

Crystal Structure of a Papain-fold Protein Without the Catalytic Residue: A Novel Member in the Cysteine Proteinase Family

Min Zhang^{1,2,3}, Zhiyi Wei^{1,2}, Shaojie Chang², Maikun Teng²
and Weimin Gong^{1,2*}

¹National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101 P.R.China

²School of Life Sciences, University of Science and Technology of China, Hefei Anhui 230026, P.R.China

³School of Life Sciences, Anhui University, Hefei, Anhui P.R.China

A 31 kDa cysteine protease, SPE31, was isolated from the seeds of a legume plant, *Pachyrhizus erosus*. The protein was purified, crystallized and the 3D structure solved using molecular replacement. The cDNA was obtained by RT PCR followed by amplification using mRNA isolated from the seeds of the legume plant as a template. Analysis of the cDNA sequence and the 3D structure indicated the protein to belong to the papain family. Detailed analysis of the structure revealed an unusual replacement of the conserved catalytic Cys with Gly. Replacement of another conserved residue Ala/Gly by a Phe sterically blocks the access of the substrate to the active site. A polyethyleneglycol molecule and a natural peptide fragment were bound to the surface of the active site. Asn159 was found to be glycosylated. The SPE31 cDNA sequence shares several features with P34, a protein found in soybeans, that is implicated in plant defense mechanisms as an elicitor receptor binding to syringolide. P34 has also been shown to interact with vegetative storage proteins and NADH-dependent hydroxypyruvate reductase. These roles suggest that SPE31 and P34 form a unique subfamily within the papain family. The crystal structure of SPE31 complexed with a natural peptide ligand reveals a unique active site architecture. In addition, the clear evidence of glycosylated Asn159 provides useful information towards understanding the functional mechanism of SPE31/P34.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: crystal structure; elicitor receptor; plant defense; papain glycoprotein

*Corresponding author

Introduction

Cysteine proteases constitute an important class of proteolytic enzymes found in prokaryotes as well as eukaryotes like plants and animals. Structurally they belong to the papain family where individual members accomplish catalysis employing a common mechanism involving the highly conserved catalytic triad (Cys-His-Asn) residues.^{1–5} Although the overall topological features and mechanism of catalysis are conserved, unique active site residues

define substrate specificities for individual members.

Protein SPE31 isolated from the seeds of the legume *Pachyrhizus erosus* was designated a papain-like protein based on N-terminal amino acid sequencing results.⁶ Analysis of the sequence alignments for SPE31 identified P34, a protein commonly found in soybeans, as sharing 78% sequence identity. A striking feature of the catalytic site common to both SPE31 and P34, a glycine replacing the highly conserved catalytic cysteine, makes both these proteins unique members of the papain family.^{7,8} It was suggested that another cysteine residue close to the active site of P34 participates in the catalysis by mounting a nucleophilic attack upon deprotonation.⁷ However, so far, no enzymatic activity has been reported for P34. P34 was also recognized as a major soybean allergen protein accounting for IgE binding.⁹ Binding of P34

Abbreviations used: PEG, polyethyleneglycol; VSP, vegetative storage proteins; HPR, NADH-dependent hydroxypyruvate reductase; GP-II, ginger protease II; N-GlcNAc, N-acetyl-D-glucosamine N-linked to Asn; GlcNAc, N-acetyl-D-glucosamine; Fuc, fucose.

E-mail address of the corresponding author: wgong@sun5.ibp.ac.cn

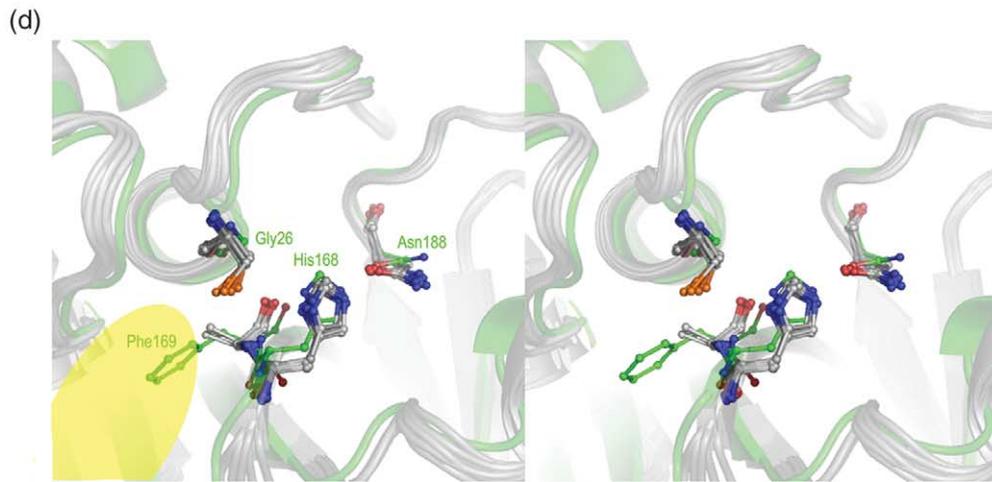
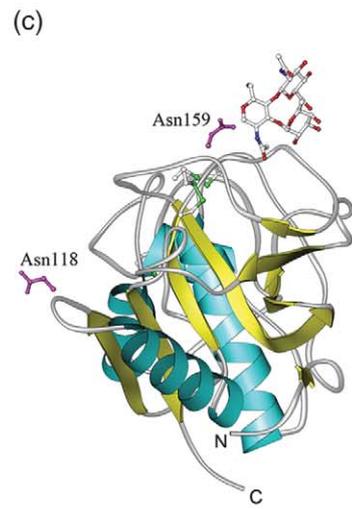
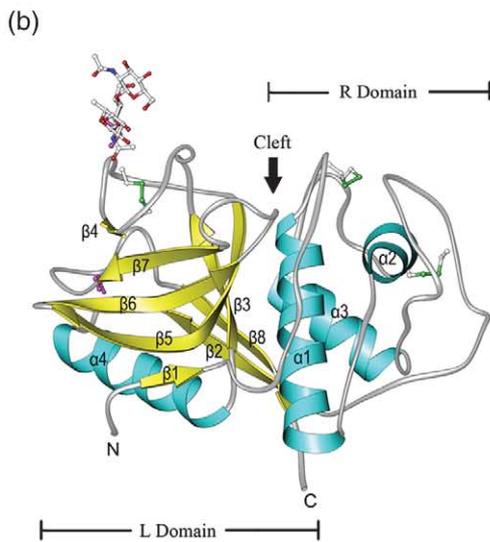
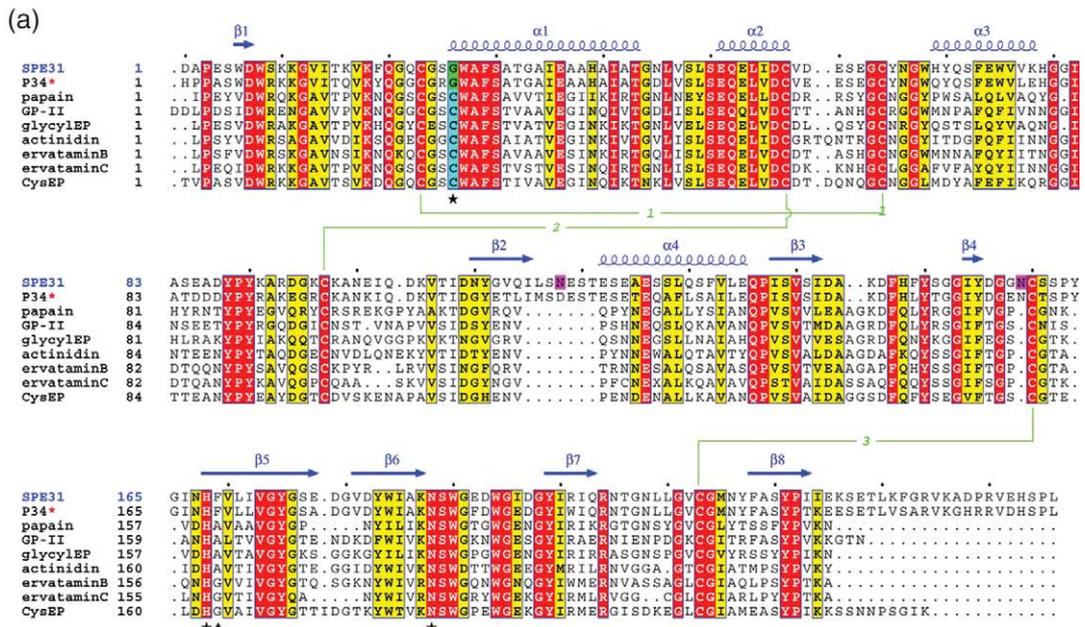


Figure 1. (a) Multiple-sequence alignment of SPE31 and P34 against cysteine proteases with known crystal structures. The sequences of SPE31 from *Pachyrhizus erosus*, P34 (Swiss-Prot accession number: P22895) from soybean, papain

to syringolide and interaction with VSP and a NADH-dependent hydroxypyruvate reductase implicated P34 as playing an important role in plant defenses against bacteria.^{10–12} These evidences suggest SPE31 and P34 may represent a new subfamily of cysteine proteinases.

The novel replacement of conserved catalytic cysteine by a glycine prompted us to undertake a 3D structural investigation of the active site architecture. Here we report the crystal structures of SPE31 bound with a polyethylene glycol (PEG) molecule and a natural peptide. The structures are named SPE31_PEG and SPE31_PEP, respectively. Detailed analysis of the structures clearly shows the absence of cysteine residues around the active site. The ligand-bound structure solutions suggest possible sites for the interaction of P34 and syringolide or other proteins.

Results

The cDNA sequence of SPE31

The cDNA encoding SPE31 is 951 base-pairs in length with a poly(A)⁺ tail. The open reading frame consists of 738 base-pairs encoding 246 amino acid residues and is predicted to express a 27 kDa protein with two N-glycosylation sites (Asn118 and Asn159). The SPE31 amino acid sequence alignment results indicated the protein to be papain-like (Figure 1(a)). SPE31 retains most of the conserved residues found in and around the active site of cysteine proteases. One striking feature that sets SPE31 apart from other cysteine proteases is the replacement of catalytic Cys25 with Gly26. The conserved catalytic triad Cys25-His159-Asn175 (papain numbering) found in normal cysteine proteases is replaced by Gly26-His168-Asn188 in SPE31. This modified catalytic

triad has previously been reported for another cysteine protease, P34.

Overall structure of SPE31

The crystal structures of SPE31_PEG and SPE31_PEP were refined with good crystallographic values (Table 1). The final model for both the structures contains one monomer per crystallographic asymmetric unit. There are no obvious differences in the fold between the two structures with a root mean square (r.m.s.) difference of 0.23 Å for all C α atom pairs. The overall fold is identical to those previously reported for papain-like proteins. SPE31 is made up of two domains (R domain and L domain) with a cleft between them (Figure 1(b)). The C-terminal region seems disordered with no clear electron density for residues 227–246 in SPE31_PEG and for residues 228–246 in SPE31_PEP. This suggests that the C terminus could be very flexible. The three disulfide bonds (Cys23–Cys64, Cys57–Cys97, Cys160–Cys213) conserved in the papain family are also seen in SPE31. In comparison to other members of the papain family, there is a seven residue insertion between β 2 and α 4 in SPE31 and P34 (Figure 1(a)). However, the insertion has little effect on the overall fold and is not involved in the formation of the active site.

Oligosaccharide chains of SPE31

SPE31 was reported as a glycoprotein previously.⁶ Two N-glycosylation sites, Asn118 and Asn159, were predicted using NetNGlyc[†]. Electron density for three glycosyl residues (N-GlcNAc, GlcNAc, Fuc) bound to Asn159 was observed (Figure 2). Similar glycosylation has been reported for cysteine protease GP-II; however, the site of glycosylation is Asn156. Asn159 in SPE31 and

[†] <http://www.cbs.dtu.dk/services/ne>

(PDB id: 1POP) and glycylEP (PDB id: 1GEC) from *Carica papaya*, actinidin (PDB id: 1AEC) from *Actinidia chinensis*, ervataminB (PDB id: 1IWD) and ervataminC (PDB id: 1O0E) from *Ervatamia coronaria*, CysEP (PDB id: 1S4V) from castor bean (*Ricinus communis*), GP-II (PDB id: 1CQD) from ginger (*Zingiber officinale*) are aligned using T-coffee.²¹ The secondary structure of SPE31, which is mainly defined by the analysis of the SPE31 structure using DSSP program,²² is indicated above the alignment. Identical and chemically similar residues in the alignment are boxed in red and yellow, respectively. The catalytic cysteine and the substituted glycine are emphasized in cyan and green box, respectively. The three conserved disulfide bonds are linked with a green line. The two predicted N-glycosylation sites in SPE31 are boxed in purple. The three residues of the catalytic triad are marked with black stars under the alignment. Phe169, which may be an obstacle for substrate binding, is indicated with a triangle. This picture was prepared using ESPrpt.²³ *It is worth mentioning that residue 64 of P34 is changed from a serine in the original P34 sequence (Swiss-Prot accession number: P22895) to a cysteine as seen in case of protein Bd 30k from soybean (Swiss-Prot accession number: O64458). (b) and (c) The overall structure of SPE31. SPE31 contains two domains, L domain and R domain, which form a cleft between them. L domain consists of eight β strands (β 1– β 8) and one α helix (α 4) while R domain is an all- α helix domain. All structural elements are labeled. The residues involved in the formation of the three conserved disulfide bonds are represented as stick-and-ball with disulfide bonds colored in green. The residues of the two N-glycosylation sites (stick-and-ball model) are colored in purple. The oligosaccharide chain linked to Asn159 is shown in stick-and-ball model. The view in (c) is rotated 90° around a vertical axis from the view in (b). These two pictures were prepared using Ribbons.²⁴ (d) Stereo view of superposition of the active sites of SPE31 and papain family members shown in (a) excluding P34. The SPE31 structure is colored in green while remaining structures are colored in white. The four residues that are indicated with black stars and triangle in (a) are shown as stick-and-ball model and labeled. The yellow shadow indicates the region occupied by the substrate in normal papain family members. Figures 1–3 were prepared using PyMOL (<http://pymol.sourceforge.net>).

Table 1. Data collection and structure refinement statistics

	SPE31_PEG	SPE31_PEP
<i>Data collection</i>		
Crystal name		
Space group	P4 ₁ 2 ₁ 2	P4 ₁ 2 ₁ 2
Unit cell parameters (Å)	$a=b=61.52, c=142.19$	$a=b=61.36, c=143.99$
Resolution range (Å)	50–2.0 (2.05–2.0)	50–2.4 (2.46–2.4)
No. of total reflections	240,100	100,736
No. of unique reflections	19,291	11,433
Completeness (%)	99.9 (99.9)	99.8 (99.5)
I/σ	34.8 (7.2)	26.5 (4.6)
R_{merge} (%) ^a	7.1 (38.6)	8.0 (50.4)
<i>Structure refinement</i>		
Resolution (Å)	22.12–2.0 (2.13–2.0)	22.35–2.4 (2.55–2.4)
$R_{\text{cryst}}/R_{\text{free}}$ (%) ^b	18.2/20.9 (19.5/24.4)	20.6/26.7 (25.8/32.9)
No. of reflections		
Working set	18,229	10,799
Test set	977	567
r.m.s.d. from ideal values		
Bond length (Å)	0.005	0.006
Bond angles (deg.)	1.1	1.2
Average B -factor (Å ²)		
Main-chain	21.5	32.3
Side-chain	24.5	33.7
Ligand	44.9	43.8
No. of atoms		
Protein	1735	1733
Carbohydrate	38	38
Ligand	20	32
Solvent	271	68
Ramachandran plot		
Most favored regions (%)	88.0	89.3
Additionally allowed (%)	12.0	10.7

Numbers in parentheses are for the highest resolution shell.

^a $R_{\text{merge}} = \sum |I_i - I_m| / \sum I_i$, where I_i is the intensity of the measured reflection and I_m is the mean intensity of all symmetry-related reflections.

^b $R_{\text{cryst}} = \sum ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum |F_{\text{obs}}|$, where F_{obs} and F_{calc} are observed and calculated structure factors. $R_{\text{free}} = \sum_T ||F_{\text{obs}}| - |F_{\text{calc}}|| / \sum_T |F_{\text{obs}}|$, where T is a test data set of about 5% of the total reflections randomly chosen and set aside prior to refinement.

Asn156 in GP-II are both located in similar regions, but are not correspondent to each other structurally. In addition, the glycosyl chain in GP-II was shown to participate in the crystal lattice arrangement,⁵ while no such involvement of the carbohydrate in intermolecular bond formation was observed for SPE31. While comparing the glycosylation of other papain family members, the sequence N¹⁵⁹C(S/T)SPYG¹⁶⁵ seems unique to SPE31 and P34 (Figure 1(a)), which indicates that P34 may have the same glycosylation site as SPE31.

The physiological role of this oligosaccharide needs further investigation.

Although Asn118 is also predicted to be a glycosylation site, the electron density for the carbohydrate is not well defined to identify the exact nature of the glycosyl structure (Figure 2). Another reason for the poor density could be that the oligosaccharide is disordered. The corresponding residue in P34 is not conserved, with an aspartic acid replacing the asparagine.

The active cleft of SPE31

In SPE31, Gly26 of the L domain replaces the conserved catalytic cysteine (Figure 1(d)). The substitution has no effect on the position of the remaining two residues (His168 and Asn188) of the catalytic triad (Figure 1(d)). Detailed analysis of the active site failed to locate any cysteine, serine or aspartic acid, in close proximity that could act as a potential nucleophile. In addition, Phe169 of SPE31 (also found in P34) replaces the highly conserved Ala/Gly found in papain and normal cysteine proteinases (Figure 1(a)). The aromatic side-chain of Phe169 extends into the active cleft and obstructs substrate binding (Figure 1(d)). The absence of catalytic cysteine and the substitution of conserved Ala/Gly with Phe169, prevents SPE31 from exhibiting a normal proteinase activity.

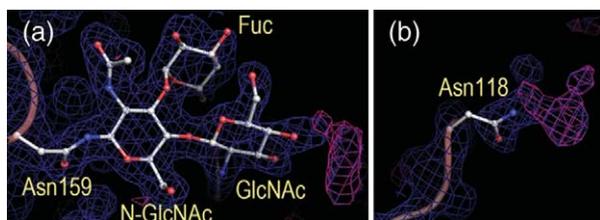


Figure 2. The N-linked oligosaccharide chains of SPE31. The electron density ($2F_o - F_c$ maps contoured at 1σ , blue) and the difference Fourier maps ($F_o - F_c$ maps contoured at 3σ , purple) surrounding the oligosaccharide chain linked to Asn159 and the possible disordered oligosaccharide chain linked to Asn118 are showed in (a) and (b), respectively. The structure of the oligosaccharide chain linked to Asn159 is depicted as stick-and-ball model.

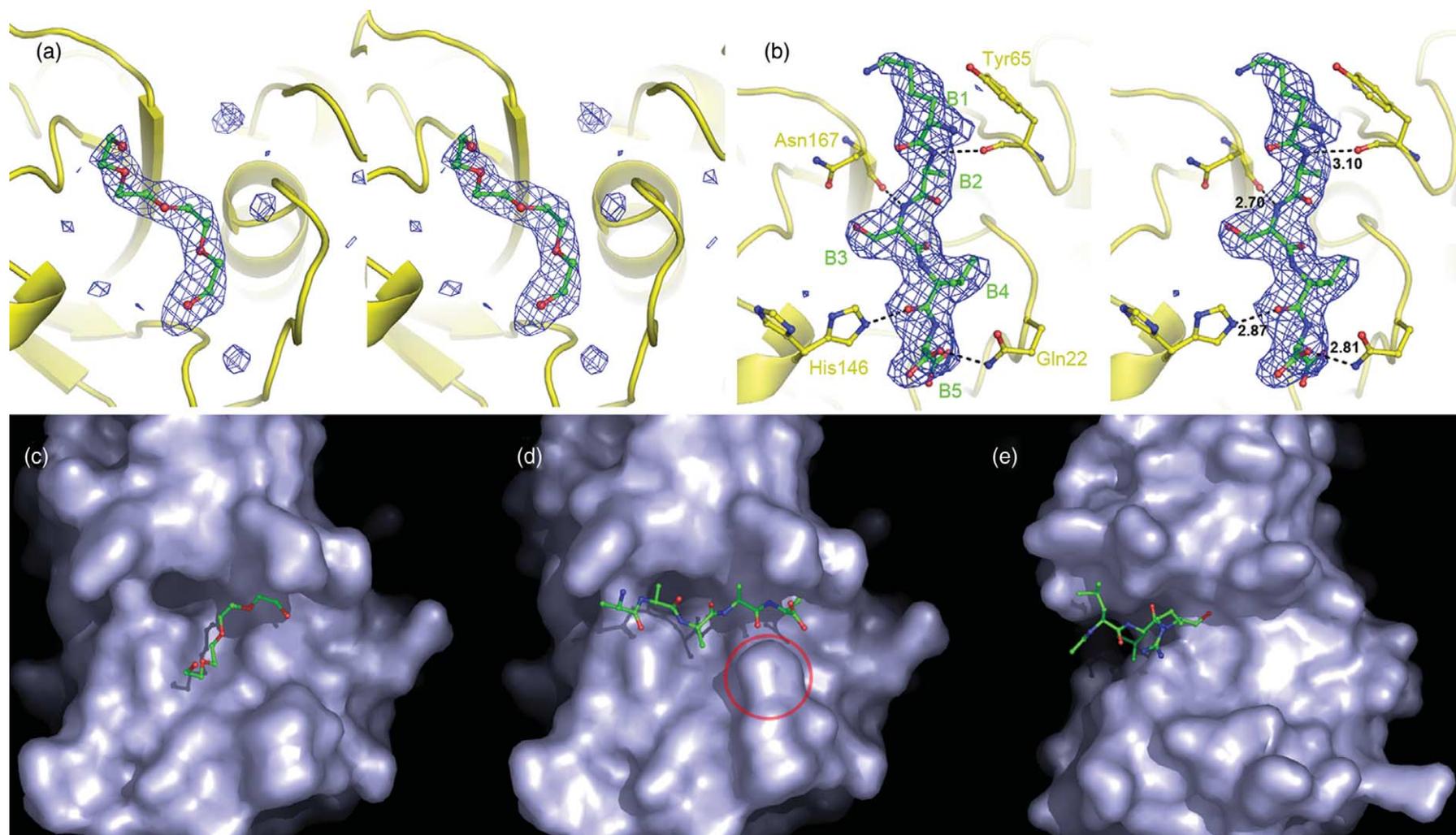


Figure 3. Omit maps of ligands bound to SPE31 and the mode of binding of the ligands to SPE31 and papain. The $F_o - F_c$ omit map was constructed by omitting the PEG molecule (a) and the peptide fragment (b) and was contoured at 3σ . (a) The stereo view of the PEG molecule binding to SPE31. (b) The stereo view of the peptide fragment binding to SPE31. B1, B2, B3, B4, and B5 indicate the five residues (KASVG) of the peptide fragment. The hydrogen bonds between the peptide fragment and SPE31 are indicated by broken lines and bond lengths (Å) are labeled. (c)–(e) Binding surfaces of the PEG, the peptide fragment in SPE31 and leupeptin (*N*-acetyl-L-leucyl-L-leucyl-L-argininal) in papain (PDB id: 1POP) are shown from left to right. The increase in the binding surface of the peptide fragment compared to that of the PEG molecule is indicated by a red circle, which is caused by the double-conformation of the side-chain of His148.

Binding of the ligands to the active cleft of SPE31

The structure of SPE31_PEG showed a PEG molecule bound to the surface of the “active cleft” (Figure 3(a)). A much longer electron density was observed for the structure of SPE31_PEP at the same location, when the crystals were grown in the absence of PEG. The density was later identified as a peptide fragment bound to the active cleft (Figure 3(b)). The intensity of the density of the peptide fragment was comparable to that of the protein, suggesting a 1:1 binding of the peptide to SPE31. Since no ligands were added during the purification or crystallization of the protein, the bound peptide seems to have originated from the seeds from where the protein was isolated. On the other hand, high concentrations of PEG used for crystallization of SPE31_PEG seem to have displaced the peptide from the active cleft. The peptide fragment was identified as KASVG at 2.4 Å resolution with the C-terminal glycine adopting a double-conformation (Figure 3(b)).

In case of typical cysteine proteinases, the substrate or the peptide-like inhibitor binds deeply inside the active cleft and has direct interaction with the catalytic cysteine. However, the SPE31 structure shows the PEG and the peptide, both bound at the surface and not able to stretch into the cleft (Figure 3(c)–(e)). Phe169 with a large side-chain in the substrate-binding pocket seems to sterically hinder the access of the ligands to the active site cleft. (Figure 1(d)). The binding of the peptide fragment to SPE31 was mediated primarily through hydrogen bonding and hydrophobic interactions (Figure 3(b)). In addition, the side-chain of His148 adopts a double-conformation in the SPE31_PEP structure. One conformation is similar to that found in SPE31_PEG, while in another conformation, a hydrogen bond is formed between the N^ε atom of the imidazole ring of the histidine and the oxygen atom of Val4 in the peptide fragment (Figure 3(b) and (d)). This interaction seems to be crucial for the tight binding of the peptide fragment. However, no hydrogen bonding was found between the PEG molecule and SPE31. The PEG bound to SPE31 through Van der Waals interactions.

Discussion

Comparison with P34

SPE31 shares 78% amino acid sequence identity with P34, a protein commonly found in the seeds of soybeans. In both the proteins the catalytic cysteine is replaced by a glycine. SPE31 and P34 are seed proteins of leguminous plants. Besides the unusual substitution of the crucial cysteine, additional differences in the primary amino acid sequence sets SPE31 and P34 apart from the other members of the papain family. The fact that the SPE31 cDNA could only be isolated from the mid maturation

seeds and not mature seeds or seedlings is consistent with similar observations reported for P34, which is also synthesized primarily during seed maturation. This expression pattern is different from that of other cysteine proteases of plant seed origin that are usually synthesized during seedling growth.⁷ In addition, SPE31 co-exists in seeds with another protein, SPE32, which is 11 amino acid residues longer at the N terminus when compared to SPE31,⁶ while P34 co-exists in soybeans with P32 which is 11 amino acid residues shorter than P34.⁸ Both the 11 amino acid residue sequences are very similar, with seven residues being identical. These unique features shared by SPE31 and P34 and not found amongst other cysteine proteases suggest that they form a separate group distinct from the other members of the papain family.

In P34, the catalytic cysteine is replaced by a glycine (Gly26). A neighboring cysteine residue, Cys23, was assumed to play the catalytic role, since no other Cys residue was found nearby that could form a disulfide linkage with Cys23.⁷ In SPE31, Cys23 forms a disulfide bridge with Cys64 as in papain; while in P34, the position 64 is occupied by a serine. This difference could possibly be attributed to an error in the cDNA sequencing results for P34. The coding triplets for a cysteine are UGU or UGC, while UCX codes for a serine. A misinterpretation of the second base of the coding triplet could have likely resulted in serine being reported for cysteine. In case of protein Bd 30k (the same protein as P34), the sequence reported by another group (T. Takano, & M. Yamada, Swiss-Prot accession number: O64458), shows position 64 to be occupied by a cysteine. The X-ray crystal structure of SPE31 rules out the possibility of the role of Cys23 as a nucleophile.

Binding sites for syringolide elicitor or other proteins

Syringolides are elicitors produced by pathogens, which upon contact elicit a hypersensitive response specifically in soybean plants carrying the *Rpg4* disease resistance gene.¹³ P34 shows specific syringolide binding activity and could very well function as a syringolide receptor in the signal perception of plant defenses.¹⁰ P34 could also associate with VSP proteins in the presence of syringolides¹⁰ and interact with NADH-dependent hydroxypyruvate reductase (HPR), which could be a secondary messenger to invoke defense response.¹¹ HPR is an important enzyme of the plant photorespiration system, which catalyzes the conversion of hydroxypyruvate to glycerate. It is assumed that the hypersensitive response results from the inhibition of the HPR activity due to interaction with syringolide-P34,¹¹ but the mechanism of the signal recognition and transmission is still unknown. Since P34 and SPE31 are highly homologous, the structure of SPE31 can be used to predict the possible binding sites for syringolide

elicitor or other proteins for the unique group composed by SPE31 and P34.

The structure of SPE31 provides a means to identify the possible binding sites for syringolide elicitors or other proteins. The results of sequence alignment and comparison of the structure suggest that SPE31 and P34 both have retained the original papain-fold but lost the catalytic activity by replacing the catalytic cysteine with glycine. The substitution, however, seems to have no effect on the protein binding activity, which has been retained during the course of evolution. Thus, the first site for binding of other proteins appears to be the surface on the cleft where the peptide was seen bound. The second binding site might be the conserved Asn159, which is glycosylated in SPE31, and may also be glycosylated in P34. Oligosaccharides are commonly involved in protein-protein or protein-ligand recognition. Structurally syringolides resemble a tri-ribose-like head with a long hydrophobic tail. Thus, the glycosylation site can be assumed to provide a reasonable binding site for syringolide. In addition, compared to other papain family members, the seven residue insertion between $\beta 2$ and $\alpha 4$ and the longer and flexible C terminus seen in both SPE31 and P34 might provide additional binding sites for syringolide elicitors or other proteins.

The sequence of the peptide fragment bound to SPE31, KASVG, is identical to the N-terminal sequence of the plant enzyme, ribulose 1,5-bisphosphate carboxylase/oxygenase (Rubisco). Like HPR, Rubisco is also a plant photorespiration enzyme, whose amino acid sequences from different origins are almost identical. It has been reported that the downstream products (glycerate and 3-phosphoglycerate) of HPR can inhibit the hypersensitive response.¹¹ 3-Phosphoglycerate happens to be one of the products of Rubisco catalysis. All the evidence above suggests that SPE31/P34 together with syringolide interact with photorespiration enzymes, Rubisco or HPR, and initiate a hypersensitive response in plants.

The crystal structure of SPE31 reported here and the functional studies on P34, strongly suggest that SPE31 and P34 represent a new subfamily within the papain family and play an important role in plant defense mechanism. This subfamily of proteins seems to have lost the proteinase activity by two key substitutions Cys/Gly and Gly(Ala)/Phe in the active site, but retain the papain fold as a scaffold for partner protein(s) binding.

Conclusions

We have determined the crystal structure of SPE31, which provides the first structural details for a novel group of proteins that possess a papain fold but lack the catalytic cysteine in the active site. Although SPE31 shows considerable homology both in sequence and structure to other members of the papain family, the structural differences are

quite remarkable. Besides the substitution of the catalytic cysteine by a glycine, the substrate-binding cleft occupied by an aromatic residue (Phe169) interferes with the enzyme activity. By analyzing the natural peptide fragment bound to the surface of SPE31 and its comparison with P34, we suggest the role of SPE31/P34 in HR and provide some plausible modes of binding of SPE31/P34 with other proteins and small molecules (especially syringolides).

Materials and Methods

cDNA cloning and sequence analysis of SPE31

Total RNA was isolated from the mid maturation seeds of *Pachyrhizus erosus*. The seeds were ground to powder in liquid nitrogen and the total RNA was extracted using RNeasy Plant Mini Kit (QIAGEN). The sequence of the forward primer 5'-gat(c) gcl ccl gag(a) tc(g)I tgg gat(c) tgg-3' was designed according to the N-terminal amino acid sequence DAPESWDW previously obtained. Oligo(dT) primers were used as reverse primer. 25 μ l RT reaction mixtures consisted of 5 μ l total RNA, 4 μ l oligo(dT) (20 μ M), 2000 units of M-MLV reverse transcriptase, 1.5 μ l 2.5 mM dNTP, 1 μ l RNase inhibitor (40 units/ μ l) and 5 \times buffer. After completion of the RT reaction, 8 μ l of the product was used for PCR amplification in 50 μ l reaction mixtures containing 10 \times PCR buffer, five units of Taq DNA polymerase, 0.25 mM dNTP, and 0.8 mM each of forward and reverse primers. The PCR program consisted of an initial cycle of denaturation (95 $^{\circ}$ C, 3 min), annealing (52 $^{\circ}$ C, 1 min) and prolonged extension (72 $^{\circ}$ C, 15 min), followed by 30 cycles of denaturation (94 $^{\circ}$ C, 1 min), annealing (52 $^{\circ}$ C, 40 s) and extension (72 $^{\circ}$ C, 1 min). A final 10 min extension step at 72 $^{\circ}$ C ended the amplification. The RT-PCR product was gel purified and subcloned into the pGEM T vector (TaKaRa Biotechnology Co., Ltd). Positive cDNA clones were identified by subjecting the plasmids extracted from transformants to PCR amplification using the conditions described above. The positive clones were further verified by sequencing (TaKaRa Biotechnology Co., Ltd).

The similarity searches for the amino acid sequence were performed using BLAST.¹⁴ NetNGlyc 1.0 Server[†] was used for the prediction of N-glycosylation sites of SPE31.

Purification, crystallization and data collection

Protein purification and crystallization of SPE31_PEG was carried out as described before.⁶ SPE31_PEG crystals were grown at 281 K using the hanging-drop vapor-diffusion method. 2 μ l hanging drops consisted of 1 μ l of 15 mg/ml protein solution and 1 μ l of reservoir solution (2.0 M ammonium sulfate, 5% isopropyl alcohol). Diffraction data were collected at 100 K with MarResearch (for SPE31_PEG) and Rigaku R-Axis IV⁺⁺ (for SPE31_PEP) imaging-plate detectors. Data were processed using DENZO and SCALEPACK.¹⁵ Data collection statistics are listed in Table 1.

[†] <http://www.cbs.dtu.dk/services/NetNGlyc/>

Structure determination

The SPE31_PEG structure was solved by molecular replacement (AMoRe) using papain (PDB id: 1POP) as a search model.¹⁶ The program O¹⁷ was used to rebuild the initial model. The model was refined at 50–2.0 Å resolution range by using CNS.¹⁸ PEG, oligosaccharide and water molecules were added to the model and individual atomic *B*-factors were refined at latter stage of refinement. The refined structure of SPE31_PEG was used as the initial model for the SPE31_PEP structure refinement. A five residue peptide fragment (KASVG) in the active cleft was built into the model with $F_o - F_c$ map as a reference by using Coot.¹⁹ The stereochemical qualities of the final models were checked by PROCHECK.²⁰ The final refinement statistics are listed in Table 1.

GenBank and Protein Data Bank accession codes

The cDNA sequence for SPE31 has been deposited in the GenBank database and assigned GenBank accession number DQ152924. The atomic coordinates of SPE31_PEG (PDB id: 2B1M) and SPE31_PEP (PDB id: 2B1N) are available at the RCSB Protein Data Bank†.

Acknowledgements

This work is supported by National Foundation of Talent Youth (grant no. 30225015), the 973 program (grant no. 2004CB720000 and no. 2004CB520801) the 863 Special Program of China (grant no. 2002BA711A13), the Key Important Project and other projects from the National Natural Science Foundation of China (grant no. 10490913, no. 30121001, and no. 30130080) and Chinese Academy of Sciences (KSCX2-SW-224). We thank Professors Peng Liu and Yuhui Dong at the Institute of High Energy Physics, and Yi Han from the Institute of Biophysics for diffraction data collection.

References

- Kamphuis, I. G., Drenth, J. & Baker, E. N. (1985). Thiol proteases comparative studies based on the high-resolution structures of papain and actinidin, and on aminoacid sequence information for cathepsins B and H, and stem bromelain. *J. Mol. Biol.* **182**, 317–329.
- Thakurta, P. G., Biswas, S., Chakrabarti, C., Sundd, M., Jagannadham, M. V. & Dattagupta, J. K. (2004). Structural basis of the unusual stability and substrate specificity of ervatamin C, a plant cysteine protease from *Ervatamia coronaria*. *Biochemistry*, **43**, 1532–1540.
- Sivaraman, J., Nagler, D. K., Zhang, R., Menard, R. & Cygler, M. (2000). A cysteine protease with the proregion covalently linked to the active site cysteine. *J. Mol. Biol.* **295**, 939–951.
- Wenig, K., Chatwell, L., von Pawel-Rammingen, U., Bjorck, L., Huber, R. & Sondermann, P. (2004). Structure of the streptococcal endopeptidase IdeS, a cysteine proteinase with strict specificity for IgG. *Proc. Natl Acad. Sci. USA*, **101**, 17371–17376.
- Choi, K. H., Laursen, R. A. & Allen, K. N. (1999). The 2.1 Å structure of a cysteine protease with proline specificity from ginger rhizome, *Zingiber officinale*. *Biochemistry*, **38**, 11624–11633.
- Chang, S., Song, X., Yan, M., Zhou, Z., Wu, F. & Gong, W. (2004). Purification, characterization and preliminary crystallographic studies of a cysteine protease from *Pachyrrhizus erosus* seeds. *Acta Crystallog. sect. D*, **60**, 187–189.
- Kalinski, A., Weisemann, J. M., Matthews, B. F. & Herman, E. M. (1990). Molecular cloning of a protein associated with soybean seed oil bodies that is similar to thiol proteases of the papain family. *J. Biol. Chem.* **265**, 13843–13848.
- Kalinski, A., Melroy, D. L., Dwivedi, R. S. & Herman, E. M. (1992). A soybean vacuolar protein (P34) related to thiol proteases is synthesized as a glycoprotein precursor during seed maturation. *J. Biol. Chem.* **267**, 12068–12076.
- Ogawa, T., Tsuji, H., Bando, N., Kitamura, K., Zhu, Y. L., Hirano, H. & Nishikawa, K. (1993). Identification of the soybean allergenic proteins, Gly m Bd 30K, with the soybean seed 34 kDa oil-body-associated protein. *Biosci. Biotechnol. Biochem.* **57**, 1030–1033.
- Ji, C., Boyd, C., Slaymaker, D., Okinaka, Y., Takeuchi, Y., Midland, S. L. *et al.* (1998). Characterization of a 34-kDa soybean binding protein for the syringolide elicitors. *Proc. Natl Acad. Sci. USA*, **95**, 3306–3311.
- Okinaka, Y., Yang, C. H., Herman, E., Kinney, A. & Keen, N. T. (2002). The P34 syringolide elicitor receptor interacts with a soybean photorespiration enzyme, NADH-dependent hydroxypyruvate reductase. *Mol. Plant Microbe Interact.* **15**, 1213–1218.
- Ji, C., Okinaka, Y., Takeuchi, Y., Tsurushima, T., Buzzell, R. I., Sims, J. J. *et al.* (1997). Specific binding of the syringolide elicitors to a soluble protein fraction from soybean leaves. *Plant Cell*, **9**, 1425–1433.
- Midland, S. L., Keen, N. T., Sims, J. J., Midland, M. M., Stayton, M. M., Burton, V. *et al.* (1993). The structures of syringolides 1 and 2, novel C-glycosidic elicitors from *Pseudomonas syringae* pv. tomato. *J. Organ. Chem.* **58**, 2927–3222.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389–3402.
- Otwinowski, Z. & Minor, W. (1997). Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326.
- Navaza, J. (1994). AMoRe: an atomated package for molecular replacement. *Acta Crystallog. sect. A*, **50**, 157–163.
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
- Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.

† <http://www.rcsb.org/>

19. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallog. sect. D*, **60**, 2126–2132.
20. Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). PROCHECK—a program to check the stereochemical quality of protein structures. *J. Appl. Crystallog.* **26**, 283–291.
21. Notredame, C., Higgins, D. G. & Heringa, J. (2000). T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217.
22. Kabsch, W. & Sander, C. (1983). Dictionary of protein secondary structure: pattern recognition of hydrogen bonded and geometrical features. *Biopolymers*, **22**, 2577–2637.
23. Gouet, P., Courcelle, E., Stuart, D. I. & Metoz, F. (1999). ESPript: multiple sequence alignments in PostScript. *Bioinformatics*, **15**, 305–308.
24. Carson, M. (1997). Ribbons. *Methods Enzymol.* **277**, 493–505.

Edited by R. Huber

(Received 26 October 2005; received in revised form 27 December 2005; accepted 17 January 2006)
Available online 3 February 2006