Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

Structural analysis of actinidin and a comparison of cadmium and sulfur anomalous signals from actinidin crystals measured using in-house copper- and chromium-anode X-ray sources

Manickam Yogavel,[a] Nirmal Nithya,[b] Atsuo Suzuki,[c] Yasuo Sugiyama,[d] Takashi Yamane,[c] Devadasan Velmurugan[b] and Amit Sharma[a]*

[a]Structural and Computational Biology Group, International Centre for Genetic Engineering and Biotechnology, Aruna Asaf Ali Road, New Delhi 110 067, India, [b]Centre of Advanced Study in Crystallography and Biophysics, University of Madras, Guindy Campus, Chennai 600 025, India, [c]Department of Biotechnology, School of Engineering, Nagoya University, Chikusa-ku, Nagoya 464-8603, Japan, and [d]Center for Gene Research, Nagoya University, Chikusa-ku, Nagoya 464-8602, Japan

Correspondence e-mail: amit.icgeb@gmail.com

Received 6 May 2010
Accepted 8 October 2010

PDB References: actinidin, 3p5u; 3p5v; 3p5w; 3p5x.

The structure of the 24 kDa cysteine protease *saru*-actinidin from the fruit of *Actinidia arguta* Planch. (sarunashi) was determined by the cadmium/sulfur-SAD method with X-ray diffraction data collected using in-house Cu $K\alpha$ and Cr $K\alpha$ radiation. The anomalous scatterers included nine sulfurs and several cadmium ions from the crystallization solution. The high quality of the diffraction data, the use of chromium-anode X-ray radiation and the substantial anomalous signal allowed structure determination and automated model building despite both a low solvent content (<40%) and low data multiplicity. The amino-acid sequence of *saru*-actinidin was deduced from the cDNA and was modified based on experimental electron-density maps at 1.5 Å resolution. The active site of *saru*-actinidin is occupied by a cadmium ion and the active-site cysteine is found to be in an unmodified, cysteine sulfenic acid or cysteine sulfinic acid form. The cadmium sites, coordination geometries and polygonal water structures on the protein surface have also been extensively analyzed. An analysis and comparison of the sulfur/cadmium anomalous signals at the Cu $K\alpha$ and Cr $K\alpha$ wavelengths was carried out. It is proposed that the inclusion of cadmium salts in crystallization solutions coupled with chromium-anode radiation can provide a convenient route for structure determination.

1. Introduction

Actinidia arguta Planch. (sanusashi; also known as hardy kiwi) produces a small kiwi-like fruit from which a novel cysteine protease, *saru*-actinidin, has been isolated. *Saru*-actinidin shares ∼50–81% sequence identity with previously structurally determined plant cysteine proteases, including actinidin from *A. chinensis* (*chins*-actinidin; Baker & Dodson, 1980), papain, caricain, chymopain and glycyl endopeptidases from *Carica papaya* (Pickersgill *et al.*, 1991, 1992; Maes *et al.*, 1996; Groves *et al.*, 1996), ervatamins A, B and C from *Tabernaemontana divaricata* (Biswas *et al.*, 2003; Thakurta *et al.*, 2004), ginger protease II (GPII) from *Zingiber officinale* (Choi *et al.*, 1999), mexicain from *Jacaratia mexicana* and endoprotease B isoform 2 (EP-B2) from *Hordeum vulgare* (Bethune *et al.*, 2006). Unprotected active-site cysteine residues can undergo various covalent modifications which may regulate protein activity (Reddie & Carroll, 2008). Cysteine can form three different species by reacting with oxygen: cysteine sulfenic (–SOH), cysteine sulfinic ($-SO_2H$) and cysteine sulfonic

(–SO_3H) acids. The cysteine proteases were purified in inactive forms in which the active-site cysteines were protected by adducts (Choi *et al.*, 1999; Maes *et al.*, 1996; Biswas *et al.*, 2003; Thakurta *et al.*, 2004; Ghosh *et al.*, 2008). Unprotected forms and modified cysteine sulfinic acid and cysteine sulfonic acid forms have also been reported (Kamphuis *et al.*, 1984; Baker & Dodson, 1980).

The use of chromium-anode X-ray radiation (λ = 2.29 Å) is attractive for single-wavelength anomalous diffraction (SAD) experiments. For example, in the case of sulfur the anomalous scattering signal is more than doubled (f'' = 1.14) compared with conventional Cu $K\alpha$ X-rays (f'' = 0.56). Furthermore, naturally bound metals and atoms from crystallization solutions (*e.g.* Cl, P or Cd) tend to show a significant increase in anomalous scattering with Cr radiation. The anomalous contribution (f'') of cadmium at the Cu $K\alpha$ and Cr $K\alpha$ wavelengths is ∼4.7 and 8.97, respectively. Cadmium may replace calcium and is expected to have similar chelation by acidic protein residues (Naismith *et al.*, 1993; Harrop *et al.*, 1996). It has been shown previously that some divalent cations can induce or improve the crystallization of proteins (Trakhanov *et al.*, 1998; Trakhanov & Quiocho, 1995). For example, the addition of ∼200 m*M* cadmium ion improved the diffraction quality of a variety of protein crystals (Yao *et al.*, 1994; Nickitenko *et al.*, 1995; Trakhanov & Quiocho, 1995; Trakhanov *et al.*, 1998; Liu *et al.*, 2006; Stegmann *et al.*, 2010). In most of these the bound cadmium was located at the molecular interface/protein surface and there was no evidence of significant changes in the three-dimensional structure regardless of cadmium concentration (Yao *et al.*, 1994; Adams & Jia, 2005; Boraston *et al.*, 2006; Stegmann *et al.*, 2010). So far, few structures have been solved by the cadmium-SAD or SIRAS methods (Harrop *et al.*, 1996; Liu *et al.*, 2006; Boraston *et al.*, 2006; Kaus-Drobek *et al.*, 2007; Kajander *et al.*, 2007; Xu *et al.*, 2008; Grimshaw *et al.*, 2008; van Bueren *et al.*, 2009; Gangelhoff *et al.*, 2009; Stegmann *et al.*, 2010) and in most of

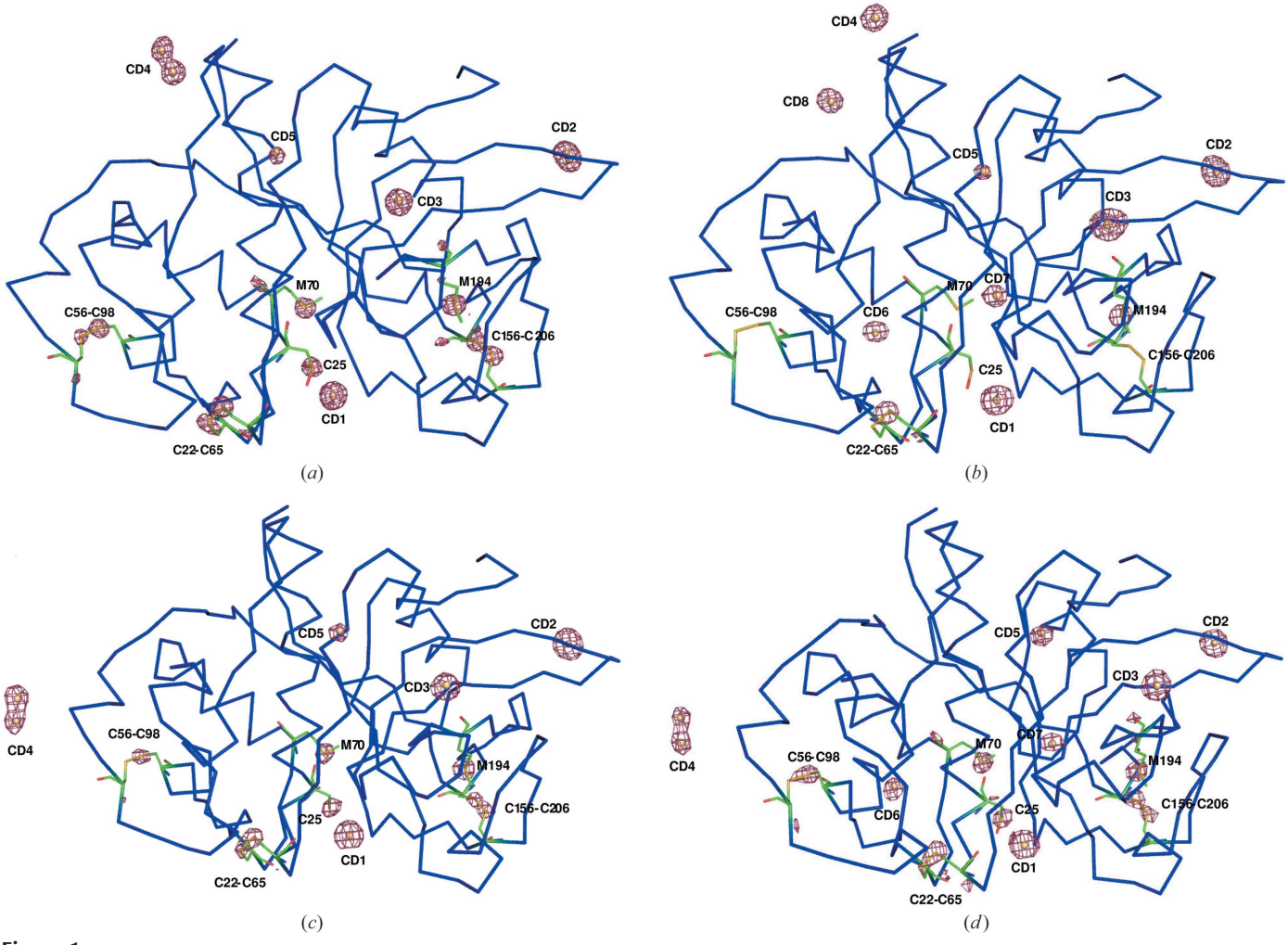

Figure 1
Anomalous difference Fourier maps around S atoms and bound cadmium ions at 4σ (magenta). (*a*) A native crystal grown in the presence of 30 m*M* $CdCl_2$ was used to collect data to 1.5 Å resolution using Cu $K\alpha$ radiation (Cu30m*M*$CdCl_2$ data), (*b*) Cu300m*M*$CdCl_2$ data were collected to 1.9 Å resolution, (*c*) Cr30m*M*$CdCl_2$ data were collected to 2.2 Å resolution and (*d*) Cr300m*M*$CdCl_2$ data were collected to 2.2 Å resolution. Met and Cys residues are shown as green sticks and bound cadmium ions are highlighted as brown spheres. Two closely related cadmium ions were located in the CD4 site in all of the data sets apart from Cu300m*M*$CdCl_2$.

those cases the cadmium ions arose from the crystallization solution. Here, we report the structure determination of *saru*-actinidin by the cadmium/sulfur-SAD technique using four different data sets collected at in-house Cu $K\alpha$ and Cr $K\alpha$ radiation sources and the anomalous signals were analyzed. The *saru*-actinidin crystals were grown in the presence of 30 m*M* $CdCl_2$ (part of the crystallization solution) and data sets, namely Cu30m*M*$CdCl_2$ and Cr30m*M*$CdCl_2$, were collected at the Cu $K\alpha$ and Cr $K\alpha$ wavelengths. In order to compare and utilize the potentially enhanced anomalous signal for SAD phasing, we increased the concentration of $CdCl_2$ from 30 to 300 m*M* using the quick cryosoaking method and two further data sets (Cu300m*M*$CdCl_2$ and Cr300m*M*$CdCl_2$) were collected.

2. Materials and methods

2.1. Crystallization

The *saru*-actinidin crystals were grown by the hanging-drop vapour-diffusion method at 288 K. The crystallization droplet consisted of 2 µl protein solution (15 mg ml^{-1} in buffer consisting of 20 m*M* sodium phosphate pH 6.0) and 2 µl reservoir solution [30 m*M* $CdCl_2$ and 30%(*v*/*v*) PEG 400 in 0.1 *M* sodium acetate pH 4.6] and was equilibrated against 300 µl well solution. Rod-shaped crystals appeared after 3 d and belonged to the orthorhombic space group $P2_12_12_1$ with one molecule per asymmetric unit. The Matthews coefficient and solvent content are 1.96 $Å^3$ Da^{-1} and ∼37%, respectively.

2.2. Data collection and processing

The X-ray diffraction data sets were collected at the High Intensity X-ray Laboratory, Nagoya University at 100 K using Cu $K\alpha$ radiation (λ = 1.54 Å) and Cr $K\alpha$ radiation (λ = 2.29 Å) generated by a Rigaku FR-E rotating-anode X-ray generator operated at 45 kV and 45 mA with Osmic mirrors. The diffraction images were recorded using a Rigaku R-AXIS VII detector. The crystals were cryosoaked by adding 20% glycerol to the mother liquor prior to data collection. Two data sets (Cu30m*M*$CdCl_2$ and Cr30m*M*$CdCl_2$) were collected using Cu $K\alpha$ and Cr $K\alpha$ radiation, respectively. The $CdCl_2$ concentration in the mother liquor was increased from 30 to 300 m*M* in the cryosolution and two further data sets (Cu300m*M*$CdCl_2$ and Cr300m*M*$CdCl_2$) were collected using Cu $K\alpha$ and Cr $K\alpha$ radiation, respectively. For each data set, a total of 360 frames were collected using 1° oscillation steps with 180 s exposure per frame in each case. The crystal-to-detector distance was set to 70, 100, 86 and 86 mm for

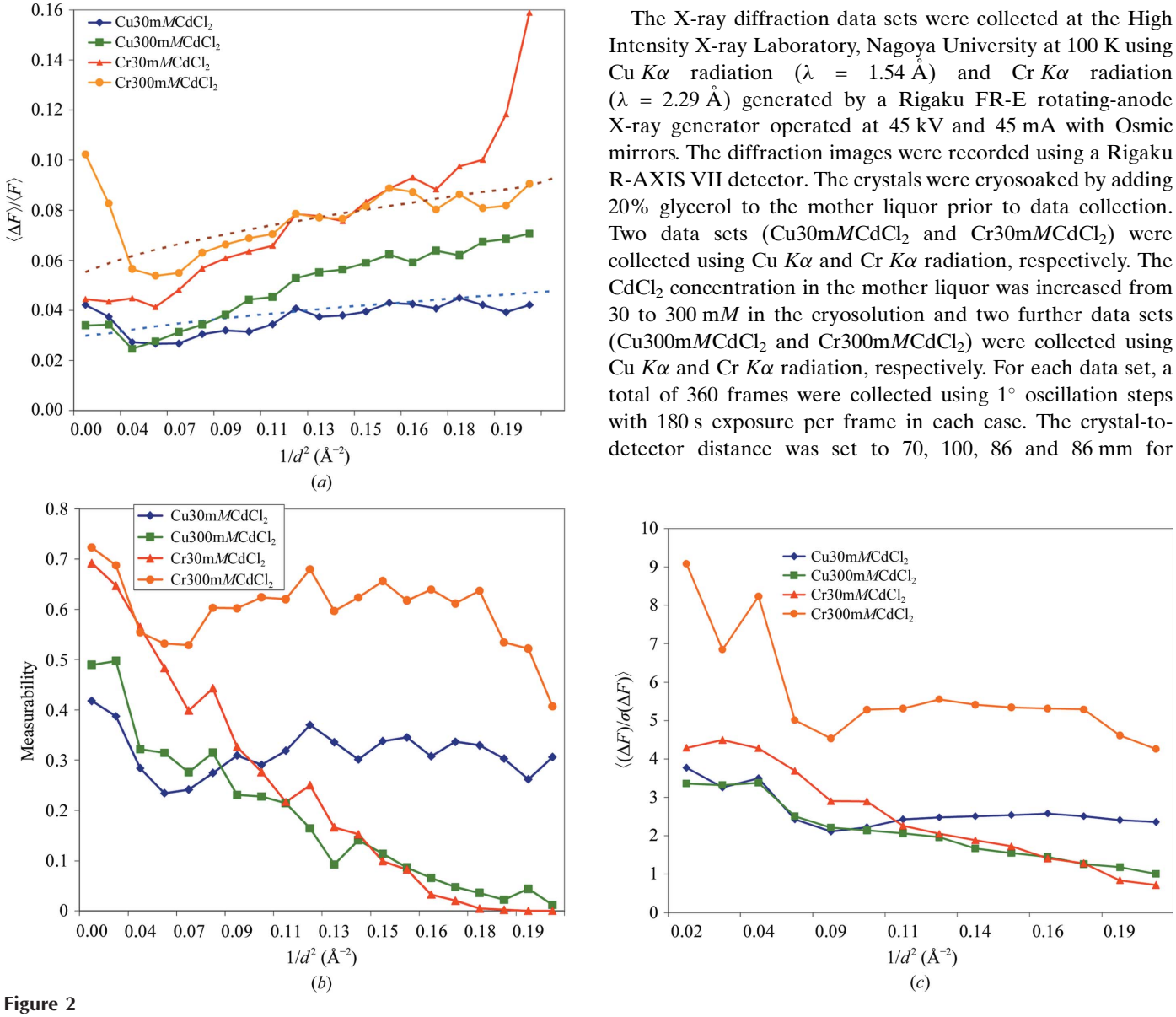

Figure 2
Plots of the Bijvoet ratio $\langle\Delta F\rangle/\langle F\rangle$, measurability $|\Delta I|/\sigma(\Delta I)$ and anomalous signal-to-noise ratio $\langle(\Delta F)/\sigma(\Delta F)\rangle$ as a function of resolution. (*a*) Experimentally measured Bijvoet ratio together with the theoretical curve. Dotted blue and red lines represent the theoretically expected Bijvoet ratio for Cu $K\alpha$ and Cr $K\alpha$ wavelengths for nine sulfurs and six cadmium ions, respectively, calculated using the modified Hendrickson formula. (*b*) Measurability curve calculated from experimental data using *phenix.reflection* statistics. It is defined as the fraction of Bijvoet-related intensity differences for which $|\Delta I|/\sigma(\Delta I) > 3.0$, $\min[I^+/\sigma(I^+), I^-/\sigma(I^-)] > 3.0$ holds. (*c*) Anomalous signal-to-noise ratio *versus* resolution calculated using *HKL2MAP*.

Table 1
Data-collection statistics.

Values in parentheses are for the last shell.

	Cu30m*M*CdCl$_2$	Cu300m*M*CdCl$_2$	Cr30m*M*CdCl$_2$	Cr300m*M*CdCl$_2$
Wavelength (Å)	1.54 [Cu $K\alpha$]	1.54 [Cu $K\alpha$]	2.29 [Cr $K\alpha$]	2.29 [Cr $K\alpha$]
Crystal-to-detector distance (mm)	70	100	86	86
Unit-cell parameters (Å)	a = 48.79, b = 56.18, c = 70.88	a = 47.27, b = 55.81, c = 70.23	a = 48.74, b = 56.10, c = 70.76	a = 48.52, b = 56.04, c = 70.92
Space group	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$	$P2_12_12_1$
Mosaicity (°)	0.8	1.2	1.1	0.6
Resolution (°)	110–1.50 (1.55–1.50)	90.0–1.90 (1.97–1.90)	90.0–2.20 (2.28–2.20)	110.0–2.20 (2.28–2.20)
Unique reflections	60045 (5964)	15226 (1494)	10270 (932)	10212 (914)
Multiplicity	7.0 (6.6)	12.8 (13.0)	11.9 (7.9)	6.8 (4.3)
Completeness (%)	99.5 (98.4)	99.8 (100)	99.2 (92.3)	99.3 (94.2)
$\langle I/\sigma(I)\rangle$	45.40 (9.64)	41.28 (6.13)	41.15 (5.11)	51.25 (33.81)
$R_{\rm merge}$	0.038 (0.251)	0.083 (0.275)	0.071 (0.216)	0.046 (0.077)
$\langle\Delta F/\sigma(\Delta F)\rangle$	2.5 (1.31)	2.1 (0.62)	2.8 (0.7)	5.9 (4.3)

Table 2
Summary of phasing and model building.

	Cu30m*M*CdCl$_2$	Cu300m*M*CdCl$_2$	Cr30m*M*CdCl$_2$	Cr300m*M*CdCl$_2$
AutoSol				
Bijvoet pairs	27805	12646	8395	8631
Mean anomalous signal ($\langle\Delta F\rangle/\langle F\rangle$)	0.04	0.05	0.06	0.07
Solvent fraction	0.39	0.36	0.38	0.38
Bayesian CC	56.9 ± 14.4	43.6 ± 19.8	54.3 ± 15.3	58.8 ± 14.1
FOM	0.53	0.43	0.51	0.58
Sites used for phasing	7	8	5	7
Refined sites	18	11	16	32
Located sulfur sites	9	3	9	9
AutoBuild				
Residues built	211	205	218	219
Side chains	200	197	194	198
$R_{\rm work}/R_{\rm free}$	0.21/0.23	0.22/0.26	0.19/0.25	0.19/0.22
Model–map CC	0.77	0.75	0.80	0.81

the Cu30m*M*CdCl$_2$, Cu300m*M*CdCl$_2$, Cr30m*M*CdCl$_2$ and Cr300m*M*CdCl$_2$ data sets, respectively. The data sets were processed and scaled with *DENZO* and *SCALEPACK* (Otwinowski, 1990). The anomalous pairs were merged separately as I^+ and I^-. The relevant data-collection statistics are summarized in Table 1.

2.3. Software packages used

The programs *HKL2MAP* (Pape & Schneider, 2004) and *PHENIX* (Adams *et al.*, 2002) were used to prepare and analyze the anomalous signal from the scaled intensity data. Substructure solution, phasing and model building were carried out for different data sets using *AutoSol* and *AutoBuild* in *PHENIX*. *Coot* (Emsley & Cowtan, 2004) was used to build missing residues and side chains into the autobuilt models. *REFMAC*5 (Murshudov *et al.*, 1997) and *phenix.refine* (Afonine *et al.*, 2005) were used for refinement. The theoretical Bijvoet ratio was calculated using the modified Hendrickson formula (Hendrickson & Teeter, 1981; Dauter *et al.*, 2002) and the measurability was calculated using *phenix.reflections_statistics* (Zwart *et al.*, 2005). The quality of all atomic models was assessed with *PROCHECK* (Laskowski *et al.*, 1993) and *MolProbity* (Chen *et al.*, 2010). The figures were generated using *Chimera* (Pettersen *et al.*, 2004) and *PyMOL* (http://www.pymol.org).

3. Results and discussion

3.1. Substructure determination of *saru*-actinidin

Initially, diffraction data were collected to 1.5 Å resolution (Cu30m*M*CdCl$_2$) using a native crystal and Cu $K\alpha$ radiation. *Saru*-actinidin contains 220 amino acids, including seven Cys and two Met residues. From the crystal parameters, it was clear that there was one molecule per asymmetric unit, with a Matthews coefficient of 1.96 Å^3 Da^{-1} and a solvent content of ∼37%. The mean anomalous signal ($\langle\Delta F\rangle/\langle F\rangle$) was 0.04 and was higher than the expected value (0.009) as calculated using the Hendrickson formula (Hendrickson & Teeter, 1981), $\langle\Delta F^{\pm}\rangle/\langle F\rangle_{\rm calc} = 2^{1/2} \times (f''_{\rm A}N_{\rm A}^{1/2})/[f_{\rm eff}(\theta)N_{\rm P}^{1/2}]$, where $N_{\rm A}$ and f'' are the number and imaginary scattering contribution of the anomalous scatterer, $N_{\rm P}$ is the total number of non-H protein atoms and $f_{\rm eff}(\theta)$ is the average number of electrons for protein atoms. Apart from nine sulfurs, the numbers and atom types of extra bound anomalous scatterers were initially unknown. The structure was determined by the sulfur-SAD technique using the *PHENIX* package (Adams *et al.*, 2002). Initially, seven sites were located and these sites were used for phasing. After phasing, a total of 18 sites were refined (Fig. 1*a* and Table 2). At this stage, a preliminary model was built automatically and the obtained model was fed into *AutoBuild* for iterative model building and refinement. A total of 211 resides were built and 190 side chains were placed with $R_{\rm work}$ and $R_{\rm free}$ values of 0.21 and 0.23, respectively. The amino-acid sequence information deduced from the cDNA was verified and modified based on the electron density at 1.5 Å resolution. Preliminary structure analysis was carried out to identify the unknown bound anomalous scatterers (there was a choice between chloride and cadmium) since the *saru*-actinidin crystals were grown in 30 m*M* CdCl$_2$ and cryosoaked with mother liquor supplemented with 20% glycerol. During substructure solution it was clear that the top six peaks belonged to unknown scatterers

Table 3
Summary of refinement statistics.

	Cu30m*M*CdCl$_2$	Cu300m*M*CdCl$_2$	Cr30m*M*CdCl$_2$	Cr300m*M*CdCl$_2$
Refinement				
Resolution (Å)	19.0–1.50	21.0–1.90	32.0–2.20	35.0–2.20
No. of reflections in work set/test set	29524/1992	13496/1501	9164/1018	9190/1022
R_{work}/R_{free} (%)	18.64/20.86	16.42/21.23	17.12/23.92	16.84/22.21
No. of residues/waters	220/288	220/216	220/153	220/141
Cd ions	5	8	5	7
Stereochemistry				
Bond lengths (Å)	0.001	0.007	0.010	0.009
Bond angles (°)	0.887	0.983	1.066	1.032
Ramachandran plot, residues in (%)				
Most favoured regions	89.5	87.3	88.6	87.8
Additionally allowed regions	10.5	12.7	11.4	12.2
Mean *B* factors (Å^2)				
Protein atoms	15.1	19.0	22.0	18.2
Cd ions	19.9	32.3	36.8	27.7
Water molecules	28.9	32.8	27.8	23.1
MolProbity				
MolProbity score	1.42	1.46	1.57	1.51
All-atom clash score	3.08	3.99	4.61	4.61
Bad rotomers (%)	1.7 (3/174)	2.3 (4/174)	2.3 (4/174)	2.3 (4/174)
Ramachandran favoured/outliers (%)	97.2/0.0	98.6/0.0	97.7/0.0	98.6/0.0

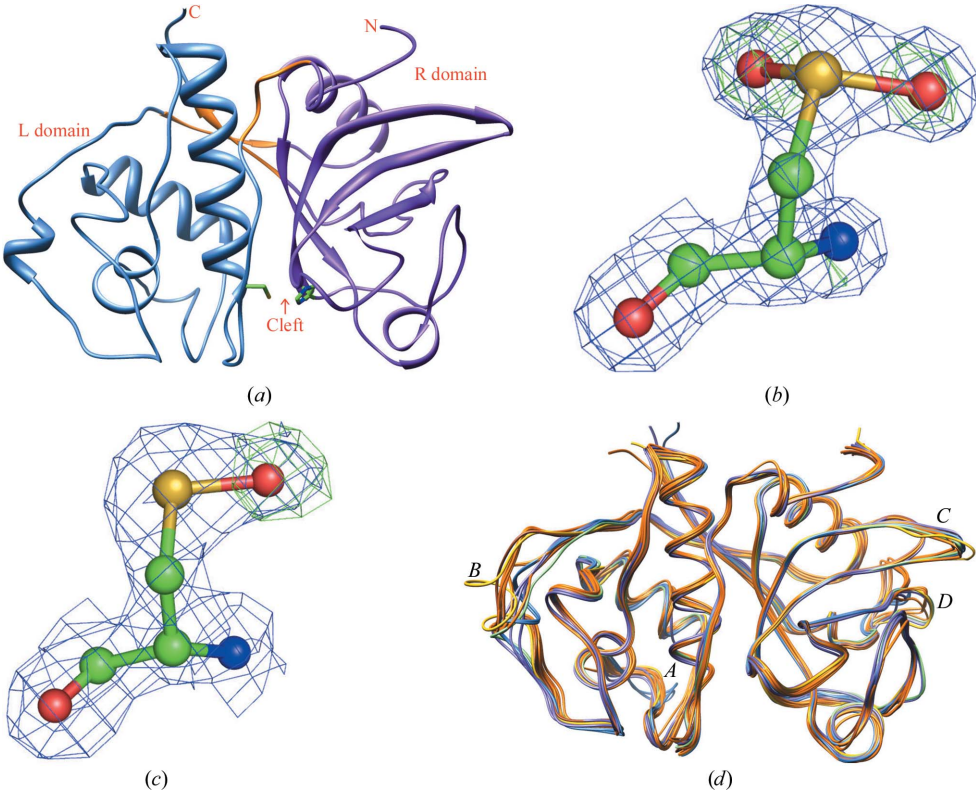

Figure 3
Overall structure of *saru*-actinidin. (*a*) Ribbon diagram of *saru*-actinidin. Two domains (L and R) and a hinge region are represented by blue, purple and orange colours, respectively. (*b*, *c*) The modified active-site Cys25 in the cysteine sulfinic acid and cysteine sulfenic forms is shown together with difference Fourier ($F_o - F_c$, green) and final ($2F_o - F_c$, blue) maps contoured at the 3.0σ and 1.5σ levels, respectively. (*d*) Superposition of the C$^\alpha$ atoms of *saru*-actinidin from *A. arguta* Planch. and *chins*-actinidin from *A. chinensis* (blue), cysteine proteases (papain, caricain, chymopain and glycyl endopeptidase) from *C. papaya* (orange), ervatamin A, B and C from *T. divaricata* (purple), ginger protease II from *Z. officinale* (green), mexicain from *J. mexicana* (brown) and EP-B2 from *H. vulgare* (yellow). Loop regions with deviations are indicated as follows: *A* (residues 59–62), *B* (residues 100–104), *C* (residues 172–175) and *D* (residues 199–201).

and most of these were on the protein surface. The interacting residues for these were Asp and Glu with distances between 2.0 and 2.5 Å. These parameters allowed us to assign the bound ions as cadmiums. In order to compare cadmium sites with sulfurs and cadmium/sulfur anomalous signals, we collected three further data sets (Cu300m*M*CdCl$_2$, Cr30m*M*CdCl$_2$ and Cr300m*M*CdCl$_2$) from a further three crystals. All of these structures were also determined using the cadmium/sulfur-SAD technique. Initially, *AutoSol* located eight, five and seven sites; after phasing, additional sites were located and refined (11, 16 and 32 sites for Cu300m*M*CdCl$_2$, Cr30m*M*CdCl$_2$ and Cr300m*M*CdCl$_2$, respectively). Only two sulfur substructures (Fig. 1*b*) were located in the Cu300m*M*CdCl$_2$ data set collected to 1.9 Å resolution with 5% anomalous signal (in the presence of eight bound cadmiums) and 12.8-fold overall redundancy. All S atoms could be located in the Cr30m*M*CdCl$_2$ and Cr300m*M*CdCl$_2$ data sets collected to 2.2 Å resolution using Cr $K\alpha$ radiation (Figs. 1*c* and 1*d*). The multiplicity required for the latter data was lower since the f'' values of sulfur and cadmium are doubled when using Cr $K\alpha$ radiation (λ = 2.29 Å). Substructure solution, phasing and model-building details are listed in Table 2. In all cases the figure of merit (FOM) was greater than 0.4 and the Bayesian CC was greater than 40. More than 95% of the residues and 90% of the side chains were placed automatically. Model completion, addition of waters and occupancy refinement of bound cadmium ions were carried out using *Coot* (Emsley & Cowtan, 2004) and *REFMAC*5 (Murshudov *et al.*, 1997). The final refinement parameters are listed in Table 3. Anomalous difference Fourier maps were computed and have been superimposed on the

(*a*)

(*b*)

Figure 4

(*a*) Structure-based sequence alignment of some plant cysteine proteases. The proteins are *saru*-actinidin from *A. arguta* Planch. and *chins*-actinidin from *A. chinensis* (PDB code 2act; Baker & Dodson, 1980), papain (1ppo, 9pap and 1ppn; Pickersgill *et al.*, 1991, 1992; Kamphuis *et al.*, 1984), caricain (1meg and 1pci; Katerelos *et al.*, 1996; Groves *et al.*, 1996), chymopain (1yal; Maes *et al.*, 1996) and glycyl endopeptidase (1gec; O'Hara *et al.*, 1995) from *C. papaya* (orange), ervatamin A (3bcn; Ghosh *et al.*, 2008), ervatamin B (1iwd; Biswas *et al.*, 2003) and ervatamin C (1o0e; Thakurta *et al.*, 2004) from *T. divaricata*, ginger protease II from *Z. officinale* (1cqd; Choi *et al.*, 1999), mexicain from *J. mexicana* (2bdz; Gavira *et al.*, 2007) and EP-B2 from *H. vulgare* (2fo5; Bethune *et al.*, 2006). Identical/well conserved, conserved and semi-conserved residues are marked with asterisks, semicolons and dots, respectively. Active-site residues (red), L-domain and R-domain interface interaction residues (green) and cadmium-interacting residues (blue) are highlighted. (*b*) Residue conservation in plant cysteine proteases is shown as a schematic diagram. Identical, conserved, semi-conserved and weakly conserved residues are rendered in red, pink, grey and blue, respectively.

Table 4
Substructure solution from *AutoSol* (*PHENIX*), with each peak annotated with the corresponding atom.

Cu30m*M*CdCl$_2$		Cu300m*M*CdCl$_2$		Cr300m*M*CdCl$_2$		Cr300m*M*CdCl$_2$	
Site	Atom	Site	Atom	Site	Atom	Site	Atom
1	Cd1	1	Cd1	1	Cd1	1	Cd1
2	Cd2†	2	Cd2	2	Cd2	2	Cd2
3	Cd3	3	Cd3	3	Cd3	3	Cd3
4	Cd4‡	4	Cd4‡	4	Cd4‡	4	Cd4‡
5	Cd5	5	Cd5	5	Cd4‡	5	Cd5
6	Cd4‡	6	Cd6	6	Cd5	6	Cd4‡
7	Met194	7	Cd7	7	Met194	7	Cd6
8	Met70	8	Cd8	8	Met70	8	Cd7
9	Cys156§	9	Met194	11	Cys156§	9	Met194
10	Cys206§	10	Cd4‡	12	Cys22¶	10	Met70
11	Cys22¶	11	Cys22–Cys65††	13	Cys206§	11	Cys22¶
12	Cys56‡‡			14	Cys25	14	Cys206§
13	Cys25§§			15	Cys65¶	15	Cys156§
14	Cys65¶			16	Cys56–Csy98††	16	Cys65¶
15	Cys98‡‡					17	Cys56‡‡
						18	Cys25§§
						20	Cys98‡‡

† Alternate positions were refined. ‡ Two closely located cadmium sites. § Disulfide linkage Cys156–Cys206. ¶ Disulfide linkage Cys22–Cys65. †† For this disulfide bond, a super-sulfur peak was located. ‡‡ Disulfide linkage Cys56–Cys98. §§ Free Cys.

Table 5
Sequence identity and r.m.s.d. of *saru*-actinidin with other known plant cysteine protease structures.

PDB code	Identity (%)	R.m.s.d. (Å)
2act	81.19	0.39
1ppn	49.53	0.81
1gec	51.39	0.79
1meg	51.85	0.85
1yal	53.21	0.85
1cqd	62.50	0.62
2bdz	51.42	0.80
2fo5	56.82	0.78
3bcn	58.37	0.75
9pap	49.06	0.79
1ppo	52.31	0.87
1o0e	60.10	0.70
1iwd	62.79	0.58
1pci	51.85	0.90

final atomic model to verify the assigned cadmium sites (Fig. 1).

3.2. Analysis of anomalous signals

The X-ray diffraction data were collected in a standard manner, in which the crystals were cryomounted in an arbitrary orientation and data sets were collected without using the inverse-beam technique. The expected anomalous signal ($\langle\Delta F^{\pm}\rangle/\langle F\rangle$) was calculated using the modified Hendrickson formula (Hendrickson & Teeter, 1981; Dauter *et al.*, 2002), $\langle\Delta F^{\pm}\rangle/\langle F\rangle_{\rm calc} = 2^{1/2}[(f''_{1A}N_{1A}^{1/2}) + (f''_{2A}N_{2A}^{1/2})]/[f_{\rm eff}(\theta)N_{P}^{1/2}]$, where f_1'' and f_2'' are the imaginary scattering contributions of sulfur and cadmium ions, respectively, and N_{1A}, N_{2A} and N_P are the number of sulfurs and cadmium ions and the total number of non-H atoms in the molecule, respectively. The average number of electrons for protein atoms [$f_{\rm eff}(\theta)$] is 6.7. The present *saru*-actinidin molecule contains nine sulfurs, six cadmium ions and 1690 non-H atoms. The expected value of the anomalous signal ($\langle\Delta F^{\pm}\rangle/\langle F\rangle$) is about 0.07 and 0.13 for the Cu $K\alpha$ and Cr $K\alpha$ wavelengths, respectively. The observed Bijvoet ratio, measurability and anomalous signal-to-noise ratio in each data set are illustrated in Fig. 2. Generally, the experimentally measured Bijvoet ratios agree with the theoretical values in the low-resolution bins and the deviation increases with increasing resolution, probably owing to a decrease in the measurement accuracy (Dauter *et al.*, 2002). In the cases of the Cu30m*M*CdCl$_2$, Cu300m*M*CdCl$_2$ and Cr300m*M*CdCl$_2$ data sets the observed Bijvoet ratio mostly agrees with the calculated values except in the low-resolution bins (Fig. 2*a*) and this indicates that there were additional heavy atoms (*i.e.* cadmiums) in these crystals. This is so in the cases of the Cu300m*M*CdCl$_2$ and Cr300m*M*CdCl$_2$ data sets, in which two or additional cadmium sites were located (Figs. 1*b* and 1*d*). There were no additional sites in the case of the Cu30m*M*CdCl$_2$ data set; however, the located sites had high occupancy values compared with other data sets. In the case of the Cu30m*M*CdCl$_2$ and Cr30m*M*CdCl$_2$ data sets the Bijvoet ratio is significantly higher in the high-resolution bins and this is a consequence of reduced accuracy in the estimation of the high-resolution reflection intensities [*i.e.* relatively low $\langle I/\sigma(I)\rangle$ for high-resolution reflections; Table 1]. The increment from 2.7 Å resolution onwards in the observed Bijvoet ratio is a consequence of poorer quality data in these data sets. In contrast to the anomalous signal, a sudden drop in measurability $|\Delta I|/\sigma(\Delta I)$ and anomalous signal-to-noise ratio from 2.7 Å resolution onwards was observed for the Cu300m*M*CdCl$_2$ and Cr30m*M*CdCl$_2$ data sets (Figs. 2*b* and 2*c*). The mean anomalous signal-to-noise ratio was significant only to <2.7 Å resolution for the Cu300m*M*CdCl$_2$ and Cr30m*M*CdCl$_2$ data sets (Fig. 2*c*), which suggested that 2.7 Å resolution is an appropriate cutoff choice for substructure solution.

3.3. Comparison of anomalous scatterers

In order to verify the located atom types in the substructure-solution peaks, we superimposed solution peaks on the corresponding final protein models. In all cases, the initially located sites are cadmium ions and were most likely to have originated from the crystallization solution (Fig. 1, Table 4). To verify that all anomalous scatterer sites had been located, an anomalous difference map was calculated using phases from the refined model, but no strong peaks were found. Of the top six peaks (including two sites located in proximity to each other; CD4 in Fig. 1), the first three sites showed high occu-

pancy. The occupancies (0.5–0.6) of other cadmium sites are in the same range as that of the Met194 S atom. In substructure solutions, only the Met194 S atom and Cys22–Cys65 (single peak, super-sulfur; Table 4 and Fig. 1*b*) were located for the Cu300m*M*$CdCl_2$ data set, whereas in the other three data sets all sulfurs (Met and Cys) could be located (Table 4, Fig. 1). The cadmium ions and Met and free Cys sulfurs showed spherical densities, whereas the disulfide bond showed elliptical density (Fig. 1).

3.4. Description of the protease structure

Saru-actinidin has a papain-like fold and consists of two domains (L and R). The L domain is mostly α-helical, while the R domain is built around a twisted β-sheet (Fig. 3). The L domain consists of residues 14–111 and 217–220 and the R domain contains residues 1–9 and 116–212. There are three portions of the polypeptide chain (residues 10–13, 112–115 and 213–216) that act as linkers between the two domains and both the amino- and carboxyl-terminal ends of the polypeptide cross over to the other domain. The catalytic dyad residues Cys25 (L domain) and His162 (R domain) are situated on either side of the inter-domain cleft. The active-site Cys25 exists in the cysteine sulfenic acid form in the Cu300m*M*$CdCl_2$ and Cr300m*M*$CdCl_2$ data sets, but in the cysteine sulfinic acid form in the Cu30m*M*$CdCl_2$ data set (Figs. 3*b* and 3*c*).

3.5. Sequence and structural comparison with known plant cysteine proteases

The structure of *saru*-actinidin has been superimposed on other plant cysteine protease structures such as those of *chins*-actinidin, various cysteine proteases (papain, caricain, chymopain and glycyl endopeptidase) from *C. papaya*, ervatamin A, B and C from *T. divaricata*, GPII from *Z. officinale*, mexicain from *J. mexicana* and EP-B2 from *H. vulgare*. The low r.m.s.d. values for C^{α} atoms (<0.9 Å; Table 5) indicate a high similarity in overall conformation. Deviations are mainly observed in four surface-loop regions, which are designated regions *A*–*D* in Fig. 3(*d*). The *saru*-actinidin primary sequence proposed here compares well with the amino-acid sequence of plant cysteine proteases, as illustrated in Fig. 4. The sequence identity of *saru*-actinidin is 81% to *chins*-actinidin, 63% to ginger protease II (GPII), 59–63% to ervatamin A, B and C, and 50–53% to various cysteine proteases from *C. papaya* and mexicain. In comparison to other members of the plant cysteine protease family, there is a two-residue insertion between α-helix 1 and α-helix 2 in *saru*-actinidin and *chins*-

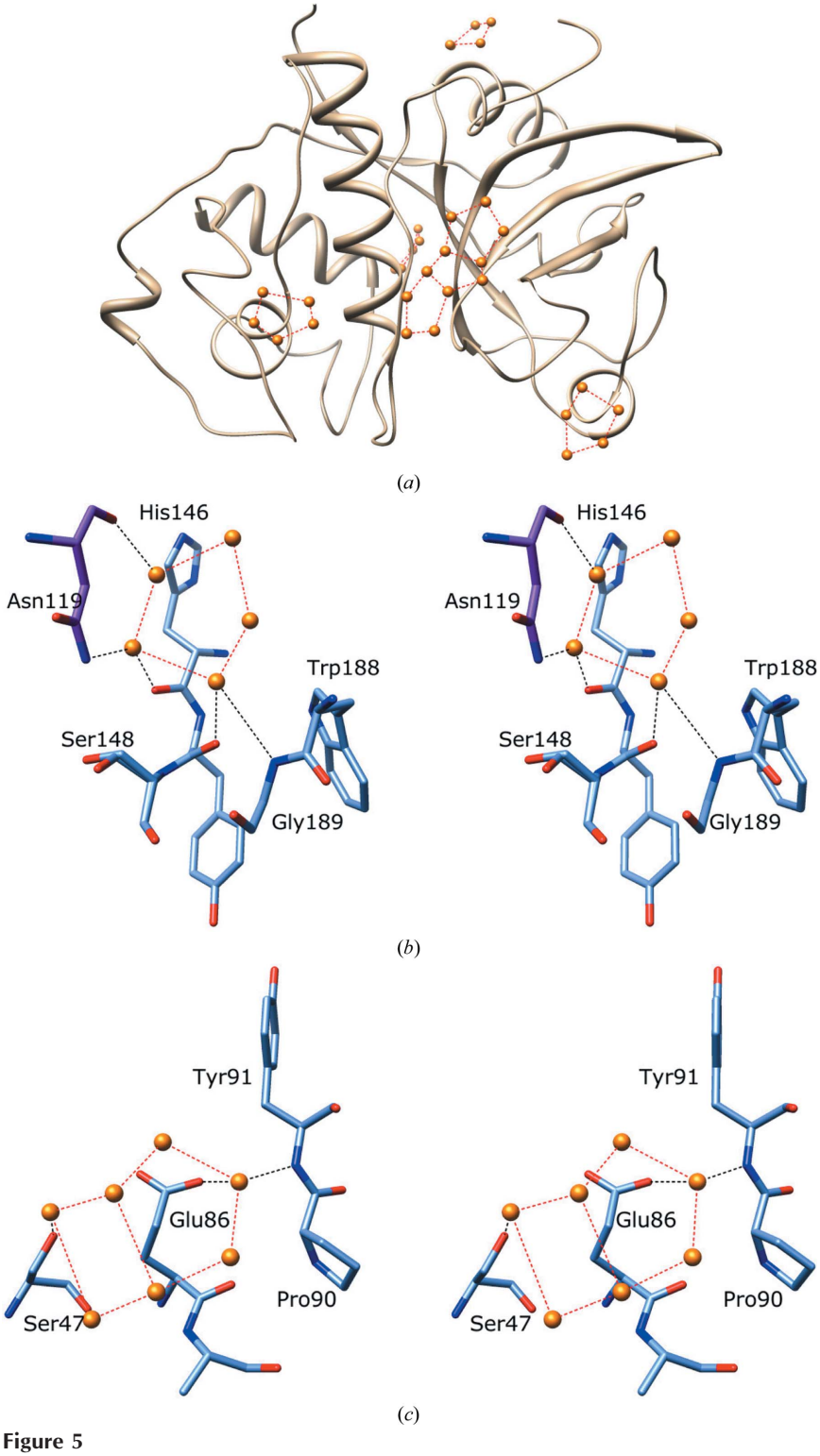

Figure 5
Polygonal water structures. (*a*) View of polygonal arrangements of hydration water molecules around the protein surface. Close-up stereoviews showing (*b*) a pentagonal ring, (*c*) fused tetragonal and pentagonal rings.

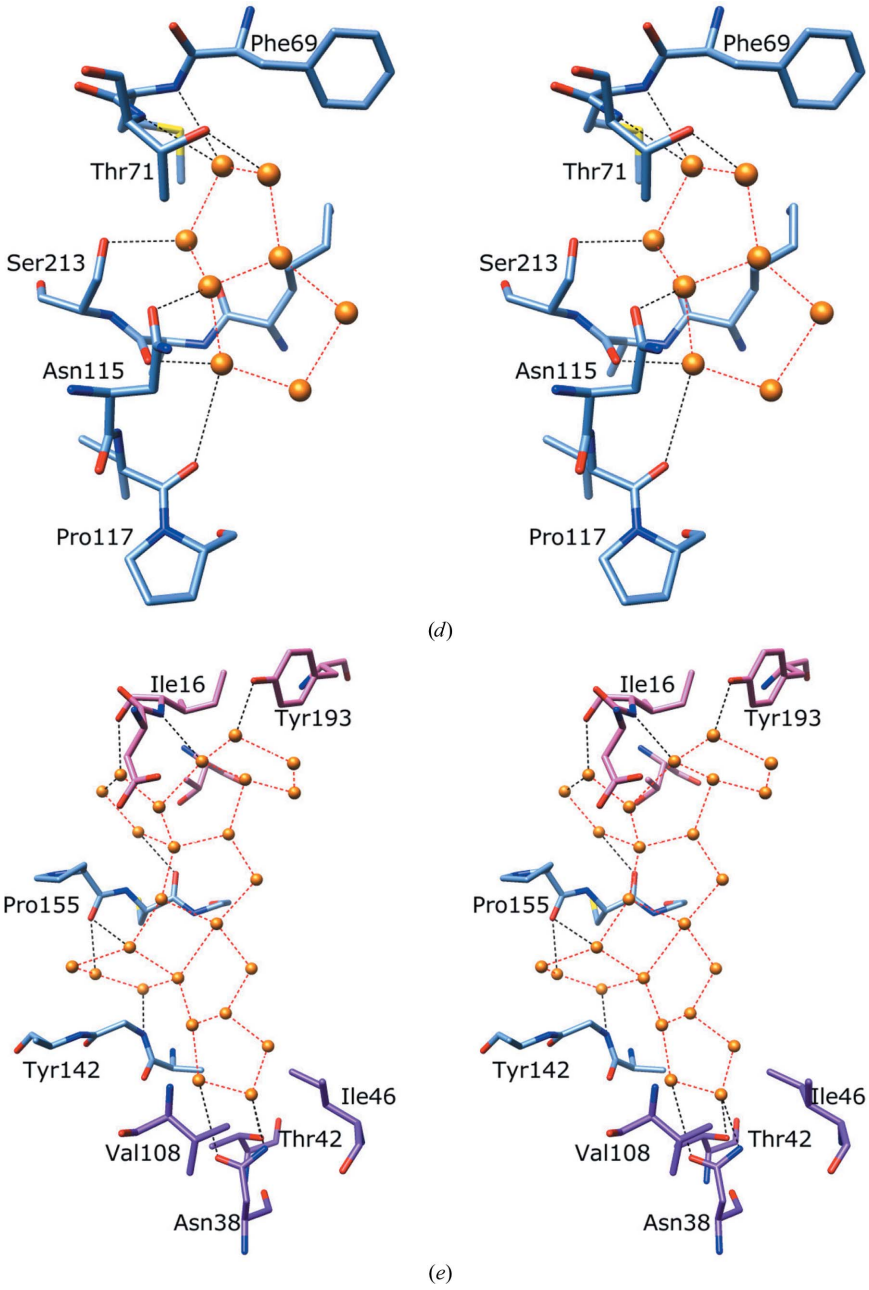

Figure 5 (continued)
Polygonal water structures. (*d*) Two fused pentagonal rings and (*e*) fused tetragonal and pentagonal rings forming a polygonal structure. Tetragonal and pentagonal rings are found suspended above the hydrophobic portions of side chains and are anchored by hydrogen-bonding interactions with the surrounding polar atoms and water molecules (not shown). The reference and neighbouring molecules are shown in blue or pink/violet, respectively.

actinidin (Figs. 2*d* and 3*a*). This insertion does not affect the overall fold and is not involved in active-site formation. However, it is located near the bottom of the active-site cavity.

3.6. Pentagonal arrays of water molecules

Water molecules play an important role in maintaining the structural stability of proteins and can control protein folding, structure and overall activity (Chaplin, 2001, 2006). Pentagonal rings of water molecules have previously been identified on the surface of proteins (Teeter, 1984; Kumaraswamy *et al.*, 1996; Nakasako, 2004; Narayana, 2006; Britton *et al.*, 2006). Water pentagons have also been found in the crystal structures of DNA fragments (Neidle *et al.*, 1980). It is interesting to note that in the high-resolution *saru*-actinidin structure water molecules form a network of hydrogen bonds that resemble fused pentagonal rings (Fig. 5). A total of 13 such rings have been identified at six distinct places on the protease surface and most of these are constructed around a hydrophobic residue (Figs. 5*b*–5*e*). The average interaction distance between these arranged waters is ∼2.8 Å and the mean *B* value is ∼26 Å^2 (which is lower than the value of ∼29 Å^2 for all waters).

3.7. Cadmium-binding sites

The residues involved in cadmium coordination are generally not conserved amongst plant cysteine proteases, indicating that the analyzed enzymes may not be cadmium-dependent proteins (Fig. 4*a*). Analysis of the four data sets that we have collected indicates that five cadmium sites are present in all data sets (Fig. 6*a* and Table 4). The average coordination distance between cadmium ions and protein atoms and water molecules is ∼2.4 Å and this value is comparable with previously reported values (Naismith *et al.*, 1993; Dokmanić *et al.*, 2008) and those for small molecules from the Cambridge Structural Database (CSD; Allen & Kennard, 1993). Like zinc, cadmium possesses a filled d^{10} orbital shell and is therefore able to accommodate different coordination geometries of similar energy (Pye *et al.*, 2006; Dokmanić *et al.*, 2008). The active-site-bound CD1 (cadmium site 1) displays distorted octahedral geometry and its coordination ligands are Cys25, His162, #Asn220 and waters W1 and W2 (where # indicates a symmetry-related molecule). The sulfenic/sulfinic acid O atom, water molecules W1 and W2 and #OXT of the C-terminal Asn220 form an octahedral base plane, whereas His162 $N^{\delta1}$ and #Asn220 O are located at the vertices of the octahedron (Figs. 6*b* and 6*c*). The distance between CD1 and the active-site S^{γ} atom of Cys25 is ∼3.2 and 3.5 Å for the Cu30m*M*$CdCl_2$ and Cu300m*M*$CdCl_2$ data sets, respectively. The remaining cadmium sites are located on the protein surface, where they are involved in less elaborate inter-

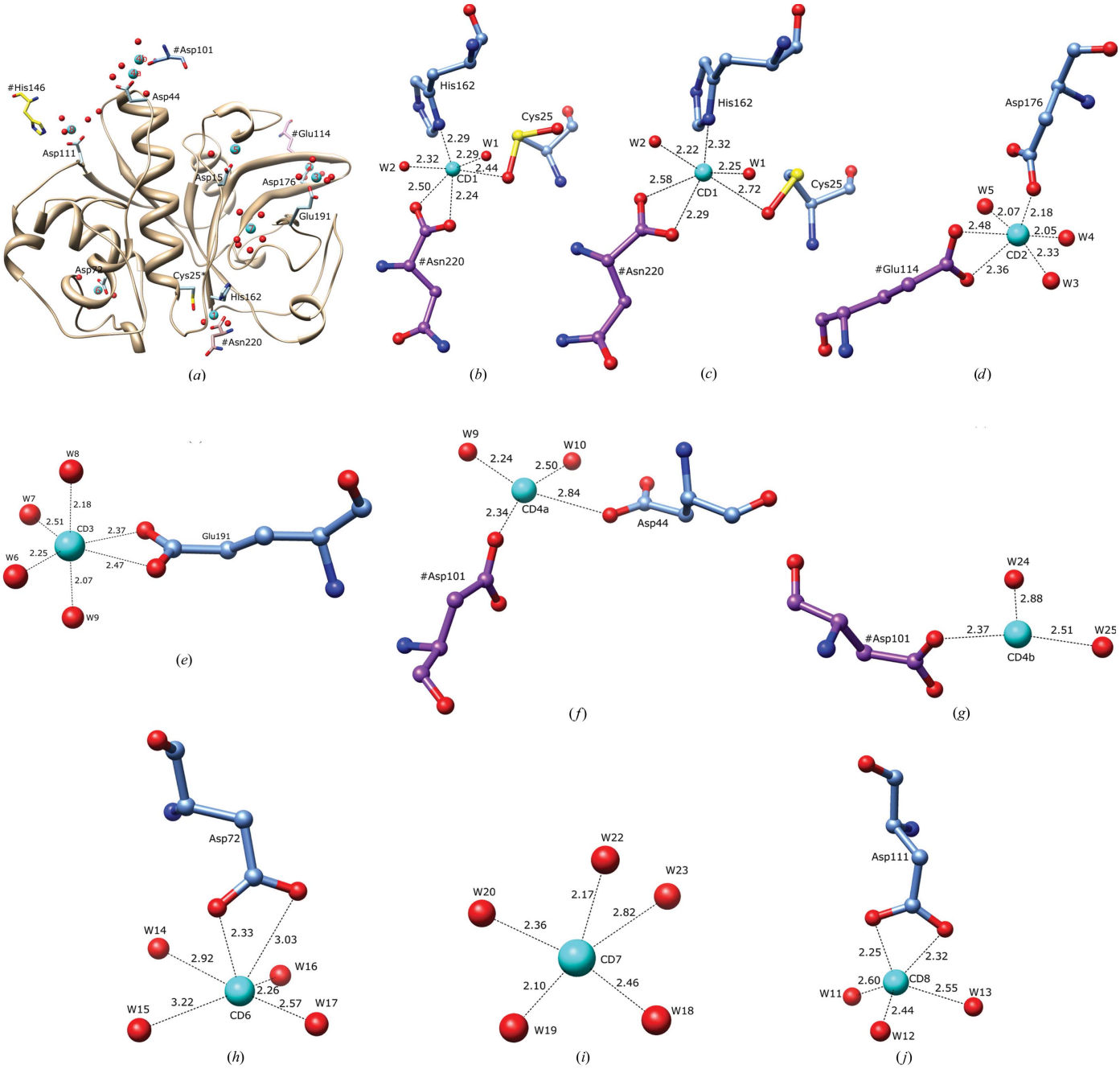

Figure 6
Cadmium sites. (*a*) Cd ions in the protein involved in various coordination geometries such as (*b*, *c*) distorted octahedral, (*d*) distorted octahedral or square pyramidal, (*e*) trigonal bipyramidal, (*f*) square planar, (*g*) trigonal, (*h*, *i*) distorted square pyramidal and (*j*) tetrahedral. The reference and symmetry-related molecules are shown in blue and purple, respectively. Bound cadmiums and waters are shown as cyan and red spheres, respectively. The marked coordination distances are in Å.

molecular bridging and generally display a variety of coordinations (Figs. 6*d*–6*j*). The carboxylate O atoms of Glu114, Glu191 and Asp111 act as symmetrical bidentate ligands for CD2, CD3 and CD8, respectively.

For sulfur-SAD phasing using a Cu $K\alpha$ radiation source, the redundancy (>∼20) and the resolution (<2.0 Å) can be very important factors in substructure solution. However, additional anomalous signal enhancers such as solvent Cl, P and naturally bound metals can contribute and allow easier substructure solution. This can be seen in the Cu300m*M*$CdCl_2$ data we collected, in which two sulfurs were located together with cadmium sites. In the data sets collected using Cr $K\alpha$ radiation all the S atoms were located, revealing that access to an in-house source of chromium-generated X-ray radiation will make it easier to solve protein structures using bound lighter atoms and metals. We propose that a dual source of

X-rays at in-house facilities coupled with the spiking of crystallization screens with atoms that are capable of anomalous scattering will be of considerable benefit.

MY was supported by FP7 EU grant PREMALSTRUCT and AS was supported by a grant from the Department of Biotechnology, Government of India. Work in the Sharma laboratory is supported by FP7 EU grants MALSIG, PREMALSTRUCT and MEPHITIS. We thank Mr Tatsuo Hikage of the High Intensity X-ray Laboratory, Nagoya University for his kind help with data collection.

References

Adams, P. D., Grosse-Kunstleve, R. W., Hung, L.-W., Ioerger, T. R., McCoy, A. J., Moriarty, N. W., Read, R. J., Sacchettini, J. C., Sauter, N. K. & Terwilliger, T. C. (2002). *Acta Cryst.* D**58**, 1948–1954.
Adams, M. & Jia, Z. (2005). *J. Biol. Chem.* **280**, 28675–28682.
Afonine, P. V., Grosse-Kunstleve, R. W. & Adams, P. D. (2005). *CCP4 Newsl. Protein Crystallogr.* **42**, contribution 8.
Allen, F. H. & Kennard, O. (1993). *Chem. Des. Autom. News*, **8**, 31–37.
Baker, E. N. & Dodson, E. J. (1980). *Acta Cryst.* A**36**, 559–572.
Bethune, M. T., Strop, P., Tang, Y., Sollid, L. M. & Khosla, C. (2006). *Chem. Biol.* **13**, 637–647.
Biswas, S., Chakrabarti, C., Kundu, S., Jagannadham, M. V. & Dattagupta, J. K. (2003). *Proteins*, **51**, 489–497.
Boraston, A. B., Healey, M., Klassen, J., Ficko-Blean, E., Lammerts van Bueren, A. & Law, V. (2006). *J. Biol. Chem.* **281**, 587–598.
Britton, K. L., Baker, P. J., Fisher, M., Ruzheinikov, S., Gilmour, D. J., Bonete, M. J., Ferrer, J., Pire, C., Esclapez, J. & Rice, D. W. (2006). *Proc. Natl Acad. Sci. USA*, **103**, 4846–4851.
Bueren, A. L. van, Ghinet, M. G., Gregg, K., Flury, A., Brzezinski, R. & Boraston, A. B. (2009). *J. Mol. Biol.* **385**, 131–139.
Chaplin, M. (2006). *Nature Rev. Mol. Cell Biol.* **7**, 861–866.
Chaplin, M. F. (2001). *Biochem. Mol. Biol. Educ.* **29**, 54–59.
Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Cryst.* D**66**, 12–21.
Choi, K. H., Laursen, R. A. & Allen, K. N. (1999). *Biochemistry*, **38**, 11624−11633.
Dauter, Z., Dauter, M. & Dodson, E. J. (2002). *Acta Cryst.* D**58**, 494–506.
Dokmanić, I., Šikić, M. & Tomić, S. (2008). *Acta Cryst.* D**64**, 257–263.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Gangelhoff, T. A., Mungalachetty, P. S., Nix, J. C. & Churchill, M. E. A. (2009). *Nucleic Acids Res.* **37**, 3153–3164.
Gavira, J. A., González-Ramírez, L. A., Oliver-Salvador, M. C., Soriano-García, M. & García-Ruiz, J. M. (2007). *Acta Cryst.* D**63**, 555–563.
Ghosh, R., Chakraborty, S., Chakrabarti, C., Dattagupta, J. K. & Biswas, S. (2008). *FEBS J.* **275**, 421–434.
Grimshaw, J. P. A., Stirnimann, C. U., Brozzo, M. S., Malojcic, G., Grütter, M. G., Capirani, G. & Glockshuber, R. (2008). *J. Mol. Biol.* **380**, 667–680.
Groves, M. R., Taylor, M. A., Scott, M., Cummings, N. J., Pickersgill, R. W. & Jenkins, J. (1996). *Structure*, **4**, 1193–1203.
Harrop, S. J., Helliwell, J. R., Wan, T. C. M., Kalb (Gilboa), A. J., Tong, L. & Yariv, J. (1996). *Acta Cryst.* D**52**, 143–155.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
Kajander, T., Cortajarena, A. L., Mochrie, S. & Regan, L. (2007). *Acta Cryst.* D**63**, 800–811.
Kamphuis, I. G., Kalk, K. H., Swarte, M. B. & Drenth, J. (1984). *J. Mol. Biol.* **179**, 233–256.
Katerelos, N. A., Taylor, M. A., Scott, M., Goodenough, P. W. & Pickersgill, R. W. (1996). *FEBS Lett.* **392**, 35–39.
Kaus-Drobek, M., Czapinska, H., Sokolowska, M., Tamulaitis, G., Szczepanowski, R. H., Urbanke, C., Siksnys, V. & Bochtler, M. (2007). *Nucleic Acids Res.* **35**, 2035–2046.
Kumaraswamy, V. S., Lindley, P. F., Slingsby, C. & Glover, I. D. (1996). *Acta Cryst.* D**52**, 611–622.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Liu, Y., Wu, J., Song, J., Sivaraman, J. & Hew, C. L. (2006). *J. Virol.* **80**, 10419–10427.
Maes, D., Bouckaert, J., Poortmans, F., Wyns, L. & Looze, Y. (1996). *Biochemistry*, **35**, 16292–16298.
Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Cryst.* D**53**, 240–255.
Naismith, J. H., Habash, J., Harrop, S. J., Helliwell, J. R., Hunter, W. N., Wan, T. C. M., Weisgerber, S., Kalb (Gilboa), A. J. & Yariv, J. (1993). *Acta Cryst.* D**49**, 561–571.
Nakasako, M. (2004). *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **359**, 1191–1206.
Narayana, N. (2006). *Acta Cryst.* D**62**, 695–706.
Neidle, S., Berman, H. M. & Shieh, H. S. (1980). *Nature (London)*, **288**, 129–133.
Nickitenko, A. V., Trakhanov, S. & Quiocho, F. A. (1995). *Biochemistry*, **34**, 16585–19595.
O'Hara, B. P., Hemmings, A. M., Buttle, D. J. & Pearl, L. H. (1995). *Biochemistry*, **34**, 13190–13195.
Otwinowski, Z. (1990). *DENZO Data Processing Package.* Yale University, New Haven, Connecticut, USA.
Pape, T. & Schneider, T. R. (2004). *J. Appl. Cryst.* **37**, 843–844.
Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C. & Ferrin, T. E. (2004). *J. Comput. Chem.* **25**, 1605-1612.
Pickersgill, R. W., Harris, G. W. & Garman, E. (1992). *Acta Cryst.* B**48**, 59–67.
Pickersgill, R. W., Rizkallah, P., Harris, G. W. & Goodenough, P. W. (1991). *Acta Cryst.* B**47**, 766–771.
Pye, C. C., Tomney, M. R. & Rudolph, W. W. (2006). *Can. J. Anal. Sci. Spectrosc.* **51**, 141–146.
Reddie, K. G. & Carroll, K. S. (2008). *Curr. Opin. Chem. Biol.* **12**, 746–754.
Stegmann, C. M., Lührmann, R. & Wahl, M. C. (2010). *PLoS ONE*, **5**, e10013.
Teeter, M. M. (1984). *Proc. Natl Acad. Sci. USA*, **81**, 6014–6018.
Thakurta, P. G., Biswas, S., Chakrabarti, C., Sundd, M., Jagannadham, M. V. & Dattagupta, J. K. (2004). *Biochemistry*, **43**, 1532–1540.
Trakhanov, S., Kreimer, D. I., Parkin, S., Ames, G. F. & Rupp, B. (1998). *Protein Sci.* **7**, 600–604.
Trakhanov, S. & Quiocho, F. A. (1995). *Protein Sci.* **4**, 1914–1919.
Xu, Y., Feng, L., Jeffery, P. D., Shi, Y. & Morel, F. M. M. (2008). *Nature (London)*, **452**, 56–62.
Yao, N., Trakhanov, S. & Quiocho, F. A. (1994). *Biochemistry*, **33**, 4769–4779.
Zwart, P. H., Grosse-Kunstleve, R. W. & Adams, P. D. (2005). *CCP4 Newsl. Protein Crystallogr.* **43**, contribution 7.