallographic phasing from weak anomalous signals. Acta Crystallogr D Biol Crystallogr 71, 1095–1101.

McSweeney, S., Liu, Q., 2019. Synchrotron microcrystal native-SAD phasing at a low energy. IUCrJ 7, 532–542.

Mueller-Dieckmann, C., Panjikar, S., Schmidt, A., Mueller, S., Kuper, J., Geerlof, A., Olieric, V., Weinert, T., Finke, A.D., Senda, M., Senda, T., 2000. The dsRNA binding protein family: critical roles, diverse cellular functions. Faseb. J. 17, 961–983.

Olieric, V., Schneider, T.R., Sheldrick, G.M. 2002. Substructure solution with SHELXD. Acta Crystallogr D Biol Crystallogr 58, 1772–1779.

Sheldrick, G.M., 2002. Macromolecular phasing with SHELXLE. Z. Kristallogr. 217, 644–650.

Sheldrick, G.M., 2005. From bacterial to human dihydrouridine synthase: automated structure determination. Acta Crystallogr D Biol Crystallogr 71, 1564–1572.

Sarma, G.N., Karplus, P.A., 2006. In-house sulfur SAD phasing: a case study of the effects of data quality and resolution cutoffs. Acta Crystallogr D Biol Crystallogr 62, 707–716.

Saunders, L.R., Barber, G.N., 2003. The antiRNA binding protein family: critical roles, diverse cellular functions. Faseb. J. 17, 961–983.

Thompson, A., 2015. S-SAD phasing of monoclinic histidine kinase from Chlamydia pneumoniae. J. Struct. Biol. 189, 1443–1451.

Usov, I., Leonarski, F., Geng, T., Rappas, M., Dor, A., Knopp, G., Cirelli, C., Martiel, I., Casadei, C., Weinert, T., Olieric, V., Waltersperger, S., Panepucci, E., Chen, L., Zhang, H., Zhou, D., Park, F., Gajiwala, K., Noland, B., Wu, L., He, D., Molinari, J., Loomis, K., Pagirigan, B., Kearsin, P., Christoper, J., Peat, T., Badger, J., Hendle, J., Lin, J., Buchanan, S., 2015. The 1.59 Å resolution crystal structure of TM6006, a flavin mononucleotide binding protein from Thermotoga maritima. Proteins 55, 772–774.

Weinert, T., Finke, A.D., Senda, M., Senda, T., 2000. The dsRNA binding protein family: critical roles, diverse cellular functions. Faseb. J. 17, 961–983.