Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

De novo calcium/sulfur SAD phasing of the human formylglycine-generating enzyme using in-house data

Dirk Roeser, Achim Dickmanns, Kathrin Gasow and Markus G. Rudolph*

Department of Molecular Structural Biology, University of Göttingen, D-37077 Göttingen, Germany

Correspondence e-mail: markus.rudolph@bio.uni-goettingen.de

Received 31 March 2005
Accepted 2 May 2005

PDB Reference: FGE, 1z70, r1z70sf.

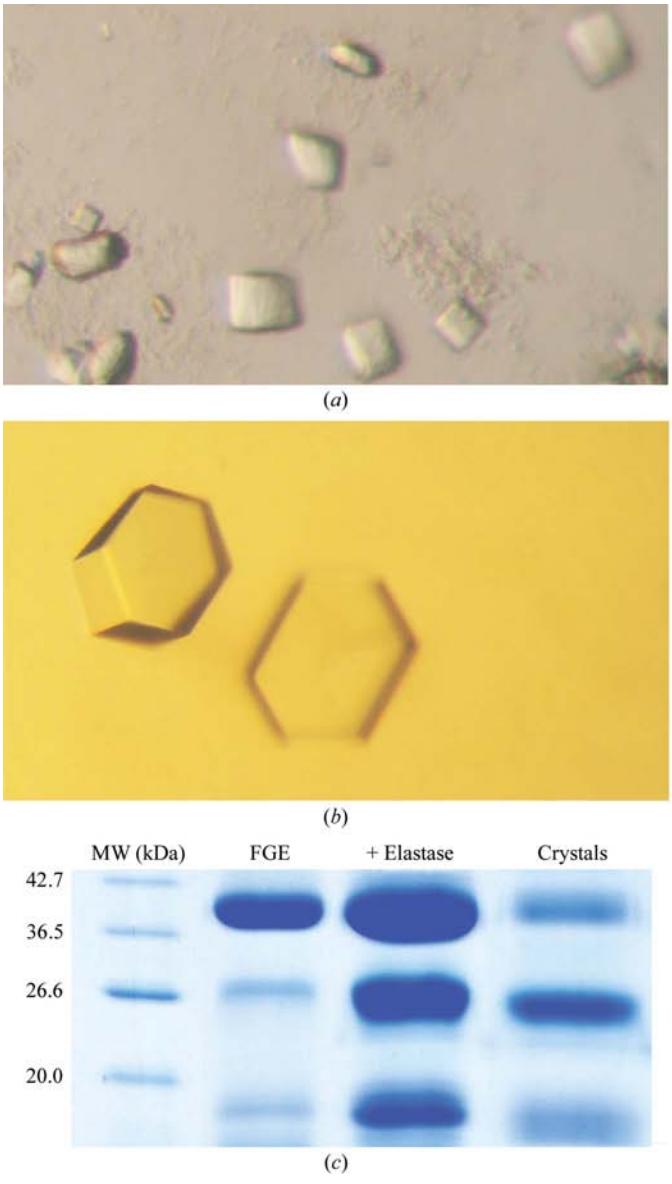
Sulfatases are a family of enzymes essential for the degradation of sulfate esters. Formylglycine is the key catalytic residue in the active site of sulfatases and is generated from a cysteine residue by FGE, the formylglycine-generating enzyme. Inactivity of FGE owing to inherited mutations in the FGE gene results in multiple sulfatase deficiency (MSD), which leads to early death in infants. Human FGE was crystallized in the presence of traces of the protease elastase, which was absolutely essential for crystal growth, and the structure of FGE was determined by molecular replacement. Before this model was completed, the FGE structure was re-determined by SAD phasing using in-house data based on the anomalous signal of two calcium ions bound to the native enzyme and intrinsic S atoms. A 14-atom substructure was determined at 1.8 Å resolution by *SHELXD*; *SHELXE* was used for density modification and phase extension to 1.54 Å resolution. Automated model building with *ARP*/*wARP* and refinement with *REFMAC*5 yielded a virtually complete model without manual intervention. The minimal data requirements for successful phasing and the relative contributions of the Ca and S atoms are discussed and compared with the related FGE paralogue, pFGE. This work emphasizes the usefulness of *de novo* phasing using weak anomalous scatterers and in-house data.

1. Introduction

Hydrolysis of sulfate esters such as glycosaminoglycans, sulfolipids and steroid sulfates in eukaryotic cells is mediated by the sulfatase enzyme family. All 13 human sulfatases described to date (Hopwood & Ballabio, 2001) contain formylglycine as a unique catalytic residue that is generated from a cysteine precursor. This post-translational modification is catalyzed by the formylglycine-generating enzyme (FGE) in the endoplasmic reticulum (Cosma *et al.*, 2003; Dierks *et al.*, 2003). A severe reduction or lack of all sulfatase activity, termed multiple sulfatase deficiency, originates from mutations in the gene coding for FGE. The oxidation of a cysteine to a formylglycine residue involves a unique mechanism that uses molecular oxygen and a reducing agent in an oxygenase-like fashion. In contrast to other oxygenases, FGE does not depend on any oxygen-activating cofactors such as flavines or nicotinamides, but employs a redox-active disulfide bond for catalysis. This disulfide bond is missing in a closely related protein, the FGE paralogue pFGE, which also resides in the endoplasmic reticulum. The function of pFGE is ill-defined at present, but may be coupled to FGE activity (Mariappan *et al.*, 2005). The crystal structure of pFGE has been determined recently and revealed FGE and pFGE to adopt very similar structures. The two proteins exhibit a novel fold and constitute

the first members of a large family of proteins harbouring a so-called domain of unknown function (DUF323; Dickmanns *et al.*, 2005). The crystal structure of FGE is described in detail elsewhere, including a putative catalytic mechanism (Dierks *et al.*, 2005). The present work details the *de novo* FGE structure determination by combined Ca^{2+}/sulfur SAD phasing and its refinement at 1.15 Å resolution. The phasing results are compared with the case of pFGE, which is highly homologous to FGE but crystallizes in a different space group.

During the last 15 y, progressively weaker anomalous scatterers, most notably sulfur, have been successfully employed for SAD and MAD phasing (Ramagopal *et al.*, 2003). Since sulfur is a component of most proteins, crystal derivatization is obviated and extraction of the weak anomalous signal from S atoms is an elegant and potentially time-efficient approach to phasing. Several guidelines have been compiled to maximize the anomalous signal during data collection, such as large counting statistics (*i.e.* high multiplicity), minimal radiation damage and optimal choice of wavelengths and data-collection strategy (Dauter, 1999; González, 2003; González *et al.*, 1999). Most sulfur-SAD applications have used synchrotron radiation, often at longer wavelengths where f'' of sulfur is appreciably large (Brown *et al.*, 2002; Gordon *et al.*, 2001; Liu *et al.*, 2000; Micossi *et al.*, 2002; Ramagopal *et al.*, 2003; Weiss *et al.*, 2001, 2004), but sometimes also at a wavelength of 1.54 Å to demonstrate the feasibility of the sulfur-SAD approach for Cu $K\alpha$ radiation (Dauter *et al.*, 1999; Sekar *et al.*, 2004). However, with the much higher brilliance and beam quality of synchrotron radiation compared with a rotating-anode generator, a full comparison of synchrotron data with in-house data may not be warranted in all cases. Few examples have been described where phasing relied entirely on the anomalous scattering by sulfur from in-house data, such as



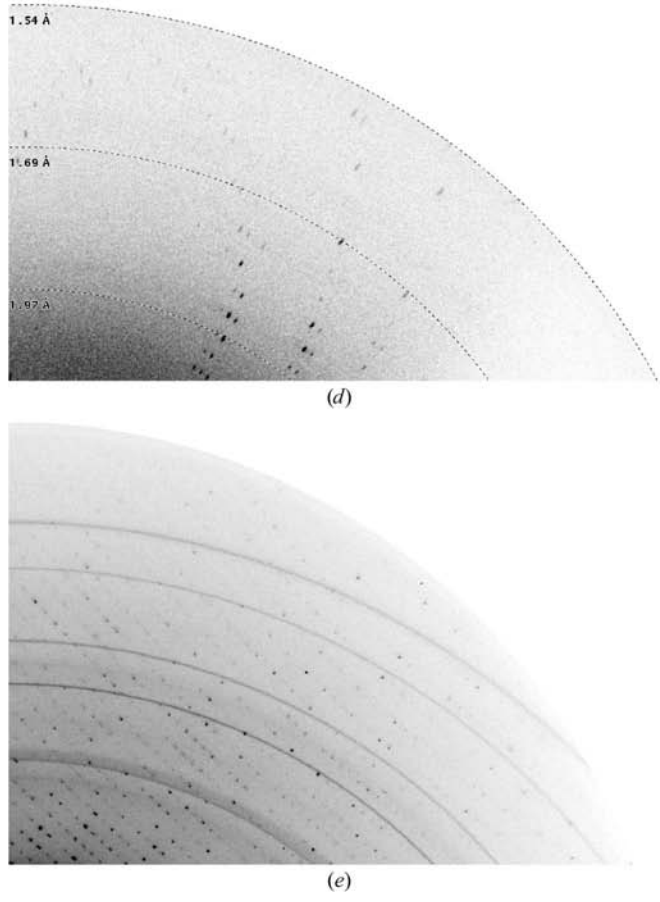


Figure 1
Crystallization and diffraction of FGE. (*a*) Initial crystals of FGE. The largest dimension is ∼0.1 mm. (*b*) Hexagonal-shaped crystals from refined conditions with a diameter of ∼0.4 mm. (*c*) SDS–PAGE of FGE before (lane 2) and after 1 h treatment with elastase (lane 3) and of dissolved FGE crystals (lane 4) showing the presence of two fragments. The intact FGE in lane 4 is a consequence of incomplete washing of the crystal, as elastase is indispensable for crystallization. FGE displays a higher apparent molecular weight than 33.7 kDa owing to glycosylation. (*d*) Section of a diffraction pattern collected in-house on a MAR345 imaging-plate detector at a distance of 100 mm after 2 min exposure. The outer resolution is 1.54 Å. (*e*) Section of a diffraction pattern collected from a different crystal at BESSY beamline BL2 with $2\theta = 10^{\circ}$. The outer resolution limit at the top left corner is 0.99 Å and the weak outermost ice ring is at 1.08 Å resolution.

cubic insulin (Debreczeni, Bunkoczi, Ma *et al.*, 2003), thaumatin (Debreczeni, Bunkoczi, Ma *et al.*, 2003; Yang & Pflugrath, 2001) and lima bean trypsin inhibitor (Debreczeni, Bunkoczi, Girmann *et al.*, 2003). In other cases employing sulfur-SAD, successful phasing did not solely rely on the anomalous signal from sulfur but was in part a consequence of the presence of other anomalous scatterers such as Cl^- in lysozyme (Dauter *et al.*, 1999; Yang & Pflugrath, 2001) and Ca^{2+} in trypsin (Debreczeni, Bunkoczi, Ma *et al.*, 2003; Yang & Pflugrath, 2001). A common theme in successful sulfur-SAD phasing using Cu $K\alpha$ radiation seems to be the requirement for enormously high multiplicity (in the range from 19.4 for thaumatin to 85.3 for lima bean trypsin inhibitor) or a very strongly scattering crystal which may not have reached its diffraction limit at the resolution cutoff used for phasing [$I/\sigma(I)$ in the high-resolution shell of 18.5 for thaumatin and 16.6 for lima bean trypsin inhibitor].

It is demonstrated here for the case of FGE that the SAD approach for *de novo* phasing using the weak anomalous signals from sulfur and Ca^{2+} and standard in-house equipment is possible with 'average' quality data and is not restricted to high crystal symmetry, special crystal orientation during data collection, extremely high multiplicity, thin-slice data collection, very low mosaicity or high solvent content. While the major contribution to phasing arises from the Ca^{2+} ions, the additional information from the sulfur sites increases the quality of the electron-density maps significantly, even at a comparatively low data multiplicity. These results are compared with the phasing attempts on pFGE, which successfully yielded a partial substructure but which did not result in interpretable electron-density maps.

2. Materials and methods

2.1. Protein purification, crystallization and data collection

FGE was cloned and produced as described by Preusser-Kunze *et al.* (2005). As extensive crystallization trials with the full-length FGE protein were unsuccessful, limited proteolysis using elastase and thermolysin was employed to generate a protease-stable FGE entity consisting of 24 and 10 kDa fragments. A mixture of elastase and 9.6 mg ml^{-1} FGE [1:1000(*w*:*w*)] was directly subjected to crystallization trials at 293 K and yielded irregularly shaped crystals overnight from condition No. 6 of Crystal Screen I (Hampton Research; 30% PEG 4000, 0.1 *M* Tris–HCl pH 8.5, 0.2 *M* $MgCl_2$; Fig. 1*a*). Refinement of these conditions yielded large apparently single crystals of hexagonal habit from 20–25% PEG 4000, 0.1 *M* Tris–HCl pH 8.0–9.0, 0.2–0.3 *M* $CaCl_2$ (Fig. 1*b*). SDS–PAGE analysis proved the presence of the two fragments in the crystals (Fig. 1*c*). Crystals were directly vitrified in a 100 K N_2-stream without additional cryoprotectant and mounted in an arbitrary orientation. Data were collected in-house on a MAR345 imaging-plate detector mounted on a MicroMax 007 generator operating with a copper target (Table 1) and reduced with the *HKL* suite of programs (Otwinowski & Minor, 1997). The exposure time was 2 min, the oscillation angle was 0.5° and the crystal-to-detector distance was 100 mm (Fig. 1*d*). No radiation damage was observed as judged by comparison of the overall $R_{\rm sym}$ value with the R values for each image and by the constant B value and scale factors per image. Based on systematic absences, the space group is $P2_12_12$, with one molecule (302 residues) in the asymmetric unit, a Matthews coefficient (Matthews, 1968) of 2.4 Å^3 Da^{-1} and a solvent content of 49%. The 1.15 Å resolution data set was collected at BESSY beamline BL2 with an oscillation angle of 0.2°, a distance of 170 mm and a 2θ offset of 10° (Fig. 1*e*). The small angular oscillation was necessary to avoid overlapping reflections in the high-resolution shell but, probably as a result of the prolonged crystal exposure, radiation damage became evident during the later stages of data collection. The pFGE protein purification and crystallization has been reported previously (Dickmanns *et al.*, 2005; Mariappan *et al.*, 2005). A high-multiplicity data set of 2894 images was collected in-house on a single crystal with an exposure time of 5 min at a distance of 150 mm and an oscillation angle of 0.5°. Data-collection statistics are summarized in Table 1.

2.2. Structure determination and refinement

The data were analyzed for anomalous signal content with *SHELXC* by comparing the anomalous signal-to-noise ratio based on the variances of F^+ and F^- [$\Delta F/\sigma(\Delta F)$]. Symmetry-related reflections were kept separate during data reduction in *SCALEPACK* to allow local scaling in *XPREP* prior to calculation of anomalous $F_{\rm A}$ values. However, little differences were found between substructure searches using data with and without merging of symmetry-related reflections. All substructure-determination calculations were performed with *SHELXD* (Schneider & Sheldrick, 2002) using data truncated at a resolution of 1.8 Å (FGE) or 2.6 Å (pFGE). For FGE, *SHELXE* was used to resolve the phase ambiguity inherent in a SAD experiment and for density modification and phase extension to 1.54 Å resolution (Sheldrick, 2002). *ARP*/*wARP* and *REFMAC*5 (Collaborative Computational Project, Number 4, 1994) were used for automatic FGE model building and refinement using default parameters. The 1.15 Å resolution FGE structure was refined anisotropically with *SHELXL* (Sheldrick & Schneider, 1997). Refinement statistics are collected in Table 3.

3. Results and discussion

3.1. Combined *in situ* limited proteolysis and crystallization of FGE

As all attempts to crystallize full-length FGE were unsuccessful, limited proteolysis was employed to truncate FGE to a form amenable to crystallization. Two fragments of apparent molecular weight of 24 and 10 kDa were obtained by elastase treatment (Preusser-Kunze *et al.*, 2005) but could not be purified to homogeneity. This suggested that only an internal loop was removed by the protease and that the two protease-resistant fragments were still attached to each other in a

Table 1
Crystallographic data collection and analysis.

Values in square brackets and parentheses correspond to the lowest and highest resolution shell, respectively.

	In-house FGE	Synchrotron FGE	In-house pFGE
Wavelength (Å)	1.54	0.918	1.54
Source/detector	MicroMax 007/MAR345	BL2/MAR345	MicroMax 007/MAR345
Temperature (K)	100	100	100
Resolution range (Å)	[30–3.32] 30–1.54 (1.60–1.54)	[41.2–3.12] 41.2–1.15 (1.17–1.15)	[45.7–6.0] 45.7–2.18 (2.24–2.18)
Space group	$P2_12_12$	$P2_12_12$	$P2_1$
Unit-cell parameters (Å,°)	a = 61.8, b = 109.6, c = 43.4	a = 62.3, b = 109.8, c = 43.5	a = 47.2, b = 101.0, c = 62.3, β = 101.4
Measured reflections	528118 (20378)	278944 (5899)	578546 (27460)
Unique reflections	43568 (3439)	93768 (4230)	28529 (1309)
Multiplicity	[13.1] 12.2 (5.9)	[4.8] 3.0 (1.4)	[20.2] 19.7 (9.4)
Completeness (%)	[99.8] 97.8 (78.6)	[90.3] 88.2 (80.8)	[99.6] 97.3 (62.2)
Anomalous completeness (%)	[99.2] 97.6 (54.8)	—	[99.5] 96.9 (57.7)
Mosaicity (°)	0.52	0.44	0.86
R_{sym}† (%)	[4.4] 6.1 (22.8)	[5.5] 6.1 (12.9)	[3.3] 4.1 (14.1)
Average $I/\sigma(I)$	[54.6] 41.6 (5.6)	[30.4] 23.9 (5.4)	[98.5] 59.4 (18.8)

† $R_{sym} = 100\sum_h \sum_i |I_i(h) - \langle I(h)\rangle|/\sum_h \sum_i I_i(h)$, where $I_i(h)$ is the ith measurement of reflection h and $\langle I(h)\rangle$ is the average reflection intensity.

folded form of FGE. The yield of the elastase-treated FGE was insufficient for crystallographic purposes, which prompted an attempt of combined *in situ* hydrolysis and crystallization (see §2). This approach was highly successful, yielding diffraction-quality crystals overnight. Combined *in situ* proteolysis and crystallization, *i.e.* deliberately contaminating a protein with trace amounts of protease prior to crystallization, requires a crystal-growth rate equal to or faster than the hydrolysis reaction in order to avoid complete degradation of the protein. The method may prove to be of general applicability in cases where a flexible surface loop is detrimental to formation of an ordered crystal lattice and where the proteolytic products cannot be purified.

3.2. S/Ca^{2+} SAD phasing of FGE using in-house data

The processed form of FGE is a monomeric protein with a molecular weight of 33.7 kDa (302 residues). The crystal structure of FGE was initially determined by molecular replacement using an unrefined model of the FGE paralogue pFGE (Dickmanns *et al.*, 2005) as the search model. Before the FGE structure was rebuilt, a high-resolution (1.54 Å) data set of high multiplicity was collected in-house. Motivated by the requirement for higher (0.2–0.3 *M*) $CaCl_2$ concentrations in the mother liquor to generate large FGE crystals and by the relatively low merging *R* values of the data (Table 1), SAD phasing was attempted, initially based on the anomalous signal of calcium. Although the overall multiplicity of the data was only 12.2 (anomalous multiplicity of 6.4), a number of significant peaks were apparent in the anomalous difference Patterson map (Fig. 2). *SHELXC* was used to estimate the anomalous signal-to-noise ratio (1.80 overall to 1.8 Å resolution and 1.50 in the highest resolution shell) and to derive anomalous differences from data scaled with *SCALEPACK*. Ca^{2+} substructure searches with *SHELXD* to a maximum resolution of 1.8 Å were successful in space group $P2_12_12$, but not in other primitive orthorhombic space groups, also proving the correct space-group assignment. A set of 14 positions was obtained, two of which are stronger than the other 12 (Table 2). Phase extension to 1.54 Å was carried out with *SHELXE* and the initial maps were readily interpretable (Figs. 3*a* and 3*b*).

While the highly occupied sites concur with the presence of two Ca^{2+} ions, the less occupied sites represent S atoms, a third Ca^{2+} ion (site 12) and a Cl^- ion (site 13). The lower occupancy of the sulfur positions is an expression of the smaller dispersive signal of sulfur compared with Ca^{2+} ($f'' = 0.56\ e^-$ *versus* $f'' = 1.29\ e^-$) for Cu $K\alpha$ radiation. Indeed, hydrolysis of FGE with trypsin and mass-spectrometric analysis of the resulting peptide fragments established the presence of three disulfide bonds in FGE. Sites 3 and 5 (Table 2) are 2.5 Å apart, thus indicating a disulfide bond. When the substructure determination was repeated with a smaller minimum allowed distance between sites in order to resolve disulfide bonds, this distance

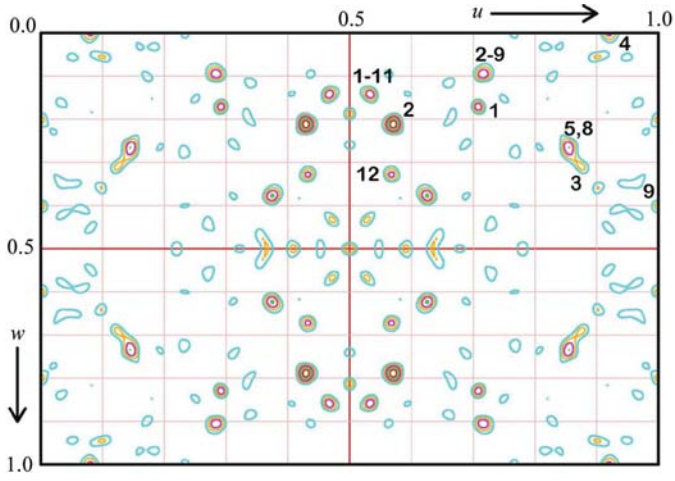


Figure 2
Anomalous difference Patterson map. Sharpened map of the section v = 0.5 calculated to a resolution of 2 Å. Contours are in steps of 1σ and are coloured differently for each level. Not all sites are represented in the Patterson section and some peaks that are present are not Harker peaks. In addition to the Harker peaks marked, two cross-peaks between sites having almost the same y coordinate (1–11 and 2–9) are also visible.

reduced to 2.0 Å for the same sites, which is very close to the canonical distance of a disulfide bond of 2.03 Å. The nature of the search atom (calcium, sulfur, potassium, bromine and chlorine were tried) was irrelevant for successful phasing and all attempts resulted in the same set of heavy-atom positions but with different relative occupancies and the quality of the

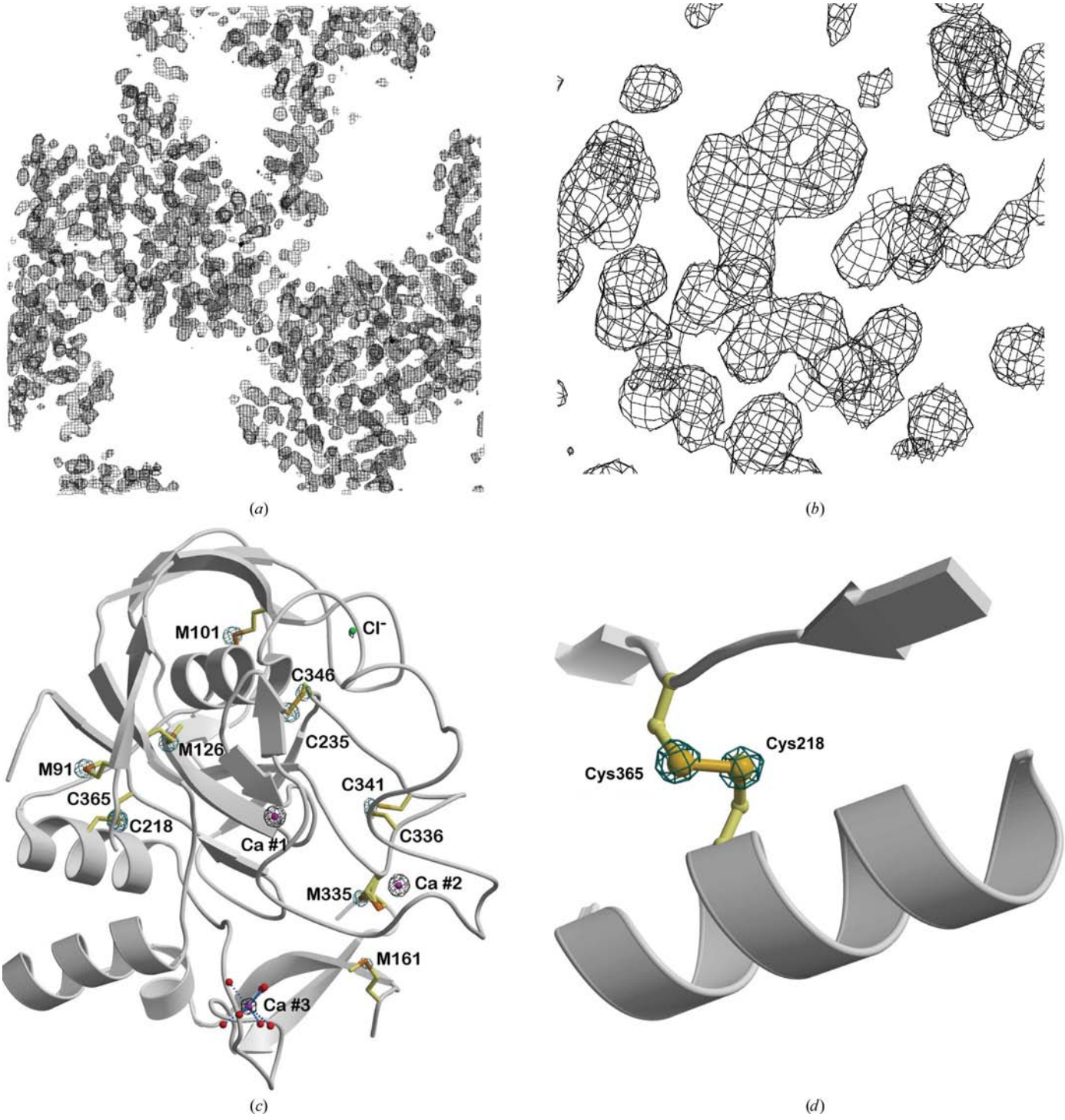


Figure 3
Experimental electron-density maps after density modification and phase extension in *SHELXE*. (*a*) Section encompassing two molecules and showing a clear solvent boundary. (*b*) Close-up of a tryptophan side chain with a hole in the benzene moiety of the indole group. The maps are contoured at the 1σ level. (*c*) Ribbon plot of the refined FGE model superimposed onto the anomalous map showing the Ca^{2+} (magenta), sulfur (orange) and Cl^{-} (green) sites. The map was calculated using the unbiased phases from (*a*) and is contoured at 10σ for the protein-bound Ca^{2+} ions and at 5σ for S atoms, Cl^{-} and the third Ca^{2+}, which is complexed by seven water molecules (red spheres) and located between two symmetry-related molecules. Met335 adopts two conformations (occupancies 0.7 and 0.3), only one of which is resolved in the map. Cys341 is the only S atom for which no anomalous peak is observed. (*d*) Close-up of the disulfide bond between Cys218 and Cys365 showing the anomalous difference peaks at 8σ for the individual S atoms.

Table 2
Fractional coordinates of the heavy atoms in the FGE substructure.

The relative occupancies of the heavy atoms from *SHELXD* show two highly occupied and 12 less occupied sites.

Site	*x*	*y*	*z*	Occupancy	Residue
1	0.396	0.321	0.418	1.00	Ca 1
2	0.535	0.308	0.608	0.99	Ca 2
3	0.310	0.284	0.155	0.43	Cys365
4	0.283	0.355	0.506	0.40	Cys235
5	0.321	0.304	0.126	0.40	Cys218
6	0.266	0.363	0.539	0.39	Cys346
7	0.212	0.249	0.201	0.35	Met91
8	0.186	0.424	0.364	0.35	Met101
9	0.243	0.311	0.302	0.35	Met126
10	0.576	0.371	0.430	0.31	Met335
11	0.634	0.323	0.439	0.29	Met161
12	0.039	0.534	0.590	0.24	Ca 3
13	0.239	0.421	0.614	0.23	Cl^-
14	0.461	0.387	0.520	0.25	Cys336

resulting maps was almost identical (data not shown). The robustness of the substructure-determination and phasing process and the fact that *SHELXD* searches for a few more heavy atoms than given in the input allows the finding of additional (weaker) anomalous scatterers that may be of different chemical type and can significantly contribute to the phasing (see below). In fact, all but one of the S atoms (Cys341) were represented by $>5\sigma$ peaks in an anomalous $F^+ - F^-$ difference map calculated with the initial unbiased phases after density modification (Figs. 3*c* and 3*d*). The absence of an anomalous signal of Cys341 is reflected by the elevated *B* value of its S^γ atom (33.2 Å^2) in the refined structure (Dierks *et al.*, 2005) compared with the mean *B* value of all S atoms (19.9 $\pm$ 6.6 Å^2).

3.3. Minimal data requirements for phasing

Data collection on this FGE crystal was compromised by a machine failure, which left a data set of 697 images ($\Delta\varphi$ = 348.5°) with an overall multiplicity of 12.2 (Table 1). After SAD phasing of the FGE crystal, the question arose what minimal multiplicity of data would have sufficed for successful substructure determination. To delineate the minimal anomalous signal required, the data multiplicity was successively reduced by scaling the images in bunches of 50 starting from the first image. The overall anomalous signal of these data sets was estimated using *SHELXC* (Fig. 4), substructure searches were performed with *HKL2MAP* (Pape & Schneider, 2004), density modification and phase extension was performed in *SHELXE* and *ARP/wARP* was used for automated model building. Identical program parameters were used for all data sets to ensure a valid comparison of the results. The first substructure including the two Ca^{2+} positions was obtained with data from 450 images ($\Delta\varphi$ = 225°, multiplicity 7.9, mean anomalous signal to 1.8 Å of 1.60). However, automated model building was not possible into electron-density maps generated from this substructure. By contrast, including only 50 more images ($\Delta\varphi$ = 250°, multiplicity 8.8, mean anomalous signal to 1.8 Å of 1.64) improved the map quality such that *ARP/wARP* returned a model with 266 residues out of 287 built manually in the final model. In all cases using data with higher multiplicity, the results were essentially identical, yielding auto-built models with 260–266 residues. Thus, there is quite a sharp borderline (within 25° $\Delta\varphi$) between success and failure in phasing the FGE data.

In order to test the relative contributions of the Ca^{2+} and sulfur substructures to phasing, separate phasing attempts with one atom type only were tried. For this, the substructures obtained from the data sets of different multiplicity were divided into the Ca^{2+} (two) and sulfur (ten) sites and used separately for phase extension in *SHELXE* and automated model building in *ARP/wARP*. When the Ca^{2+} ions were eliminated from the substructure and only the remaining sulfur positions were used for phasing, no interpretable electron-density maps were obtained, clearly showing that the main phasing power arises from the Ca^{2+} ions. In line with this result, interpretable electron-density maps could be generated using the two Ca^{2+} sites alone, further establishing their importance for phasing. The quality of the final Ca^{2+}-phased maps was slightly poorer than those calculated from the complete substructure: the minimal data multiplicity necessary for successful auto-building of the model was higher when the S-atom positions were removed from the substructure. While the first almost complete model (266 residues) was obtained from data with a multiplicity of 8.8, the minimal multiplicity for Ca^{2+}-only phasing was 9.6 and automated model building using these phases returned a model with only 245 residues built. Thus, while the sulfur sites alone do not lead to interpretable electron-density maps, they nevertheless significantly contribute to the phasing and reduce the required data multiplicity.

Sr^{2+} ions can often replace Ca^{2+} in proteins. This was also the case for FGE and the question arose whether phasing was also feasible with Sr^{2+}, possibly with a lesser amount of data as the theoretical f'' value of strontium at the Cu $K\alpha$ wavelength is 1.825, 40% larger than that of calcium. FGE crystals were grown in the presence of 0.2 *M* $SrCl_2$ and a good quality data set (Dierks *et al.*, 2005) was collected to 1.66 Å resolution that had an anomalous multiplicity of 2.0 (3.6 with Friedel mates averaged) and an $I/\sigma(I)$ in the high-resolution shell of 4.5 with a concomitant $R_{\rm sym}$ value of 13.0% in this shell (4.9% overall). Substructure searches to 2.1 Å resolution using *SHELXD* returned both Sr^{2+} sites, which were subsequently confirmed and discriminated against Ca^{2+} by their higher electron density and reasonable *B* values in refinement. However, *de novo* SAD phasing using these two Sr^{2+} sites was unsuccessful as judged by uninterpretable electron-density maps after density modification and phase extension to 1.66 Å. To find additional weaker sites that could help in the phasing, an anomalous Fourier map using these initial phases and the anomalous differences of the intensities was calculated. Unfortunately, this map showed no significant peaks other than those for the Sr^{2+} ions. In the Ca^{2+} case (see above), the minimal multiplicity for successful substructure determination was 7.9, whereas for Sr^{2+} it is 3.6. Thus, even if it is not sufficient for phasing, a 40% increase in theoretical f'' value allows halving

the amount of data while still rendering substructure determination successful.

As pointed out by a reviewer, the isomorphous difference of 18 electrons between Sr^{2+} and Ca^{2+} might be exploited for structure determination by treating the Ca^{2+} and Sr^{2+} data as a two-wavelength MAD or SIRAS experiment. Since the SAD approach using the Ca^{2+} data alone was already successful but Sr^{2+}-SAD failed, SIRAS was tested by merging all symmetry-equivalent reflections in the Ca^{2+} data set, thus eliminating the anomalous signal and using it as native data. The same Sr^{2+} data set as above was used as the isomorphous derivative containing anomalous information. As expected, substructure searches at resolutions of 1.8–2.0 Å returned both Sr^{2+} sites as in the SAD case, with only one other site of occupancy 0.29 that was located between the S^{γ} atoms of a disulfide bridge and could be viewed as a super-S atom. However, using density modification, the substructure consisting of the two Sr^{2+} ions alone was sufficient for phasing of FGE and *ARP*/*wARP* returned a model of 262 residues that was indistinguishable from the Ca^{2+}-SAD derived models (see above). Thus, while Sr^{2+}-SIRAS and Sr^{2+}-SAD returned both Sr^{2+} positions, only SIRAS enabled phasing during density modification.

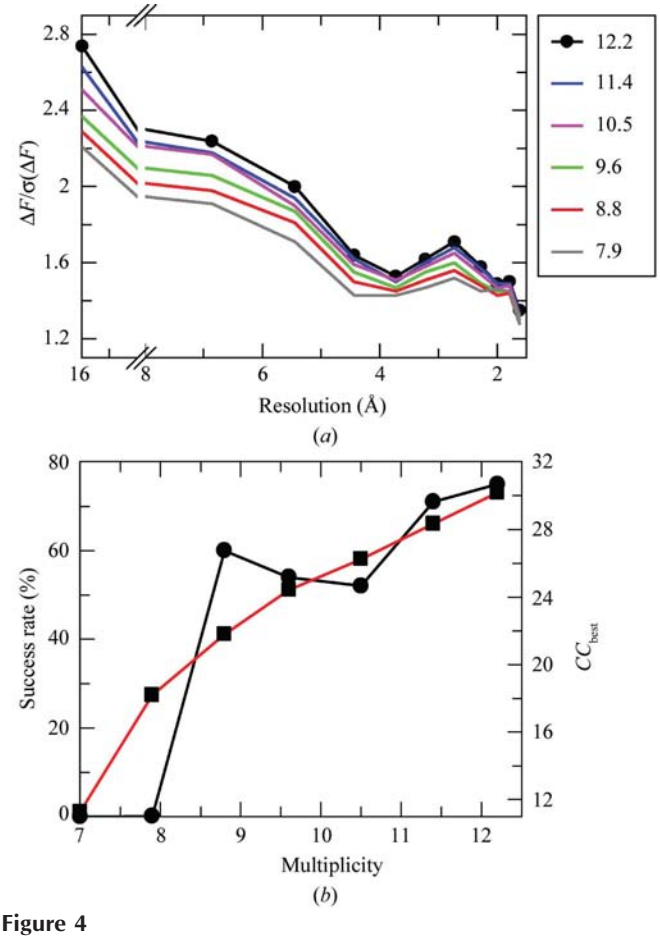


Figure 4
Minimal data requirements for successful FGE phasing. (*a*) Anomalous signal-to-noise ratios calculated as $\Delta F/\sigma(\Delta F)$ (1.0 is random) for data sets scaled with decreasing multiplicity, as given in the legend. The unusually low anomalous signal from reflections in the 3.7 Å resolution range probably arises from the presence of a slight ice ring in the diffraction patterns. The lowest multiplicity that still led to successful substructure determination and interpretable electron-density maps after density modification and phase extension is 8.8 (plotted in red). Data with a multiplicity of 7.9 allowed approximate substructure determination, but did not yield interpretable electron-density maps (plotted in grey). (*b*) Success rate and overall quality of the solutions as a function of multiplicity. 100 trial runs (1000 in case of multiplicities of <8.8) were performed with *SHELXD* and the percentage of successful substructure determination extracted based on CC values of >17% (circles). Note the steep rise in success rate for multiplicities of >8. The best overall correlation coefficient (squares) increases almost linearly for multiplicities of >8.

Similar results to the Sr^{2+}-SAD approach, namely finding the strongest anomalous scatterers but failing to phase the entire crystal with them, were obtained using data collected from a crystal of pFGE, the paralogue of FGE. The proteins share 50.6% sequence identity over 283 residues and have very similar structures, with a root-mean-square deviation of 1.3 Å over 249 common C^{α} atoms. pFGE crystallizes in space group $P2_1$ with two molecules in the asymmetric unit. This lower symmetry compared with the orthorhombic FGE and the lower resolution limit of the pFGE data (Table 1) prompted the question whether the pFGE structure could also be determined based on the anomalous signal of the Ca^{2+} ions and S atoms. A data set of high multiplicity was collected (Table 1) and revealed several peaks in the anomalous difference Patterson map (Fig. 5). The average anomalous signal-to-noise ratio of 2.3 (1.4 in the high resolution shell 2.4–2.2 Å) is even larger than that for FGE (1.8). Substructure searches for the expected 20 sites (four Ca^{2+} ions and 16 S

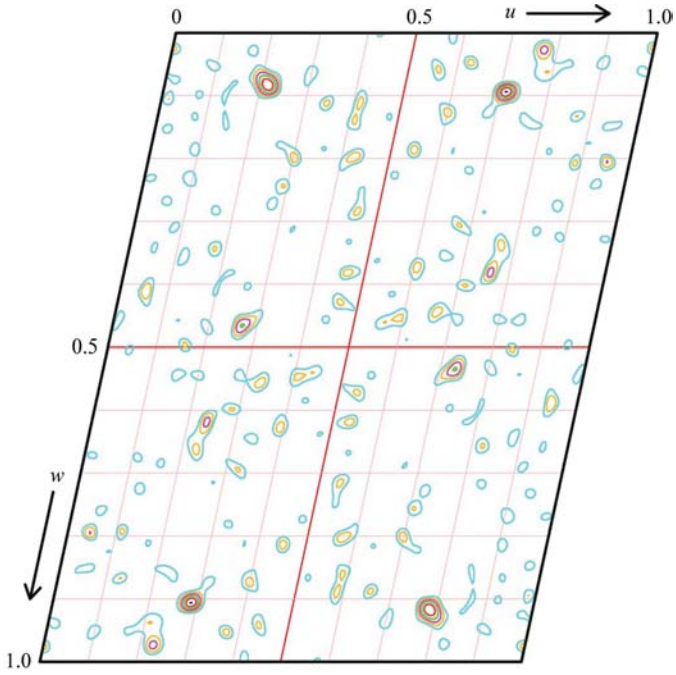


Figure 5
Anomalous difference Patterson map and fractional coordinates of the heavy atoms in the pFGE substructure. The sharpened map of the section $v = 0.5$ was calculated to a resolution of 2.2 Å. Contours are in steps of 1σ and are coloured differently for each level. The strongest two peaks at the top of the section ($w \simeq 0.1$) correspond to the Ca^{2+}-ion positions with the lowest B value.

atoms) returned two highly occupied sites, which were verified *a posteriori* as the Ca^{2+} ions at site 1 of molecules *A* and *B* of the pFGE dimer using the coordinates of the refined structure (Dickmanns *et al.*, 2005). Weaker sites corresponding to the other two Ca^{2+} ions and the S atoms of methionine residues 30, 249 and 275 as well as a super-S atom at a disulfide bridge (Cys156–Cys290) were identified as $>5\sigma$ peaks in an anomalous Fourier map. These sites were present in both molecules and using *RESOLVE* a twofold NCS-operation was detected, which was expected given that pFGE crystallizes as a dimer related by a twofold rotation (Dickmanns *et al.*, 2005). However, phase improvement by NCS averaging was also unsuccessful and the substructure did not yield useful phases as judged by uninterpretable electron-density maps. The same negative result was obtained from *SHARP*/*SOLOMON*, another phasing/density-modification program combination needed to resolve the phase ambiguity in SAD. Thus, although part of the pFGE substructure could be determined, the quality of the derived phases was not high enough to yield interpretable maps in this case. Possible explanations for the failure to phase the pFGE crystal include the larger mosaicity and lower resolution limit of the pFGE data (Table 1) and the higher mean *B* value of the final model (31.5 ± 7.5 Å^2 for all protein atoms) compared with FGE (18.8 ± 5.2 Å^2). In fact, the pFGE substructure found by *SHELXD* contains the two Ca^{2+} ions with the smallest *B* values (23.8 and 27.5 Å^2), while the other two ions exhibit higher *B* values (30.0 and 31.7 Å^2). The phasing power from the two Ca^{2+} ions may have been too low given the presence of double the number of protein atoms in the monoclinic pFGE crystal compared with the orthorhombic FGE crystal. In addition, while the overall statistics (Table 1) do not seem to support this, the FGE data quality is much higher than that from the pFGE crystal. The R_{sym} value for the FGE data in the resolution range 2.30–2.09 Å, which roughly corresponds to the high-resolution shell of the pFGE data, is only 6.9%, with an $I/\sigma(I)$ of 41.6. By contrast, R_{sym} for the high-resolution shell in pFGE (2.24–2.18 Å) is 14.4% with an $I/\sigma(I)$ of 18.8 (Table 1). These large differences in R_{sym} and $I/\sigma(I)$ may have led to the differences in the substructure determinations of FGE and pFGE, which stresses the importance of data quality for the Ca^{2+}-based SAD structure-determination approach.

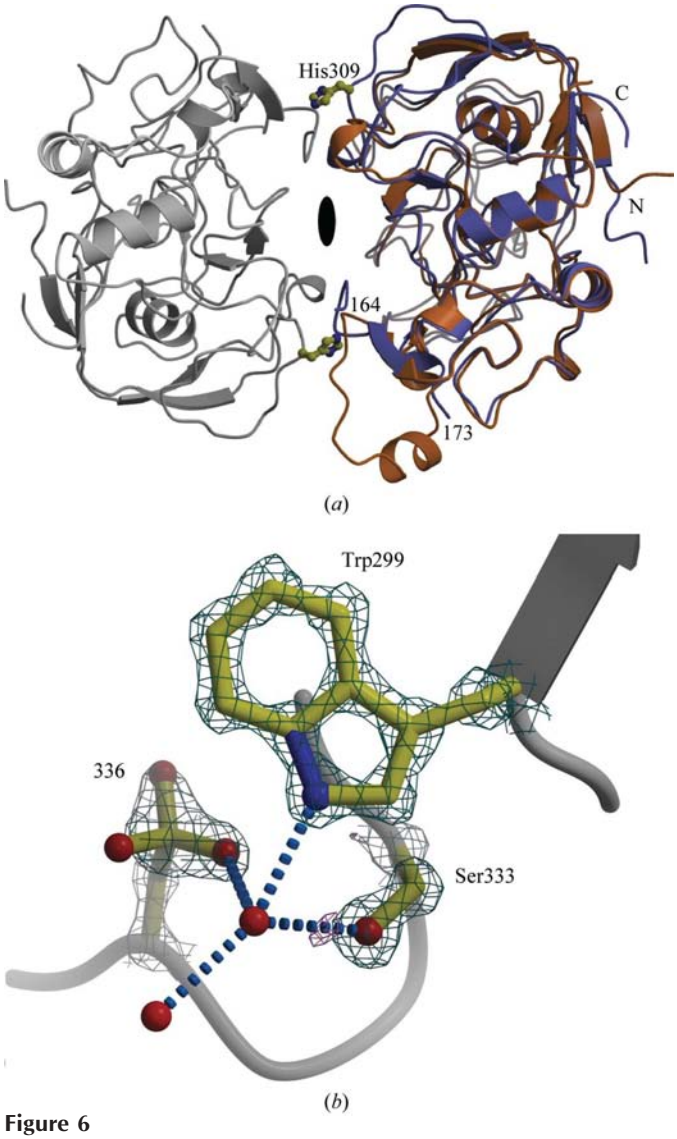


Figure 6
Crystal packing and electron density of FGE. (*a*) Overlay of pFGE (orange) onto FGE (blue) to highlight a possible conformation of the loop region that was removed by elastase during FGE crystallization. A symmetry-related FGE molecule is drawn in grey. His309, shown as a ball-and-stick model in both FGE molecules, from the symmetry-related FGE molecule clashes with the loop in this conformation and explains why it is impossible to crystallize intact FGE in this lattice. The twofold axis relating the two FGE molecules is marked in the centre. (*b*) Difference density in the 1.15 Å resolution map with the putative H-atom position at Ser333 O^{γ}. The $mF_o - DF_c$ map (magenta) is contoured at 2.2σ, the $2mF_o - DF_c$ map (blue) is contoured at 3σ and the refined model is superimposed. Cys336 was oxidized in this structure to the sulfonic acid. A water molecule is bound between Trp299, Ser333 and residue 336 that would act as a hydrogen-bond acceptor from Ser333.

3.4. Refinement and analysis of FGE

FGE is a unique enzyme that post-translationally converts a cysteine residue in all known human sulfatases to formylglycine. In its hydrated form as a geminal diol, formylglycine acts as a nucleophile for transfer of the sulfuryl group of sulfate esters to the sulfatase. Elimination of sulfate and addition of water regenerates the resting state of the sulfatase with a non-hydrated formylglycine residue (Boltes *et al.*, 2001). Because sulfatases absolutely require formylglycine for their activity, FGE forms a bottleneck for their activation and FGE inactivity leads to the severe syndrome of MSD.

The crystallization of FGE required the presence of elastase in the crystallization setup. No electron density was observed for the region Glu164–Ala173, which corresponds to the sequence removed by the protease (Dierks *et al.*, 2005). Because the first cysteine residue in FGE is Cys218, the fragments are not connected by disulfide bonds but held together by the hydrophobic core of FGE. In the crystal structure of intact pFGE, this region has the same length with a slightly different sequence and forms a loop protruding from the body of pFGE (Dickmanns *et al.*, 2005). Superposition of pFGE onto FGE and analysis of the crystallographic FGE

Table 3
Refinement statistics.

Values in parentheses are for the highest resolution shell.

Resolution range (Å)	41.2–1.15 (1.18–1.15)
No. of reflections	89044
No. in test set	5967
R_{cryst}† (%)	13.8 (15.0)
R_{free}† (%)	17.0 (17.1)
No. of residues	278
No. of protein atoms	2333
No. of water O atoms	546
No. of sugar residues	2
No. of Ca^{2+} ions	2
No. of Cl^- ions	1
No. of other atoms	57
Coordinate error‡ (Å)	0.027
R.m.s.d. bonds (Å)	0.014
R.m.s.d. angles (°)	1.77
R.m.s.d. dihedrals (°)	6.8
Ramachandran plot§	
Core (%)	88.9
Allowed (%)	10.3
Generously allowed (%)	0
Disallowed (%)	0.9
Average *B* values (Å^2)	
Protein	13.2
Water	31.0
Ca^{2+}	6.5
Others	30.4

neighbours confirm the steric incompatibility of this loop conformation with the packing in this crystal form (Fig. 6*a*). Surprisingly, only a single residue was found to interfere with the packing of the intact FGE. His309 of a symmetry-related molecule would clash with the loop not only if it adopted the same conformation as in pFGE but also in all other possible conformations owing to space restrictions around position Glu164. Thus, in order to crystallize the full-length FGE, either a different crystal form or crystal engineering by mutagenesis will be required.

Based on six high-resolution crystal structures of FGE in different redox environments, it could be demonstrated that two cysteine residues in FGE, Cys336 and Cys341, form parts of the active site. A catalytic mechanism for the oxidation of a substrate cysteine to formylglycine has been proposed (Dierks *et al.*, 2005), which included as key proposals the presence of a cysteine sulfenic acid intermediate and a catalytic base, possibly Ser333. Ultrahigh-resolution data on FGE promised the localization of H atoms on these residues that would reveal more details of the catalytic mechanism. FGE diffraction data to 1.67 Å resolution were collected in-house from a different crystal to that used for phasing. These data were extended to 1.15 Å using the same crystal and synchrotron radiation (BESSY beamline BL2). The electron-density maps revealed the two critical cysteine residues to be fully oxidized to the sulfonic acids (Dierks *et al.*, 2005; Fig. 6*b*). While oxidation to sulfonic acids is irreversible and clearly a side-reaction of FGE, it nevertheless demonstrates the high redox activity of these cysteine residues, which are involved in catalysis. Although S atoms are particularly sensitive to radiation-induced damage (Nave & Garman, 2005, and references therein; Weik *et al.*, 2000), oxidation of these cysteine residues in FGE was not a consequence of radiation damage as the in-house data also revealed the presence of sulfonic acids. Thus, these cysteine residues were probably oxidized by air after crystallization. However, during synchrotron data collection, radiation damage was evident. The starting resolution of the crystal was 0.99 Å (Fig. 1*e*) but fell to about 1.15 Å towards the end of the experiment. Also, ice rings prevented collection of a complete data set (Table 1), which explains the somewhat elevated R_{free} value of the refined model (Table 3). Nevertheless, the first electron-density maps calculated to 1.15 Å resolution after initial isotropic refinement with polar H atoms excluded showed difference peaks near some residues that based on geometric criteria could represent H atoms, including the proposed catalytic base Ser333 (Fig. 6*b*). A conserved water molecule is bound between Ser333 and residue 336 and may also partake in the catalytic mechanism, as it is seen in this position in several other FGE crystal structures (Dierks *et al.*, 2005). However, any new insight into the molecular mechanism of FGly formation at ultrahigh resolution, particularly with respect to H-atom positions, requires reduction of the radiation damage seen with this structure.

4. Conclusions

Several phasing techniques exploiting the anomalous signal contained in data collected with in-house generators have emerged during recent years. These include scatterers such as halides, sulfur, calcium and manganese. As sulfur is naturally present in most proteins, it is the least invasive approach for deriving phase information as no derivatization is necessary. The presented case of FGE not only adds to the growing list of structures determined by the use of only single-wavelength anomalous scattering data, but also serves as a testimony for rapid phasing using weak anomalous scatterers and average-quality in-house data. The failure to use SAD on the highly related pFGE structure may indicate that the FGE case marks a borderline with respect to data quality and space-group symmetry at which phasing with such weak anomalous scatterers becomes impossible. Future ultrahigh-resolution structures of FGE will reveal atomic details of intermediates and conformational changes during catalysis, provided that irreversible oxidation of the catalytically important cysteine residues and radiation damage can be prevented.

We would like to thank Martin Fieber-Erdmann at BESSY beamline BL2 for guidance during synchrotron data collection, Ralf Ficner for continuous support and Dagmar Klostermeier for very helpful discussions. This work was supported by grants from the Deutsche Forschungsgemeinschaft to MGR.

References

Boltes, I., Czapinska, H., Kahnert, A., von Bulow, R., Dierks, T., Schmidt, B., von Figura, K., Kertesz, M. A. & Uson, I. (2001). *Structure*, **9**, 483–491.

Brown, J., Esnouf, R. M., Jones, M. A., Linnell, J., Harlos, K., Hassan, A. B. & Jones, E. Y. (2002). *EMBO J.* **21**, 1054–1062.
Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* D**50**, 760–763.
Cosma, M. P., Pepe, S., Annunziata, I., Newbold, R. F., Grompe, M., Parenti, G. & Ballabio, A. (2003). *Cell*, **113**, 445–456.
Dauter, Z. (1999). *Acta Cryst.* D**55**, 1703–1717.
Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G. & Sheldrick, G. M. (1999). *J. Mol. Biol.* **289**, 83–92.
Debreczeni, J. É., Bunkoczi, G., Girmann, B. & Sheldrick, G. M. (2003). *Acta Cryst.* D**59**, 393–395.
Debreczeni, J. É., Bunkoczi, G., Ma, Q., Blaser, H. & Sheldrick, G. M. (2003). *Acta Cryst.* D**59**, 688–696.
Dickmanns, A., Schmidt, B., Rudolph, M. G., Mariappan, M., Dierks, T., von Figura, K. & Ficner, R. (2005). *J. Biol. Chem.* **280**, 15180–15187.
Dierks, T., Dickmanns, A., Preusser, A., Schmidt, B., Mariappan, M., von Figura, K., Ficner, R. & Rudolph, M. G. (2005). *Cell*, **121**, 541–552.
Dierks, T., Schmidt, B., Borissenko, L. V., Peng, J., Preusser, A., Mariappan, M. & von Figura, K. (2003). *Cell*, **113**, 435–444.
González, A. (2003). *Acta Cryst.* D**59**, 315–322.
González, A., Pedelacq, J., Sola, M., Gomis-Ruth, F.-X., Coll, M., Samama, J. & Benini, S. (1999). *Acta Cryst.* D**55**, 1449–1458.
Gordon, E. J., Leonard, G. A., McSweeney, S. & Zagalsky, P. F. (2001). *Acta Cryst.* D**57**, 1230–1237.
Hopwood, J. J. & Ballabio, A. (2001). *The Metabolic and Molecular Bases of Inherited Diseases*, edited by C. R. Scriver, A. L. Beaudet, W. S. Sly, D. Valle, B. Childs, K. W. Kinzler & B. Vogelstein, pp. 3725–3732. New York: McGraw–Hill.
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.
Mariappan, M., Preusser-Kunze, A., Balleininger, M., Eiselt, N., Schmidt, B., Gande, S. L., Wenzel, D., Dierks, T. & von Figura, K. (2005). *J. Biol. Chem.* **280**, 15173–15179.
Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
Micossi, E., Hunter, W. N. & Leonard, G. A. (2002). *Acta Cryst.* D**58**, 21–28.
Nave, C. & Garman, E. F. (2005). *J. Synchrotron Rad.* **12**, 257–260.
Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
Pape, T. & Schneider, T. R. (2004). *J. Appl. Cryst.* **37**, 843–844.
Preusser-Kunze, A., Mariappan, M., Schmidt, B., Gande, S. L., Mutenda, K., Wenzel, D., von Figura, K. & Dierks, T. (2005). *J. Biol. Chem.* **280**, 14900–14910.
Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003). *Acta Cryst.* D**59**, 1020–1027.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Sekar, K., Rajakannan, V., Velmurugan, D., Yamane, T., Thirumurugan, R., Dauter, M. & Dauter, Z. (2004). *Acta Cryst.* D**60**, 1586–1590.
Sheldrick, G. M. (2002). *Z. Kristallogr.* **217**, 644–650.
Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol.* **277**, 319–343.
Weik, M., Ravelli, R. B., Kryger, G., McSweeney, S., Raves, M. L., Harel, M., Gros, P., Silman, I., Kroon, J. & Sussman, J. L. (2000). *Proc. Natl Acad. Sci. USA*, **97**, 623–628.
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.