research papers

De novo sulfur SAD phasing of the lysosomal 66.3 kDa protein from mouse

Kristina Lakomek,[a] Achim Dickmanns,[a] Uwe Mueller,[b] Katrin Kollmann,[c] Florian Deuschl,[c] Annette Berndt,[a] Torben Lübke[c] and Ralf Ficner[a]*

[a]Department of Molecular Structural Biology, Institute of Microbiology and Genetics, Georg-August University Göttingen, Justus-von-Liebig-Weg 11, D-37077 Göttingen, Germany, [b]BESSY GmbH, Macromolecular Crystallography Group, Albert-Einstein-Strasse 15, D-12489 Berlin, Germany, and [c]Center of Biochemistry and Molecular Cell Biology, Department of Biochemistry II, Georg-August University Göttingen, Heinrich-Dueker-Weg 12, D-37073 Göttingen, Germany

Correspondence e-mail: rficner@gwdg.de

The 66.3 kDa protein from mouse is a soluble protein of the lysosomal matrix. It is synthesized as a glycosylated 75 kDa preproprotein which is further processed into 28 and 40 kDa fragments. Despite bioinformatics approaches and molecular characterization of the 66.3 kDa protein, the mode of its maturation as well as its physiological function remained unknown. Therefore, it was decided to tackle this question by means of X-ray crystallography. After expression in a human fibrosarcoma cell line, the C-terminally His-tagged single-chain 66.3 kDa variant and the double-chain form consisting of a 28 kDa fragment and a 40 kDa fragment were purified to homogeneity but could not be separated during the purification procedure. This mixture was therefore used for crystallization. Single crystals were obtained and the structure of the 66.3 kDa protein was solved by means of sulfur SAD phasing using data collected at a wavelength of 1.9 Å on the BESSY beamline BL14.2 of Freie Universität Berlin. Based on the anomalous signal, a 22-atom substructure comprising 21 intrinsic S atoms and one Xe atom with very low occupancy was found and refined at a resolution of 2.4 Å using the programs *SHELXC/D* and *SHARP*. Density modification using *SOLOMON* and *DM* resulted in a high-quality electron-density map, enabling automatic model building with *ARP/wARP*. The initial model contained 85% of the amino-acid residues expected to be present in the asymmetric unit of the crystal. Subsequently, the model was completed and refined to an R_{free} factor of 19.8%. The contribution of the single Xe atom to the anomalous signal was analyzed in comparison to that of the S atoms and was found to be negligible. This work should encourage the use of the weak anomalous scattering of intrinsic S atoms in SAD phasing, especially for proteins, which require both expensive and time-consuming expression and purification procedures, preventing extensive screening of heavy-atom crystal soaks.

1. Introduction

Lysosomes are membrane-bordered organelles in eukaryotic cells that contain a set of about 60 acid hydrolases and associated proteins that are responsible for the digestion of various macromolecules and even whole organelles derived from various sources by endocytosis, autophagy and other trafficking pathways (Sleat *et al.*, 2008; reviewed by Sleat *et al.*, 2007; Lübke *et al.*, 2008). Typically, the lack of function of lysosomal proteins causes severe pathogenic phenotypes collectively referred to as 'lysosomal storage diseases' which are associated with the accumulation of undigested molecules in the lysosomal compartment (reviewed by Scriver *et al.*,

2001). In addition to their degradative function, lysosomal enzymes are involved in many pathophysiological processes such as cancer (Fehrenbacher & Jaattela, 2005), tumour metastasis and propagation (Garcia *et al.*, 1996; Kos & Lah, 1998), neuroprotection (Cravatt *et al.*, 2001) and Alzheimer's disease (Nixon & Cataldo, 2006), as well as in the modulation of hormones and bioactive lipids implicated, for example, in tissue homeostasis, inflammation (Capasso *et al.*, 2001; Izzo *et al.*, 2001; Feulner *et al.*, 2004; reviewed by Hansen *et al.*, 2000) and anorexia (Saftig *et al.*, 1995).

Most lysosomal enzymes are directed to lysosomes by a mannose 6-phosphate (M6P) residue that is attached to their *N*-glycans post-translationally, recognized by M6P receptors (MPRs) at the trans-Golgi network and subsequently targeted to the lysosomes. Based on the composition of the oligosaccharide, the *N*-glycans can be subdivided in three types, namely the 'high-mannose', 'hybrid' and 'complex' types, all of which exhibit a common pentasaccharide core that consists of three different mannose residues and two *N*-acetylglucosamine (NAG) moieties. In order to reveal novel lysosomal proteins and thus novel lysosomal functions, several subproteomic studies dealing with soluble lysosomal proteins have recently been carried out (Sleat, Wang *et al.*, 2006; Sleat, Zheng *et al.*; 2006; Kollmann *et al.*, 2005; Sleat *et al.*, 2008; reviewed by Sleat *et al.*, 2007). Among others, the 66.3 kDa protein has been identified as a putative soluble lysosomal protein in several lysosomal proteomic studies from mouse, rat and human (Lübke *et al.*, 2008). Subsequently, the murine 66.3 kDa protein and its human orthologue p76 were characterized in more detail regarding their lysosomal localization, processing and glycosylation (Deuschl *et al.*, 2006; Jensen *et al.*, 2007). A C-terminally RGS-His$_6$-tagged derivative of the mouse 66.3 kDa protein was stably expressed at high levels in the human fibrosarcoma cell line HT1080, which secreted the protein into the medium. It was then purified by a combination of affinity and ion-exchange chromatography (Deuschl *et al.*, 2006). The 66.3 kDa protein is synthesized at the rough ER as a glycosylated precursor with an apparent molecular weight of 75 kDa and is further processed into a 28 kDa N-terminal fragment and a 40 kDa C-terminal fragment (Deuschl *et al.*, 2006). Such an extensive endosomal/lysosomal maturation by limited proteolysis is a common step towards the final activation of lysosomal hydrolases (Hasilik, 1992). However, despite bioinformatics studies and molecular characterization, the physiological function of the 66.3 kDa protein so far remains unknown. In order to obtain initial insights into its activity, we set out to determine the crystal structure of the 66.3 kDa protein from mouse. Since no structure of a protein with sufficient similarity at the amino-acid sequence level is available, the phases were determined experimentally.

Currently, several high-brilliance third-generation synchrotron beamlines are available worldwide that support the required energy range for long-wavelength phasing applications (Djinovic Carugo *et al.*, 2005). Thus, the use of SAD phasing using only weak anomalous scatterers has increased (Ramagopal *et al.*, 2003) since the first protein structure was determined by sulfur SAD in 1981 (Hendrickson & Teeter,

1981). Most sulfur SAD data have been collected on tunable synchroton beamlines in order to make use of the appreciably larger f'' of sulfur at longer wavelengths in the range between 1.7 and 2.5 Å (Brown *et al.*, 2002; Gordon *et al.*, 2001; Liu *et al.*, 2000; Ramagopal *et al.*, 2003; Weiss, 2001; Weiss *et al.*, 2004), although Cu $K\alpha$ is also feasible for sulfur SAD phasing (Dauter *et al.*, 1999; Sekar *et al.*, 2004; Roeser *et al.*, 2005). However, successful phasing often relies on the additional presence of several different weak anomalous scatterers such as Ca^{2+} and Cl^- ions (Dauter *et al.*, 1999; Yang & Pflugrath, 2001; Debreczeni, Bunkóczi, Ma *et al.*, 2003). Only rarely are intrinsic S atoms the sole source of the anomalous signal (Debreczeni, Bunkóczi, Girmann *et al.*, 2003; Debreczeni, Girmann *et al.*, 2003; Yang & Pflugrath, 2001). When further heavy-atom sites were determined in addition to the sulfur sites, the former refined to significantly higher occupancies; for example, occupancies of 1.0 and 0.99 for two calcium ions compared with 0.43 for the highest occupancy for a sulfur site in the formylglycine-generating enzyme (Roeser *et al.*, 2005). To our knowledge, the largest structure solved by true sulfur SAD phasing to date has been that of TT0570 from *Thermus thermophilus*, with a molecular weight of 69 kDa. However, this protein was crystallized in space group $P2_12_12$ with two molecules in the asymmetric unit (Watanabe *et al.*, 2005). Like TT0570, the majority of crystals suitable for sulfur SAD phasing experiments belong to high-symmetry space groups. In contrast, only a limited number of successful sulfur SAD phasings of proteins that crystallize in low-symmetry space groups have been described for monoclinic crystals; for example, xylanase from *Thermoascus aurantiacus* (molecular weight 33 kDa; Ramagopal *et al.*, 2003) and the hypothetical protein AF1432 (molecular weight 20 kDa; PDB code 1ynb; A. Dong, T. Skarina, A. Savchenko, E. F. Pai & A. Edwards, unpublished work) crystallized in space groups $P2_1$ and $C2$, respectively. In this work, we demonstrate on the basis of the *de novo* phasing of the 66.3 kDa protein that sulfur SAD phasing is also feasible for a larger protein that crystallizes in a low-symmetry space group such as $C2$.

2. Materials and methods

2.1. Protein purification and crystallization

The 66.3 kDa protein was purified as described by Deuschl *et al.* (2006). For crystallization, size-exclusion chromatography was added as a final purification step (10 m*M* Tris–HCl pH 8.0; Superdex 200 HR 10/300, GE Healthcare). The molecular-weight markers were purchased from Fermentas (SM0431; Fermentas, St Leon-Rot, Germany). The purified protein, which consisted of a mixture of the 66.3 kDa full-length protein and the 40 kDa fragment as well as the 28 kDa fragment, was crystallized using the sitting-drop vapour-diffusion method at 293 K. The 66.3 kDa protein was crystallized in a drop composed of 1.6 µl protein solution with a concentration of 23 mg ml^{-1} and 2.0 µl reservoir solution containing 12%(*w/v*) PEG 4000, 200 m*M* ammonium acetate

and 100 m*M* sodium acetate/acetic acid pH 4.6. The final pH of the reservoir solution was determined to be 5.0.

2.2. Data collection and processing

For data collection, the crystal was transferred into a cryoprotecting solution that consisted of 16%(*w*/*v*) PEG 4000, 130 m*M* ammonium acetate, 60 m*M* sodium acetate/acetic acid pH 4.6 and 10%(*v*/*v*) glycerol. Subsequently, the 66.3 kDa protein crystal was derivatized with xenon in a gas chamber for 4 min under a gas pressure of 2.8 MPa and then directly flash-cooled in liquid nitrogen before mounting in the cryostream. The data set was collected on BESSY beamline BL14.2, which was equipped with an SX165 detector (Rayonics LLC, Illinois, USA). The crystal was mounted at a distance of 50 mm and the beam stop was adjusted to a distance of 14 mm from the detector. 1120 images were recorded at a wavelength of 1.900 Å from a single crystal in 1.0° oscillations with 3.2 s exposure time at 100 K using 0.03 mm aluminium foil between the X-ray beam and the crystal in order to reduce the beam intensity and radiation damage. The images were processed with *DENZO* and *SCALEPACK* as implemented in *HKL*-2000 (HKL Research, Inc., Charlottesville, Virginia, USA). The scaled data were analysed with *XPREP* (Bruker AXS Inc., Madison, Wisconsin, USA). R_{anom} and $R_{p.i.m.}$ were determined with *SCALA* (Collaborative Computational Project, Number 4, 1994) based on the data scaled using *SCALEPACK* without merging the original indices.

2.3. Structure determination and model building

The anomalous completeness listed in Table 1 corresponds to the output from *SHELXC* (Sheldrick, 2008). *autoSHARP* (de La Fortelle & Bricogne, 1997) served as the platform for structure determination using data in the resolution range 32.11–2.40 Å. *SHELXC/D* (Sheldrick, 2008) and *SHARP* (de La Fortelle & Bricogne, 1997) were used for substructure determination and for positional refinement and phase calculations, respectively. The phases obtained were further improved by solvent flattening and histogram matching using *SOLOMON* and *DM* (Collaborative Computational Project, Number 4, 1994) as implemented in *SHARP*. A free-atom model was built into the electron-density map by *ARP/wARP* (Perrakis *et al.*, 1999) as implemented in *SHARP* using data to a resolution of 2.40 Å and a solvent content of 56.4%. Subsequent auto-tracing of the amino-acid chain using the *warpNtrace* procedure of *ARP/wARP* resulted in an initial model of high quality containing 475 amino acids (of 559 expected amino acids) in 20 chains with a connectivity index of 0.92. Subsequently, the model was completed manually with *Coot* (Emsley & Cowtan, 2004) and refined using *REFMAC*5 from the *CCP*4 suite (Murshudov *et al.*, 1997; Collaborative Computational Project, Number 4, 1994) to *R* factors of R_{work} = 15.6% and R_{free} = 19.8% (see Table 1), with r.m.s. deviations of 0.013 Å and 1.44° for bond lengths and angles, respectively. Stereochemical analysis of the refined structure was performed with *PROCHECK* (Laskowski *et al.*, 1993). 520

Table 1
Crystallographic data and refinement statistics.

Values in square brackets and parentheses are for the lowest and highest resolution shells, respectively.

Wavelength (Å)	1.900
X-ray source/detector	BL14.2/SX165
Resolution range (Å)	[50–5.17] 50.0–2.40 (2.49–2.40)
Space group	*C*2
No. of molecules in ASU	1
Unit-cell parameters (Å, °)	a = 148.80, b = 89.67, c = 64.95, α = 90.0, β = 98.7, γ = 90.0
Measured reflections	751575
Unique reflections	32026
Rejected reflections	195 (~0.01%)
Multiplicity of reflections	[23.4] 23.0 (21.2)
Completeness (%)	[99.7] 98.9 (97.5)
Anomalous completeness† (%)	[98.3] 98.7 (97.0)
Mosaicity (°)	0.76
R_{merge}‡ (%)	[3.8] 8.3 (37.3)
Average $I/\sigma(I)$	[102.9] 48.8 (10.2)
Maximum $\Delta F''/\sigma(\Delta F)$	1.81
f'/f''§ (electrons)	0.37/0.82
R_{Cullis} (anomalous)¶	0.95
Anomalous phasing power†	0.474
Expected no. of S atoms	22
Heavy-atom sites found in ASU	22 (21 S, 1 Xe)
Amino acids in ASU (expected/placed)	559/520
R_{work} (%)	15.6
R_{free} (%)	19.8

† According to *SHELXC* with resolution bins ∞–8.0 and 2.6–2.4 Å. ‡ R_{merge} = 100 × $\sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$. § Theoretical values for sulfur at a wavelength of 1.9000 Å used by *SHARP*. ¶ R_{Cullis} (acentric reflections) = $\sum E / \sum \Delta F$, where E is the residual lack-of-closure error.

of the 559 amino-acid residues (Val63–Thr238 and Cys249–Pro592), five *N*-acetylglucosamine moieties, 294 water molecules, three glycerol molecules, three polyethylene glycol molecules, one sodium ion, one acetate ion and one Xe atom with an occupancy of 10% were included in the final structure. The protein atoms exhibit average *B* factors of only 23.9 Å^2 for the amino-acid residues and of 49.2 Å^2 for the atoms of the sugar moieties attached to Asn115, Asn236, Asn441 and Asn520, while the average *B* factors of the water molecules and the other solvent molecules were 33.7 and 51.4 Å^2, respectively. Cys249 was oxidized in some molecules forming the crystal as visualized in the electron density. Therefore, this residue was modelled as a cysteine sulfonic acid with an occupancy of 0.5 for all side-chain O atoms.

The Harker section in Fig. 3 was prepared using the program *XPREP*. Figs. 2 and 4(*b*)–4(*f*) were prepared using *CCP4mg* (Potterton *et al.*, 2004), while Fig. 4(*a*) was prepared with *Coot*.

3. Results and discussion

3.1. Purification and crystallization of the 66.3 kDa protein

The glycosylated 66.3 kDa protein from mouse was produced by overexpression as a C-terminally RGS-His$_6$-tagged derivative in a human fibrosarcoma cell line (HT1080) and purified as described elsewhere (Deuschl *et al.*, 2006). Since the 66.3 kDa proprotein and the 28 and 40 kDa fragments could not be separated during the purification procedure, this mixture was used for crystallization. Initial crystals were

obtained under acidic conditions near the physiological pH of lysosomes (pH 4.6) using PEG 4000 and ammonium acetate as precipitants. Initial crystals grew in compact clusters and could not be separated (Fig. 1a). Extensive variation of the reservoir solution did not improve crystal growth, although exchanging the ammonium acetate for diammonium hydrogen citrate resulted in single crystals (Fig. 1b). However, the crystals exhibited a thin plate-like shape and were also not suitable for data collection. Only optimization of the protein concentration and variation of the protein:reservoir ratio in addition to a final size-exclusion chromatography step in the purification procedure resulted in crystals of suitable quality. The supplementary gel filtration seemed to be most important in order to obtain crystals that could be separated despite growing in clusters. Such crystals are shown in Fig. 1(c). It is most likely that this optimization is based on the removal of small amounts of aggregated protein that elute in the void volume of the column. The crystals obtained were separated from adhering precipitate in a cryoprotecting solution which contained both an increased PEG concentration and 10%(v/v) glycerol to prevent ice formation. Since the protein mixture used for crystallization contained three different polypeptide chains, the crystals were dissolved in order to gain information about their content. Owing to the sensitivity of the thin plate-like crystals towards breakage, only two washing steps in reservoir solution were performed. SDS–PAGE analysis of the dissolved crystals and staining with Coomassie Brilliant Blue R/G revealed only one prominent band corresponding to the 40 kDa fragment and an additional faint signal of about 28 kDa corresponding to the second fragment, as shown in

Fig. 1(d). However, weak staining of the 28 kDa band was also observed for the purified protein solution and the weak signal seemed to be caused by the crystal content rather than by precipitate that had remained stuck to the crystal despite washing. The single-chain form represented by the band at around 66 kDa was unambiguously not present in the crystal. The complete absence of the 66.3 kDa form was also observed after incubation for two weeks at 293 K of crystallization drops from which no crystals were obtained, suggesting that it is processed autocatalytically into the stable 28 and 40 kDa fragments over time. Both the 66.3 kDa protein and the fragments of the solution used for crystallization displayed a higher apparent molecular weight in SDS–PAGE than expected owing to the presence of five glycosylation sites which were all utilized upon expression as demonstrated by two different methods: deglycosylation by peptide:N-glycosidase F (PNGase F) treatment (Deuschl et al., 2006) and structural analysis (see below). However, the addition of PEG 4000 resulted in a lower apparent molecular weight of the proteins from the dissolved crystal, as can be seen in lane 4 of Fig. 1(d) by comparison with the solution of the purified protein in 10 mM Tris–HCl pH 8.0 (lanes 2 and 5). Therefore, this solution (lane 2) was mixed with an appropriate amount of reservoir solution (lane 3) in order to serve as a reference for the dissolved crystal.

3.2. Data collection and structure determination

After incubation in cryobuffer, the crystal containing both fragments was subjected to pressurization with xenon in a
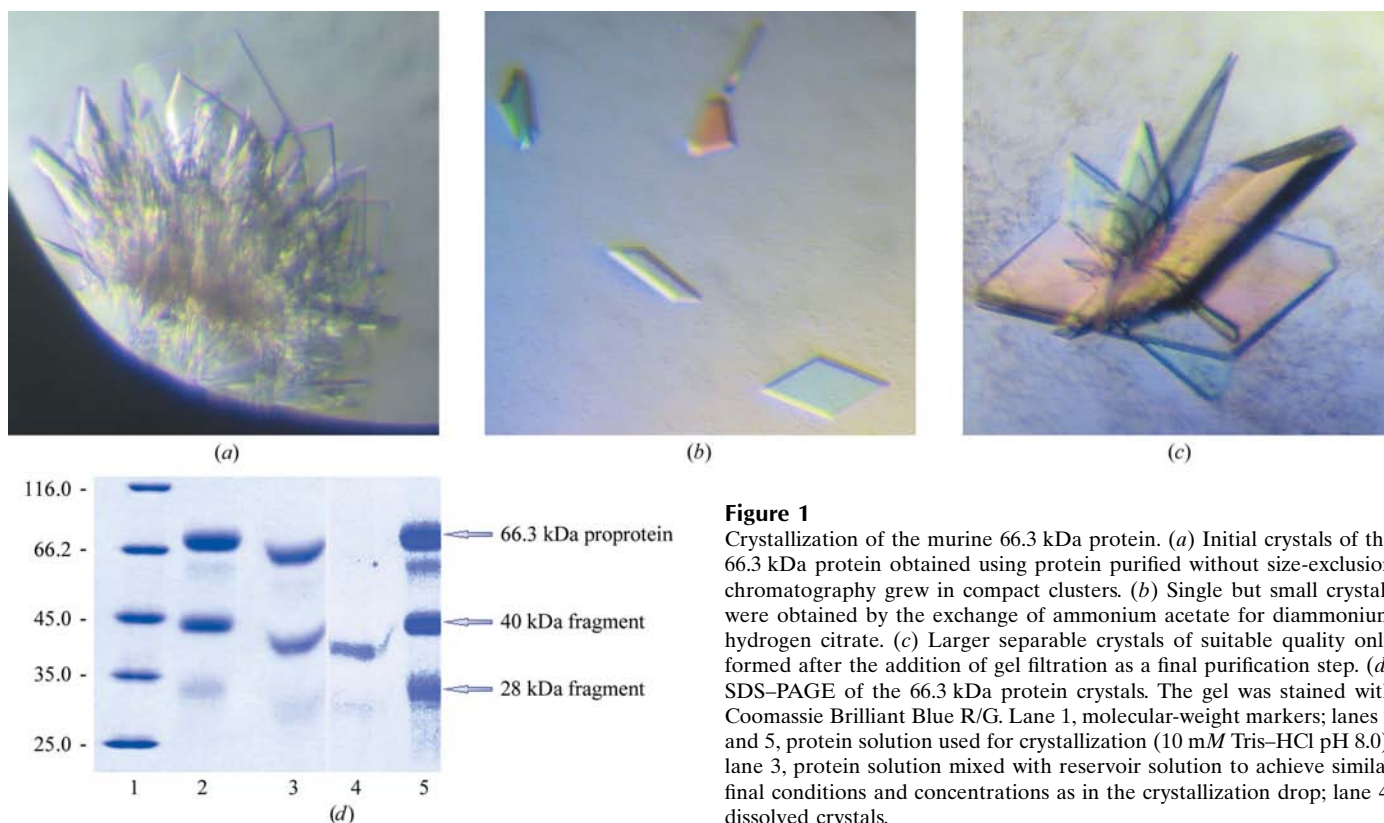

Figure 1
Crystallization of the murine 66.3 kDa protein. (a) Initial crystals of the 66.3 kDa protein obtained using protein purified without size-exclusion chromatography grew in compact clusters. (b) Single but small crystals were obtained by the exchange of ammonium acetate for diammonium hydrogen citrate. (c) Larger separable crystals of suitable quality only formed after the addition of gel filtration as a final purification step. (d) SDS–PAGE of the 66.3 kDa protein crystals. The gel was stained with Coomassie Brilliant Blue R/G. Lane 1, molecular-weight markers; lanes 2 and 5, protein solution used for crystallization (10 mM Tris–HCl pH 8.0); lane 3, protein solution mixed with reservoir solution to achieve similar final conditions and concentrations as in the crystallization drop; lane 4, dissolved crystals.

xenon gas chamber (Hampton Research, Aliso Viejo, USA) at BESSY, Berlin. An initial data set was collected on beamline BL14.2 using a wavelength of 1.700 Å in order to use the anomalous signal of Xe atoms for experimental phasing. However, no significant anomalous signal was detected by *SHELXC* (data not shown). Therefore, a second data set was collected at a wavelength of 1.900 Å using the same crystal in order to determine the phases by means of sulfur SAD (see Table 1). 1120 images of 1.0° were taken to ensure high multiplicity of the data. The crystals belonged to the mono-
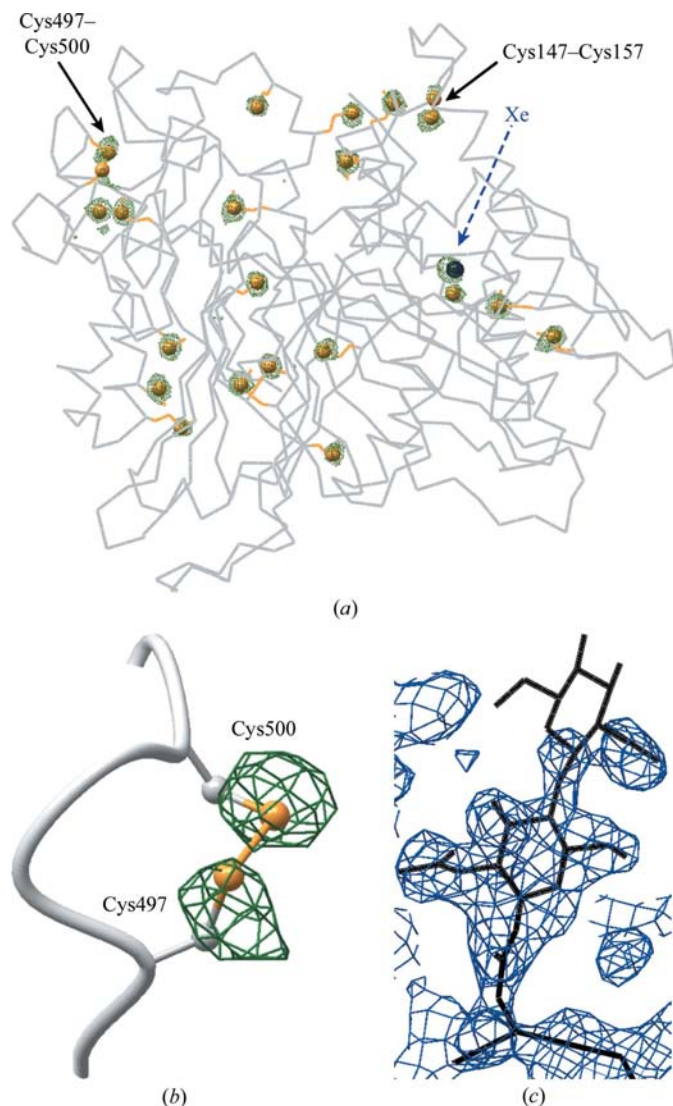

Figure 2
(*a*) Anomalous difference map contoured at 4.0σ. The map is coloured dark green. One peak represents the Xe atom, which is indicated as a blue sphere. For orientation, the C^α trace of the refined 66.3 kDa protein structure is shown with cysteine and methionine side chains in stick mode and S atoms highlighted as yellow spheres. The two disulfide bonds are marked with black arrows. (*b*) One of the two internal disulfide bridges is represented by two individual peaks in the anomalous difference map and enabled both cystine S atoms to be found. The anomalous map is contoured at the 4.0σ level. (*c*) Electron-density map after density modification. One *N*-acetylglucosamine moiety of the glycan attached to Asn115 (for example) is already clearly visible in the experimental map at a level of 1.5σ before the inclusion of any model phases and manual intervention.

clinic space group *C*2, with unit-cell parameters *a* = 148.8, *b* = 89.7, *c* = 65.0 Å, β = 98.7° and one molecule in the asymmetric unit. An average redundancy of 23.0 was obtained after 1120° of rotation, before radiation damage became a serious problem as indicated by an increase in the *R* factor. The damage was confirmed during model building, when negative peaks became visible in the $F_o - F_c$ difference map for the carboxyl groups of some glutamate and aspartate side chains. The data were processed with *HKL*-2000 and cut off at a resolution of 2.40 Å. The scaled data were analyzed using *XPREP*. As indicated by the $R_{p.i.m.}$ value of 2.4%, the data were of reasonable quality and enabled us to solve the structure by means of sulfur/Xe SAD phasing using the xenon-derivatized crystal. Data statistics are summarized in Table 1. The crystal displayed a mosaicity of 0.76° and a unit-cell volume of about 857 000 $Å^3$. The $d''/\mathrm{sig}(d'')$ value output by *SHELXC* as an estimate of the anomalous signal amounted to a maximum of 1.81 in the inner resolution shell (∞–8.0 Å) and dropped below the threshold (0.80) at 3.0 Å, indicating the presence of only moderate anomalous signal. Nevertheless, using *SHELXC/D* and *SHARP* as implemented in *auto-SHARP*, 22 heavy-atom sites could be identified in the asymmetric unit and refined to reliable occupancies (0.36–0.75; see §2). As suggested by SDS–PAGE analysis of dissolved crystals, structure determination proved that the crystals were composed of the 40 kDa and the 28 kDa fragments in a 1:1 molar ratio. Thus, all cysteine residues of the processed 66.3 kDa protein except for Cys157 and all methionines were represented by significant peaks in the anomalous difference map (six and 15 heavy-atom sites, respectively). The heavy-atom sites are shown in Fig. 2(*a*) together with the C^α ribbon as well as the cysteine and methionine side chains of the final model. Two disulfide bridges are formed in the 66.3 kDa protein, namely between Cys147 and Cys157 and between Cys497 and Cys500. Since a peak for Cys157 only appears in the anomalous map at a contour level below 3.5σ and thus the respective S atom has not been detected by *SHELXD*, the disulfide bond between Cys147 and Cys157 only corresponds to a single significant peak around Cys147 in the anomalous map and to one heavy-atom position. In contrast, the second disulfide bond (between Cys497 and Cys500) is represented by individual heavy-atom sites refined to similar occupancies (0.36 and 0.42, respectively) and by two separate peaks in the anomalous map as shown in Fig. 2(*b*). At the resolution of 2.4 Å used for finding the heavy-atom sites, disulfide bridges with a typical distance of 2.1 Å between the S^γ atoms are commonly not represented by individual peaks but appear together as so-called 'super-sulfurs' (Sheldrick, 2008). The detection of individual peaks for the Cys497–Cys500 disulfide bridge and a peak corresponding to only one S^γ atom of the other cysteines suggest that both S—S bonds have been partially reduced during protein purification and crystallization owing to the absence of a reducing agent or by radiation damage during data collection.

In addition to the peaks representing the intrinsic S atoms of the protein in the anomalous difference map, one further

Table 2
Comparison of the phasing statistics with and without the use of the anomalous signal of the Xe atom as output by *SHARP*.

Automatic model building was performed in 100 cycles of *ARP/wARP* and *REFMAC*5. For comparison, the statistics from *autoSHARP* (using the Xe atom, but treating it like the S atoms) are also given.

Phasing statistics	*autoSHARP* (Xe as S)	Including the Xe atom in *SHARP*	*SHARP* without the Xe atom as a heavy-atom site
Phasing power (*SHARP*)	0.47	0.47	0.47
R_{Cullis}	0.95	0.95	0.95
FOM$_{\text{acentric}}$	0.25	0.25	0.25
R_{work} from *REFMAC*5 (*ARP/wARP* cycling) (%)	25.0	27.3	27.7

peak indicates the binding of an Xe atom at a specific site. The xenon had been caught in a large hydrophobic pocket during a xenon pressurization performed prior to data collection. In the anomalous map it is represented by a peak at the 14.8σ level, while the S atom of Cys347 nearby gives rise to a peak of only 9.3σ. However, a comparison with the anomalous densities of the S$^{\delta}$ atoms of Met175 (14.3σ) and Met171 (17.4σ; the strongest peak of all) also located in the close vicinity shows that the Xe atom does not contribute excessively to the anomalous signal. Additionally, there are no peaks of outstanding intensity in the $v = 0$ Harker section shown in Fig. 3.

For successful sulfur SAD phasing, the error in the observed intensities as described by the merging R factor $R_{\text{p.i.m.}}$ (Weiss, 2001) has to be significantly smaller than the observed signal described by the anomalous R factor R_{anom} [Weiss *et al.*, 2001; $R_{\text{anom}} = 100 \times \sum_{hkl}|I(hkl) - I(-h-k-l)|/\sum_{hkl}\langle I(hkl)\rangle$, $R_{\text{p.i.m.}} = 100 \times \sum_{hkl}[1/(N-1)]^{1/2}\sum_i|I_i(hkl)-\langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)]$. Accordingly, Weiss and coworkers consider an $R_{\text{anom}}/R_{\text{p.i.m.}}$ ratio of 1.5 to be favourable. For the data presented here, the $R_{\text{anom}}/R_{\text{p.i.m.}}$ ratio is only 1.8%/1.7% = 1.1 and thus at first would seem to prevent sulfur SAD phasing ($R_{\text{p.i.m.}}$ for all $I+$ and $I-$). However, as described in this paper, these data were successfully used to determine the structure of the 66.3 kDa protein, thus extending the previous lower limit for the $R_{\text{anom}}/R_{\text{p.i.m.}}$ ratio of 1.5 significantly to 1.1. In contrast, the second frequently used indicator of the amount of anomalous signal, the estimated Bijvoet intensity ratio χ, had been promising from the beginning. It can be calculated using the equation $\chi = \langle\Delta F^{\text{anom}}\rangle/\langle F\rangle = (2N_A/N_P)^{1/2}/(f_A''/f_{\text{eff}}^{\text{O}})$, where
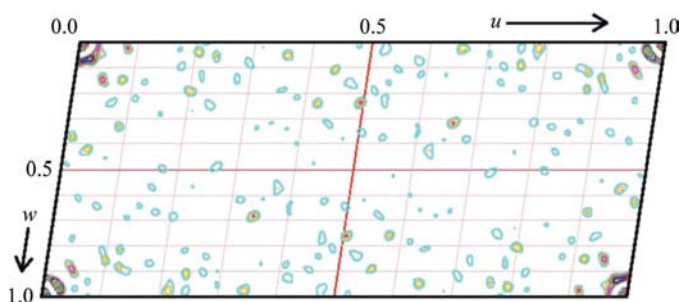

Figure 3
Anomalous difference Patterson map. Sharpened map of the Harker section $v = 0$ calculated at a resolution of 2.4 Å. Contours are in increments of 1.3σ and are coloured differently for each level.

Table 3
Occupancies and temperature factors of the heavy-atom sites.

A comparison of the occupancies of the xenon site and the S atoms as output by *SHARP* clearly reveals that the Xe atom detected as sulfur does not contribute greatly to the anomalous signal.

Heavy-atom site	Occupancy	B factor (Å^2)	Residue
1	0.75	64.9	Cys249
2	0.72	78.6	Met384
3	0.69	70.1	Cys147
4	0.68	44.1	Met140
5	0.65	33.3	Met171
6	0.61	45.8	Met551
7	0.60	34.4	Met275
8	0.59	38.0	Met575
9	0.59	54.1	Met585
10	**0.57†**	**41.8**	**Xe**
11	0.56	43.6	Met578
12	0.56	35.6	Met175
13	0.52	36.9	Met137
14	0.49	35.9	Met480
15	0.48	45.3	Met477
16	0.47	33.3	Met484
17	0.46	35.7	Cys562
18	0.42	23.2	Cys500
19	0.37	27.8	Met412
20	0.37	41.1	Cys347
21	0.36	20.2	Met549
22	0.36	26.7	Cys497

† The value of the occupancy of the Xe atom does not reflect its real occupancy, since the Xe atom was treated as an S atom during refinement of the heavy-atom sites (see text for details).

N_A is the number of anomalous scatterers and N_P is the total number of protein atoms and with the atomic scattering factor $f_{\text{eff}}^{\text{O}} = 6.7$ of an 'average' protein atom (Hendrickson & Teeter, 1981; $\Delta F^{\text{anom}} = \Delta F^{\pm}$). For the 66.3 kDa protein with 22 expected anomalously scattering S atoms of about 4400 atoms in the whole structure with a theoretical value of $f_S'' = 0.82$ at 1.900 Å, the expected Bijvoet ratio amounts to about 1.8%. Bijvoet ratios in the same range have been reported, for example, for lima bean trypsin inhibitor (1.96%) with data collected at 1.54 Å (Debreczeni, Bunkóczi, Girmann *et al.*, 2003) as well as for the protein TT0570 (1.1%), the structure of which was determined by Watanabe and coworkers using a recently developed loopless free crystal mounting as well as a Cu/Cr dual-wavelength system (Watanabe, 2006). Previously, with commonly used cryoloops, a Bijvoet ratio of only 0.6% had been predicted by Wang to be sufficient for successful SAD phasing (Wang, 1985).

The experimental map obtained for the 66.3 kDa protein was of excellent quality. Some glyosylations were already visible in the electron-density map prior to manual intervention, *e.g.* an *N*-acetylglucosamine moiety at Asn115 as shown in Fig. 2(*c*).

3.3. Relative contribution of the Xe and the S atoms to the anomalous scattering: requirements for phasing

The Xe atom, which was caught in a hydrophobic pocket during pressurization with xenon gas, was not essential for phase determination. A comparison of the occupancies of the 22 sulfur sites with that of the single xenon site showed that

the Xe atom only contributed about 5% of the overall occupancies of all anomalous scatterers and thus can be regarded as not being a prerequisite for successful phasing. In order to prove the negligible contribution of the Xe atom with the available data, phase determination was carried out explicitly omitting the xenon site in calculations using *SHARP*. A comparison of the initial electron-density maps clearly shows that there are no significant differences between the maps calculated in the absence or the presence of the Xe atom as a heavy-atom site (see Fig. 4). The phasing statistics for both

procedures are summarized in Table 2. According to the program *MAPMAN* (Kleywegt & Jones, 1996), the correlation coefficient for the two maps amounts to 0.75. The quality of both maps enabled automatic model building with *ARP/wARP* (Perrakis *et al.*, 1999).

The minor contribution of the Xe atom to the anomalous signal reflects its quite low occupancy in the crystal, which can be explained as follows. During *autoSHARP* calculations for the determination of the 66.3 kDa protein structure, the Xe atom was assumed to display the same properties as the 21 S
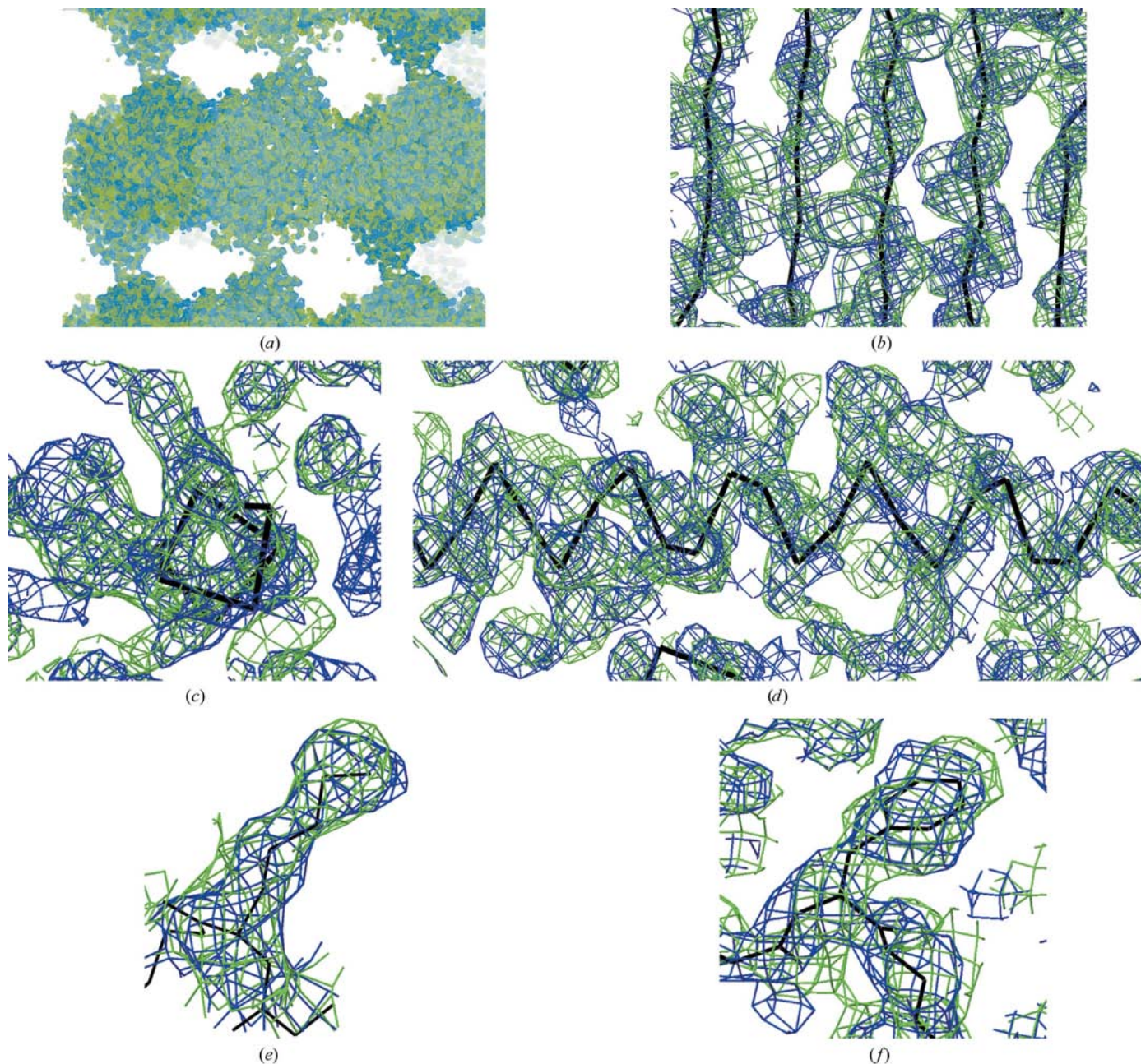

Figure 4
Comparison of the experimental electron-density maps after density modification providing *SHARP* with input parameters with heavy-atom sites including and lacking the Xe atom, respectively. The maps are coloured blue and green, respectively, and are contoured at the 1.5σ level. (*a*) The section encompassing several neighbouring molecules shows a clear solvent boundary. (*b*) The β-strands forming a β-sheet are well defined in the electron-density map. (*c*, *d*) Density for an α-helical structure is also visible, viewed from the top (*c*) and from the side (*d*) of the helix, respectively. (*e*, *f*) Close-up views of the side chains of Lys340 at the surface (*e*) and Phe371 in the hydrophobic core of the protein (*f*), respectively.

atoms used for phasing. Therefore, the occupancy of the Xe atom listed in Table 3 (0.57) was refined based on the assumption that the site was occupied by an S atom. In order to obtain the actual occupancy of the Xe atom, a correction factor has to be applied to the value listed in Table 3. This correction factor has to take into account the relationship between the anomalous signals of S and Xe atoms. At the wavelength of 1.9000 Å at which the SAD data were collected, an f'' value of about 10.29 is expected for an Xe atom, while sulfur only exhibits a value of $f'' \simeq 0.82$. Therefore, a fully occupied xenon site commonly provides a signal with an intensity 12.5 times as high as that of a fully occupied sulfur site. Thus, the correction factor for calculation of the occupancy of the xenon site amounts to 12.5 and the occupancy of 0.57 for the xenon site (Table 3) has to be divided by 12.5, resulting in an actual occupancy of only 0.046. However, the mean occupancy of the S atoms used for phasing amounts to only 0.54 (see Table 3) instead of the theoretical value of 1.0 for an intrinsic protein atom, indicating that the occupancy of all heavy-atom sites is likely to be underestimated by a factor of about 1.85. Even upon application of this further correction factor, the Xe atom seems to be bound in a maximum of 8.5% of all molecules of the 66.3 kDa protein forming the crystal. Thus, an occupancy of 0.1 has been assigned to the Xe atom in the structure (PDB code 3fbx).

3.4. Model building

Starting with an initial model from *ARP/wARP* (Perrakis *et al.*, 1999), the structure of the 66.3 kDa protein was completed manually by cycling between *Coot* and *REFMAC*5. 520 residues (Val63–Thr238 and Cys249–Pro592) were included in the final structure, which was refined to *R* factors of $R_{\mathrm{work}} = 15.6\%$ and $R_{\mathrm{free}} = 19.8\%$. In the Ramachandran plot, 90.7% of the residues are located in the core region and 8.9% lie in allowed regions, while only 0.4% of the residues belong to the generously allowed region. There are no residues in the disallowed region. Stereochemical analysis with *PROCHECK* (Laskowski *et al.*, 1993) detected *cis*-peptide conformations for two prolines (Pro502 and Pro592) and for three nonproline residues: Gly76, Gly155 and Asp316. For the first N-terminal residues Leu47–Pro62, the last C-terminal residues Trp593 and Asp594 and the 11 residues of the C-terminal affinity tag, no electron density was visible in the final map. Amino acids Asn239–Ser248 were also omitted from the model owing to a lack of unambiguous density. However, there is some remaining electron density near Arg531 which might represent some of the missing residues. At the five putative glycosylation sites of the 66.3 kDa protein (Deuschl *et al.*, 2006), Asn93, Asn115, Asn236, Asn441 and Asn520, some additional density is present that indicates the presence of *N*-acetyl-glucosamine (NAG) moieties attached to the asparagine side chains. Of these modifications, two NAG moieties at Asn115 as well as one NAG moiety each at Asn236, Asn441 and Asn520 were clearly defined in the electron-density map and thus were included in the final model. The electron density for the second sugar moiety at Asn115 was not directly visible

after density modification (Fig. 2c) but improved during refinement. The 66.3 kDa protein is a rigid structure with 17 β-strands, 13 α-helices and six 3_{10}-helices. The β-strands are arranged in two stacked antiparallel β-sheets, which are flanked by helices on both sides.

4. Conclusions

In this work, the lysosomal 66.3 kDa protein from mouse was crystallized in the monoclinic space group *C*2 and its structure was determined by means of sulfur/Xe SAD phasing. However, the contribution of the Xe atom to the overall scattering and therefore to the phasing power was not required for successful phase determination. The structure obtained is to our knowledge one of the largest structures that has been shown to be feasible for structure determination by sulfur SAD to date and belongs to a small group of proteins crystallized in monoclinic space groups that have been solved successfully using this method.

During phase improvement in *autoSHARP*, different values of the solvent content were systematically tested in steps of 3.0% using *SOLOMON* (Abrahams & Leslie, 1996). Valuable phase information for determining the 66.3 kDa protein structure was only provided for a solvent content of 56.4%. Thus, as previously stated by Watanabe *et al.* (2005), the solvent content and consequently density-improvement procedures such as solvent flattening and solvent flipping seem to play an essential role in sulfur SAD phasing.

Especially for the 66.3 kDa protein, the use of the anomalous signal of intrinsic S atoms was a valuable alternative to standard experimental phasing procedures for the following reasons. The expression system, a human fibrosarcoma cell line, allowed neither a high yield, which is required for extensive screening of heavy-atom derivatives, nor efficient incorporation of selenomethionine. The latter issue may also have been the reason that we could not obtain crystals in conditions containing the appropriately modified 66.3 kDa protein.

To the best of our knowledge, the only structure of similar molecular weight that has been solved by sulfur SAD phasing is that of the 69 kDa protein TT0570 from *T. thermophilus* (Watanabe *et al.*, 2005). In contrast to the procedure described here, Watanabe and coworkers did not use synchroton radiation at a wavelength of 1.9000 Å and a standard loop and mounting system, but applied longer wavelength Cr *Kα* radiation and a recently developed mounting technique that reduces the absorption by the cryobuffer and cryoloop (Kitago *et al.*, 2005). The most important difference concerns the respective space group: the 66.3 kDa protein formed monoclinic crystals, whereas TT0570 crystallized in *P*2$_1$2$_1$2, a space group of higher symmetry. The challenge in the use of sulfur SAD phasing for monoclinic or even triclinic crystals also became obvious in a broad study of 23 different crystal forms by Mueller-Dieckmann *et al.* (2007). In this study, neither the structures of three monoclinic crystal forms nor that of a triclinic crystal form could be solved automatically by sulfur SAD, while submission to the *AutoRickshaw* pipeline

research papers

(Panjikar *et al.*, 2005) was successful for the majority of the higher symmetry examples. The determination of the 66.3 kDa protein structure is an extraordinary example of successful sulfur SAD phasing in that the protein is not only larger than most structures solved by this method so far, but was also crystallized in a low-symmetry space group (*C*2). Hence, this work is encouraging for the wider application of this experimental phasing procedure, since it uses only the anomalous signal of intrinsic protein atoms, obviating the need for crystal derivatization with heavy atoms.

References

Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* D**52**, 30–42.
Brown, J., Esnouf, R. M., Jones, M. A., Linnell, J., Harlos, K., Hassan, A. B. & Jones, E. Y. (2002). *EMBO J.* **21**, 1054–1062.
Capasso, R., Izzo, A. A., Fezza, F., Pinto, A., Capasso, F., Mascolo, N. & Di Marzo, V. (2001). *Br. J. Pharmacol.* **134**, 945–950.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cravatt, B. F., Demarest, K., Patricelli, M. P., Bracey, M. H., Giang, D. K., Martin, B. R. & Lichtman, A. H. (2001). *Proc. Natl Acad. Sci. USA*, **98**, 9371–9376.
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.* D**59**, 393–395.
Debreczeni, J. É., Bunkóczi, G., Ma, Q., Blaser, H. & Sheldrick, G. M. (2003). *Acta Cryst.* D**59**, 688–696.
Debreczeni, J. É., Girmann, B., Zeeck, A., Krätzner, R. & Sheldrick, G. M. (2003). *Acta Cryst.* D**59**, 2125–2132.
Deuschl, F., Kollmann, K., von Figura, K. & Lubke, T. (2006). *FEBS Lett.* **580**, 5747–5752.
Djinović Carugo, K., Helliwell, J. R., Stuhrmann, H. & Weiss, M. S. (2005). *J. Synchrotron Rad.* **12**, 410–419.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Fehrenbacher, N. & Jaattela, M. (2005). *Cancer Res.* **65**, 2993–2995.
Feulner, J. A., Lu, M., Shelton, J. M., Zhang, M., Richardson, J. A. & Munford, R. S. (2004). *Infect. Immun.* **72**, 3171–3178.
Garcia, M., Platet, N., Liaudet, E., Laurent, V., Derocq, D., Brouillet, J. P. & Rochefort, H. (1996). *Stem Cells*, **14**, 642–650.
Gordon, E. J., Leonard, G. A., McSweeney, S. & Zagalsky, P. F. (2001). *Acta Cryst.* D**57**, 1230–1237.
Hansen, H. S., Moesgaard, B., Hansen, H. H. & Petersen, G. (2000). *Chem. Phys. Lipids*, **108**, 135–150.
Hasilik, A. (1992). *Experimentia*, **48**, 130–151.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
Izzo, A. A., Fezza, F., Capasso, R., Bisogno, T., Pinto, L., Iuvone, T., Esposito, G., Mascolo, N., Di Marzo, V. & Capasso, F. (2001). *Br. J. Pharmacol.* **134**, 563–570.

Jensen, A. G., Chemali, M., Chapel, A., Kieffer-Jaquinod, S., Jadot, M., Garin, J. & Journet, A. (2007). *Biochem. J.* **402**, 449–458.
Kitago, Y., Watanabe, N. & Tanaka, I. (2005). *Acta Cryst.* D**61**, 1013–1021.
Kleywegt, G. J. & Jones, T. A. (1996). *Acta Cryst.* D**52**, 826–828.
Kollmann, K., Mutenda, K. E., Balleininger, M., Eckermann, E., von Figura, K., Schmidt, B. & Lubke, T. (2005). *Proteomics*, **5**, 3966–3978.
Kos, J. & Lah, T. T. (1998). *Oncol. Rep.* **5**, 1349–1361.
La Fortelle, E. de & Bricogne, G. (1997). *Methods Enzymol.* **276**, 472–494.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Liu, Z.-J., Vysotski, E. S., Chen, C.-J., Rose, J. P., Lee, J. & Wang, B.-C. (2000). *Protein Sci.* **9**, 2085–2093.
Lübke, T., Lobel, P. & Sleat, D. E. (2008). *Biochim. Biophys. Acta*, doi:10.1016/j.bbamcr.2008.09.018.
Mueller-Dieckmann, C., Panjikar, S., Schmidt, A., Mueller, S., Kuper, J., Geerlof, A., Wilmanns, M., Singh, R. K., Tucker, P. A. & Weiss, M. S. (2007). *Acta Cryst.* D**63**, 366–380.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Nixon, R. A. & Cataldo, A. M. (2006). *J. Alzheimers Dis.* **9**, 277–289.
Panjikar, S., Parthasarathy, V., Lamzin, V. S., Weiss, M. S. & Tucker, P. A. (2005). *Acta Cryst.* D**61**, 449–457.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Potterton, L., McNicholas, S., Krissinel, E., Gruber, J., Cowtan, K., Emsley, P., Murshudov, G. N., Cohen, S., Perrakis, A. & Noble, M. (2004). *Acta Cryst.* D**60**, 2288–2294.
Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003). *Acta Cryst.* D**59**, 1020–1027.
Roeser, D., Dickmanns, A., Gasow, K. & Rudolph, M. G. (2005). *Acta Cryst.* D**61**, 1057–1066.
Saftig, P., Hetman, M., Schmahl, W., Weber, K., Heine, L., Mossmann, H., Koster, A., Hess, B., Evers, M., von Figura, K. & Peters, C. (1995). *EMBO J.* **14**, 3599–3608.
Scriver, C. R., Sly, W. S., Childs, B., Beaudet, A. L., Kinzler, K. W. & Vogelstein, B. (2001). Editors. *The Metabolic and Molecular Bases of Inherited Disease*, 8th ed. New York: McGraw–Hill.
Sekar, K., Rajakannan, V., Velmurugan, D., Yamane, T., Thirumurugan, R., Dauter, M. & Dauter, Z. (2004). *Acta Cryst.* D**60**, 1586–1590.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Sleat, D. E., Della Valle, M. C., Zheng, H., Moore, D. F. & Lobel, P. (2008). *J. Proteome Res.* **7**, 3010–3021.
Sleat, D. E., Jadot, M. & Lobel, P. (2007). *Proteomics Clin. Appl.* **1**, 1134–1146.
Sleat, D. E., Wang, Y., Sohar, I., Lackland, H., Li, Y., Li, H., Zheng, H. & Lobel, P. (2006). *Mol. Cell. Proteomics*, **5**, 1942–1956.
Sleat, D. E., Zheng, H., Qian, M. & Lobel, P. (2006). *Mol. Cell. Proteomics*, **5**, 686–701.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
Watanabe, N. (2006). *Acta Cryst.* D**62**, 891–896.
Watanabe, N., Kitago, Y., Tanaka, I., Wang, J., Gu, Y., Zheng, C. & Fan, H. (2005). *Acta Cryst.* D**61**, 1533–1540.
Weiss, M. S. (2001). *J. Appl. Cryst.* **34**, 130–135.
Weiss, M. S., Mander, G., Hedderich, R., Diederichs, K., Ermler, U. & Warkentin, E. (2004). *Acta Cryst.* D**60**, 686–695.
Weiss, M. S., Sicker, T., Djinovic-Carugo, K. & Hilgenfeld, R. (2001). *Acta Cryst.* D**57**, 689–695.
Yang, C. & Pflugrath, J. W. (2001). *Acta Cryst.* D**57**, 1480–1490.