# Purification, characterization and preliminary crystallographic studies of a cysteine protease from *Pachyrrhizus erosus* seeds

**Shaojie Chang,[a] Xiaomin Song,[a] Ming Yan,[a] Zhaocai Zhou,[a] Fang Wu[a] and Weimin Gong[b]***

[a]School of Life Sciences, University of Science and Technology of China, Key Laboratory of Structural Biology, Hefei, Anhui 230026, People's Republic of China, and [b]National Laboratory of Biomacromolecules, Institute of Biophysics, Chinese Academy of Sciences, Beijing 100101, People's Republic of China

Correspondence e-mail: wgong@sun5.ibp.ac.cn

The proteins Spe31 and Spe32, named after their respective molecular weights of about 31 and 32 kDa, were purified simultaneously from the seeds of *Pachyrrhizus erosus*. They cannot be separated from each other by column chromatography. N-terminal sequence analysis indicated that they belonged to the papain family of cysteine proteases. An in-gel activity assay revealed that Spe31 possesses proteolytic activity while Spe32 only displays very weak activity for protein degradation. Both of them are glycoproteins as detected by the periodic acid and Schiff's reagent method. Crystals were obtained from the protein mixture by the hanging-drop vapour-diffusion method; they diffracted to a resolution of 2.61 Å on an in-house X-ray source. The crystals belong to space group $P4_{1(3)}2_12$, with unit-cell parameters $a = b = 61.96$, $c = 145.61$ Å. Gel electrophoresis under non-denaturing conditions showed that the protein crystallized was Spe31.

## 1. Introduction

Cysteine proteases (CPs; EC 3.4.22) are peptidyl hydrolases that depend on the free thiol group of a cysteine residue for their activity. In plants, CPs have been detected in various species and play major roles in diverse physiological processes, such as the development and ripening of fruits (Brady, 1985), specific propolypeptide processing during seed maturation (Muramatsu & Fukazawa, 1993), degradation of storage proteins in germinating seeds (Callis, 1995), control of developmental and pathogen-activated programmed cell deaths (Lam *et al.*, 1999) and activation of proenzymes and degradation of defective proteins (Rudenskaya *et al.*, 1998). Plant CPs are also involved in defence against foreign substrates (Boller, 1986) and in processes that lead to organogenesis (Pompe-Novak *et al.*, 2002).

CPs have been classified into various families according to similarities in their primary sequences (Rawlings & Barrett, 2000). The papain family is the largest family identified so far and most of the plant CPs for which crystal structures have been elucidated belong to this family (Kamphuis *et al.*, 1984; Pickersgill *et al.*, 1991; Maes *et al.*, 1996; O'Hara *et al.*, 1995; Baker, 1980; Heinemann *et al.*, 1982; Choi *et al.*, 1999; Chakrabarti *et al.*, 2003). They usually have acidic or neutral pH optima, high thermal stabilities and exhibit their highest activities in the presence of reducing agents and substances that are able to bind metal impurities, which are strong inhibitors of these enzymes (Forsthoefel *et al.*, 1998; Blackburn, 1976).

CPs from legumes have been widely investigated, but no crystal structure has been reported for any of them so far. Here, we present for the first time the purification, characterization and preliminary crystallographic studies of a moderately abundant protein named Spe31 from the seeds of *Pachyrrhizus erosus*, a legume found throughout the tropics of Asia and America. N-terminal sequence analysis indicated that Spe31 belongs to the CP papain family. Spe31 is also a glycoprotein, as revealed by the periodic acid and Schiff's reagent method.

## 2. Experimental methods

### 2.1. Protein purification

100 g of *P. erosus* seeds was homogenized in 800 ml buffer *A* (20 m*M* sodium phosphate buffer pH 7.2, 0.1 *M* NaCl) and left overnight. The homogenate was then filtered and the pH of the filtrate was adjusted to 4.0 with 50%(*v/v*) acetic acid. The mixture was left for 4 h, centrifuged and the supernatant brought to 100% ammonium sulfate saturation. The resulting precipitated proteins were recovered by centrifugation and dissolved in 100 ml buffer *B* (20 m*M* Tris–HCl pH 7.2). The solution was desalinized by ultrafiltration with frequent changes of buffer *B* and loaded onto a DEAE-Sepharose column (Amersham Pharmacia) pre-equilibrated with the same buffer. After extensive washing, Spe31 was eluted with a linear gradient of 0–0.5 *M* NaCl. The subsequent Superdex 75 HR 10/30 size-exclusion chromatography was performed with buffer *B* containing 0.15 *M* NaCl. The first elution peak

**187**

containing Spe31 was pooled, concentrated and kept at 253 K. The presence of Spe31 was monitored by SDS–PAGE electrophoresis. The protein concentration was determined by Bio-Rad Protein Assay (Bradford, 1976) using BSA as a standard. Non-denaturing (native) polyacrylamide gel (PAGE) and electrophoresis buffer were prepared as for standard SDS–PAGE, but with the omission of SDS. In two-dimensional electrophoresis, isoelectric focusing in the first dimension was achieved using a 18 cm long IPG strip gel with a pH range 3–10 on a IPGphor Isoelectric Focusing System (Amersham Biosciences; Gorg *et al.*, 1988); the second-dimension separation was performed by 12% SDS–PAGE. Native IEF was performed on a gel plate as described by Li *et al.* (1994). A mixed solution of 7%(*w/v*) acrylamide containing 2.5%(*v/v*) ampholine pH 3.5–10, 0.05%(*w/v*) ammonium persulfate and 0.01%(*v/v*) TEMED was used to prepare the gel plate and the electrode solutions were 1 *M* NaOH (cathode) and 1 *M* H$_3$PO$_4$ (anode). The purified protein was transblotted from an SDS–PAGE gel to a PVDF membrane. Protein bands of interest were excised and their N-terminal sequences were determined at Hunan Normal University, People's Republic of China. A search for homologous sequences was conducted using *BLAST* 2.0 (Altschul *et al.*, 1997).

## 2.2. In-gel protease assay

An in-gel protease assay was conducted using a combination of the methods described by Gomes *et al.* (1997) and Tanaka *et al.* (1993). A protein sample was run on a 15% PAGE gel at 200 V in cold buffer for 1 h. At the same time, a 12% PAGE gel containing 0.2%(*w/v*) gelatin was treated with 10 m*M* sodium phosphate buffer pH 6.5 plus 1%(*v/v*) Triton X-100 with gentle shaking. The two gels were equilibrated in 50 m*M* sodium phosphate buffer pH 6.5 containing 5 m*M* EGTA and 2 m*M* DTT at 298 K for 15 min. The 15% gel was placed tightly on the 12% gel and incubated at 298 K for 24 h and the 12% gel was then stained and destained as for SDS–PAGE gels. Protease activities in the gel were visualized as clear bands against a blue background.

## 2.3. Crystallization and X-ray diffraction

The protein sample was crystallized using the hanging-drop vapour-diffusion method at 291 K. The initial crystallization condition containing 30%(*w/v*) PEG 4000 was obtained using Crystal Screen kits I and II

from Hampton Research (Jancarik & Kim, 1991). Crystals of larger dimensions were obtained by adjusting the concentration of PEG 4000. X-ray diffraction data were collected at room temperature using an in-house X-ray source of 2.0 kW at a wavelength of 1.5418 Å equipped with a MAR Research image-plate system. The crystal-to-detector distance was set to 150 mm and the exposure time was 15 min for each image, covering 1° of oscillation. The data were processed using *DENZO* and *SCALEPACK* (Otwinowski & Minor, 1997).

## 2.4. Glycoprotein detection

Protein sample was mixed with an equal volume of non-denaturing sample buffer (*i.e.* sample buffer without SDS and DTT) and run on a 15% PAGE gel. Subsequently, the gel was subjected to glycoprotein staining using periodic acid and Schiff's reagent (PAS) as described by Zacharous *et al.* (1969). Glycoproteins in the gel were revealed as magenta bands. BSA and ovalbumin were used as negative and positive controls, respectively.

## 3. Results

### 3.1. Protein purification and characterization

The purified protein sample appeared as two distinct bands that run very close as revealed by SDS–PAGE (Fig. 1*a*), with the amount of protein in the lower band being less than that in the upper band. They were named Spe31 and Spe32 according to their
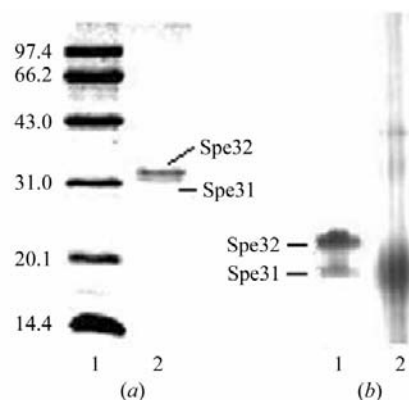
apparent molecular weights of around 31 and 32 kDa, respectively. Attempts to separate Spe31 and Spe32 using a Superdex 75 HR 10/30 gel-filtration column and other column chromatographies were unsuccessful. About 10 mg of a mixture of the two proteins was obtained from 100 g of dry seeds. Spe31 and Spe32 were also revealed as two distinct bands in native PAGE gels, which suggested that they are not associated (Fig. 1*b*; lane 1). The presence of DTT decreased the mobilities of Spe31 and Spe32 in SDS–PAGE, indicating that they are both likely to contain intramolecular disulfide bonds. Each of the two polypeptides is glycosylated as detected by the PAS method (Fig. 2). Native IEF showed that Spe31 and Spe32 have pI values of 4.0 and 4.2, but did not indicate which protein has the lower pI. However, two-dimensional electrophoresis showed that Spe31, which has a lower molecular weight than Spe32, is more acidic than Spe32. Therefore, it can be ascertained that the pI of Spe31 is 4.0 and that of Spe32 is 4.2. The sequence of the 20 N-terminal amino acids of Spe31 is DAPESWDWS-KKGVITEVKFQ and that of the first 17 amino acids of Spe32 is EKKK(C/K)EQP-SSDDAPESW; the residue at position 5 cannot be unambiguously determined. The two sequences share a six-residue overlap consisting of a DAPESW motif. A search for homologous sequences in the non-redundant sequence database suggested that both of them belong to the papain family of cysteine proteases. An in-gel protease assay



**Figure 2**
Glycoprotein detection in 15% native PAGE gel by the PAS method. Lane 1, 150 µg BSA; lane 2, 150 µg ovalbumin; lane 3, 16 µg purified Spe31 and Spe32.
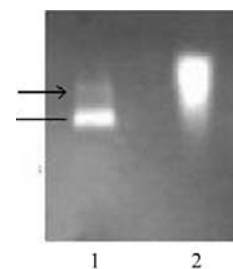


**Figure 1**
Polyacrylamide gel electrophoresis. (*a*) 15% SDS–PAGE. Lane 1, molecular-weight standards; lane 2, 1 µg of purified Spe31 and Spe32. (*b*) 15% native PAGE. Lane 1, 4 µg of purified Spe31 and Spe32. Lane 2: a crystal was picked out from the hanging drop, washed three times in the well solution and then transferred into 2 µl of double-distilled water to dissolve. The solutions containing dissolved crystals were pooled and loaded onto the gel.



**Figure 3**
In-gel protease assay. Lane 1, 25 µg of Spe31 and Spe32. The line and the arrow indicate the digestion bands corresponding to the activities of Spe31 and Spe32, respectively. Lane 2, 20 µg papain as a control.

**Table 1**
Summary of diffraction data collection and processing.

Values in parentheses refer to the highest resolution shell (2.67–2.61 Å).

| Space group | $P4_{1(3)}2_12$ |
|---|---|
| Unit-cell parameters (Å) | $a = b = 61.96$, $c = 145.61$ |
| Temperature (K) | 298 |
| Resolution limit (Å) | 50–2.61 |
| Total reflections | 139863 (5652) |
| Unique reflections | 9224 (576) |
| Unique reflections [$I/(I) > 3$] | 7661 (376) |
| Completeness (%) | 99.7 (98.3) |
| Completeness [$I/(I) > 3$] (%) | 82.8 (64.2) |
| $R_{\mathrm{merge}}$† | 0.098 (0.359) |
| Multiplicity | 15.2 (9.8) |
| $I/\sigma(I)$ | 22.6 (6.4) |

† $R_{\mathrm{merge}} = \sum_{hkl}\sum_{j} |I(hkl)_j - \langle I(hkl)\rangle| / \sum_{hkl}\sum_{j} I(hkl)_j$, where $I(hkl)_j$ is the $j$th measurement of the intensity of reflection $hkl$ and $\langle I(hkl)\rangle$ is the mean intensity of reflection $hkl$.

showed that Spe31 had obvious proteolytic activity against gelatin copolymerized in the gel, while Spe32 only displayed very weak activity for protein degradation (Fig. 3).

### 3.2. Crystallization and preliminary crystallographic studies

Using a protein concentration of 10 mg ml$^{-1}$, the best crystals grew with a well solution containing 32.5%($w/v$) PEG 4000, 0.2 $M$ NH$_4$OAc and 0.1 $M$ HOAc/NaOAc buffer pH 4.6 and were good enough for X-ray data collection. Non-denaturing gel electrophoresis of dissolved crystals revealed that the crystallized polypeptide was Spe31, not Spe32 (Fig. 1b; lane 2). The crystal of Spe31 diffracted to 2.61 Å resolution and belongs to space group $P4_{1(3)}2_12$, with unit-cell parameters $a = b = 61.96$, $c = 145.61$ Å (Table 1). Assuming one Spe31 molecule per asymmetric unit, the $V_{\mathrm{M}}$ value (Matthews, 1968) is 2.22 Å$^3$ Da$^{-1}$, corresponding to a solvent content of 45%.

### 4. Discussion

The similarity of the amino-terminal sequences of Spe31 and Spe32 and the similarity of their molecular weights and isoelectric points indicate that the two polypeptides are likely to be encoded by the same gene, with Spe31 derived from Spe32 by post-translational cleavage of a N-terminal 11-residue peptide. Essentially, all CPs of the papain family are synthesized as inactive zymogens with N-terminal proregions of around 60 residues (or longer) (Cygler & Mort, 1997). Thus, Spe32 is unlikely to be an 'intact' proenzyme of Spe31 but rather an intermediate in the process from a putative proenzyme, with Spe31 being the final mature product. The proregions of CPs are inhibitors of their enzymatic activities (Cygler & Mort, 1997). Although the 11-residue extension to the N-terminus of Spe31 is much shorter than the lengths of proregions of CPs from the papain family, it may also have some inhibitory effect. This is consistent with the observation that Spe32 showed extremely weak proteolytic activity, while Spe31 gave rise to very efficient protein degradation. CPs in seeds are usually induced during germination, while the active Spe31 is present in the mature seeds of *P. erosus*. This indicates that Spe31 may have other roles than the general function of storage-protein degradation. Some plant CPs such as bromelain (Takahasi *et al.*, 1973), calotropin FI and FII (Abraham & Joshi, 1979) are glycoproteins, but the biological significance of their carbohydrate moieties is not well understood. Of the plant CPs of known crystal structure, only zingipain from *Zingiber officinale* is a glycoprotein (Choi *et al.*, 1999), so Spe31 provides another model for the observation of the impact of glycosyl chains on the conformation and function of these enzymes.

## References

Abraham, K. I. & Joshi, P. N. (1979). *Biochim. Biophys. Acta*, **568**, 111–119.

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.

Baker, E. N. (1980). *J. Mol. Biol.* **141**, 441–484.

Blackburn, S. (1976). *Enzyme Structure and Function*, Vol. 3, pp. 267–302. New York & Basel: Marcel Dekker Inc.

Boller, T. (1986). *Plant Proteolytic Enzymes*, pp. 67–96. Boca Raton, FL, USA: CRC Press.

Bradford, M. M. (1976). *Anal. Biochem.* **72**, 248–254.

Brady, C. J. (1985). *Annu. Rev. Plant Physiol.* **38**, 155–178.

Callis, J. (1995). *Plant Cell*, **7**, 845–857.

Chakrabarti, C., Biswas, S. & Dattagupta, J. K. (2003). *Proteins Struct. Funct. Genet.* **51**, 489–497.

Choi, K. H., Laursen, R. A. & Allen, K. N. (1999). *Biochemistry*, **38**, 11624–11633.

Cygler, M. & Mort, J. S. (1997). *Biochimie*, **79**, 645–652.

Forsthoefel, N. R., Cushman, M. A. F., Ostrem, J. A. & Cushman, J. C. (1998). *Plant Sci.* **136**, 195–206.

Gomes, A. V., Sirju-Charran, G. & Barnes, J. A. (1997). *Phytochemistry*, **46**, 185–193.

Gorg, A., Postel, W., Domscheit, A. & Gunther, S. (1988). *Electrophoresis*, **9**, 681–692.

Heinemann, U., Pal, G. P., Hilgenfeld, R. & Saenger, W. (1982). *J. Mol. Biol.* **161**, 591–606.

Jancarik, J. & Kim, S.-H. (1991). *J. Appl. Cryst.* **24**, 409–411.

Kamphuis, I. G., Kalk, K. H., Swarte, M. B. & Drenth, J. (1984). *J. Mol. Biol.* **179**, 233–256.

Lam, E., Pontier, D. & Pozo, O. (1999). *Curr. Opin. Plant Biol.* **2**, 502–507.

Li, J. W., Xiao, N., Yu, R. Y., Yuan, M. X., Chen, L. R., Chen, Y. H. & Chen, L.T. (1994). *Biochemical Theory and Methods*, pp. 197–202. Peking University Press.

Maes, D., Bouckaert, J., Poortmans, F., Wyns, L. & Looze, Y. (1996). *Biochemistry*, **35**, 16292–16298.

Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.

Muramatsu, M. & Fukazawa, C. (1993). *Eur. J. Biochem.* **215**, 123–132.

O'Hara, B. P., Hemmings, A. M., Buttle, D. J. & Pearl, L. H. (1995). *Biochemistry*, **34**, 13190–13195.

Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.

Pickersgill, R. W., Rizkallah, P., Harris, G. W. & Goodenough, P. W. (1991). *Acta Cryst.* B**47**, 766–771.

Pompe-Novak, M., Poljsak-Prijatelj, M., Popovic, T., Strukelj, B. & Ravnikar, M. (2002). *Physiol. Mol. Plant Pathol.* **60**, 71–78.

Rawlings, N. D. & Barrett, A. J. (2000). *Nucleic Acids Res.* **28**, 323–325.

Rudenskaya, G. N., Bogacheva, A. M., Preusser, A., Kuznetsova, A. V., Dunaevsky, Y. E., Golovkin, B. N. & Stepanov, V. M. (1998). *FEBS Lett.* **437**, 237–240.

Takahasi, N., Yasuda, Y., Goto, K., Miyake, T. & Murachi, T. (1973). *J. Biochem.* **74**, 355–373.

Tanaka, T., Minamikawa, T., Yamauchi, D. & Ogushi, Y. (1993). *Plant Physiol.* **101**, 421–428.

Zacharous, R. M., Zell, T. E., Morrison, J. H. & Woodlock, J. J. (1969). *Anal. Biochem.* **30**, 148–152.