Acta Crystallographica Section D
Biological Crystallography
ISSN 0907-4449

High-resolution structure of a papaya plant-defence barwin-like protein solved by in-house sulfur-SAD phasing

Joëlle Huet,[a] Emmanuel Jean Teinkela Mbosso,[a,b] Sameh Soror,[c,d] Franck Meyer,[a] Yvan Looze,[a] René Wintjens[a,e]‡ and Alexandre Wohlkönig[c]‡

[a]Laboratoire des Biopolymères et des Nanomatériaux Supramoléculaires (CP206/04), Faculté de Pharmacie, Université Libre de Bruxelles, B-1050 Brussels, Belgium, [b]Département des Sciences Biologiques, Faculté de Médecine et des Sciences Pharmaceutiques (FMSP), Université de Douala, BP 2701, Douala, Cameroon, [c]Structural Biology Brussels, Vrije Universiteit Brussel, B-1050 Brussels, Belgium, [d]Helwan Structural Biology Research, Faculty of Pharmacy, Helwan University, Cairo 11795, Egypt, and [e]Unité de Biologie Structurale (CP300), Département de Biologie Moléculaire, Université Libre de Bruxelles, B-6041 Gosselies, Belgium

‡ RW and AW contributed equally.

Received 22 March 2013
Accepted 29 June 2013

PDB References: carwin, 4jp6; 4jp7

The first crystal structure of a barwin-like protein, named carwin, has been determined at high resolution by single-wavelength anomalous diffraction (SAD) phasing using the six intrinsic S atoms present in the protein. The barwin-like protein was purified from *Carica papaya* latex and crystallized in the orthorhombic space group $P2_12_12_1$. Using in-house Cu $K\alpha$ X-ray radiation, 16 cumulative diffraction data sets were acquired to increase the signal-to-noise level and thereby the anomalous scattering signal. A sequence-database search on the papaya genome identified two carwin isoforms of 122 residues in length, both containing six S atoms that yield an estimated Bijvoet ratio of 0.93% at 1.54 Å wavelength. A systematic analysis of data quality and redundancy was performed to assess the capacity to locate the S atoms and to phase the data. It was observed that the crystal decay was low during data collection and that successful S-SAD phasing could be obtained with a relatively low data multiplicity of about 7. Using a synchrotron source, high-resolution data (1 Å) were collected from two different crystal forms of the papaya latex carwin. The refined structures showed a central β-barrel of six strands surrounded by several α-helices and loops. The β-barrel of carwin appears to be a common structural module that is shared within several other unrelated proteins. Finally, the possible biological function of the protein is discussed.

1. Introduction

A common characteristic of laticiferous plants consists of a particular system of tube-like cells containing a white fluid referred to as latex. This fluid consists of an emulsion of a mixture of organic compounds (small molecules and proteins) combining several physiological functions in the plant such as storage, conduction, defence and secretion. The latex of laticiferous plants has traditionally been used to produce several materials, and consequently many laticiferous plant species are cultivated around the world.

In the course of the proteomic study of *Carica papaya* latex, a product with great economic importance that is used in brewing and wine making and in the textile and tanning industries, we identified a new protein homologous to barwin, a 125-amino-acid protein isolated from aqueous extracts of barley seeds (Svensson *et al.*, 1992). The *Carica* protein was here denoted carwin by analogy to the barwin protein. The structure of barwin has previously been solved by NMR spectroscopy (Ludvigsen & Poulsen, 1992*a*,*b*). It contains three disulfide bonds, a four-stranded antiparallel β-sheet, two parallel β-sheets packed antiparallel to each other and four short α-helices.

Table 1
Summary of data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

Data set	700-image	1450-image
X-ray source	SLS	SLS
Detector	PILATUS 2M-F	PILATUS 2M-F
Crystal-to-detector distance (mm)	112.49	120.0
Wavelength (Å)	0.80	0.80
No. of images	700	1450
Oscillation width (°)	0.15	0.25
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters		
a (Å)	54.91	27.69
b (Å)	56.96	54.97
c (Å)	74.20	73.51
Resolution range (Å)	50.00–1.05 (1.11–1.05)	50.00–1.00 (1.06–1.00)
Completeness (%)	98.4 (97.1)	99.9 (99.2)
No. of unique reflections	107418 (16975)	61603 (9747)
Multiplicity	3.72 (3.47)	12.28 (11.69)
$R_{\rm meas}$† (%)	4.1 (54.4)	4.3 (15.4)
$\langle I/\sigma(I)\rangle$‡	16.74 (2.64)	36.79 (13.01)
Refinement data statistics		
Resolution limits (Å)	39.53–1.05	36.75–1.00
Reflections used	101999	58439
No. of refined atoms		
Protein atoms	1924	962
Water atoms	214	160
Other atoms	6	0
Final $R_{\rm work}$§/$R_{\rm free}$¶ (%)	14.2/16.4	15.3/17.9
R.m.s.d., bond lengths (Å)	0.028	0.025
R.m.s.d., angles (°)	2.152	2.105
Overall mean B factor (Å^2)	14.08	11.76
Ramachandran plot analysis††		
Most favoured region (%)	93.3	92.3
Disallowed regions (%)	0.5	0.0
MolProbity score	1.01	0.91

† $R_{\rm meas} = \sum_{hkl}\{N(hkl)/[N(hkl)-1]\}^{1/2}\sum_i |I_i(hkl) - \langle I(hkl)\rangle|/\sum_{hkl}\sum_i I_i(hkl)$ is the multiplicity-independent $R_{\rm merge}$, where $N(hkl)$ is the multiplicity, $I_i(hkl)$ is the intensity of the ith observation of reflection hkl and $\langle I(hkl)\rangle$ is the mean intensity of reflection hkl. ‡ Mean $I/\sigma(I)$ of unique reflections. § $R_{\rm work} = \sum_{hkl}\big||F_{\rm obs}| - |F_{\rm calc}|\big|/\sum_{hkl}|F_{\rm obs}|$, where $F_{\rm calc}$ and $F_{\rm obs}$ are the calculated and the observed structure-factor amplitudes, respectively. ¶ $R_{\rm free} = \sum_{hkl}\big||F_{\rm obs}| - |F_{\rm calc}|\big|/\sum_{hkl}|F_{\rm obs}|$, where all reflections belong to a test set of randomly selected data. †† Ramachandran analysis was performed with *PROCHECK*.

Similarities were found between the sequences of barwin and other plant proteins, such as, for example, the C-terminal regions of the products of two wound-induced genes (*win1* and *win2*) from potato (Stanford *et al.*, 1989) and the C-terminal product of the pro-hevein gene from rubber trees (Broekaert *et al.*, 1990). Barwin and other related wound-induced plant proteins are classified into pathogenesis-related protein family 4 (PR-4 proteins). Since the first two PR-4 proteins, WIN1 and WIN2 from potato (Stanford *et al.*, 1989), were described, many other members have been found in different plants, such as tomato (Linthorst *et al.*, 1991), tobacco (Jamet & Fritig, 1986), *Arabidopsis* (Potter *et al.*, 1993), wheat (Caruso *et al.*, 1999), pepper (Lee *et al.*, 2001), cabbage (Ryang *et al.*, 2002), maize (Chevalier *et al.*, 1995) and, of course, barley (Gregersen *et al.*, 1997). By similarity to the classification of the plant chitinase family, it has been proposed that PR-4 proteins should be segregated into two subclasses according to the presence (class I) or the absence (class II) of a conserved N-terminal cysteine-rich domain homologous to hevein, a small chitin-binding structural module isolated from the latex of the rubber plant *Hevea brasiliensis* (Van Parijs *et al.*, 1991). Finally, although PR-4 proteins have been identified in several plants, they have been poorly described to date and it remains difficult to define a biological function that is common to all of them.

In the present study, we isolated, purified and determined the crystal structure of a barwin-like protein from papaya latex, here referred to as carwin, at high resolution (1.0 Å) by single-wavelength anomalous diffraction (SAD) phasing utilizing the six intrinsic S atoms in the native protein. The use of SAD phasing was necessary as all of our attempts to solve the phase by the molecular-replacement method using the NMR structure of the barwin protein as a search model were unsuccessful.

Since the first protein structure determined by sulfur-SAD (S-SAD) was reported in 1981 (Hendrickson & Teeter, 1981), few structures have been solved using solely their intrinsic S atoms as a source of anomalous signal (Yang & Pflugrath, 2001; Liu *et al.*, 2013). Successful SAD phasing often relies on the additional presence of other weak anomalous scatterers in addition to sulfur (Dauter, 1999; Debreczeni *et al.*, 2003; Goulet *et al.*, 2010). At the present moment, about 60 structures have been solved by SAD phasing using anomalous signals from atoms not heavier than atomic number 20 (Liu *et al.*, 2013), a number that is insignificant in comparison with the tens of thousands of structures solved by other experimental phasing methods. This is unfortunate given the advantage of the S-SAD phasing method, which is based on the natural presence of S atoms from methionines and cysteines: S-SAD phasing does not require additional crystals of the protein of interest. However, the method requires a higher wavelength for the incident X-ray radiation in order to maximize the anomalous contribution f'' of S atoms; the main drawback of the method lies in the low strength of the anomalous scattering signal from S atoms. The use of S-SAD phasing as a routine method in macromolecular crystallography using modern laboratory equipment offers great potential, but unfortunately remains underexploited (Doutch *et al.*, 2012).

2. Materials and methods

2.1. Crystallization and data collections

Details of the purification procedure are given in the Supplementary Material[1]. The protein solution consisted of 12.5 mg ml^{-1} carwin in water. Crystallization screening was carried out in hanging drops using the vapour-diffusion method with commercially available sparse-matrix screens from Hampton Research and Sigma–Aldrich. Drops consisting of 1 µl protein solution and 1 µl crystallization solution were equilibrated against 500 µl crystallization solution at 293 K. Crystals of carwin appeared in 1 d in a crystallization solution consisting of 2 *M* ammonium sulfate. The

[1] Supplementary material has been deposited in the IUCr electronic archive (Reference: TZ5031). Services for accessing this material are described at the back of the journal.

Table 2
S-SAD data-collection and phasing statistics.

Common crystal and instrumental parameters: wavelength, 1.5418 Å; oscillation width, 0.5°; space group, $P2_12_12_1$; unit-cell parameters, a = 27.54, b = 54.86, c = 73.40 Å; resolution range, 20.0–1.81 (1.92–1.81) Å.

	Set *A*	Set *B*	Set *C*	Set *D*	Set *E*	Set *F*
No. of frames	360	720	1080	1340	2160	2880
Total reflections	67786 (5886)	135416 (11642)	202893 (17468)	270287 (23191)	405184 (34746)	540297 (46444)
Unique reflections	9875 (1095)	9952 (1134)	10182 (1193)	10256 (1228)	10259 (1229)	10388 (1293)
Anomalous multiplicity	3.7 (3.8)	7.4 (5.4)	10.9 (7.9)	14.3 (10.1)	21.3 (14.9)	27.8 (18.9)
Anomalous completeness (%)	93.3 (72.6)	94.3 (76.7)	94.9 (77.7)	96.6 (81.3)	96.6 (81.4)	97.6 (85.0)
$R_{\rm meas}$ (%)	1.9 (1.2)	2.7 (4.8)	3.0 (4.6)	2.9 (5.3)	3.1 (4.8)	3.3 (5.8)
$R_{\rm anom}$ (%)	1.2 (2.4)	1.0 (1.9)	1.0 (1.8)	0.9 (1.8)	0.9 (1.6)	0.9 (1.6)
$R_{\rm p.i.m.}$ (%)	1.2 (2.5)	0.9 (1.9)	0.8 (1.7)	0.7 (1.5)	0.6 (1.3)	0.6 (1.2)
$\langle I/\sigma(I)\rangle$	68.1 (34.4)	82.0 (41.7)	99.51 (69.27)	104.7 (51.2)	126.2 (61.8)	139.4 (66.1)
d''/σ†	0.99–2.08	1.04–2.62	1.10–2.94	1.18–3.23	1.31–3.84	1.38–4.30
$R_{\rm anom}/R_{\rm p.i.m.}$	1.00 (0.96)	1.11 (1.00)	1.25 (1.06)	1.28 (1.20)	1.50 (1.23)	1.50 (1.33)
Phasing statistics‡						
SHELXD $CC_{\rm all}$	33.15	44.49	43.22	43.86	35.82	46.34
SHELXD $CC_{\rm weak}$	17.08	23.41	24.28	25.66	19.18	25.36
SHELXE pseudo-free CC	54.75	84.67	85.12	85.42	83.40	85.03
Traceable residues§ (%)	0.0	98.4	98.4	97.5	97.5	97.5
$R_{\rm work}$¶ (%)	30.56	23.42	23.33	23.24	23.96	23.85

† d''/σ is the anomalous signal indicator provided by *SHELXC*. The values indicated are for the high-resolution (1.81–8 Å) range for each data set, with a d''/σ value of about 0.8 indicating zero signal (Sheldrick, 2008). ‡ The procedure used in *SHELXC*/*D*/*E* was the following: Patterson seeding for a vector of length 2.06 Å and direct methods were used to find six S atoms. 100 random starts were used and the S-SAD data resolution was not truncated (maximum resolution of 1.81 Å). For *SHELXE*, the high-resolution data displaying similar unit-cell parameters as the S-SAD data (*i.e.* the 1450-image data set) were used as native input data. All other parameters were set to default values. § Automated model building was performed in 50 cycles of *ARP*/*wARP* and *REFMAC*5. ¶ $R_{\rm work}$ obtained from *REFMAC*5.

carwin protein crystallized in the orthorhombic space group $P2_12_12_1$.

Prior to data collection, the crystals were soaked for 1 min in cryoprotectant solution [crystallization solution prepared with 15%(*v*/*v*) glycerol] and were then flash-cooled in liquid nitrogen. Data were collected at cryogenic temperature (100 K). The sulfur anomalous diffraction data were collected in-house using a Cu $K\alpha$ (1.5418 Å, 8.05 keV) rotating-anode X-ray source (Rigaku MicroMax-007 HF operating at 40 kV, 30 mA) and a Saturn944+ detector with a 41.5 mm crystal-to-detector distance and 10 s exposure time per frame. To improve the signal-to-noise ratio of the sulfur anomalous signal, the data-collection strategy consisted of 16 data sets of 180 images collected from the same crystal with an oscillation range of 0.5° per frame to give a total of 2880 images. Each data set was characterized by the use of different combinations of φ (0, 90, 180 and 270°) and κ (0, 5, 10 or 15°) angles in order to maximize the data redundancy on Bijvoet pairs of reflections (Dauter, 1999). The complete acquisition lasted for ∼22 h of total exposure. Data-collection and processing statistics for the 16 S-SAD data sets are provided in Supplementary Table S1. The raw S-SAD diffraction data are available on request.

Two high-resolution data sets were collected using two different protein crystals. The first data set (the 700-image data set) was acquired with an oscillation range per image of 0.15° using 0.8 Å wavelength X-rays (15.5 keV), an exposure time of 1 s and a crystal-to-detector distance of 112.49 mm (Table 1). The second data set (the 1450-image data set) consists of 1450 images with an oscillation range per image of 0.25°, an incident X-ray wavelength of 0.8 Å, an exposure time per image of 0.5 s and a crystal-to-detector distance of 120.0 mm (Table 1). All high-resolution data were collected on the X06DA-PXIII beamline at the Swiss Light Source (SLS; Villigen, Switzerland) equipped with a PILATUS 2M-F detector. Note that all synchrotron data were collected using an attenuated X-ray beam. The data-collection statistics are summarized in Tables 1 and 2 and Supplementary Table S1.

All diffraction data were processed with the *XDS* program package (Kabsch, 2010). $R_{\rm anom}$ and $R_{\rm p.i.m.}$ were determined using *SCALA* (Winn *et al.*, 2011) based on the data scaled using *XSCALE* (Kabsch, 2010) without merging the original indices. The Matthews coefficient and solvent content were calculated using programs from the *CCP*4 suite (Winn *et al.*, 2011).

2.2. Phasing and refinement

The structure was solved by the sulfur-SAD method (Wang, 1985) with the aid of the *HKL2MAP* graphical interface (Pape & Schneider, 2004) for the *SHELX* programs (Sheldrick, 2008). Searches for sulfur positions were carried out with *SHELXD* (Schneider & Sheldrick, 2002) using the complete S-SAD data collected in-house. The redundancy computed on the complete data was ∼52. Initial phasing was performed using *SHELXE* (Sheldrick, 2002). Details of the *SHELXC*/*D*/*E* procedures used can be found in Table 2.

Most of the polypeptide backbone could be traced automatically using the *ARP*/*wARP* web server (http://www.embl-hamburg.de/ARP/; Perrakis *et al.*, 1999) and the high-resolution data set (the 1450-image data set), with similar unit-cell parameters to those of the S-SAD data set. Refined structures were obtained using the two high-resolution data sets (the 700-image and 1450-image data sets) starting with the solution obtained by molecular replacement in *MOLREP* (Vagin & Teplyakov, 2010) and the unrefined structure from

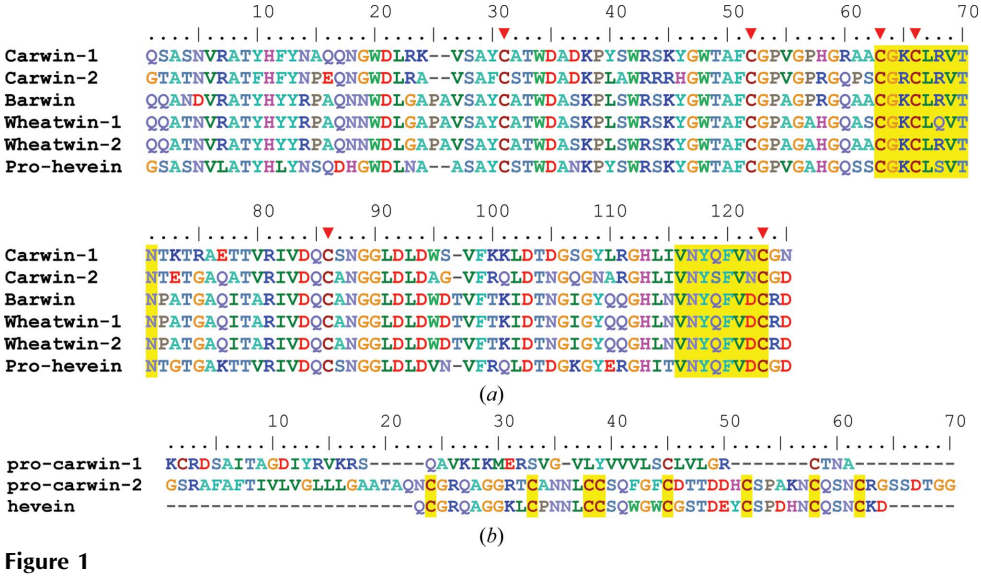


Figure 1
Sequence alignment. (*a*) Multiple sequence alignment between the two carwin isoforms and four PR-4 proteins. The two identified carwin proteins are referred to as carwin-1 and carwin-2. The PR-4 proteins are barwin, the wheatwin-1 and wheatwin-2 proteins and pro-hevein. The two barwin-domain signatures from the PROSITE database (C-G-[KR]-C-L-*x*-V-*x*-N and V-[DN]-Y-[EQD]-F-V-[DN]-C) are highlighted in yellow. The red arrows above the sequences indicate the positions of the six cysteine residues forming the three disulfide bridges. (*b*) N-terminal pro-sequences of the two carwin isoforms aligned with the hevein sequence. The eight cysteines forming the typical hevein pattern of four disulfide bridges are highlighted.

ARP/*wARP* as the search model. Revisions to the models were manually performed using *Coot* (Emsley & Cowtan, 2004) and refinement was carried out using *REFMAC*5 from the *CCP*4 suite (Murshudov *et al.*, 2011; Winn *et al.*, 2011) against the higher resolution data sets. The refinement converged to give the statistics presented in Table 1. The stereochemical quality of the final model was evaluated using *PROCHECK* (Laskowski *et al.*, 1993) and *MolProbity* (Chen *et al.*, 2010).

2.3. Sequence and structure analysis

The genome of the plant *C. papaya* has recently been deciphered, although it has only partially been annotated (Ming *et al.*, 2008). From the EMBL database (release 113, August 2012), 123 366 nucleic acid sequences of *C. papaya* genes were extracted. Using the barwin sequence (Swiss-Prot entry P28814) as a query, a *TBLASTN* search (Altschul *et al.*, 1997) was performed against our local papaya database. The retrieved sequences were automatically aligned with *ClustalW* (Thompson *et al.*, 1994) and were visualized with *BioEdit* (Hall, 1999). Structural analysis was performed using several programs such as *DALI* (Holm & Rosenström, 2010), *DaliLite* (Holm & Park, 2000), *PROMOTIF* (Hutchinson & Thornton, 1996) and *PISA* (Krissinel & Henrick, 2007). Ligand-binding site predictions were performed using two different web-server programs: *FTSite* (Ngan *et al.*, 2012) and *MetaPocket* (Huang, 2009). Graphical pictures were generated using *Coot* (Emsley & Cowtan, 2004), *PyMOL* (Schrödinger LLC) or *MolScript* (Kraulis, 1991) and were rendered with *Raster3D* (Merritt & Bacon, 1997).

3. Results and discussion

3.1. Amino-acid sequences of the papaya barwin-like protein

On studying the proteome of *C. papaya* latex, a new protein homologous to the barwin protein was identified on the basis of in-gel trypsin digestion and mass-spectrometric sequencing. In order to obtain the complete sequence of the papaya barwin-like protein, as the draft genome of the plant is available (Ming *et al.*, 2008), a *BLAST* search was performed on a local database containing all nucleic sequences from the *C. papaya* species. The sequence of the barwin protein was used as a query sequence. The search yielded several hits that corresponded to two protein isoforms, referred to here as carwin-1 and carwin-2 (see Supplementary Fig. S1).

From their similarity to the mature form of the barwin protein, both carwin proteins showed different N-terminal pro-sequences. They share only 50.0% sequence identity considering the pro-sequence and up to 72.7% upon removal of the pro-sequence. A multiple alignment of the carwin proteins with other PR-4 proteins indicated perfect conservation of the six cysteine residues that characterize the barwin-domain signature (PROSITE documentation entry PDOC00619; Fig. 1*a*). Additionally, the two barwin-domain signatures referenced by the PROSITE database (Sigrist *et al.*, 2013) were perfectly found in the sequences of both carwin isoforms.

All PR-4 proteins share a common C-terminal barwin-like domain but differ in the presence (class I) or the absence (class II) of an N-terminal domain similar to hevein, a small chitin-binding domain (Neuhaus *et al.*, 1996). We found here that the first isoform does not contain a hevein domain in its N-terminal part, unlike the second isoform (Fig. 1*b*), thus revealing that the papaya genome contains both class I and class II PR-4 proteins.

3.2. Purification and crystallization

The isolation and purification of proteins from the latex of the tropical plant *C. papaya* has been the focus of our laboratory for several years (El Moussaoui *et al.*, 2001; Looze *et al.*, 2009). In this study, applying a double SP Sepharose–FF chromatography purification step followed by thiophilic interaction chromatography yielded a homogenous protein sample that migrated as a 15 kDa protein on an SDS–PAGE (Fig. 2*a*). The purification procedure is described in the Supplementary Material. Crystals of papaya carwin were

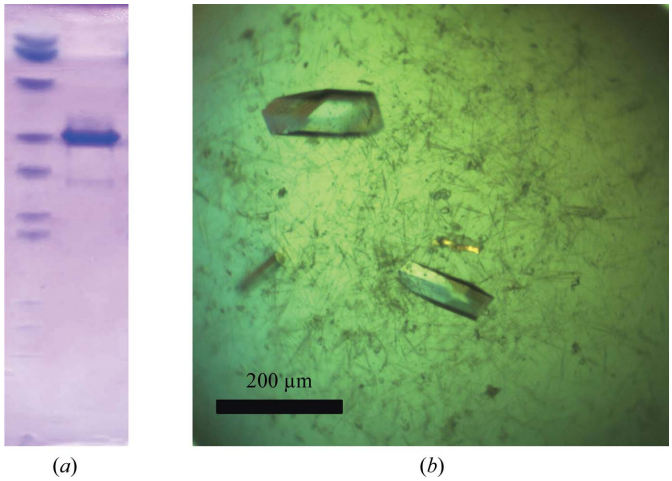
obtained using 2 *M* ammonium sulfate as a crystallization solution (Fig. 2*b*).

Several crystals of carwin were utilized for structure determination. All of the crystals produced belonged to space group $P2_12_12_1$ and diffracted to a high resolution of around 1 Å. The first diffraction data set collected, referred to as the 700-image data set, gave unit-cell parameters a = 54.91, b = 56.96, c = 74.20 Å (Table 1). According to the calculated Matthews coefficient of 2.11 Å^3 Da^{-1}, a solvent content of 41.8% is estimated assuming two molecules per asymmetric unit. The second crystal, although obtained under identical conditions and having an identical space group, presented a unit-cell a axis that was almost halved compared with the first crystal, resulting in an estimated solvent content of 39.7% and one molecule per unit cell.

3.3. In-house S-SAD phasing

Initially, we attempted phasing by molecular replacement using the NMR structure of the homologous barwin protein as a template (Ludvigsen & Poulsen, 1992*b*). Unfortunately, all of our attempts were unsuccessful. As the carwin protein possesses six cysteine residues, the anomalous signal of S atoms was used to determine the phases. Assuming a protein of 212 residues in length (Fig. 1) and using 1.54 Å wavelength X-rays, the Bijvoet ratio ($\langle\Delta F^{\pm}\rangle/\langle F\rangle$) was estimated to be 0.93% (Hendrickson & Teeter, 1981), a promising value that was significantly above the so-called Wang limit of 0.6% expected for successful S-SAD (Wang *et al.*, 2006).

An in-house Cu $K\alpha$ X-ray radiation source was then used to collect a highly redundant data set consisting of 16 data-collection runs from the same crystal with an oscillation range of 0.5° per frame to give a total of 2880 images (statistics are given in Supplementary Table S1). Despite more than 22 h of exposure time, analysis of the complete data using *XDSSTAT* (Diederichs, 2006) showed that the crystal did not appear to suffer from radiation damage, as indicated by the monotonic behaviour of the decay R factor (R_d) over the course of data collection (Fig. 3*a*). Consistent with this analysis, the indicators of data quality, $I/\sigma(I)$ and R_{meas}, remained nearly constant over the 16 data-collection runs (Supplementary Table S1). Thus, the crystal used seemed to be appropriate for this long exposure time.



(*a*) (*b*)

Figure 2
Purification and crystallization of the carwin protein. (*a*) SDS–PAGE of the carwin protein. The gel was stained with Coomassie Brilliant Blue. Lane 1, molecular-weight markers (molecular weights of 2.5, 3.5, 6, 14.4, 21.5, 30.0, 36.5, 55.4, 66.3, 97.4, 116.3 and 200 kDa); lane 2, purified solution of papaya barwin-like protein or carwin. (*b*) Typical crystals of carwin.

In order to systematically study the effects of data quality on substructure solution and phasing, the 2880 oscillation images were divided into six cumulative data sets labelled *A*–*F*; the anomalous multiplicity increased from 3.7 to 27.8 in data sets *A*–*F* (Table 2). Analysis of the data revealed excellent statistics (Table 2 and Supplementary Table S1).

For SAD phasing, the quality of the anomalous signal is an important factor; however, estimating this quality is not trivial (Fu *et al.*, 2004; Zwart, 2005; Dauter, 2006). We monitored several indicators throughout the cumulative data sets. The anomalous scattering signal $\langle d''/\mathrm{sig}(d'')\rangle$, an estimate of the anomalous signal calculated by *SHELXC* (Sheldrick, 2010), was enhanced by increasing redundancy (Fig. 3*b*). For data set *F* (the complete S-SAD data set), the $\langle d''/\mathrm{sig}(d'')\rangle$ value rose to a maximum of 4.3 in the inner resolution shell (∞–8.0 Å) and dropped to 1.38 at 1.82 Å resolution (Fig. 3*b*), indicating the presence of significant anomalous signal. Similar results were obtained in the plot of SigAno, another indicator of anomalous signal computed by *XDS* (Kabsch, 2010), *versus* resolution (Fig. 3*c*). SigAno is defined as the absolute anomalous differences $|\Delta F^{\pm}|$ divided by their standard deviation [*i.e.* $|\Delta_{ano}|/\sigma(\Delta_{ano})$]. However, a value of above 1.5 in all resolution shells is considered to be a necessary requirement for successful S-SAD experiments (Cianci *et al.*, 2008). For the data analyzed here, the SigAno value for data set *F* was found to be above the threshold of 1.5 in all resolution shells except for the highest shell, while for the other smaller data sets the values dropped below the threshold in all shells of higher resolution (Fig. 3*c*).

The ratio of the anomalous R factor (R_{anom}) to the merging R factor $R_{p.i.m.}$ has also often been used as an indicator of the level of anomalous signal (Mueller-Dieckmann *et al.*, 2004, 2005; Weiss, 2001; Weiss *et al.*, 2001, 2004). An $R_{anom}/R_{p.i.m.}$ ratio of above 1.5 is empirically considered to be favourable for successful S-SAD phasing (Mueller-Dieckmann *et al.*, 2005). As seen in Table 2, the ratio reached a value of 1.5 only when 2160 images (*i.e.* data set *E*) were considered.

The *SHELXC*/*D*/*E* programs (Sheldrick, 2010), through the graphical interface *HKL2MAP* (Pape & Schneider, 2004), were used for sulfur-substructure determination, initial phasing and phase improvement by density modification. The programs were applied to data sets *A*–*F* followed by 50 cycles of *ARP*/*wARP* (Langer *et al.*, 2008) and *REFMAC*5 (Murshudov *et al.*, 2011) using the high-resolution data set displaying similar unit-cell parameters to those of the S-SAD data set. Table 2 gives the obtained phasing statistics and Fig. 3(*d*) shows a plot of the *SHELXE* pseudo-free CC parameter as a function of anomalous multiplicity. Data set *B*, consisting of 720 images with a multiplicity of only 7.4, gave a

clear solution allowing 120 amino-acid residues to be automatically fitted with an R_{work} of 23.42%. Indeed, the difference between successful and unsuccessful data sets was quite evident from Table 2. With data set *B*, *SHELXD* clearly found three pairs of positions, thus showing a promising result for phasing (Fig. 4*a*). Note that we did not use a Patterson seeding parameter adapted to the length of a disulfide bridge in *SHELXC*, but a search for six individual sites (Table 2).

The $R_{anom}/R_{p.i.m.}$ ratio for data set *B* was only 1.11, which is significantly below the empirically estimated threshold of 1.5 for successful S-SAD (Weiss *et al.*, 2001; Mueller-Dieckmann *et al.*, 2005). As previously noted by others (Sarma & Karplus, 2006; Lakomek *et al.*, 2009), the $R_{anom}/R_{p.i.m.}$ threshold may be adjusted to a lower limit, *i.e.* 1.1, as found in the present study and as indicated previously in the S-SAD solution of the lysosomal mouse protein (Lakomek *et al.*, 2009). Similarly, the SigAno parameter, or $|\Delta_{ano}|/\sigma(\Delta_{ano})$, may also require adjustment. Data set *B* showed a SigAno value greater than the cutoff of 1.5 (Cianci *et al.*, 2008) only for resolution shells above 3.8 Å. The value fell to lower than 1 in the highest resolution shell of data set *B*. Finally, the solvent content is known to be an important factor in successful S-SAD phasing with weak anomalous signal (Ramagopal *et al.*, 2003). The solvent content of data set *B* was 39.6%. Obviously, the crystal quality is also another important factor and in our case it was certainly a factor that contributed greatly to the success of S-SAD. Finally, it can be noted that the use of crystal re-orientation by the kappa angle was not needed in our S-SAD phasing as data set *B* with a constant kappa angle equal to 0° generated a clear solution (Fig. 4*a*, Table 2 and Supplementary Table S1).

3.4. Model building and description

Starting with the initial unrefined model from *ARP/wARP* (Langer *et al.*, 2008) as a search model, the two high-resolution data sets (the 700-image and 1450-image data sets) were solved by molecular replacement. The structures of papaya barwin-like protein were further completed manually and refined by cycling between *Coot* (Emsley & Cowtan, 2004) and *REFMAC5* (Murshudov *et al.*, 2011). The final refinement statistics are summarized in Table 1. The 700-image data set gave a model with two molecules of carwin per asymmetric unit cell, while a monomeric model was refined with the 1450-image set. The dimeric model was refined to 1.05 Å resolution and to a final *R* factor of 14.2% (free *R* factor of 16.4%) and the monomeric model to 1.00 Å resolution with an *R* factor of 15.3% (free *R* factor of 17.9%).

Interestingly, analysis with *PISA* (Krissinel & Henrick, 2007) indicated that the observed dimeric complex of carwin appears to be the result of crystal packing. The *PISA* complex significance score (CSS) is 0.00 and the interface surface is

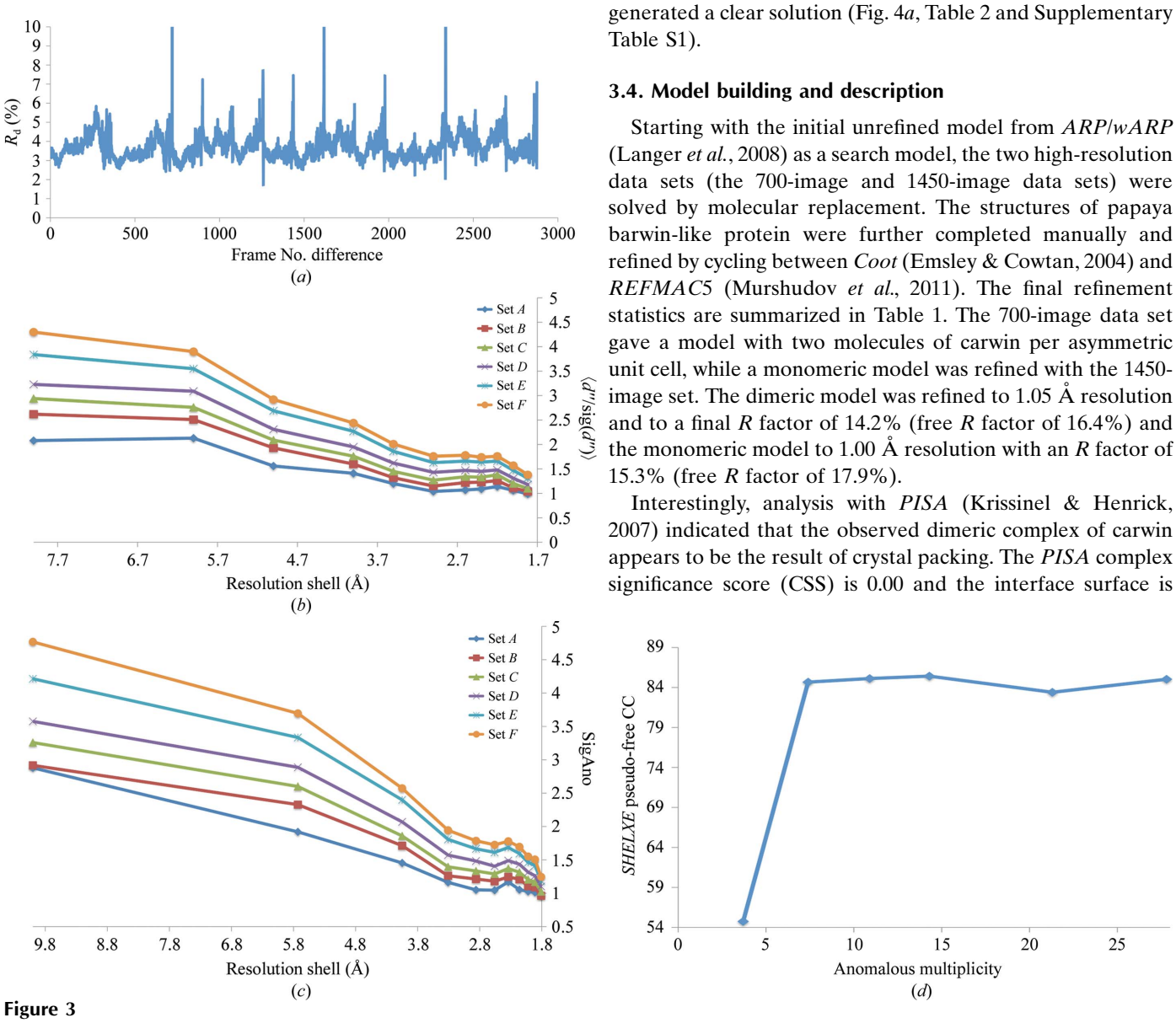


Figure 3
S-SAD data statistics. (*a*) Radiation damage monitored by *XDSSTAT* (as a percentage; Diederichs, 2006). (*b*) A comparison of the *SHELXC* anomalous scattering signal $d''/\mathrm{sig}(d'')$ as a function of resolution for data sets *A*–*F*. (*c*) SigAno as a function of resolution for data sets *A*–*F*. SigAno describes the absolute anomalous differences divided by their standard deviation, *i.e.* $|F(+) - F(-)|/\sigma$, where $F(+)$ and $F(-)$ are structure-factor estimates obtained from the merged intensity observations in each parity class. SigAno values were computed by *XDS* (Kabsch, 2010). (*d*) *SHELXE* pseudo-free CC parameter as a function of anomalous multiplicity.

about 250 Å^2, a value that is considered to be too low to have biological significance (Krissinel & Henrick, 2007). In addition, neither carwin nor its homologues were found to form dimers in solution. Indeed, the two crystals have different packing (Supplementary Fig. S2). Despite our efforts, we could not elaborate a clear explanation for the fact that the carwin protein could be crystallized in two different forms under the same crystallization condition. However, this does not appear to be an isolated case: a protein from *Streptococcus* was reported to give three different crystal forms in the same droplet under identical conditions (Meinhart *et al.*, 2001).

The central core of the carwin structure consists of a six-stranded β-barrel surrounded by four α-helices, two 3_{10}-helices and several loops (Figs. 5 and 6). The shear number of the β-barrel, a measure of the stagger of the strands in the closed β-sheet, is 10 (Fig. 5), and the topology of the β-barrel was found to be [2, 2, −1, −2, −2] (Richardson, 1981), thereby generating four interpenetrated ψ loops. A pseudo-twofold axis can topologically relate β1, β2 and β3 to β4, β5 and β6, respectively. Therefore, the four ψ loops can be regarded as a pair of symmetrical ψ loops. The first ψ loop consists of the two antiparallel strands β1 (β4) and β2 (β5) with the strand β5 (β2) in between, whereas the second ψ loop consists of the strands β2 (β5) and β3 (β6) and the strand β4 (β1) (Fig. 5). This particular fold has previously been called a 'double-ψ β-barrel' and has been observed in several other protein structures (Castillo *et al.*, 1999). The central pores of the β-barrel are sealed off at the N-terminal side by the loop Cys50–His56 and, on the other side, by the α3 helix (Trp93–Asp100). The three observed disulfide bonds (Cys2–Cys29, Cys50–Cys84 and Cys64–Cys120) are also likely to enhance structural stability.

Finally, the excellent quality of the electron-density map indicated that the N-terminal residue, Gln1, was cyclized into pyroglutamic acid or 5-oxoproline (Fig. 4*b*) and that carwin does not appear to be glycosylated. An N-terminal glutamine residue is well known to spontaneously cyclize into pyroglutamic acid. However, under physiological conditions this cyclization is rather slow (Arii *et al.*, 1999) and therefore may require enzymatic catalysis. Interestingly, the papaya latex contains a glutaminyl cyclase (Wintjens *et al.*, 2006) that could be responsible for the cyclization observed in the carwin protein.

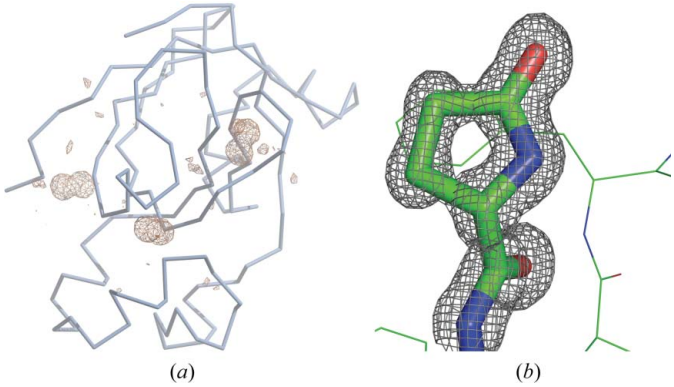


Figure 4
(*a*) Anomalous difference Fourier map contoured at 2.5σ and calculated at 1.81 Å resolution using data set *B* from the S-SAD data collection and the program *FFT*. The map is coloured brown and the C$^\alpha$ trace of the refined carwin structure is shown in blue. (*b*) $2F_o - F_c$ electron-density map of the first N-terminal residue Gln1 obtained after refinement at 1.0 Å resolution and contoured at the 1.5σ level. The pyroglutamic acid modification is clearly visible and was modelled.

3.5. Biological function of papaya carwin

The high-resolution structure unequivocally revealed the exact amino-acid sequence of the carwin protein extracted from papaya latex. The sequence corresponds perfectly to the isoform carwin-1, identified as the class II PR-4 protein of *C. papaya* (Fig. 1). The biological function of PR-4 proteins of either class I or II remains elusive; however, these proteins share the capacity to inhibit the proliferation of phytopathogenic fungi (Hejgaard *et al.*, 1992; Ponstein *et al.*, 1994; Niderman *et al.*, 1995; Caruso *et al.*, 1996; Li *et al.*, 2010).

Chitinase and 1,3-β-glucanase activities have been reported for some PR-4 proteins (Kauffmann *et al.*, 1987; Brunner *et al.*, 1998; Lu *et al.*, 2012). Such activities can be easily explained by the fact that the PR-4 proteins of subclass I possess the hevein chitin-binding domain. However, this does not apply to carwin-1, which is a class II PR-4 protein. Furthermore, neither chitinase nor 1,3-β-glucanase activities were detected for a purified solution of papaya latex carwin (J. Huet, unpublished results).

Ribonuclease activity was found for the wheatwin-1 protein, a protein isolated from wheat kernels (Caporale *et al.*, 2004; Bertini *et al.*, 2009), and dual RNase/DNase activities have been described for a PR-4 protein from chili pepper

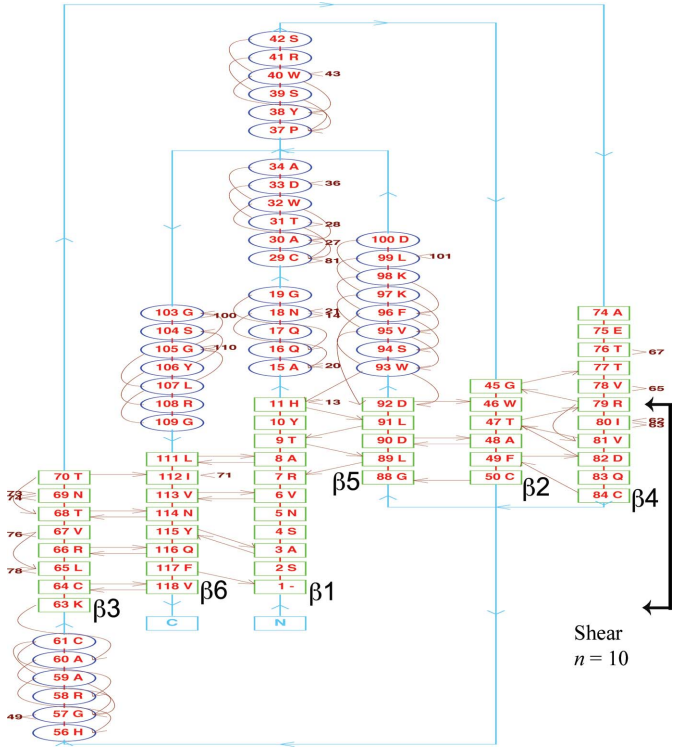


Figure 5
Topology diagram of the carwin structure generated by *PROMOTIF*.

(Guevara-Morato *et al.*, 2010). More recently, RNase and chitinase activities were found for a class I PR-4 protein from jelly fig (Lu *et al.*, 2012), and a class I PR-4 vacuolar protein from *Arabidopsis* was described to possess RNase activity (Bertini *et al.*, 2012). Furthermore, ribonuclease and antifungal activities are linked, as revealed by mutagenesis studies on wheatwin-1 protein (Bertini *et al.*, 2009).

Overall, RNase/DNase activity thus appears to be the most promising assumption for the biological function of papaya carwin. Possible binding sites on the carwin protein surface were evaluated (Supplementary Fig. S3 and Table S2). Three binding sites were identified: a large pocket lined by Trp20, Asp92 and Lys98 (site 1), a site centred around the tetrad His11, Tyr13, Asp82 and Asp90 (site 2) and a positively charged cavity surrounded by Trp32, Lys36, Trp40, Arg58 and Arg79 (site 3). With the assumption that the enzymatic activity of carwin uses the classical acid–base mechanism common to ribonucleases, binding site 2 could harbour such an activity, with a histidine ideally located in close vicinity to a couple of acidic residues (Supplementary Fig. S3). Interestingly, the tetrad of residues involved in binding site 2 was highly conserved among the sequences of the PR-4 protein family. Furthermore, His11 was revealed to be a key residue for ribonuclease activity as its mutation partially impaired the activity (Bertini *et al.*, 2009).

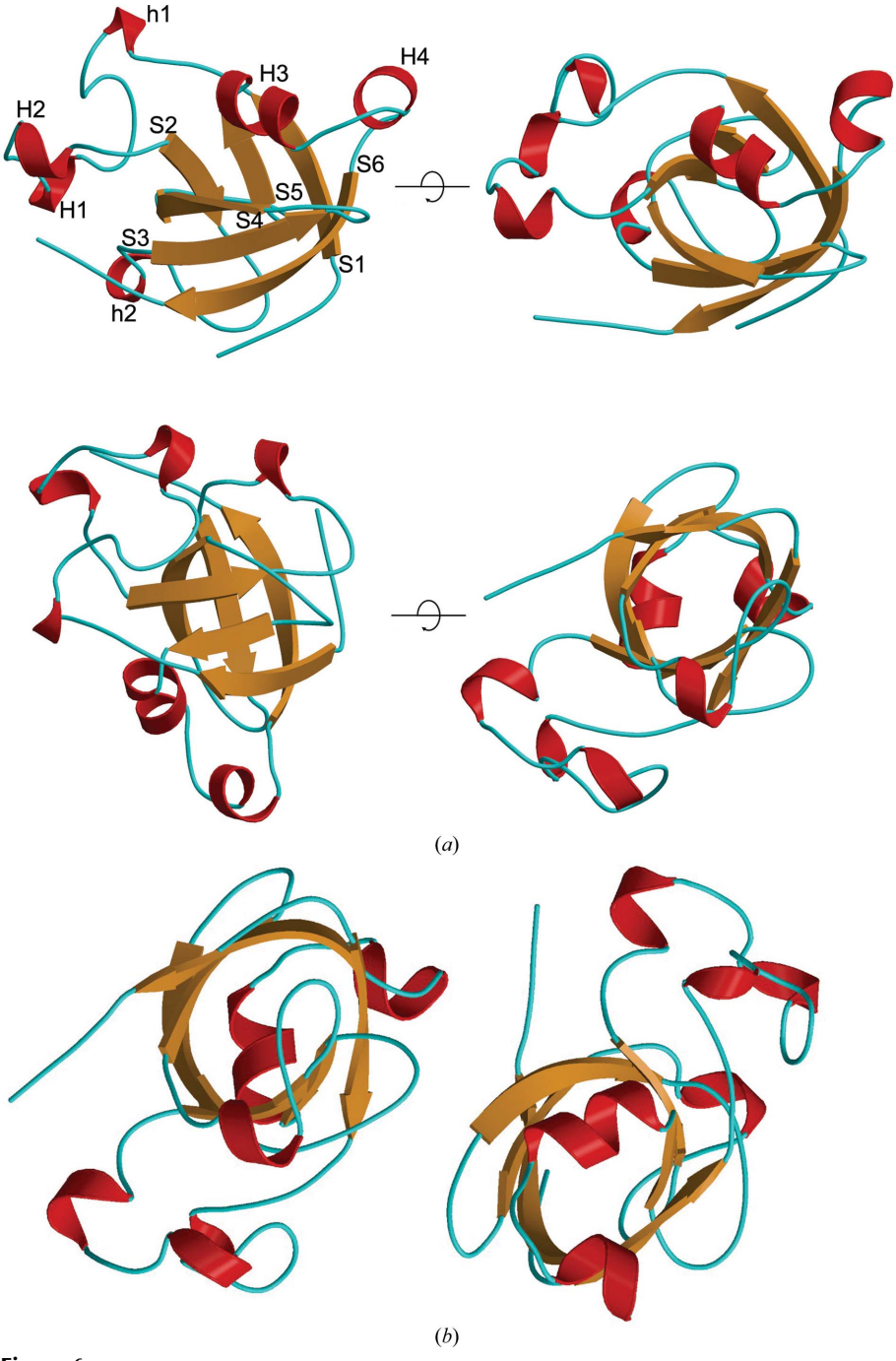


Figure 6
The refined high-resolution structure of the carwin protein. (*a*) Four different views of the model shown as coloured ribbon representations, with secondary structures labelled in the upper left view. The secondary-structure limits are as follows: strands S1–S6, 2–11, 45–49, 64–69, 75–83, 89–92 and 111–118; helices H1–H4, 29–34, 37–42, 93–100 and 103–109; 3_{10}-helices h1 and h2, 15–19 and 56–61. (*b*) Ribbon representation of the dimeric form of the carwin protein.

3.6. A common structural module in the Protein Data Bank

A PDB search with the *DALI* web server (Holm & Rosenström, 2010) using the carwin structure identified several proteins containing a double-ψ β-barrel domain (Supplementary Table S3). The closest structural relatives of carwin are barwin (*DALI* *Z*-score = 14.4, r.m.s.d. = 2.6 Å), several cerato-platanin-like proteins (mean *Z*-score = 12.1), the elicitor Sm1 of plant defence responses from *Trichoderma virens* (*Z*-score = 13.0), the β-expansin YoaJ (*Z*-score = 10.9) and the two protein allergens Phl p 1 and Zea m 1. A structural genomics program on *Pseudomonas aeruginosa* recently solved the structure of a 125-residue PA4485 protein with the same fold as carwin with an r.m.s.d. of 2.3 Å (Moynie *et al.*, 2013). Endoglucanases of glycosyl hydrolase family 45 (GH45) are also known to contain a similar structural module. Finally, several unrelated enzymes were identified that included a double-ψ β-barrel domain, such as, for instance, the N-domain of ATPase p97, the C-domain of formate dehydrogenase H and domains from arsenite oxidase and RNA polymerase II.

Note that most of the identified protein structures display the canonical β-barrel topology [2, 2, −1, −2, −2] and show the four ψ loops typical of the carwin fold. In contrast, the number of disulfide bonds does not appear to be at all essential to maintain the structural fold. Although the barwin protein is used as a reference protein for the double-ψ β-barrel superfamily, its NMR structure deviates from the common structural module: the β-barrel topology is [4, −2, 1, −2x] and only one ψ loop is detected using *PROMOTIF*.

Despite the poor sequence identity, the double-ψ β-barrel module is found in diverse protein structures, thereby suggesting the existence of a common structural ancestor of these protein families (Castillo *et al.*, 1999). A common functional role for the ψ loop has been proposed (Castillo *et al.*, 1999). The ψ loops of the double-ψ β-barrel module could be seen as fairly rigid surface loops, thereby representing a good scaffold for a catalytic or a cofactor-binding site.

4. Concluding remarks

The barwin-like protein isolated from papaya latex was found to belong to the class II PR-4 protein family and its amino-acid sequence was inferred from the genome-sequence database and validated by the X-ray crystal structure. The protein was predicted to have ribonuclease biological activity and a putative binding site for the activity was proposed. However, this last statement requires experimental confirmation.

Furthermore, the present study used two common approaches for increasing the anomalous scattering signal that are crucial for the success of S-SAD: the use of soft and longer wavelength X-rays and the collection of highly redundant data. Here, S-SAD phasing of the papaya barwin-like protein, displaying an estimated Bijvoet ratio of 0.93% with 1.54 Å wavelength X-rays, showed that high anomalous multiplicity is not necessarily required and that although indicators of successful S-SAD predict an unfavourable outcome this phasing method can still be fruitful. Obviously, successful S-SAD phasing strongly depends on the crystal quality. Indeed, the crystal quality and resistance to optical damage are probably the limiting factors for the method. Given that the S-SAD method is largely underutilized in X-ray structural determination, it may deserve a more systematic use for good-quality diffraction crystals and for low-molecular-weight proteins that contain a relatively high content of cysteine/methionine residues.

Diffraction data were collected at the Swiss Light Source (Paul Scherrer Institute, Villigen, Switzerland). We are grateful to the machine and beamline groups whose outstanding efforts made these experiments possible. EJTM thanks the Belgium program (IN–Wallonie–Bruxelles International) for the fellowship. We also thank Professor Han Remaut for helpfully assistance in the in-house S-SAD data collection. RW is Research Associate at the National Fund for Scientific Research FNRS-FRS (Belgium), which is gratefully acknowledged for support.

References

Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
Arii, K., Kobayashi, H., Kai, T. & Kokuba, Y. (1999). *Eur. J. Pharm. Sci.* **9**, 75–78.
Bertini, L., Caporale, C., Testa, M., Proietti, S. & Caruso, C. (2009). *FEBS Lett.* **583**, 2865–2871.
Bertini, L., Proietti, S., Aleandri, M. P., Mondello, F., Sandini, S., Caporale, C. & Caruso, C. (2012). *Biol. Chem.* **393**, 1533–1546.
Broekaert, I., Lee, H.-I., Kush, A., Chua, N.-H. & Raikhel, N. (1990). *Proc. Natl Acad. Sci. USA*, **87**, 7633–7637.
Brunner, F., Stintzi, A., Fritig, B. & Legrand, M. (1998). *Plant J.* **14**, 225–234.
Caporale, C., Di Berardino, I., Leonardi, L., Bertini, L., Cascone, A., Buonocore, V. & Caruso, C. (2004). *FEBS Lett.* **575**, 71–76.
Caruso, C., Bertini, L., Tucci, M., Caporale, C., Leonardi, L., Saccardo, F., Bressan, R. A., Veronese, P. & Buonocore, V. (1999). *DNA Seq.* **10**, 301–307.
Caruso, C., Caporale, C., Chilosi, G., Vacca, F., Bertini, L., Magro, P., Poerio, E. & Buonocore, V. (1996). *J. Protein Chem.* **15**, 35–44.
Castillo, R. M., Mizuguchi, K., Dhanaraj, V., Albert, A., Blundell, T. L. & Murzin, A. G. (1999). *Structure*, **7**, 227–236.
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.
Chevalier, C., Bourgeois, E., Pradet, A. & Raymond, P. (1995). *Plant Mol. Biol.* **28**, 473–485.
Cianci, M., Helliwell, J. R. & Suzuki, A. (2008). *Acta Cryst.* D**64**, 1196–1209.
Dauter, Z. (1999). *Acta Cryst.* D**55**, 1703–1717.
Dauter, Z. (2006). *Acta Cryst.* D**62**, 867–876.
Debreczeni, J. É., Bunkóczi, G., Ma, Q., Blaser, H. & Sheldrick, G. M. (2003). *Acta Cryst.* D**59**, 688–696.
Diederichs, K. (2006). *Acta Cryst.* D**62**, 96–101.
Doutch, J., Hough, M. A., Hasnain, S. S. & Strange, R. W. (2012). *J. Synchrotron Rad.* **19**, 19–29.
El Moussaoui, A., Nijs, M., Paul, C., Wintjens, R., Vincentelli, J., Azarkan, M. & Looze, Y. (2001). *Cell. Mol. Life Sci.* **58**, 556–570.
Emsley, P. & Cowtan, K. (2004). *Acta Cryst.* D**60**, 2126–2132.
Fu, Z.-Q., Rose, J. P. & Wang, B.-C. (2004). *Acta Cryst.* D**60**, 499–506.
Goulet, A., Vestergaard, G., Felisberto-Rodrigues, C., Campanacci, V., Garrett, R. A., Cambillau, C. & Ortiz-Lombardía, M. (2010). *Acta Cryst.* D**66**, 304–308.
Gregersen, P., Thordal-Christensen, H., Förster, H. & Collinge, D. (1997). *Physiol. Mol. Plant Pathol.* **51**, 85–97.
Guevara-Morato, M. A., de Lacoba, M. G., García-Luque, I. & Serra, M. T. (2010). *J. Exp. Bot.* **61**, 3259–3271.
Hall, T. A. (1999). *Nucleic Acids Symp. Ser.* **41**, 95–98.
Hejgaard, J., Jacobsen, S., Bjørn, S. E. & Kragh, K. M. (1992). *FEBS Lett.* **307**, 389–392.
Hendrickson, W. A. & Teeter, M. M. (1981). *Nature (London)*, **290**, 107–113.
Holm, L. & Park, J. (2000). *Bioinformatics*, **16**, 566–567.
Holm, L. & Rosenström, P. (2010). *Nucleic Acids Res.* **38**, W545–W549.
Huang, B. (2009). *OMICS*, **13**, 325–330.
Hutchinson, E. G. & Thornton, J. M. (1996). *Protein Sci.* **5**, 212–220.
Jamet, E. & Fritig, B. (1986). *Plant Mol. Biol.* **6**, 69–80.
Kabsch, W. (2010). *Acta Cryst.* D**66**, 125–132.
Kauffmann, S., Legrand, M., Geoffroy, P. & Fritig, B. (1987). *EMBO J.* **6**, 3209–3212.
Kraulis, P. J. (1991). *J. Appl. Cryst.* **24**, 946–950.
Krissinel, E. & Henrick, K. (2007). *J. Mol. Biol.* **372**, 774–797.
Lakomek, K., Dickmanns, A., Mueller, U., Kollmann, K., Deuschl, F., Berndt, A., Lübke, T. & Ficner, R. (2009). *Acta Cryst.* D**65**, 220–228.

Langer, G., Cohen, S. X., Lamzin, V. S. & Perrakis, A. (2008). *Nature Protoc.* **3**, 1171–1179.
Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.
Lee, S. C., Kim, Y. J. & Hwang, B. K. (2001). *Plant Cell Physiol.* **42**, 1321–1330.
Li, X., Xia, B., Jiang, Y., Wu, Q., Wang, C., He, L., Peng, F. & Wang, R. (2010). *Mol. Biol. Rep.* **37**, 995–1001.
Linthorst, H. J. M., Danhash, N., Brederode, F. T., Van Kan, J. A. L., De Wit, P. J. G. M. & Bol, J. F. (1991). *Mol. Plant Microb. Interact.* **4**, 586–592.
Liu, Q., Liu, Q. & Hendrickson, W. A. (2013). *Acta Cryst.* D**69**, 1314–1332.
Looze, Y., Boussard, P., Huet, J., Vandenbusche, G., Vandenbussche, G., Azarkan, M., Raussens, V. & Wintjens, R. (2009). *Phytochemistry*, **70**, 970–978.
Lu, H.-C., Lin, J.-H., Chua, A. C. N., Chung, T.-Y., Tsai, I.-C., Tzen, J. T. C. & Chou, W.-M. (2012). *Plant Physiol. Biochem.* **56**, 1–13.
Ludvigsen, S. & Poulsen, F. M. (1992*a*). *Biochemistry*, **31**, 8771–8782.
Ludvigsen, S. & Poulsen, F. M. (1992*b*). *Biochemistry*, **31**, 8783–8789.
Meinhart, A., Alings, C., Sträter, N., Camacho, A. G., Alonso, J. C. & Saenger, W. (2001). *Acta Cryst.* D**57**, 745–747.
Merritt, E. A. & Bacon, D. J. (1997). *Methods Enzymol.* **277**, 505–524.
Ming, R. *et al.* (2008). *Nature (London)*, **452**, 991–996.
Moynie, L., Schnell, R., McMahon, S. A., Sandalova, T., Boulkerou, W. A., Schmidberger, J. W., Alphey, M., Cukier, C., Duthie, F., Kopec, J., Liu, H., Jacewicz, A., Hunter, W. N., Naismith, J. H. & Schneider, G. (2013). *Acta Cryst.* F**69**, 25–34.
Mueller-Dieckmann, C., Panjikar, S., Tucker, P. A. & Weiss, M. S. (2005). *Acta Cryst.* D**61**, 1263–1272.
Mueller-Dieckmann, C., Polentarutti, M., Djinovic Carugo, K., Panjikar, S., Tucker, P. A. & Weiss, M. S. (2004). *Acta Cryst.* D**60**, 28–38.
Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Cryst.* D**67**, 355–367.
Neuhaus, J.-M., Fritig, B., Linthorst, H. J. M., Meins, F., Mikkelsen, J. D. & Ryals, J. (1996). *Plant Mol. Biol. Rep.* **14**, 102–104.
Ngan, C.-H., Hall, D. R., Zerbe, B., Grove, L. E., Kozakov, D. & Vajda, S. (2012). *Bioinformatics*, **28**, 286–287.
Niderman, T., Genetet, I., Bruyère, T., Gees, R., Stintzi, A., Legrand, M., Fritig, B. & Mösinger, E. (1995). *Plant Physiol.* **108**, 17–27.
Pape, T. & Schneider, T. R. (2004). *J. Appl. Cryst.* **37**, 843–844.
Perrakis, A., Morris, R. & Lamzin, V. S. (1999). *Nature Struct. Biol.* **6**, 458–463.
Ponstein, A. S., Bres-Vloemans, S. A., Sela-Buurlage, M. B., van den Elzen, P. J., Melchers, L. S. & Cornelissen, B. J. (1994). *Plant Physiol.* **104**, 109–118.
Potter, S., Uknes, S., Lawton, K., Winter, A. M., Chandler, D., DiMaio, J., Novitzky, R., Ward, E. & Ryals, J. (1993). *Mol. Plant Microb. Interact.* **6**, 680–685.
Ramagopal, U. A., Dauter, M. & Dauter, Z. (2003). *Acta Cryst.* D**59**, 1020–1027.
Richardson, J. S. (1981). *Adv. Protein Chem.* **34**, 167–339.
Ryang, S.-H., Chung, S.-Y., Lee, S.-H., Cha, J.-S., Kim, H. Y. & Cho, T.-J. (2002). *Biochem. Biophys. Res. Commun.* **299**, 352–359.
Sarma, G. N. & Karplus, P. A. (2006). *Acta Cryst.* D**62**, 707–716.
Schneider, T. R. & Sheldrick, G. M. (2002). *Acta Cryst.* D**58**, 1772–1779.
Sheldrick, G. M. (2002). *Z. Kristallogr.* **217**, 644–650.
Sheldrick, G. M. (2008). *Acta Cryst.* A**64**, 112–122.
Sheldrick, G. M. (2010). *Acta Cryst.* D**66**, 479–485.
Sigrist, C. J., de Castro, E., Cerutti, L., Cuche, B. A., Hulo, N., Bridge, A., Bougueleret, L. & Xenarios, I. (2013). *Nucleic Acids Res.* **41**, D344–D347.
Stanford, A., Bevan, M. & Northcote, D. (1989). *Mol. Gen. Genet.* **215**, 200–208.
Svensson, B., Svendsen, I., Højrup, P., Roepstorff, P., Ludvigsen, S. & Poulsen, F. M. (1992). *Biochemistry*, **31**, 8767–8770.
Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
Vagin, A. & Teplyakov, A. (2010). *Acta Cryst.* D**66**, 22–25.
Van Parijs, J., Broekaert, W. F., Goldstein, I. J. & Peumans, W. J. (1991). *Planta*, **183**, 258–264.
Wang, B.-C. (1985). *Methods Enzymol.* **115**, 90–112.
Wang, J., Dauter, M. & Dauter, Z. (2006). *Acta Cryst.* D**62**, 1475–1483.
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.
Winn, M. D. *et al.* (2011). *Acta Cryst.* D**67**, 235–242.
Wintjens, R., Belrhali, H., Clantin, B., Azarkan, M., Bompard, C., Baeyens-Volant, D., Looze, Y. & Villeret, V. (2006). *J. Mol. Biol.* **357**, 457–470.
Yang, C. & Pflugrath, J. W. (2001). *Acta Cryst.* D**57**, 1480–1490.
Zwart, P. H. (2005). *Acta Cryst.* D**61**, 1437–1448.