Expression, purification and crystallization of pecan (*Carya illinoensis*) vicilin

BoRam Lee,^{a,†} Renhao Zhang,^{a,§}
Wen-Xian Du,^a Larry J. Grauke,^b
Tara H. McHugh^a and Yu-Zhu
Zhang^{a,*}

^aHealthy Processed Foods Research Unit,
USDA–ARS–PWA, Western Regional Research
Center, 800 Buchanan Street, Albany, CA
94710, USA, and ^bCrop Germplasm Research,
USDA–ARS–SPA, Southern Plains Agricultural
Research Center, 2881 F&B Road, College
Station, TX 778450, USA

† Present address: Department of Biological
Sciences, Korea Advanced Institute of Science
and Technology (KAIST), Daejeon 305-701,
Republic of Korea.

§ Present address: Department of Radiation
Oncology, University of California, San
Francisco, San Francisco, CA 94143-1331,
USA.

Correspondence e-mail:
yuzhu.zhang@ars.usda.gov

Received 14 March 2014

Accepted 27 May 2014

Tree nuts are responsible for many cases of severe food allergies. The 7S seed storage protein vicilin has been identified as a food allergen in many kinds of tree nuts. The vicilin protein consists of an N-terminal low-complexity region with antimicrobial activity and a C-terminal domain that forms a trimeric structure that belongs to the cupin superfamily. In this study, vicilin from pecan (*Carya illinoensis*) was isolated and was expressed in bacteria for the first time. The cupin structural core of the protein, residues 369–792, was purified by metal-affinity and gel-filtration chromatography to high purity. Vicilin crystals were obtained and the best crystal diffracted to 2.65 Å resolution in space group $P2_12_12_1$.

1. Introduction

Pecan (*Carya illinoensis*) is a commercially important tree nut. The US alone produced pecans valued at more than US \$670 million in 2011 (United States Department of Agriculture Economic Research Service, 2012). Pecans are nutritious nuts and can be consumed safely by most people. However, pecans can pose serious risk to people who have allergies (Sharma *et al.*, 2011a,b). Pecan and other tree nuts are recognized to be among the eight food groups that cause the majority of food allergies (US Food and Drug Administration, 2010) and are required to be labeled in the USA by the Food Allergen Labeling and Consumer Protection Act of 2004 when they are used as ingredients in preparing a food product. Failing to inform the consumer of an allergen source on a product label is one of the leading reasons for recent food recalls (Gendel & Zhu, 2013). Both allergenic patients and the food industry may become victims of foods containing unlabeled allergen sources.

Seed storage proteins such as 2S, 7S and 11S globular proteins in many species are known allergens, including all three proteins in peanuts, hazelnuts, cashews, pistachios and walnuts. While the 2S albumin and 11S legumin in pecan have also been identified as food allergens (Sharma *et al.*, 2011a,b), the 7S protein from pecan has not been studied. The 7S globulins are called vicilins and they are present in both monocotyledonous and dicotyledonous species. Vicilins are typically trimeric proteins with a molecular mass of ~150–190 kDa. The subunit structure of vicilins revealed by SDS–PAGE is similar in the absence or presence of reducing agents, with a typical subunit molecular mass of ~50 kDa. No disulfide bond is found in vicilins, but proteolytic processing and glycosylation may occur (Gatehouse *et al.*, 1982; López-Torrejón *et al.*, 2003). Vicilin consists of an N-terminal low-complexity region and a C-terminal domain. The N-terminal is made up of 50-amino-acid repeats that each contains a C-X₃-C-X_(10–12)-C-X₃-C motif and is known to have antimicrobial activities (Marcus *et al.*, 1999). This low-complexity region belongs to the vicilin-N superfamily (Pfam 04702). The C-terminal domain of vicilin constitutes the structure core and belongs to the cupin protein superfamily which contains both monocupins and bicupins, as well as multicupins. The cupin protein superfamily is known to be one of the most functionally diverse protein superfamilies (Dunwell *et al.*, 2004) and includes various proteins with enzymatic functions, non-enzymatic transcription factors and seed storage proteins. The 7S

vicilin has no known function other than preserving nutrients for the development of young seedlings. The crystal structure of the peanut allergen Ara h 1 (the 7S vicilin from peanut; Cabanos *et al.*, 2011; Chruszcz *et al.*, 2011) has recently been determined and we recently reported the crystal structure of Korean pine vicilin, which revealed a copper center (Jin *et al.*, 2013). Based on sequence alignment, the copper center is conserved in a number of tree species, but it does not exist in the vicilins of legumes. To date, the Korean pine vicilin is the only 7S seed storage protein that has had its copper center revealed by X-ray crystallography. Here, we report the cloning, expression, purification and crystallization of the structural core of pecan 7S vicilin.

2. Materials and methods

2.1. Total RNA isolation and reverse transcription

Fresh Pawnee pecan nuts at the gel stage were collected from verified inventories at the USDA ARS Pecan Breeding and Genetics program, Burleson County, Texas, USA. An RNeasy Plus Mini Kit (Qiagen, Germany) was used to extract total RNA from the young nut kernels. The manufacturer's protocol was followed with some modifications. Briefly, 100 mg fresh seed kernels were ground in liquid nitrogen and collected in a microcentrifuge tube. As soon as the liquid nitrogen had evaporated, 450 μ l buffer RLT (Qiagen) was added and mixed by inversion. After centrifugation, genomic DNAs were removed using a gDNA Eliminator Mini Spin Column (Qiagen). The sample was then mixed with ethanol and loaded onto an RNeasy Mini Spin column. After spinning, the column was washed with the buffers that came with the kit and the total RNA was eluted with 50 μ l RNase-free water into a fresh tube containing 2 μ l (80 units) of RNase inhibitor (Invitrogen). The fresh total RNA was used for reverse transcription (RT).

The RT reaction was carried out with a Qiagen Omniscript Reverse Transcription Kit using the total RNA as template and a 20-nucleotide poly-dT oligo as the primer following the manufacturer's protocols. The RT reaction was allowed to proceed for 1 h and the reverse transcriptase was inactivated by heating the reaction mixture at 95°C for 5 min. The RT product was stored at -80°C for further use.

2.2. Plasmid construction and protein expression

Primers I (5'-CCAAATCCAGGATCCATGGTGACCAAAGC-CAAGATT-3', forward) and II (5'-CCGGTACCTTAGAAGAA-ACCTGCAAAGTCCAAAATTG-3', reverse) were designed based on the available pecan vicilin mRNA sequence in GenBank (EF689893.1). The coding sequence of pecan vicilin was amplified using the RT product as the template. The PCR product was inserted into a pBlueScript II vector by blunt-end ligation to generate pBlue-Cariv. The coding sequence of vicilin was released from the pBlue-Cariv by DNA restriction digestion with *Bam*HI and *Kpn*I and was inserted into pTAP2S, a home-made vector with the backbone of pRSFduet-1 (Merck Millipore, Billerica, Massachusetts, USA) and a different multi-cloning-site sequence. The construct was named pRSF-Cariv. Primer III (5'-CCATTCATGAGCCGGGAAGAG-GAGCAACAGA-3', forward) and primer II were used to amplify the coding sequence of pecan vicilin core with pRSF-Cariv as the template. The PCR product was inserted into pBlueScript II by blunt-end ligation to generate pBlue-Carivc. The coding sequence for vicilin core was released from the pBlue-Carivc by DNA restriction digestion with *Bsp*HI and *Kpn*I and inserted into pRSFduet-1 to

produce pRSF-Carivc. All constructs were confirmed by DNA sequencing.

Competent *Escherichia coli* BL21(DE3)* cells (Invitrogen) were transformed with pRSF-Carivc for protein expression. The bacteria were grown in 1 l batches of LB medium containing kanamycin (50 mg l⁻¹; Gold Biotechnology, St Louis, Missouri, USA) at 37°C until the OD₆₀₀ of the culture was between 1.0 and 1.2. Isopropyl β -D-1-thiogalactopyranoside (Gold Biotechnology) was added to the culture at a final concentration of 1 mM. The cells were collected by centrifugation after 5 h of induction at 37°C. The cell pellets were stored at -80°C for subsequent use.

2.3. Protein purification

Cell pellets from two 1 l cultures were thawed and resuspended in 50 ml Ni²⁺-column binding buffer (buffer NB; 20 mM imidazole pH 7.9, 0.5 M NaCl) plus 1 mM benzamidine. Cells were lysed by sonication for 8 min in a beaker surrounded by a mixture of ice and water using a Model 450 Digital Sonifier (Branson, Danbury, Connecticut, USA). The cell lysate was subjected to centrifugation at 20 000g for 30 min at 20°C. The supernatant was heated to 70°C in a water bath and incubated for 5 min at 70°C. The sample was subjected to centrifugation at 20 000g for 30 min at 20°C again and the supernatant was loaded onto a 5 ml HisTrap FF Crude column (GE Healthcare, Piscataway, New Jersey, USA). The column was washed with seven bed volumes of buffer NB and then eluted with buffer NB plus 280 mM imidazole.

The main peak of the HisTrap elution was loaded onto a XK 26/70 Superdex-200 column (GE Healthcare) pre-equilibrated and eluted with 10 mM Tris-HCl pH 7.9, 500 mM NaCl at a flow rate of 1 ml min⁻¹. All chromatographic steps were carried out at room temperature using an FPLC system (GE Healthcare). Samples of pecan vicilin after each step of purification were heated to 96°C for SDS-PAGE analysis. The gels were stained using the Colloidal Coomassie G-250 Staining protocols (Dybala & Metzger, 2009). Gel images were taken with a digital camera.

The buffer of the protein sample was changed to less than 1 mM Tris pH 8.0 by repeated dilution and concentration using an Ultracel-

Figure 1
Purification of recombinant pecan vicilin. SDS-PAGE analysis of purified pecan vicilin. The protein was separated with a 4–20% reducing gel. The molecular weights (in kDa) of the protein standards contained in the marker lane (M) are (from the top) 200, 150, 120, 100, 85, 70, 60, 50, 40, 30, 25, 20, 15 and 10.

Table 1

Data collection and processing.

Values in parentheses are for the outer shell.

Wavelength (Å)	1.000
Space group	$P2_12_12_1$
a, b, c (Å)	42.696, 151.710, 343.005
α, β, γ (°)	90, 90, 90
Mosaicity (°)	0.41
Resolution range (Å)	46.60–2.65 (2.70–2.65)
Total No. of reflections	419967
No. of unique reflections	61694 (1939)
Completeness (%)	93 (59)
Multiplicity	6.8 (2.3)
$\langle I/\sigma(I) \rangle$	7.36 (1.91)
$R_{\text{merge}}^{\dagger}$	0.202 (0.495)
Overall B factor from Wilson plot (Å ²)	43.80

$\dagger R_{\text{merge}} = \sum_{hkl} \sum_i |I_i(hkl) - \langle I(hkl) \rangle| / \sum_{hkl} \sum_i I_i(hkl)$, where $I_i(hkl)$ is the i th observation of reflection hkl and $\langle I(hkl) \rangle$ is the weighted average intensity for all observations i of reflection hkl .

50k filter device (Millipore). The final protein concentration used for crystallization screening was 40 mg ml^{−1} based on the theoretical absorbance coefficient of the protein at 280 nm.

2.4. Crystallization and data collection

The crystallization screening was carried out at 20°C using the hanging-drop vapor-diffusion method. The conditions of the Crystal Screen, Crystal Screen 2 and Crystal Screen Lite kits from Hampton Research were used. Hanging drops consisting of 1 µl protein solution mixed with 1 µl crystallization solution were sealed against 0.5 ml reservoir solutions in 24-well Linbro plates. Needle-shaped crystals were obtained in the drop equilibrated over solution No. 21 of the Crystal Screen kit [CS21; 0.2 M magnesium acetate tetrahydrate, 0.1 M sodium cacodylate trihydrate pH 6.5, 30% (v/v) (±)-2-methyl-2,4-pentanediol]. To optimize the crystallization, different NaCl solutions were used as reservoir solutions. The drops were generated by mixing the protein solutions at different concentrations with CS21 or with CS21 with reduced precipitant contents. A cryoprotectant solution was prepared by mixing 40% of glycerol and 60% of CS21 (v:v).

X-ray diffraction data were collected on the HMMI-CAT 28-2ID beamline at the ALS (Advanced Light Source), Lawrence Berkeley National Laboratory. Data were collected at −163°C with a 500 mm

crystal-to-detector distance using an ADSC Q315r CCD detector. 180 1° frames were collected with an exposure time of 5 s. The diffraction data were processed with the *HKL-2000* suite of programs (Otwinowski & Minor, 1997). Molecular-replacement calculations were carried out using the program *Phaser* (McCoy *et al.*, 2005).

3. Results and discussion

The coding sequence of pecan vicilin obtained in this study was identical to that of GenBank entry EF689893.1. The full length and the structural core (amino acids 369–792) of pecan vicilin were successfully cloned and expressed in *E. coli* without a fusion tag. Two amino acids (Met-Ser) were added to the N-terminus of the core as a result of cloning. Korean pine vicilin did not denature up to 80°C (Jin *et al.*, 2008). Therefore, we tried to partially remove unwanted proteins by thermo-denaturation at 70°C and this was successful. Previously, we had experience with two proteins that had affinity for Ni²⁺ columns despite not containing a His tag (Jin *et al.*, 2007; Zhang *et al.*, 1999). We tested whether Ni²⁺ columns could be used to purify pecan vicilin and the result was positive. Apparently homogeneous recombinant pecan vicilin core was obtained after the last step of purification with gel-filtration chromatography (Fig. 1).

Numerous needle-shaped crystals appeared in the drop equilibrated over solution No. 21 of the Crystal Screen kit within two weeks. As the protein sample contained 500 mM NaCl, the crystallization drop became larger, rather than smaller. As reverse salting-out was driving the crystallization, we sought to control the speed of the equilibration between the drops and the reservoir by using different concentrations of NaCl (from 400 mM to 0 at 50 mM intervals) as the reservoir solution. The starting protein concentration was also changed to optimize the crystallization conditions. Large crystals (1 × 1 × 0.2 mm) were obtained with a starting protein concentration of 50 mg ml^{−1} mixed with the original CS21 over 200 mM NaCl reservoir (Fig. 2). Crystals were soaked briefly in the cryoprotectant solution, flash-cooled in liquid nitrogen and stored in

Figure 2

Single crystals of pecan vicilin obtained by vapor diffusion using the hanging-drop method. One unit (between adjacent major ticks) of the ruler represents one third of a millimetre.

Figure 3

A typical frame of the X-ray diffraction of the pecan vicilin core crystal.

liquid nitrogen before use. X-ray diffraction data were collected at -163°C with 1.0° oscillation per image. Most of the crystals, however, diffracted to resolutions of lower than $4.0\text{ \AA}$. The best crystal resulted in a data set at $2.65\text{ \AA}$ resolution (Fig. 3).

Data processing revealed an orthorhombic crystal system with unit-cell parameters $a = 42.70$, $b = 151.71$, $c = 343.01\text{ \AA}$ (Table 1). Based on systematic absences in specific reflections in the diffraction, the space group was determined to be $P2_12_12_1$. A solution of the phase problem was obtained with molecular-replacement calculations using *Phaser* (McCoy *et al.*, 2005). There were two doughnut-shaped native trimers of the vicilin molecule in the crystallographic asymmetric unit. Thus, the Matthews coefficient was $1.72\text{ \AA}^3\text{ Da}^{-1}$, with a corresponding solvent content of the crystal of $\sim 28.6\%$ (Matthews, 1968). Statistically, protein crystals with a V_M of $1.72\text{ \AA}^3\text{ Da}^{-1}$ give better diffraction far beyond a resolution of $2.65\text{ \AA}$ (Kantardjieff & Rupp, 2003). Currently, further optimization of the crystallization conditions is under way.

Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the US Department of Agriculture. The Berkeley Center for Structural Biology is supported in part by the National Institutes of Health, National Institute of General Medical Sciences. The Advanced Light Source is supported by the Director, Office of Science, Office of Basic Energy Sciences of the US Department of Energy under Contract No. DE-AC02-05CH11231.

References

- Cabanos, C., Urabe, H., Tandang-Silvas, M. R., Utsumi, S., Mikami, B. & Maruyama, N. (2011). *Mol. Immunol.* **49**, 115–123.
- Chruszcz, M., Maleki, S. J., Majorek, K. A., Demas, M., Bublin, M., Solberg, R., Hurlburt, B. K., Ruan, S., Mattisohn, C. P., Breiteneder, H. & Minor, W. (2011). *J. Biol. Chem.* **286**, 39318–39327.
- Dunwell, J. M., Purvis, A. & Khuri, S. (2004). *Phytochemistry*, **65**, 7–17.
- Dyballa, N. & Metzger, S. (2009). *J. Vis. Exp.*, doi:10.3791/1431.
- Gatehouse, J. A., Lycett, G. W., Croy, R. R. & Boulter, D. (1982). *Biochem. J.* **207**, 629–632.
- Gendel, S. M. & Zhu, J. (2013). *J. Food Prot.* **76**, 1933–1938.
- Jin, T., Albillos, S. M., Chen, Y.-W., Kothary, M. H., Fu, T.-J. & Zhang, Y.-Z. (2008). *J. Agric. Food Chem.* **56**, 8159–8165.
- Jin, T., Kim, S., Guo, F., Howard, A. & Zhang, Y.-Z. (2007). *Cytokine*, **40**, 115–122.
- Jin, T., Wang, Y., Chen, Y.-W., Fu, T.-J., Kothary, M. H., McHugh, T. H. & Zhang, Y. (2013). *J. Agric. Food Chem.* **62**, 222–228.
- Kantardjieff, K. A. & Rupp, B. (2003). *Protein Sci.* **12**, 1865–1871.
- López-Torrejón, G., Salcedo, G., Martín-Esteban, M., Díaz-Perales, A., Pascual, C. Y. & Sánchez-Monge, R. (2003). *J. Allergy Clin. Immunol.* **112**, 1208–1215.
- Marcus, J. P., Green, J. L., Goulter, K. C. & Manners, J. M. (1999). *Plant J.* **19**, 699–710.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- McCoy, A. J., Grosse-Kunstleve, R. W., Storoni, L. C. & Read, R. J. (2005). *Acta Cryst.* **D61**, 458–464.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Sharma, G. M., Irsigler, A., Dhanarajan, P., Ayuso, R., Bardina, L., Sampson, H. A., Roux, K. H. & Sathe, S. K. (2011a). *J. Agric. Food Chem.* **59**, 4130–4139.
- Sharma, G. M., Irsigler, A., Dhanarajan, P., Ayuso, R., Bardina, L., Sampson, H. A., Roux, K. H. & Sathe, S. K. (2011b). *J. Agric. Food Chem.* **59**, 9542–9552.
- United States Department of Agriculture Economic Research Service (2012). *Fruit and Tree Nuts Yearbook Data Tables*. <http://www.ers.usda.gov/data-products/fruit-and-tree-nut-data/yearbook-tables.aspx>.
- US Food and Drug Administration (2010). *Food Allergies: What You Need to Know*. <http://www.fda.gov/downloads/Food/ResourcesForYou/Consumers/UCM220117.pdf>.
- Zhang, Y.-Z., Gould, K. L., Dunbrack, R. L. Jr., Cheng, H., Roder, H. & Golemis, E. A. (1999). *Physiol. Genomics*, **1**, 109–118.