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

Joëlle Huet, Emmanuel Jean Teinkela Mbosso, Sameh Soror, Franck Meyer, Yvan Looze, René Wintjens, Alexandre Wohlkönig

MATERIALS AND METHODS

Materials

Methylmethanethiolsufonate (MMTS), dithiothreitol (DTT) and acrylamide were purchased from Aldrich-Chemie (Steinheim, Germany). SP-Sepharose Fast Flow was purchased from Amersham Biosciences GE Health-care (Uppsala, Sweden) and thiophilic gel from Merck (Darmstald, Germany). Samples of spray-dried *Carica papaya* latex were obtained from E.N.R.A. (Kivu, Democratic Republic of Congo, ex-Zaire).

Stabilisation of the papaya extract

Spray-dried papaya latex (5g) was dissolved at room temperature in water (50 ml) containing 223 mg EDTA (10 mM), 50 mM sodium acetate and 23 mg DTT (2.5 mM) at pH 5.0. After activation of the latex proteases for 30 min, 200 µl of MMTS (25 mM) were added and incubated for further 30 min. The solution was then submitted to an exhaustive dialysis against 50 mM sodium acetate at pH 5.0 before centrifugation (30,000 x g, 30 min, 4° C). The resulting solution contained 2 g of soluble papaya proteins

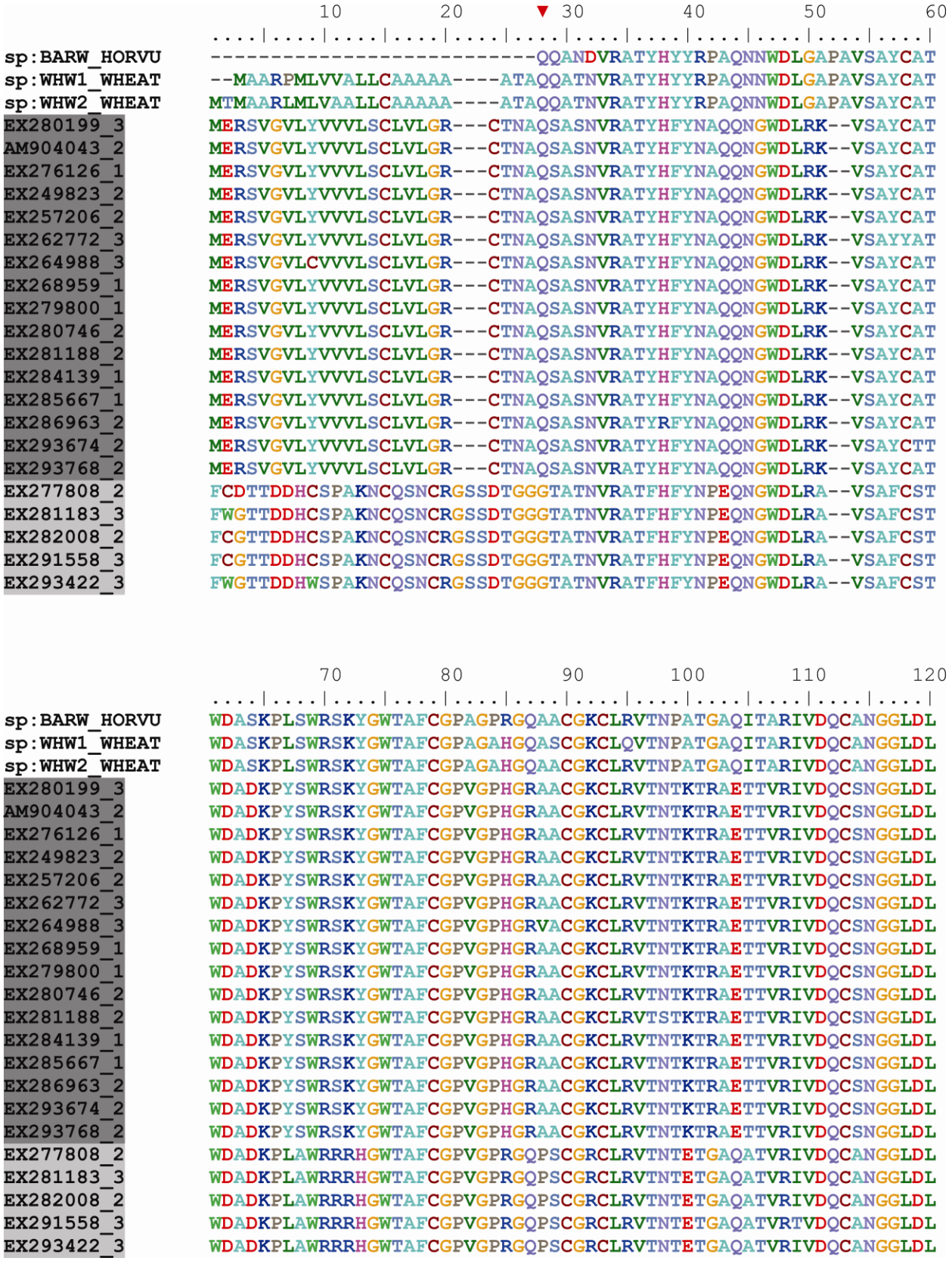
Purification of the latex papaya carwin protein

The papaya protein sample in 50 mM sodium acetate buffer at pH 5.0 (starting buffer) was applied on a (40 × 2.6 cm inner diameter (ID)) column of SP-sepharose-FF pre-equilibrated with the starting buffer. Elution was performed at a flow rate of 45 ml/h at room temperature by applying a linear gradient in conductivity (from 0 to 100 mS/cm) by addition of NaCl to the starting buffer (total volume: 2000 ml). This first chromatographic step provided four major fractions where the third one contained the protein of interest, the *papaya* barwin-like protein or carwin. Fractions containing the protein of interest were identified by SDS-page, pooled together and concentrated by ultrafiltration (cut-off 10 kDa).

The next step consisted on a second SP-sepharose-FF using similar conditions as described previously, excepted for the elution which was performed by applying a linear gradient from 20 to 120 mS/cm. The fractions containing the carwin were collected and concentrated by ultrafiltration, and subsequently submitted on the last step of purification protocol, a thiophilic interaction chromatography using T-gel support. The T-gel column (18 x 2.6 ID) was pre-equilibrated with 1.5 M ammonium sulphate and the elution was performed at flow rate of 45 ml/h at room temperature by applying a decreasing linear gradient from 1.5

to 0 M ammonium sulfate (total volume of 1600 ml). The final protein solution was dialyzed against water and concentrated to 12.5 mg/ml before crystallization screening (estimated percent solution extinction coefficient at 280 nm of 2.01 $(g/l)^{-1}$ cm^{-1}).

The chromatographic fractions were analysed by SDS-page experiments, carried out on precast gels (NuPAGE Novex 4-12% Bis-tris with MES SDS running buffer) using the X-cell SureLock Mini-cell from Invitrogen. The running conditions were 200 V, 50 mA at constant temperature 25° C during 45 minutes. A mixture of 12 pre-stained proteins from Invitrogen (Novex Sharp Protein Standard) was used as molecular weight protein standards. Protein detection was performed using Coomassie blue staining (ethanol 40% (v/v), acetic acid 10% (v/v) and Coomassie Brillant R250 0.2% (w/v)).


10 20 30 40 50 60
sp:BARW_HORVU ----------QQANDVRATYHYYRPAQNNWDLGAPAVSAYCAT
sp:WHW1_WHEAT --MAARPMLVVALLCAAAAA----ATAQQATNVRATYHYYRPAQNNWDLGAPAVSAYCAT
sp:WHW2_WHEAT MTMAARLMLVAALLCAAAAA----ATAQQATNVRATYHYYRPAQNNWDLGAPAVSAYCAT
EX280199_3 MERSVGVLYVVVLSCLVLGR---CTNAQSASNVRATYHFYNAQQNGWDLRK--VSAYCAT
AM904043_2 MERSVGVLYVVVLSCLVLGR---CTNAQSASNVRATYHFYNAQQNGWDLRK--VSAYCAT
EX276126_1 MERSVGVLYVVVLSCLVLGR---CTNAQSASNVRATYHFYNAQQNGWDLRK--VSAYCAT
EX249823_2 MERSVGVLYVVVLSCLVLGR---CTNAQSASNVRATYHFYNAQQNGWDLRK--VSAYCAT
EX257206_2 MERSVGVLYVVVLSCLVLGR---CTNAQSASNVRATYHFYNAQQNGWDLRK--VSAYCAT
EX262772_3 MERSVGVLYVVVLSCLVLGR---CTNAQSASNVRATYHFYNAQQNGWDLRK--VSAYYAT
EX264988_3 MERSVGVLCVVVLSCLVLGR---CTNAQSASNVRATYHFYNAQQNGWDLRK--VSAYCAT
EX268959_1 MERSVGVLYVVVLSCLVLGR---CTNAQSASNVRATYHFYNAQQNGWDLRK--VSAYCAT
EX279800_1 MERSVGVLYVVVLSCLVLGR---CTNAQSASNVRATYHFYNAQQNGWDLRK--VSAYCAT
EX280746_2 MERSVGVLYVVVLSCLVLGR---CTNAQSASNVRATYHFYNAQQNGWDLRK--VSAYCAT
EX281188_2 MERSVGVLYVVVLSCLVLGR---CTNAQSASNVRATYHFYNAQQNGWDLRK--VSAYCAT
EX284139_1 MERSVGVLYVVVLSCLVLGR---CTNAQSASNVRATYHFYNAQQNGWDLRK--VSAYCAT
EX285667_1 MERSVGVLYVVVLSCLVLGR---CTNAQSASNVRATYHFYNAQQNGWDLRK--VSAYCAT
EX286963_2 MERSVGVLYVVVLSCLVLGR---CTNAQSASNVRATYRFYNAQQNGWDLRK--VSAYCAT
EX293674_2 MERSVGVLYVVVLSCLVLGR---CTNAQSASNVRATYHFYNAQQNGWDLRK--VSAYCTT
EX293768_2 MERSVGVLYVVVLSCLVLGR---CTNAQSASNVRATYHFYNAQQNGWDLRK--VSAYCAT
EX277808_2 FCDTTDDHCSPAKNCQSNCRGSSDTGGGTATNVRATFHFYNPEQNGWDLRA--VSAFCST
EX281183_3 FWGTTDDHCSPAKNCQSNCRGSSDTGGGTATNVRATFHFYNPEQNGWDLRA--VSAFCST
EX282008_2 FCGTTDDHCSPAKNCQSNCRGSSDTGGGTATNVRATFHFYNPEQNGWDLRA--VSAFCST
EX291558_3 FCGTTDDHCSPAKNCQSNCRGSSDTGGGTATNVRATFHFYNPEQNGWDLRA--VSAFCST
EX293422_3 FWGTTDDHWSPAKNCQSNCRGSSDTGGGTATNVRATFHFYNPEQNGWDLRA--VSAFCST
70 80 90 100 110 120
sp:BARW_HORVU WDASKPLSWRSKYGWTAFCGPAGPRGQAACGKCLRVTNPATGAQITARIVDQCANGGLDL
sp:WHW1_WHEAT WDASKPLSWRSKYGWTAFCGPAGAHGQASCGKCLQVTNPATGAQITARIVDQCANGGLDL
sp:WHW2_WHEAT WDASKPLSWRSKYGWTAFCGPAGAHGQAACGKCLRVTNPATGAQITARIVDQCANGGLDL
EX280199_3 WDADKPYSWRSKYGWTAFCGPVGPHGRAACGKCLRVTNTKTRAETTVRIVDQCSNGGLDL
AM904043_2 WDADKPYSWRSKYGWTAFCGPVGPHGRAACGKCLRVTNTKTRAETTVRIVDQCSNGGLDL
EX276126_1 WDADKPYSWRSKYGWTAFCGPVGPHGRAACGKCLRVTNTKTRAETTVRIVDQCSNGGLDL
EX249823_2 WDADKPYSWRSKYGWTAFCGPVGPHGRAACGKCLRVTNTKTRAETTVRIVDQCSNGGLDL
EX257206_2 WDADKPYSWRSKYGWTAFCGPVGPHGRAACGKCLRVTNTKTRAETTVRIVDQCSNGGLDL
EX262772_3 WDADKPYSWRSKYGWTAFCGPVGPHGRAACGKCLRVTNTKTRAETTVRIVDQCSNGGLDL
EX264988_3 WDADKPYSWRSKYGWTAFCGPVGPHGRVACGKCLRVTNTKTRAETTVRIVDQCSNGGLDL
EX268959_1 WDADKPYSWRSKYGWTAFCGPVGPHGRAACGKCLRVTNTKTRAETTVRIVDQCSNGGLDL
EX279800_1 WDADKPYSWRSKYGWTAFCGPVGPHGRAACGKCLRVTNTKTRAETTVRIVDQCSNGGLDL
EX280746_2 WDADKPYSWRSKYGWTAFCGPVGPHGRAACGKCLRVTNTKTRAETTVRIVDQCSNGGLDL
EX281188_2 WDADKPYSWRSKYGWTAFCGPVGPHGRAACGKCLRVTSTKTRAETTVRIVDQCSNGGLDL
EX284139_1 WDADKPYSWRSKYGWTAFCGPVGPHGRAACGKCLRVTNTKTRAETTVRIVDQCSNGGLDL
EX285667_1 WDADKPYSWRSKYGWTAFCGPVGPHGRAACGKCLRVTNTKTRAETTVRIVDQCSNGGLDL
EX286963_2 WDADKPYSWRSKYGWTAFCGPVGPHGRAACGKCLRVTNTKTRAETTVRIVDQCSNGGLDL
EX293674_2 WDADKPYSWRSKYGWTAFCGPVGPHGRAACGKCLRVTNTKTRAETTVRIVDQCSNGGLDL
EX293768_2 WDADKPYSWRSKYGWTAFCGPVGPHGRAACGKCLRVTNTKTRAETTVRIVDQCSNGGLDL
EX277808_2 WDADKPLAWRRRHGWTAFCGPVGPRGQPSCGRCLRVTNTETGAQATVRIVDQCANGGLDL
EX281183_3 WDADKPLAWRRRHGWTAFCGPVGPRGQPSCGRCLRVTNTETGAQATVRIVDQCANGGLDL
EX282008_2 WDADKPLAWRRRHGWTAFCGPVGPRGQPSCGRCLRVTNTETGAQATVRIVDQCANGGLDL
EX291558_3 WDADKPLAWRRRHGWTAFCGPVGPRGQPSCGRCLRVTNTETGAQATVRTVDQCANGGLDL
EX293422_3 WDADKPLAWRRRHGWTAFCGPVGPRGQPSCGRCLRVTNTETGAQATVRIVDQCANGGLDL

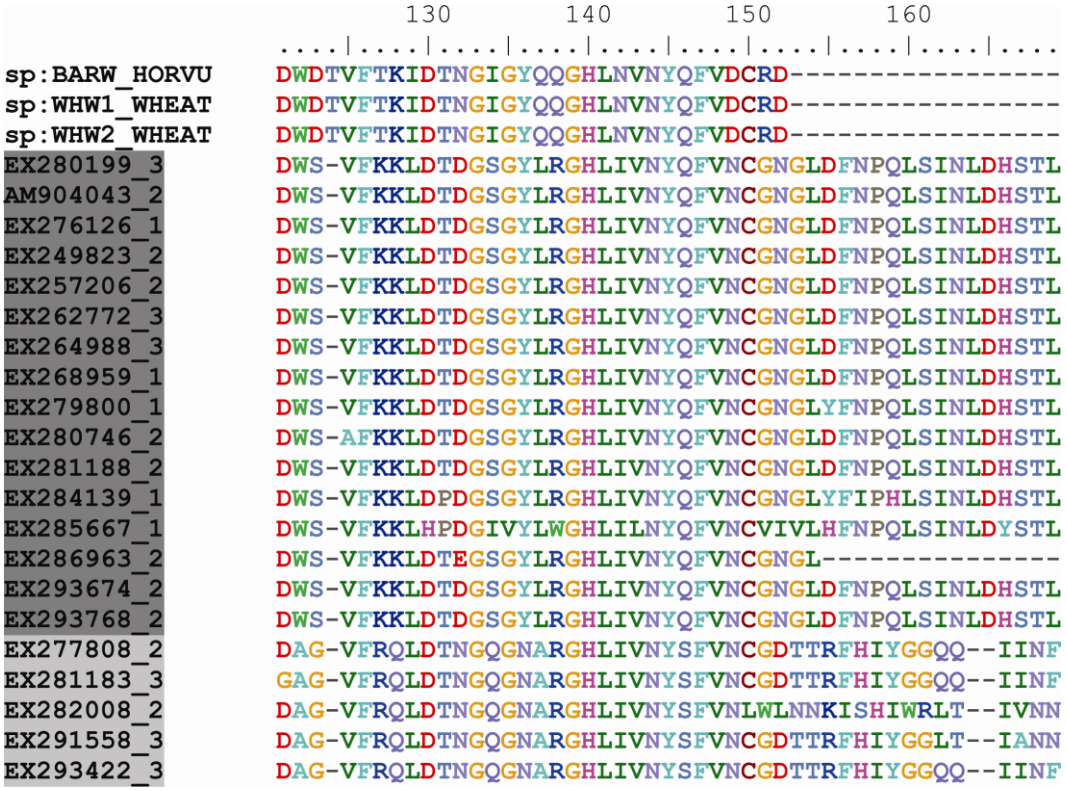


Figure S1. Multiple alignment of carwin sequences found in *Carica papaya* nucleic acid database. The three top sequences are the barwin protein (BARW_HORVU), wheatwin-1 and wheatwin-2 proteins (WHW1_WHEAT and WHW2_WHEAT). These sequences were taken from Swiss-Prot database (indicated by "sp"). The carwin sequences are labeled by their EMBL accession code with the last number giving the used open reading frame shift. Two different carwin isoforms were identified. Sequences of the first isoform (carwin-1) are highlighted in dark gray and those of the second (carwin-2) in light gray. By similarity with the barwin protein, the expected N-terminal residue is indicated by a red arrow above the alignment.

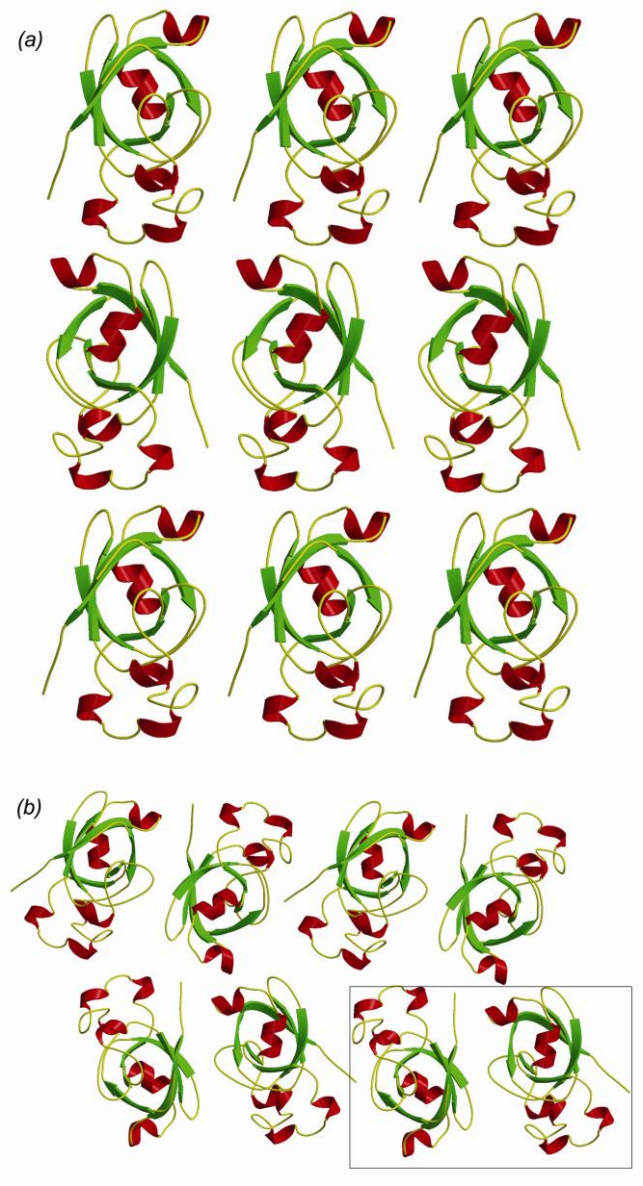


Figure S2. Pictures of the crystal packing observed in the two crystal forms of carwin. (a) crystal packing in the monomeric form. (b) crystal packing in the dimeric form. The rotation-translation matrices for the C superimposition of one chain between the two crystals (rmsd = 0.4Å) are the following.

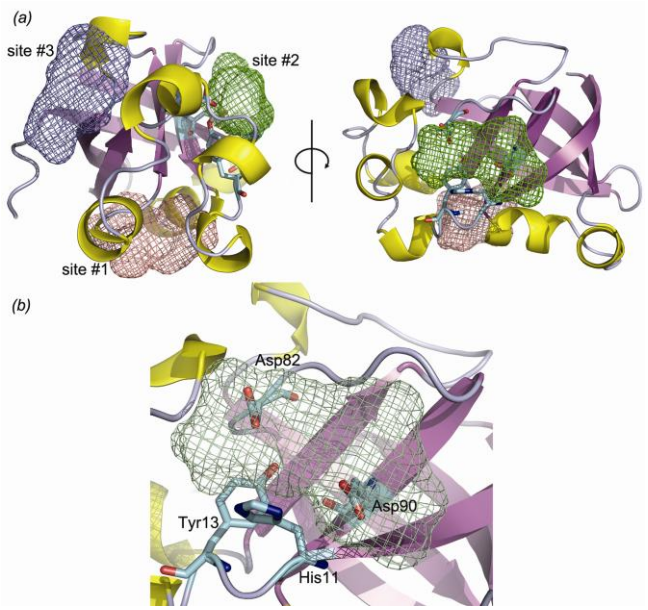


Figure S3. The top three binding sites on the carwin surface. (a) Two orientations of the protein model (stands in magenta and helices in yellow) and the top three identified binding sites as mesh surfaces coloured as following: in pink the site #1, in green site #2 and in blue the site #3 (see Table S2 for details on residues involved in the binding sites). Results were obtained with FTsite program (Ngan *et al.*, 2012). (b) Close up view of putative binding site # 2, side chains of important residues are shown in cyan.

Table S1. Data collection statistics for the 16 runs used in S-SAD phasing.

Space group number = 19

Unit-cell constants= 27.54Å, 54.86Å, 73.40Å, 90.000°, 90.000°, 90.000°

Oscillation range= 0.500000°

X-ray wavelength= 1.541800 Å

Resolution range= 20.0-1.81 (1.92-1.81)

	RUN1	RUN2	RUN3	RUN4
Number of frames	180	180	180	180
Kappa (°)	0	0	0	0
Phi (°)	0	90	180	270
Observed reflections	33656(3341)	33952(3388)	33604(3282)	33835(3361)
Multiplicity	2.3(1.8)	2.3(1.8)	2.3(1.8)	2.3(1.8)
Completeness (%)	73.8(58.9)	76.1(60.3)	73.7(57.6)	76.1(59.3)
Rmeas	2.1(3.7)	2.2(4.7)	2.3(4.3)	1.8(3.4)
I/σ(I)	42.68(24.28)	39.87(21.04)	39.73(22.74)	48.56(27.42)
Anomal Corr(%)	24(6)	27(19)	25(21)	35(28)
SigAno	1.060(0.921)	1.112(0.993)	1.060(1.025)	1.215(1.049)

	RUN5	RUN6	RUN7	RUN8
Number of frames	180	180	180	180
Kappa (°)	-10	-5	-10	-10
Phi (°)	0	90	180	270
Observed reflections	33626(3324)	337891(3352)	33683(3304)	33784(3336)
Multiplicity	2.4(1.8)	2.4(1.9)	2.4(1.8)	2.3(1.7)
Completeness (%)	72.8(59.4)	72.2(57.9)	72.4(57.8)	76.6(63.6)
Rmeas	2.1(4.2)	2.7(5.9)	2.5(4.9)	2.2(4.6)
I/σ(I)	42.22(23.16)	34.96(17.07)	35.76(19.37)	39.94(20.45)
Anomal Corr(%)	23(11)	21(13)	16(12)	24(22)
SigAno	1.040(0.972)	1.032(0.939)	0.964(0.879)	1.074(1.041)

	RUN9	RUN10	RUN11	RUN12
Number of frames	180	180	180	180
Kappa (°)	-10	-10	-5	-5
Phi (°)	0	90	180	270
Observed reflections	33709(3361)	33891(3334)	33685(3322)	33790(3318)
Multiplicity	2.4(1.8)	2.5(1.9)	2.4(1.9)	2.3(1.8)
Completeness (%)	73.0(60.0)	68.7(56.5)	71.6(56.5)	75.2(60.3)
Rmeas	2.2(4.0)	2.2(4.6)	2.5(4.8)	2.1(4.5)
I/σ(I)	41.62(22.52)	34.08(16.15)	37.12(20.14)	42.10(21.90)
Anomal Corr(%)	19(16)	18(8)	19(14)	28(20)
SigAno	1.018(0.946)	0.987(0.910)	0.985(0.933)	1.109(1.066)

	RUN13	RUN14	RUN15	RUN16
Number of frames	180	180	180	180
Kappa (°)	-15	-15	-15	-15
Phi (°)	0	90	180	270
Observed reflections	33778(3378)	33966(3393)	33768(3373)	33900(3358)
Multiplicity	2.2(1.7)	2.4(1.9)	2.3(1.8)	2.2(1.6)
Completeness (%)	77.0(63.8)	72.2(58.7)	75.4(60.9)	79.1(66.9)
Rmeas	2.2(4.4)	2.9(7.1)	2.7(5.4)	2.3(4.7)
I/σ(I)	39.62(21.04)	32.01(15.10)	33.08(17.27)	38.20(19.40)
Anomal Corr(%)	19(12)	17(11)	15(17)	25(14)
SigAno	1.013(0.918)	0.976(0.949)	0.963(0.939)	1.074(0.967)

All data were processed with the XDS program package (Kabsch, 2010).

Rmeas = redundancy independent R-factor (intensities) (Diederichs & Karplus, 1997, Nature Struct. Biol. 4, 269-27.

I/σ(I) = mean of intensity/Sigma(I) of unique reflections (after merging symmetry-related observations)

Anomal Corr = percentage of correlation between random half-sets of anomalous intensity differences. Correlation significant at the 0.1% level is marked.

SigAno = mean anomalous difference in units of its estimated standard deviation (|F(+)-F(-)|/Sigma). F(+), F(-) are structure factor estimates obtained from the merged intensity observations in each parity class.

Table S2. Binding site prediction for the *papaya* carwin protein[a]

Site #	Binding site residues with FTsite	Binding site residues with metaPocket
1	Asn14, Ala15, Gln16, Gln17, Trp20, Ser42, Lys43, Tyr44, Gly45, Trp46, Thr76, Thr77, Val78, Asp92, Ser94, Val95, Lys98	Phe12, Tyr13, Asn14, Ala15, Gln16, Gln17, Trp20, Ser42, Lys43, Tyr44, Gly45, Trp46, Asn69, Thr70, Lys71, Trh72, Thr76, Thr77, Val78, Asp92, Trp93, Ser94, Val95, Lys98, Asp102, Ser104, Arg108, Leu111, Ile112
2	Thr9, Tyr10, His11, Tyr13, Val25, Ser26, Ala27, Tyr28, Ala48, Asp82, Gln83, Cys84, Ser85, Asn86, Asp90	His11, Tyr13, Asp82, Asp90
3	Trp32, Asp33, Lys36, Trp40, Arg58, Ala60, Cys61, Gly62, Lys63, Arg79, Cys120	Ser39, Trp40, Lys43, Arg79, Gly121, Asn122

[a] The top 3 binding sites were predicted by FTsite (Ngan *et al.*, 2012) and metaPocket (Huang, 2009) programs. Residues in common for both programs are underlined.

Table S3. Results from DALI searching against PDB with the carwin structure

Protein[a]	Nres[b]	%id[c]	Z-score[d]	rmsd	β-barrel type	# ψ loop[e]	# SS bond[f]	PDB code[g]
Carwin	122	100	-	-	2,2,-1,-2,-2	4	3	4jp6, 4jp7
Barwin	125	72	14.4	2.6	4,-2,1,-2x	1	3	1bw3, 1bw4
Cerato-platanin-like	135/156	19	12.9	2.1	2,2,-1,-2,-2	4	2	3sul, 3suk, 3sum
Cerato-platanin	120	22	11.3	2.1	2,2,-1,-2,4	3	2	2kqa
EPL1	138	16	13.0	2.2	2,2,-1,-2,-2,6x	4	2	3m3g
Expansin-YoaJ (N-domain)	235 (105)	23	10.9	2.5	2,2,-1,-2,-2	4	0	4fer, 2bh0, 3d30
Phl p1 allergen (N-domain)	263 (106)	19	10.2	3.0	2,2,-1,-2,-2	4	3	1n10
PA4485	125	21	10.1	2.3	2,2,-1,-2,-2	4	0	4avr
Zea m1 (N-domain)	269 (106)	21	10.0	3.0	2,2,-1,-2,-2	4	3	2hcz
ATPase p97 (N-domain)	806 (84)	12	6.9	2.1	2,2,-1,-2,-2	4	0	3tiw, 1s3s, 3qq7
Endoglucanase	214	15	7.0	2.5	1x,1,2,-1,-2,-2	3	6	3eng, 1l8f, 1wc2
Endo-1,4-β-glucanase	181	22	6.9	3.0	1,-2,-2,1,2,2	4	6	1xc2
Formate dehydrogenase H (C-domain)	715 (115)	8	6.7	2.8	2,2,-1,-2,-2	4	0	1fdi, 2iv2, 1aa6
Arsenite oxidase (M-domain)	826 (115)	10	6.7	3.5	2,2,-1,-2,-2	4	0	1g8j, 1g6k, 1g8k
Ethylbenzene dehydrogenase (M-domain)	976 (135)	8	6.6	3.7	2,2,-1,-2,-2	4	0	2ivf
RNA polymerase II (M-domain)	1733 (75)	10	6.2	2.7	2,2,-1,-2,-2,2x	3	0	2ja5, 1nik, 3gtp
Nitrate reductase (M-domain)	1247 (143)	10	6.3	2.8	2,2,-1,-2,-2	4	0	1r27, 2jir, 1y4z

[a] protein name of matched structures. In case of a multidomain protein, the domain folded into the double-ψ β-barrel is indicated as N-terminal, C-terminal or M-domain (M- for middle).
[b] the number of amino acid residues for the complete protein and in parenthesis the number of residues for the domain folded in double-ψ β-barrel.
[c] mean sequence identity of the aligned positions between carwin and the matched structures.
[d] mean Z-score of the matches according DALI.
[e] number of ψ-loop in the matched proteins according ProMotif.
[f] number of disulfide bridges.
[g] at least three PDB codes of the matched protein if available.