Two proteins for the price of one: Structural studies of the dual-destiny protein preproalbumin with sunflower trypsin inhibitor-1

Received for publication, February 20, 2017, and in revised form, May 22, 2017 Published, Papers in Press, May 23, 2017, DOI 10.1074/jbc.M117.776955

Bastian Franke†1, Amy M. James§5, Mehdi Moblit‡1, Michelle L. Colgrave‡1, Joshua S. Mylne§2, and K. Johan Rosengren†2,3

From the †School of Biomedical Sciences and ‡Centre for Advanced Imaging, University of Queensland, St Lucia, Brisbane, Queensland 4072, Australia, the §School of Molecular Sciences and ARC Centre of Excellence in Plant Energy Biology, University of Western Australia, Crawley, Perth, Western Australia 6009, Australia, and ¶CSIRO Agriculture and Food, St Lucia, Brisbane, Queensland 4067, Australia

Edited by Joseph Jez

Seed storage proteins are both an important source of nutrition for humans and essential for seedling establishment. Interestingly, unusual napin-type 2S seed storage albumin precursors in sunflowers contain a sequence that is released as a macrocyclic peptide during post-translational processing. The mechanism by which such peptides emerge from linear precursor proteins has received increased attention; however, the structural characterization of intact precursor proteins has been limited. Here, we report the 3D NMR structure of the intact precursor protein PawS1 and provide new insights into the processing of this remarkable dual-destiny protein. In seeds, PawS1 is matured by asparaginyl endopeptidases (AEPs) into the cyclic peptide SFTI-1 (sunflower trypsin inhibitor-1) and a heterodimeric 2S albumin. The structure of PawS1 revealed that SFTI-1 and the albumin are independently folded into well-defined domains separated by a flexible linker. PawS1 was cleaved in vitro with recombinant sunflower HaAEP1 and in situ using a sunflower seed extract in a way that resembled the expected in vivo cleavages. Recombinant HaAEP1 cleaved PawS1 at multiple positions, and in situ, its flexible linker was removed, yielding fully mature heterodimeric albumin. Liberation and cyclization of SFTI-1, however, was inefficient, suggesting that specific seed conditions or components may be required for in vivo biosynthesis of SFTI-1. In summary, this study has revealed the 3D structure of a macrocyclic precursor protein and provided important mechanistic insights into the maturation of sunflower proalbumins into an albumin and a macrocyclic peptide.

Endoproteases are well known for their role in protein degradation, but are also intimately involved in protein biosynthesis, where they mature sequences by releasing flanking fragments or separating subunits by internal cleavage events (1, 2). An abundant class of proteins that are subject to substantial proteolytic maturation are seed storage proteins, which are both an important source of nutrition for humans and essential for the establishment of the seedling (3). Seed storage proteins are matured from precursors that are sent to the endoplasmic reticulum for folding and removal of their N-terminal signal sequence, before they are subject to one or more internal cleavages often followed by trimming to yield highly stable and long-lasting proteins that can withstand desiccation (4, 5). Several classes exist, with the major ones being the salt-soluble globulins, alcohol-soluble prolamin, and water-soluble albumins.

The prototypic seed albumin is Brassica napus (canola) napin, which begins as a 178-residue protein that exits the endoplasmic reticulum as a 157-amino acid protein. It is further processed through the removal of two joining peptides and the C-terminal residue by the action of endoproteases. Some asparaginyl endopeptidases (AEPs), yielding the “mature” napin. In this final form, napin exists as a heterodimer with two chains of 29 and 86 residues (6) that are cross-braced by four disulfide bonds, which are conserved throughout the 2S albumin family. Some napin-type albumins are not processed into heterodimers but remain as monomers, such as the abundant SESA3 (SFA8), SESA2-1, SESA2-2, and SESA20-2 from sunflower seeds (7, 8). Whether they are matured into a heterodimer or remain monomeric, napin-type albumins adopt a conserved five-helix tertiary structure stabilized by the characteristic disulfide array (5).

The precursor proteins for napin-type albumins in the common sunflower (Helianthus annuus) are diverse. Some precursor proteins consist of two full albumin sequences instead of one (9), but one class in particular is unusual for the fact that, in addition to an albumin, they contain a macrocyclic peptide (10). Helianthus annuus possesses two of these unusual albumin precursor

---

This work was supported by Australian Research Council Grant DP120103369 (to J. S. M. and K. J. R.). The authors declare that they have no conflicts of interest with the contents of this article.

This article contains supplemental Movies 1 and 2.

The atomic coordinates and structure factors (code5U87) have been deposited in the Protein Data Bank (http://wwpdb.org/).

The NMR resonance assignment was deposited in the Biological Magnetic Resonance Bank and given the accession code 30209.

1 Supported by Australian Research Council Grant DP120103369 (to J. S. M. and K. J. R.).

2 Supported by Australian Research Council Future Fellowships FT11000925, FT120100013, and FT130100890, respectively.

3 To whom correspondence should be addressed: School of Biomedical Sciences, University of Queensland, St Lucia, Brisbane, Queensland 4072, Australia. Tel.: 61-7-3365-1403; E-mail: j.rosengren@uq.edu.au. This is an open access article under the CC BY license.

4 The abbreviations used are: AEP, asparaginyl endopeptidase; RP, reverse-phase; SSU, small subunit; LSU, large subunit; HSGC, heteronuclear single quantum coherence; CHCA, α-cyano-4-hydroxycinnamic acid; TEV, tobacco etch virus.

5 2017 by The American Society for Biochemistry and Molecular Biology, Inc. Published in the U.S.A.
Here, we present the solution NMR structure of the PawS1 proalbumin and use both recombinant sunflower HaAEP1 and crude sunflower seed extract to illustrate the maturation of PawS1 in vitro. We provide mechanistic insights into the structures of proalbumins and the consequences of the AEP cleavages as well as the initial steps of biosynthesis of the ultrastable macrocycle SFTI-1.

**Results**

To study the structural features and the processing of the unusual albumin precursor PawS1, a synthetic gene that encoded the full SFTI-1 and albumin domains was designed. The construct consists of the C-terminal 116 residues of the 151-residue preproprotein, starting from Gly1 of SFTI-1 (Gly36 in full-length PawS1) and ending with the C-terminal Ile of the PawS1 albumin. PawS1 was expressed successfully in the Escherichia coli strain SHuffle and purified with yields of up to 150 μg/liter soluble 15N-labeled or 13C- and 15N-double-labeled protein obtained. The purity of PawS1 was assessed by reverse-phase (RP)-HPLC and LC-MS, revealing ~95% purity (Fig. 2). The average calculated mass for the oxidized double-labeled protein was 13,987 Da, and the observed average mass for the protein based on LC-MS was 13,943 Da, representing ~94% incorporation of 15N and 13C.

**Resonance assignment and structure determination**

Samples for solution NMR studies were prepared to concentrations of 3.6 mg/ml, and all NMR data were recorded at 25 °C. The 1H-15N heteronuclear single quantum coherence (HSQC) spectrum showed excellent peak dispersion and sharp lines, indicating a highly structured protein (Fig. 3). A suite of standard triple-resonance 3D experiments, including HNCA, CBCA(CO)NH, HNCO, H(CC)(CO)NH-TOCSY, and 3D (H)CC(CO)NH-TOCSY were acquired and allowed complete sequential assignment of the protein backbone, except the amide of Gln98, and the majority of side-chain atoms. The side-chain amide groups of the 26 glutamine residues were not assigned due to the severe peak overlap in the 1H-15N HSQC spectrum (Fig. 3). Assignments around the proline residues Pro9, Pro9, Pro13, Pro19, Pro50, and Pro107, which lack backbone amide protons, were confirmed using 3D 15N- and 13C-edited NOESY data.

The NMR secondary shifts (Fig. 3) of the active site loop of SFTI-1, including the characteristic downfield-shifted Hα protons of Cys2 and Cys11 resulting from the conformation of the disulfide across the β-sheet, were found to be consistent with those reported previously for native SFTI-1 (18). In contrast, some variations are seen at positions Gly1, Arg2, Phe12, Pro13, and Asp14. Structural changes around these residues are expected, because they are part of the loop where Gly1 and Asp14 are joined by the trans-peptidation reaction that leads to the head-to-tail cyclization of SFTI-1. The peptide bond preceding Pro9 was confirmed to be in a cis configuration based on NOE patterns (19) and 13C chemical shifts, like in SFTI-1, whereas all other proline residues were found to be in a trans conformation. Throughout the albumin domain, the majority of Hα resonances are upfield-shifted, and large stretches of neg-
ative secondary shift are consistent with extensive helical structure (Fig. 3).

To determine the solution structure of PawS1, structural restraints were obtained from the NMR data. Interproton distances were generated from NOE intensities in 3D $^{15}$N- and $^{13}$C-edited NOESY spectra, and backbone dihedral angles $\phi$ and $\psi$ and the side-chain torsion angle $\chi_1$ were derived from a TALOS-N analysis of the HN, H$\alpha$, Ca, C$\beta$, CO, and N chemical shifts (20). In addition, deuterium exchange experiments were conducted using $^{15}$N-labeled protein to identify hydrogen-bonded amide protons. Where hydrogen bond acceptors could be identified for slow exchanging amide protons during the structure calculations, hydrogen bond restraints were also included. Initial calculations involved automatic assignment of the NOESY data and generation of structures using torsion angle dynamics within the program CYANA (21). In the final round of structure calculation, a set of 50 structures was generated and refined in water within CNS (22), and the 20 structures with lowest energy were chosen to represent the solution structure of PawS1. Structural statistics for the ensemble demonstrate that the structure is of high quality and in good agreement with both experimental data and covalent geometry, as assessed by MolProbity (23) (Table 1).

**Structural analysis**

The 116-residue PawS1 structure is characterized by two structural entities separated by a flexible linker (Fig. 4). The N-terminal entity is the 14-residue peptide SFTI-1 in which the active site loop has retained the shape of native SFTI-1 and the small $\beta$-sheet is conserved. This is followed by the four-residue linker peptide GLDN, which connects the SFTI-1 domain to the albumin domain. The linker does not adopt a preferred single conformation in solution; thus, the relative positions of the SFTI-1 and albumin domains differ greatly within the structural ensemble (Fig. 4). The albumin structure consists of four helices that are closely packed in a tertiary fold, creating an extensive hydrophobic core. The small subunit (SSU) contains one helical segment (helix I, residues Ile$^{26}$–Thr$^{37}$), which is arranged anti-parallel to a second helix (helix II, residues Pro$^{59}$–Glu$^{65}$) located in the large subunit (LSU). The unstructured loop region comprising residues Thr$^{37}$–Asn$^{49}$ between helix I and helix II contains the seven-residue
linker peptide LRMAVEN that links the two subunits at the precursor stage. The LSU contains two additional helices (helix III (residues Gln71–Gln87) and helix IV (residues Gln94–Gln109)), which are of similar length and lie anti-parallel to each other but across the face of helix I and helix II at an angle of about 45°. The hypervariable region, which varies in length and...
structure and processing of PawS1

| Table 1 Structural statistics for the NMR structure of PawS1 |
|-----------------|-----------------|-----------------|
| Distance restraints | Total NOE | Interresidue |
|                 | 2,659 | 589 |
|                 | 2,070 | 930 |
|                 | 685 | 685 |
|                 | 455 | 455 |
| Hydrogen bond restraints | 68 |
| Dihedral-angle restraints | 208 | 94 |
|                 | 87 | 27 |
| Atomic root mean square deviation (Å) | 0.23 ± 0.06 | 0.71 ± 0.17 |
|                 | 0.53 ± 0.09 | 1.10 ± 0.11 |
| MolProbity statistics | Ramachandran favored (%) | Ramachandran outliers (%) |
|                 | 92.26 ± 1.98 | 0.53 ± 0.59 |
|                 | 2.92 ± 1.49 | 85.52 ± 2.23 |
|                 | 10.63 ± 1.65 (69%) | 3.23 ± 0.15 (57%) |
| Clash score | 1.65 (69%) | 0.09 |
| MolProbity score | 0.53 | 0.11 |

sequence among the 2S albums, is located between helix III and helix IV and in PawS1 comprises residues Gln

PawS1 contains five disulfide bonds, of which one is located in the SFTI-1 domain (Cys

To demonstrate the albumin-processing events, we used this buffer system and carried out digests of PawS1 with recombinantly produced HaAEP1. When PawS1 was incubated with HaAEP1 at 37 °C, it produced a product with a mass-to-charge ratio (m/z) of 1,930.91

NMR relaxation analysis

As is evident from the overlay of the ensemble, both domains adopt highly ordered structures throughout most of the protein sequence, but noticeable disorder is seen in the linker regions comprising residues 15–18 (GLDN) and 43–49 (LRMAVEN) (Fig. 4). Both linkers are classified as dynamic by TALOS-N based on their chemical shift pattern. To further confirm that these regions are flexible, a heteronuclear

Some non-productive events unrelated to the physiological processing were also observed under the experimental conditions. After 92 h, masses were observed at 12,712.51

In vitro digests of PawS1 with sunflower HaAEP1

Previous processing studies of the biosynthesis of macrocyclic peptides by AEPs have exclusively used small peptides as substrates (13–15). In this study, we wanted to investigate the processing of a native precursor substrate (PawS1) using a recombinant AEP. This requires a suitable buffer system that includes a redox system required for AEP activity, without reducing any of the five disulfide bonds in the protein, which would prevent successful maturation. In initial experiments using AEP-preferred conditions with DTT as a reducing agent, PawS1 disulfide bonds were found to be reduced by MS analysis. We instead chose milder conditions with an AEP activity buffer containing 0.6 mm/0.4 mm glutathione/glutathione disulfide (24), where no reduction was observed by MS. To also rule out disulfide shuffling, we compared two-dimensional

1H-15N HSQC NMR spectra of PawS1 dissolved in 90% water, 10% D₂O and PawS1 dissolved in 90% AEP activity buffer, 10% D₂O over a 2-week time frame. No significant peak shifts or appearance of new peaks were observed in the

1H-15N HSQC NMR spectra, confirming that the native structure was retained and that no substantial disulfide shuffling or degradation due to the mild redox conditions occurred.

To demonstrate the albumin-processing events, we used this buffer system and carried out digests of PawS1 with recombinantly produced HaAEP1. When PawS1 was incubated with HaAEP1 at 37 °C, it produced a product with a mass-to-charge ratio (m/z) of 1,930.91

1H-15N HSQC NMR spectra of PawS1 dissolved in 90% water, 10% D₂O and PawS1 dissolved in 90% AEP activity buffer, 10% D₂O over a 2-week time frame. No significant peak shifts or appearance of new peaks were observed in the

1H-15N HSQC NMR spectra, confirming that the native structure was retained and that no substantial disulfide shuffling or degradation due to the mild redox conditions occurred.
by HaAEP1-mediated cleavage at Asn111. The mass shift of +18 Da suggests additional HaAEP1-mediated cleavage after the linker peptide LRMAVEN between the two subunits was also occurring. Furthermore, after 5 h, a signal appeared at 1,531.61 in the low-mass MALDI TOF-MS spectrum of the no-enzyme control of PawS1, which corresponds to acyclic-SFTI-1. This reaction is a result of spontaneous hydrolysis at the Asp–Gly bond at low pH. Evidence for the reaction can also be seen in the high-mass spectrum at \( m/z \) 11,743.91, which corresponds to PawS1 still carrying the linker peptide GLDN but having lost the 14-residue SFTI-1 segment. Similar events were previously seen using the small AEP substrates (15).

**In situ digests of PawS1 with sunflower extract**

To further explore these cleavage events and any subsequent processing relying on other enzymes, PawS1 was incubated *in situ* with a sunflower seed extract, and the enzymatic reactions were monitored using mass spectrometry. After 2 h, a mixture
of intact PawS1 with an \( m/z \) of 13,255.91 and a PawS1 variant that was cleaved after the linker peptide LRMAVEN, resulting in a mass shift of +18 Da (13,274.51), was observed in the high-mass MALDI TOF-MS spectrum. After 5 h, the reaction neared completion, with the majority of the PawS1 starting material cleaved after the linker peptide LRMAVEN (Fig. 7A). A peak was observed at \( m/z \) 10,548.51 corresponding to the mature PawS1 albumin with the LRMAVEN linker peptide removed and the SFTI-GLDN peptide cleaved off at the N terminus (Fig. 7B); the latter cleavage of SFTI-GLDN from PawS1 was again, however, inefficient. In the low-mass MALDI TOF-MS spectrum, a low-intensity signal was observed at \( m/z \) 1513.91, matching cyclic SFTI-1. However, this was also present in the non-enzymatic control, consistent with SFTI-1 being present in the seed extract. It was thus not possible to confirm whether any mature SFTI-1 was produced from the PawS1 protein \textit{in situ}. If indeed SFTI-1 was produced, the amount was very limited.

To study the structural consequence of the processing, \(^{15}\text{N}\)-labeled PawS1 was incubated with sunflower \textit{in situ} extract, and the reaction was monitored by NMR spectroscopy. The initial AEP-mediated cleavage of the linker peptide LRMAVEN, which resulted in the formation of two flexible unrestrained ends, was rapid and after 7 h had gone to completion. This reaction was evident from the movement of resonances originating from residues in the region between Cys32 and Cys59, most notably Ser38, Asp40, Leu43, Ala46, and Glu48, which all reappeared in different positions in the \(^{15}\text{N}\) HSQC spectrum (Fig. 8). The appearance of Asn49 at a new downfield shift was consistent with it becoming a new C-terminal residue. The reaction continued with the removal of the entire LRMAVEN linker peptide. Previous work has suggested that aspartic
proteases are involved in the maturation of albumins and therefore could potentially be the factors responsible for this process (25). The appearance of yet another new peak with characteristics of a C terminus (Lys42) in the 15N HSQC spectra after 10 days, coupled with the disappearance of the peaks that moved initially, confirmed the removal of this peptide and that the maturation to a heterodimeric albumin PawS1 was complete. However, consistent with the mass spectrometric analysis, the second expected processing, the liberation of SFTI-GLDN from the protein, was remarkably inefficient. At the less sensitive NMR level, no evidence of any processing at the N-terminal region or consequently the formation of SFTI-1 or variants was observed. Importantly, despite the removal of the LRMAVEN linker, the appearance of the 15N HSQC spectra was largely unaffected, confirming that the overall structure was not affected by this processing event.

Discussion

PawS1, a protein with two faces and two fates

Despite the presence of five disulfide bonds, PawS1 could be expressed using the E. coli SHuffle strain into a fully oxidized and highly soluble protein in sufficient amounts for structural studies. The structure has two faces. The first is the albumin domain, which adopts a compact helical fold, rich in surface-exposed glutamine and arginine residues. Of the four helical segments in PawS1, helix I was located in the SSU, whereas the three remaining helical segments were all located in the LSU. The cysteine connectivity is conserved among seed storage albumins and features two interchain disulfide bonds connect-
Structure and processing of PawS1

Figure 9. Comparison of the NMR structure PawS1 with the sunflower albumin SESA3 and SFTI-1. A, the albumin domain of PawS1, showing SSU in green and LSU in orange, is superimposed on the sunflower albumin SESA3 shown in gray (PDB code of SESA3: 1S6D) (28). Helices are numbered with Roman numerals, and selected residues are labeled for clarity. B, the SFTI-1 domain of PawS1, shown in blue, is superimposed on the solution NMR structure of SFTI-1, shown in cyan (PDB code of SFTI-1: 1JBL) (18).

To date, seven seed storage albumin structures have been resolved and deposited in the Protein Data Bank, namely BnIb from B. napus (oilseed rape) (26), RicC3 from Ricinus communis (castor bean) (27), SESA3 (also known as SFA8) from H. annuus (sunflower) (28), rproBnIb from B. napus (29), Ara h 6 from Arachis hypogaea (peanut) (30), Mabinlin II from Capparis masaikai (31), and Ber e 1 from Bertholletia excels (Brazil nut) (32). Of these studied albumins, SESA3 is monomeric, whereas the remaining proteins are heterodimeric. Despite this difference, the fold is highly conserved. BnIb was produced in a recombinant form, rproBnIb, consisting of a single polypeptide, whereas in native BnIb, the linker peptide Ser2-Glu3-Glu4 has been removed by proteolytic cleavage during processing. Despite this difference, both proteins were shown to adopt very similar structures that are consistent with the structural architecture of RicC3 and SESA3 (29). Superimposing the albumin domain of our solution NMR structure of PawS1 on the sunflower albumin SESA3 highlights that PawS1 also adopts a similar helical bundle fold (Fig. 9A). However, a notable difference is present in the SSU. In SESA3, and all other structures, the SSU contains two helical segments, whereas PawS1 only contains one extended helix I segment. The N-terminal helix of the SSU in SESA3 is missing in PawS1, probably due to the segment between the cysteine residues in this region being three residues shorter, forcing a more extended conformation. C-terminally to the conserved SSU helical segment, both structures possess an unstructured loop of similar length but with a different amino acid composition, linking this helix to the first helix of the LSU. In the LSU, helix IV is significantly longer in PawS1, whereas helix IV of SESA3 is of similar length to the structurally equivalent counterparts in SESA3.

Despite the conserved fold, PawS1 and SESA3 have distinct characters. PawS1 is rich in nitrogen-containing amino acids, including 26 glutamines and six asparagines. In contrast, SESA3 is rich in sulfur-containing amino acids, including 16 methionines in addition to its eight cysteines. Many of the SESA3 methionine residues are buried in the protein center (28), creating a distinctly different hydrophobic core compared with PawS1, in which the core is dominated by isoleucine, leucine, and valine residues. Methionine residues in SESA3 are also found exposed on the surface, and the difference in the number of polar versus hydrophobic residues between PawS1 and SESA3 is reflected in their retention time on RP-HPLC, with SESA3 being the most late-eluting of the sunflower albumins (9). What has been referred to as the hypervariable loop region (Asn72–Met79) contains in SESA3 four hydrophobic residues (Met75, Trp76, Ile77, and Met79), which have been shown to create a solvent-accessible hydrophobic patch on the surface of the albumin. This combination of hydrophobicity and flexibility might contribute to its good emulsifying properties to form highly stable emulsions with oil/water mixtures (28). The corresponding region of PawS1 (Gln88–Gln93) is more hydrophilic and contains three glycine and three glutamine residues. The 26 glutamine residues in PawS1 are otherwise distributed throughout the entire structural surface.

The second face of PawS1 is the small N-terminal peptide region that becomes SFTI-1. Superimposing the solution structure of SFTI-1 on the SFTI-1 domain of PawS1 reveals an identical structural architecture of the β-sheet and trypsin-inhibitory binding loop (Fig. 9B). Although no trypsin inhibition assays have been performed with the full-length precursor, the identical arrangement of the binding loop in both structures suggests that PawS1 itself may also be able to inhibit trypsin like SFTI-1. The role of the cyclic backbone for the bioactivity of SFTI-1 has been investigated, and it was shown that opening the backbone in the cyclization loop causes only a small decrease in inhibitory activity (18). In acyclic-SFTI, the Gly1 and Asp14 termini remained close together, and the hydrogen-bonding network was largely intact (18). We observe that also in PawS1, the cyclization loop has adopted turn features that mimic the native SFTI-1, despite the lack of a cyclic backbone, and these features bring Gly1 and Asp14 together, which may be important for the backbone to be efficiently cyclized.

HaAEP1 can mediate post-translational processing of PawS1

The role of AEPs in albumin processing in vivo is well established. The most highly expressed AEP in sunflower seeds is HaAEP1, which can be recombinantly expressed, and thus we used this to reconstitute the first step of the reaction that produces SFTI-1 and matures the seed storage albumin. Previous studies into the processing mechanisms of this system have
been using small peptide substrates and only focused on the biosynthesis of the cyclic peptide SFTI-1 (15). Here, we show for the first time the processing events that take place in the bigger native precursor PawS1. HaAEP1 prefers Asn residues (15), and consistent with this, in our experiments it cleaves PawS1 at Asn18 and releases the SFTI-1 propeptide SFTI-GLDN. The linker peptide GLDN is located between the SFTI-1 domain and the albumin domain and is confirmed by our relaxation studies to be flexible. This flexibility is probably pivotal for enzyme access, because steric hindrance would prevent the AEP enzyme from performing the cleavage reaction. Flexibility can also be seen in the loop region that contains the linker peptide LRMAVEN between helix I (SSU) and helix II (LSU), which will allow access to the AEP cleavage site Asn49. The sunflower preproalbumins for SESA2-1, SESA2-2, and SESA20-2, unlike PawS1 and PawS2, do not contain an Asn or Asp in this loop region and consequently cannot be processed but remain monomeric (7). The sunflower albumin SES3 does contain an AEP cleavage candidate, Asn11, in this region; however, structural characterization by NMR spectroscopy revealed that it is located toward the end of helix Ib (28), where the helical structure probably prevents sunflower HaAEP1 from cleaving SES3 into a heterodimeric seed storage albumin.

Treatment with a sunflower seed extract in situ can fully mature PawS1

To further investigate the enzymatic cleavage events taking place with PawS1, a sunflower seed extract and NMR spectroscopy and mass spectrometry were used to monitor the processing in situ. The extract was able to rapidly process the linker between the SSU and LSU, and this cleavage was completed after 5–7 h, with the subsequent removal of the linker being completed over a few days. In contrast, only limited cleavage occurred after Asn18 to release SFTI-GLDN and produce the fully matured PawS1 albumin, and furthermore, the production of mature cyclic SFTI-1 was too low to be confirmed by NMR or mass spectrometry. An inefficiency of in situ macrocyclization of SFTI-1 has been reported previously (15), using seed extracts and the small enzyme substrate SFTI-GLDN. Only one in seven in situ reaction products from SFTI-GLDN was cyclic SFTI-1, with the remaining six in seven reaction products being acyclic-SFTI. Bernath-Levin et al. (15) found a degradative activity in sunflower seeds that led to the proposal of a breakdown pathway that would mask catalytic inefficiency in vivo by reducing the disulfide bonds and then degrading any acyclic-SFTI. However, the most likely limiting factor for the production of SFTI-1 in this study was the inefficient release of the SFTI-GLDN peptide segment, making the amount of smaller substrate available for further processing stoichiometrically unfavorable.

Why is SFTI-1 processing from PawS1 so inefficient in in vitro and in situ experiments?

Although we have presented evidence for the expected processing events of PawS1 using a recombinant AEP and a sunflower extract, the processes were surprisingly inefficient. This is in contrast to recent work on cyclotide processing using either extracted or recombinantly expressed AEPs, which has shown that substrates can be rapidly processed and cyclized to completion in vitro (13, 14). There are some fundamental differences between these studies. First, AEP is a cysteine protease relying on reducing conditions for function, which is not always compatible with disulfide-rich substrates. Rather than the stronger reducing conditions previously reported, we here use a milder redox system to prevent unfolding of the PawS1 protein during incubation. The stability under these conditions was found to be acceptable, but nonetheless evidence for reduction of one disulfide bond was observed, coupled with subsequent off-target processing of Asn11. Second, all previous studies have used small artificial substrates focusing on the final step of processing, the trans-peptidation reaction. Here, we use the full PawS1 processing, studying the earlier steps of processing. For the in vitro studies, we used recombinant HaAEP1. HaAEP1 mRNA expression levels in seeds are 100 times higher than HaAEP2 and HaAEP3 and 500 times higher than HaAEP4 and HaAEP5, suggesting that, although expression levels are not directly related to protein content, HaAEP1 is a dominant AEP in sunflower seeds (15). Incubation ratios of sunflower HaAEP1 and PawS1 were calculated, assuming an activity of 100% for expressed HaAEP1. However, due to the lack of an AEP enzyme inhibitor, no detailed characterization of the enzyme was performed to determine the active component of the expressed HaAEP1 fraction. Instead, a small peptide was used as an enzyme control for substrate cleavage under the same redox conditions, and it was converted to near completion within 2 h, confirming AEP activity. In contrast, less than half of PawS1 is cleaved at Asn49 after 92 h; thus, clearly the slower processing of PawS1 is substrate-specific and suggests that despite the flexibility in the key regions, the overall size of PawS1 presents an obstacle that limits AEP access. The Asn49 cleavage site between the SSU and LSU is much more efficiently targeted than the Asn18 cleavage site between SFTI-1 and the albumin. This difference can be explained from the PawS1 structure, suggesting that the conformation and protrusion from the core of the longer LRMAVEN loop make it considerably more accessible than the shorter GLDN linker, which is masked by the flanking SFTI-1 and albumin domains. Cyclic knottin precursors generally contain multiple knottin domains separated by linker segments in dedicated precursor proteins (16, 33), and these proteins may also require different conditions for efficient processing. Nonetheless, it is interesting to speculate that the processing in vivo is aided by additional co-factors in the seed compartment that ensure efficient production of SFTI-1.

Why is SFTI-1 hidden in a seed storage albumin?

Macroyclic peptides have now been described in plants, bacteria, fungi, and even mammals (17). They vary greatly in size, character, physiological function, and biosynthetic origin. Although a trend is emerging where the mature sequences are expressed as precursors with both N- and C-terminal extensions, thus requiring both cleavage and ligation events to produce mature products, the nature of the precursors, their size, and the point of cyclization are diverse (33–35). The significance of these different expression systems for the folding and production is less clear because the precursors are poorly studied. Intriguingly, there are no direct interactions between the
Structure and processing of PawS1

SFTI-1 domain and the albumin in PawS1; rather, they are separate individual structural entities. Consequently, despite their symbiotic existence from a structural perspective, it is unlikely that the albumin domain plays any role in the proper folding or processing of SFTI-1. The advantage of the genetic location of SFTI-1 may instead solely be related to the benefits of being able to hijack the albumin’s seed expression profile, folding, and processing machinery on the pathway via the endoplasmic reticulum and through the vacuolar system (10). Whether similar advantages have also driven the evolution of other classes of macrocyclic peptides to suit their physiological locations and functions remains to be seen. Recently, precursor variants of kalata B1 carrying both a short C-terminal tail and different lengths of an N-terminal sequence that is repeated before each cyclotide domain in the precursor protein were studied. The N-terminal sequence was found to be intrinsically unstructured, but appears to mediate self-association of precursors under NMR conditions, although the physiological significance of this is unclear (36).

In summary, this study provides the 3D structure of a macrocyclic precursor protein and new mechanistic insights into the maturation of sunflower proalbumins into an albumin and SFTI-1. The structural characterization of PawS1 using triple-resonance NMR experiments revealed a structure consisting of two well-defined entities connected by a flexible linker, GLDN. A second flexible linker, LRMAVEN, separates the two subunits of the albumin. These linkers can be targeted by AEPs to produce the cleaved heterodimeric albumin PawS1 and the short peptide SFTI-GLDN, the starting point for further processing into cyclic SFTI-1. The site separating the albumin subunits is significantly more accessible for processing, being much more readily cleaved by both recombinant HaAEP1 and the in situ seed extract. The extract also contains secondary processing enzymes that remove the LRMAVEN sequence, leading to the fully matured albumin observed in seeds. The formation of cyclic SFTI-1 from PawS1 in situ could not be confirmed, probably due to both the inefficiency of liberation of SFTI-GLDN and the inefficiency of its biosynthesis from the short substrate, which has been reported previously (15). This is an intriguing observation and may suggest that additional auxiliary components or enzymes that are not present or active under the in situ conditions are involved in the biosynthesis of SFTI-1.

Experimental procedures

Cloning of PawS1

A synthetic gene for PawS1 with E. coli codon optimization (GeneArt) was produced, such that it encoded the 151-residue PawS1 minus the 21-residue signal peptide and a 14-residue spacer between the signal peptide and SFTI-1. The coordinates of the PawS1 protein were PawS1(Gly36–Ile151), with PawS1 Gly36 being the Gly1 of SFTI-1. The SFTI-1 Gly36 (i.e. PawS1(Gly36)) was made the P1 cleavage site. A restriction site was included that allowed the ORF to be cloned into pQE30 (Qiagen) that adds an N-terminal sequence, including a His6 tag. The final sequence of the recombinant PawS1 protein encoded in pQE30 includes an N-terminal tag (MRGSHHHHHHH), a flexible linker coded by BamHI (GS), and a TEV cleavage site (ENLYFQ) followed by SFTI-1 (GRCTKSIPPLICFPD) and its tail (GLDN) that connects SFTI-1 to the albumin domain (Pro144–Ile151); MRGSHHHHHHH-SENLYFQGRCTKSIPPLICFDPGLDN

Recombinant expression of 12C- and 15N-labeled PawS1

The pQE30-PawS1 construct and the repressor plasmid pREP4 (Qiagen) were co-transformed into SHuffle® (New England Biolabs), an E. coli strain engineered to promote disulfide bond formation in its cytoplasm (37). SHuffle E. coli containing the pQE30-PawS1 construct was grown at 30 °C at 200 rpm in LB medium supplemented with ampicillin (100 µg/ml). When optical density reached A600 ~0.8, cells were centrifuged at 5,000 × g for 15 min at 25 °C. Cell pellets were resuspended in minimal medium (one-quarter of the volume of LB used to grow cells). The minimal medium was made as described previously and contained 15N-labeled ammonium chloride and d-glucose-13C6 (Sigma-Aldrich) (38). The labeled culture was incubated for 1 h at 30 °C at 200 rpm to promote cell recovery and then cooled to 16 °C for 1 h before adding isopropyl β-D-1-thiogalactopyranoside at a final concentration of 0.4 mM. The culture was incubated for 18 h in the dark at 16 °C at 200 rpm before being harvested by centrifugation at 6,000 × g for 15 min at 4 °C. Cell pellets were stored at −80 °C until required.

Purification of PawS1

The cell pellet was resuspended in lysis buffer 50 mM Tris, 300 mM sodium chloride, 10 mM imidazole, pH 8.0, followed by lysis by sonication. The soluble lysate was retained after centrifugation at 15,000 rpm for 30 min at 4 °C. The His₆-TEV-PawS1 fusion protein was purified by passing the supernatant over a nickel–Sepharose HisTrap HP 5-ml column (GE Healthcare Life Sciences) using an ÄKTA Start chromatography system (GE Healthcare), followed by a washing step with 50 mM tris(hydroxymethyl)aminomethane, 300 mM sodium chloride, 60 mM imidazole, pH 8.0. The PawS1 fusion protein was eluted with 50 mM tris(hydroxymethyl)aminomethane, 300 mM sodium chloride, 500 mM imidazole, pH 8.0. Flow rate was kept constant at 1 ml/min at all times. The imidazole-rich buffer was exchanged using a PD-10 column (GE Healthcare) and replaced with a buffer compatible with TEV (50 mM tris(hydroxymethyl)aminomethane, 100 mM sodium chloride, 2 mM EDTA, 0.6 mM/0.4 mM glutathione/glutathione disulfide, pH 8.0). TEV protease was added in a ratio of 1:20 (w/w), and the cleavage reaction was allowed to proceed at room temperature for 16 h with gentle agitation.

The cleaved PawS1 protein was desalted using a PD-10 column (GE Healthcare) and separated further by RP-HPLC on a Shimadzu Prominance system using an analytical Grace Vydac C18 column (250 × 4.6 mm, 5 μm, 300 Å) at a 1% gradient in which buffer A was 0.05% trifluoroacetic acid and buffer B was 90% acetonitrile, 0.05% trifluoroacetic acid. To determine the average mass of the labeled PawS1, the protein was analyzed by LC-MS on a SCIEX API 2000 LC-MS/MS electrospray mass spectrometer. A volume of 10 µl was injected at a flow rate of 0.1 ml/min with buffer A/buffer B ratio of 30:70. Buffer A consisted of 0.1% formic acid and buffer B contained 0.1% formic acid in 90% acetonitrile. Specifically, LC-MS/MS instrument settings were as follows: declustering potential, 88 V; focusing
potential, 220 V; entrance potential, 8 V; Q1 MS, Q1. The protein concentration was determined using a Direct Detect® infrared spectrometer (Merck Millipore). Purified PawS1 was lyophilized and stored at −20 °C until required.

**NMR spectroscopy**

The structure of PawS1 was determined using heteronuclear NMR. Samples for NMR contained 0.3 ml of 15N- or 13C/15N-labeled protein in 90% water and 10% D2O (v/v) at pH 4.6 at a final concentration of 3.6 mg/ml and were added to susceptibility-matched 5-mm outer-diameter microtubes (Shigemi Inc.). All spectra were acquired at 25 °C on an Avance 600 MHz spectrometer equipped with a cryoprobe (Bruker BioSpin). Resonance assignments were obtained using 2D 1H-15N HSQC, 2D 1H-13C HSQC, 2D HNCACB, 3D CBCA(CO)NH, 3D HNCO, 3D HBCA(CO)NH, 3D H(CC)(CO)NH-TOCSY, and 3D (H)(CC)(CO)NH-TOCSY. 3D 13C- and 15N-edited HSQC-3D13C- and 15N-NOESY spectra acquired using a mixing time of 120 ms were recorded for structural information, and a 2D 1H-15N-NOE relaxation experiment with saturation turned on or off during a relaxation delay of 3 s was recorded to assess flexibility. Lyophilized protein was dissolved in D2O, and deuterium exchange was monitored by recording 2D 1H-15N HSQC data sets. All data were recorded using linear sampling. 2D data were processed using Topspin version 3.0 (Bruker Biospin), and 3D data were processed with the Rowland NMR Toolkit (http://rnmrk.uchc.edu/rnmrk/RNMRTK.html).

**Spectral assignment and structure calculations**

NMR spectra were analyzed, peak picked and resonances assigned manually using CcpNmr Analysis version 2.4.2 (39). Interproton distance restraints were derived from NOE cross-peak heights from 3D 13C-aliphatic, 13C-aromatic, and 15N-NOE-NOESY-HSQC spectra acquired using a mixing time of 120 ms. Protein backbone torsion angles and side chain torsion angle ψ were predicted based on the chemical shift assignments of HN, Hα, Ca, Cβ, and N resonances using the artificial neural network-based program TALOS-N (20). For amide protons that were found to be slow exchanging, hydrogen bond restraints were included if acceptors were unambiguously identified in preliminary structures. Backbone hydrogen bonds between amides and carbonyls included Phe12–Arg20, Ala103–Val100, Gln104–Lys101, Met105–Lys102, Leu106–Ala103, Glu80–Lys76, His35–His31, Met110–Val108, Glu112–Val108, Glu110–Lys110, and Cys110–Lys110, and side chain ψ and side chain torsion angle ψ were predicted based on the chemical shift assignments of HN, Hα, Ca, Cβ, and N resonances using the artificial neural network-based program TALOS-N (20).

**Stability studies of PawS1 in AEP activity buffer**

To explore the stability of PawS1, 400 μg of 15N-labeled PawS1 was dissolved in 500 μl containing 90% AEP activity buffer (100 mM sodium acetate, 5 mM EDTA, 0.4 mM glutathione/glutathione disulfide, pH 5.0), 10% D2O (v/v), and a two-dimensional 1H-15N HSQC NMR spectrum was recorded at various time points over a 2-week period. The 1H-15N HSQC data sets of the PawS1 dissolved in 90% AEP activity buffer, 10% D2O (v/v) were then compared with the 1H-15N HSQC data sets of the PawS1 dissolved in 90% water and 10% D2O (v/v).

**In vitro digests of PawS1 with sunflower HaAEP1**

Recombinant HaAEP at a final concentration of 4 μg/ml was incubated with either 74 μM SFTI(D14N)-GL or 9.5 μM PawS1 for activity measurements in AEP activity buffer at 37 °C. SFTI(D14N)-GL at a final concentration of 37 μM in AEP activity buffer without HaAEP1 and PawS1 at a final concentration of 4.8 μM in AEP activity buffer without HaAEP1 were used as negative controls. At 0, 2, 5, 21, and 92 h, 5-μl aliquots were removed and desalted using ZipTip pipette filters (Merck Millipore) before being combined in a 1:1 ratio with α-cyano-4-hydroxycinnamic acid (CHCA) matrix (Bruker Daltonics) onto a 384 MTP polished steel target plate, ready for MALDI-TOF.
**Structure and processing of PawS1**

MS analysis. The concentration of the CHCA matrix was 0.7 mg/ml.

**In situ digests of PawS1 with sunflower seed extract**

Ten kernels of sunflower seeds (purchased from Coles, Melbourne, Australia) were frozen in liquid nitrogen and crushed using a mortar and pestle. The crushed seed meal was suspended in AEP activity buffer at a ratio of 0.1 ml of AEP activity buffer per dehulled sunflower seed kernel. The resuspended meal was vortexed for 10 min and centrifuged at 10,000 × g for 2 min. The supernatant was transferred into a clean tube and mixed with an equal amount of n-hexane to remove lipids. The mixture was spun at 14,000 × g for 10 s, the n-hexane was removed, and the extract was transferred to an Amicon Ultra 0.5-ml 10-kDa centrifugal filter (Merck Millipore) and topped up with AEP activity buffer. The sunflower seed extract was centrifuged at 14,000 × g for 10 min. Sunflower seed extract was added in a ratio of 2 parts/5 parts (volume extract/mass proteins) to unlabeled and 15N-labeled PawS1. The mixture was incubated at 37 °C for 0, 2, and 5 h, with 5–µl aliquots taken at each time point and desalted using ZipTip pipette filters (Merck Millipore) before being combined with CHCA matrix (Bruker Daltonics) onto a 384 MTP polished steel target plate, ready for MALDI-TOF MS analysis.

For 15N NMR studies, 70 kernels of sunflower seeds were crushed as described above. After supernatant was transferred into a clean tube and mixed with an equal amount of n-hexane to remove lipids, no Amicon Ultra 0.5-ml centrifugal filter was used in the 125N NMR study, but the sunflower seed extract was filtered through a 0.45-µm filter instead. 400 µg of 15N-labeled PawS1 were dissolved in 450 µl of in situ sunflower seed extract containing AEP activity buffer and 50 µl of D2O (v/v), and 1H-15N HSQC NMR measurements were undertaken at 37 °C at various time points over 10 days.

**MALDI-TOF MS**

To monitor the processing of PawS1, 1.2 µl of desalted peptides/proteins were mixed with 1.2 µl of matrix, and 2.4 µl were spotted onto a 384 MTP polished steel target plate. The matrix was prepared according to the Bruker Daltonics protocol using the CHCA dried droplet method for polished steel target plates. The Bruker peptide calibration standard (Bruker Daltonics) was used for calibrating the mass range 800–4,000 Da (low mass), and the Bruker protein calibration standard I (Bruker Daltonics) was used for calibrating the mass range 5,000–14,000 Da (high mass). Samples were analyzed on a Bruker UltrafleXtreme™ MALDI-TOF/TOF mass spectrometer (Bruker Daltonics) with laser intensity at 20% for low mass and 30% for high mass in the in vitro assay and 50% for low mass and 40% for high mass in the in situ assay. 5,000 shots were summed per MS analysis.

**Acknowledgments**—We acknowledge the Biological NMR facility at the Institute for Molecular Bioscience for access to the 600-MHz spectrometer and CSIRO Agriculture and Food for access to the Bruker UltrafleXtreme MALDI-TOF/TOF mass spectrometer.

**References**

1. Shimada, T., Yamada, K., Kataoka, M., Nakaune, S., Koumoto, Y., Kuroyanagi, M., Tabata, S., Kato, T., Shinozaki, K., Seki, M., Kobayashi, M., Kondo, M., Nishimura, M., and Hara-Nishimura, I. (2003) Vacuolar processing enzymes are essential for proper processing of seed storage proteins in Arabidopsis thaliana. J. Biol. Chem. 278, 32292–32299

2. Gruis, D., Schulze, J., and Jung, R. (2004) Storage protein accumulation in the absence of the vacuolar processing enzyme family of cysteine proteases. Plant Cell 16, 270–290

3. Shewry, P. R., and Casey, R. (1999) Seed Proteins, Kluwer Academic, Dordrecht, Netherlands

4. Otegui, M. S., Herder, R., Schulze, J., Jung, R., and Staelinlin, L. A. (2006) The proteolytic processing of seed storage proteins in Arabidopsis embryo cells starts in the multivesicular bodies. Plant Cell 18, 2567–2581

5. Mylne, J. S., Hara-Nishimura, I., and Rosengren, K. J. (2014) Seed storage albums: biosynthesis, trafficking and structures. Funct. Plant Biol. 41, 671–677

6. Ericson, M. L., Rödin, J., Lenman, M., Gilmielius, K., Josefsson, L. G., and Rask, L. (1986) Structure of the rapeseed 1.7 S storage protein, napin, and its precursor. J. Biol. Chem. 261, 14576–14581

7. Franke, B., Colgrave, M. L., Mylne, J. S., and Rosengren, K. J. (2016) Mature forms of the major seed storage albums in sunflower: a mass spectrometric approach. J. Proteomics 147, 177–186

8. Kortt, A. A., Caldwell, J. B., Lilley, G. G., and Higgins, T. J. (1991) Amino acid and cDNA sequences of a methionine-rich 2S protein from sunflower seed (Helianthus annuus L.). Eur. J. Biochem. 195, 329–334

9. Jayasena, A. S., Franke, B., Rosengren, J., and Mylne, J. S. (2016) A tripartite approach identifies the major seed albums. Theor. Appl. Genet. 129, 613–629

10. Mylne, J. S., Colgrave, M. L., Daly, N. L., Chanson, A. H., Elliott, A. G., McCallum, E. J., Jones, A., and Craik, D. J. (2011) Albums and their processing machinery are hijacked for cyclic peptides in sunflower. Nat. Chem. Biol. 7, 257–259

11. Craik, D. J., Fairlie, D. P., Liras, S., and Price, D. (2013) The future of peptide-based drugs. Chem. Biol. Drug Des. 81, 136–147

12. Elliott, A. G., Delay, C., Liu, H., Phua, Z., Rosengren, K. J., Benfield, A. H., Panero, J. L., Colgrave, M. L., Jayasena, A. S., Dunse, K. M., Anderson, M. A., Schilling, E. E., Ortiz-Barrientos, D., Craik, D. J., and Mylne, J. S. (2014) Evolutionary origins of a bioactive peptide buried within preproalbumin. Plant Cell 26, 981–995

13. Nguyen, G. K., Wang, S., Qiu, Y., Hemu, X., Lian, Y., and Tam, J. (2014) Butelase 1 is an Axx-specific ligaese enabling peptide macrocyclization and synthesis. Nat. Chem. Biol. 10, 732–738

14. Harris, K. S., Durek, T., Kaas, Q., Poth, A. G., Gilding, E. K., Conlan, B. F., Saska, I., Daly, N. L., van der Weerden, N. L., Craik, D. J., and Anderson, M. A. (2015) Efficient backbone cyclization of linear peptides by a recombinant asparaginyl endopeptidase. Nat. Commun. 6, 10199

15. Bernath-Levin, K., Nelson, C., Elliott, A. G., Jayasena, A. S., Millar, A. H., Craik, D. J., and Mylne, J. S. (2015) Peptide macrocyclization by a bifunctional endopeptidase. Chem. Biol. 22, 571–582

16. Mylne, J. S., Chan, L. L., Chanson, A. H., Daly, N. L., Schaefer, H., Bailey, T. L., Nguyenpong, C., Cascales, L., and Craik, D. J. (2012) Cyclic peptides arising by evolutionary parallelism via asparaginyl-endopeptidase-mediated biosynthesis. Plant Cell 24, 2765–2778

17. Craik, D. J. (2006) Chemistry: seamless proteins tie up their loose ends. Science 315, 1563–1564

18. Korszynczy, M. L., Schirra, H. J., Rosengren, K. J., West, J., Condie, B. A., Otvos, L., Anderson, M. A., and Craik, D. J. (2001) Solution structures by 1H NMR of the novel cyclic trypsin inhibitor SFTI-1 from sunflower seeds and an acyclic permutter. J. Mol. Biol. 311, 579–591
26. Rico, M., Bruix, M., González, C., Monsalve, R. I., and Rodríguez, R. (1996) NMR of proteins and nucleic acids, John Wiley & Sons, Inc., New York
27. Pantoja-Uceda, D., Bruix, M., Giménez-Gallego, G., Rico, M., and Santoro, J. (2004) Solution structure and stability against digestion of rproBnIb, a recombinant 2S albumin from rapeseed: relationship to its allergenic properties. J. Mol. Graph. 14, 51–55, 29–32
28. Pantoja-Uceda, D., Shewry, P. R., Bruix, M., Tatham, A. S., Santoro, J., and Rodríguez, R. I. (2005) MolProbity: all-atom structure validation for macromolecular crystallography. Acta Crystallogr. D Biol. Crystallogr. 61, 12–21
29. Pantoja-Uceda, D., Palomares, O., Bruix, M., Villalba, M., Rodriguez, R., Rico, M., and Santoro, J. (2004) Solution structure of a methionine-rich 2S albumin from Ricinus communis J. (2003) Solution structure of RicC3, a 2S albumin storage protein from Ricinus communis. Biochemistry 42, 13839–13847
30. Lehmann, K., Schweinern, K., Reese, G., Randow, S., Suhr, M., Becker, W.-M., Vieths, S., and Rösch, P. (2006) Structure and stability of 2S albumin-type peanut allergens: implications for the severity of peanut allergic reactions. Biochem. J. 395, 463–472
31. Li, D.-F., Jiang, P., Zhu, D.-Y., Hu, Y., Max, M., and Wang, D.-C. (2008) Crystal structure of Mabinlin II: a novel structural type of sweet proteins and the main structural basis for its sweetness. J. Struct. Biol. 162, 50–62
32. Rundqvist, L., Tengel, T., Zdunek, J., Björn, E., Schleucher, J., Alcocer, M. J. C., and Larsson, G. (2012) Solution structure, copper binding and backbone dynamics of recombinant Ber e 1: the major allergen from Brazil nut. PloS One 7, e46435
33. Jennings, C., West, J., Waine, C., Craik, D., and Anderson, M. (2001) Biosynthesis and insecticidal properties of plant cyclotides: the cyclic knotted proteins from Oldenlandia affine. Proc. Natl. Acad. Sci. U.S.A. 98, 10614–10619
34. Himeno, K., Rosengren, K. J., Inoue, T., Perez, R. H., Colgrave, M. L., Lee, H. S., Chan, L. Y., Henriques, S. T., Fujita, K., Ishibashi, N., Zendo, T., Wilaipun, P., Nakayama, J., Leelawatcharamas, V., Jikuya, H., Craik, D. J., and Sonomoto, K. (2015) Identification, characterization, and three-dimensional structure of the novel circular bacteriocin, enterocin NKR-5-3B, from Enterococcus faecium. Biochemistry 54, 4863–4876
35. Tang, Y.-Q., Yuan, J., Osapay, G., Osapay K. Tran, D., Miller, C. J., Ouellette, A. J., and Selsted, M. E. (1999) A cyclic antimicrobial peptide produced in primate leukocytes by the ligation of two truncated α-defensins. Science 286, 498–502
36. Daly, N. L., Gunasekera, S., Clark, R. J., Lin, F., Wade, J. D., Anderson, M. A., and Craik, D. J. (2016) The N-terminal pro-domain of the kalata B1 cyclotide precursor is intrinsically unstructured. Biopolymers 106, 825–833
37. Labbe, S., Arimura, Y., Takeda, M., Watanabe, H., Ohtsuka, M., Ueda, K., and Shiro, R. (2013) Structure and processing of PawS1 from Oldenlandia affinis Biochem. J. 450, 12398–12411