

RESEARCH ARTICLE

Crystal structure of nonspecific lipid transfer protein from *Solanum melongena*

Abha Jain^{1,2} | Dinakar M. Salunke^{1,3} 

¹Regional Centre for Biotechnology,
Structural Biology Lab, Faridabad 121001,
India

²Manipal University, Manipal, Karnataka
576104, India

³International Centre for Genetic
Engineering and Biotechnology, Structural
Immunology Lab, New Delhi 110067, India

Correspondence

Dinakar M. Salunke, International Centre
for Genetic Engineering and Biotechnology,
New Delhi 110067, India.
Email: dinakar.salunke55@gmail.com

Funding information

Departments of Science & Technology and
Biotechnology, Govt of India.

Abstract

Lipids are considered to protect protein allergens from proteolysis and are generally seen to exist in a bound form. One of the well-known plant protein families with bound lipids is non-specific lipid transfer proteins (nsLTPs). Structure-function relationships in the case of the members of non-specific lipid transfer protein family are not clearly understood. As part of exploring the seed proteome, we have analyzed the proteome of a member of Solanaceae family, *Solanum melongena* (eggplant) and a non-specific lipid transfer protein from *S. melongena*, SM80.2 was purified, crystallized and the structure was determined at 1.87 Å resolution. Overall, the tertiary structure is a cluster of α -helices forming an internal hydrophobic cavity. Absence of conserved Tyr79, known to govern the plasticity of hydrophobic cavity, and formation of hydrogen bond between Asn79 and Asn36 further reduced the pocket size. Structural analysis of SM80.2 thus gives insight about a new hydrogen bond mediated mechanism followed in closure of the binding pocket. Extra electron densities observed at two different places on the protein surface and not in the cavity could provide interesting physiological relevance. In light of allergenic properties, probably overlapping of epitopic region and ligand binding on surface could be a main reason. This work shows first crystal structure of A-like nsLTP with a close binding pocket and extra density on the surface suggesting a plausible intermediate state during transfer.

KEYWORDS

allergy, lipid, nonspecific lipid transfer proteins, *Solanum melongena*, solanaceae

1 | INTRODUCTION

Allergens are mostly water soluble proteins, stable to proteolytic degradation and the size ranging from 10 to 70 kDa. However, these characteristics alone cannot explain their allergenicity.¹ Lipid binding is one of the characteristic known in protein families that includes allergens like 2 S albumin, nonspecific lipid transfer proteins (nsLTPs) and Oleosins. While physiologically bound lipids could be critical for protein function, not much is known about the role of lipids in seed storage proteins, although it is suggested that the lipid binding can modulate stability of allergenic proteins.² Allergens have been identified in commonly consumed plants belonging to Solanaceae family.³ Patalin from potato, profilins from tomato and pathogenesis related protein P-23 from bell pepper are a few among them. In recent years, allergenic reactions due

to eggplant have been reported but they are not as common. Certain metabolites have also been identified as allergens in eggplant.⁴ Apart from Solanaceae family, allergens are also known from other plant families like Poaceae family (Rice),⁵ Anacardiaceae family (Cashew),⁶ Fabaceae family (Peanut),⁷ and Lecythidaceae (Brazil nut).⁸

Proteins causing allergic reactions are present in many parts of plants but are highly accumulated in seeds and most of them degrade during germination.⁹ Certain functional studies were reported for allergens from *Solanum lycopersicum* and *Solanum tuberosum*. So far, studies related to allergenic property of proteins have not been explored in case of *Solanum melongena*. Some reports suggested that *S. melongena* can cause allergy by means of factors associated with proteins, although such factors were not clearly identified.¹⁰

Nonspecific lipid transfer proteins (nsLTPs) are one of the major plant allergen families having lipid associated with the protein. Lipids are known to be bound in the hydrophobic cavity in these proteins.¹¹

Abbreviations: nsLTPs, nonspecific lipid transfer proteins.

The nsLTPs belong to prolamin superfamily and are characterized on the basis of conserved cysteine residues, presence of disulfide bonds, basic nature, and high content of α -helices.¹² Nonspecific lipid transfer proteins possess the ability to transfer different lipids between membranes. They are assumed to play a critical role in plant defense forming a protective hydrophobic cover on plant aerial organs. Recently these proteins are also designated as pan allergens.¹³

Thus, it becomes interesting to elucidate structure of this class of proteins and perform comparative analysis, which may shed further light on the role of lipids in causing allergy. It is possible that comparing the structure with other members of the nsLTP family will help in deciphering the importance played by these proteins in physiology processes and allergy. Here we report the identification, purification, crystallization and structure determination of SM80.2, a 9.388 kDa protein from *S. melongena* by *ab initio* phasing. SM80.2 showed very high sequence homology with A-like nsLTPs. A-like nsLTPs are one of the subclass of Type1 nsLTPs. This suggests that SM80.2 belongs to Type1 nsLTP family of seed storage proteins, members of which are identified to be intrinsically allergenic. SM80.2 showed alpha helical structure with a closed hydrophobic pocket. We also found unexplained electron densities at two places on the protein surface which we modeled as two relevant lipids. To best of our knowledge, this is the first nsLTP structure where closure of the lipid binding pocket through hydrogen bonding is observed. Unusual electron density observed on the surface at two places has been interpreted as possible lipid binding.

2 | EXPERIMENTAL PROCEDURES (MATERIALS AND METHODS)

2.1 | Isolation and purification

Protein was isolated from 50 g *S. melongena* seeds bought from National Seed Corporation, New Delhi. Isolation and purification protocol used was same as used in our earlier study.¹⁴ In brief, seeds were washed, ground and delipidified. Delipidification was done using petroleum ether. The dried extract was then homogenized in 50 mM Tris-Cl, 150 mM NaCl pH 8 along with protease inhibitor cocktail (Sigma) and stirred at 4°C for 3 h. Homogenized extract was centrifuged at 7 000 rpm (12 300 g) for 1 h at 4°C and salt fractionation was performed in the supernatant over a concentration range of 0–95% (w/v) ammonium sulfate. This resulted in solubility based separation of proteins. Pellets obtained from each fraction were dissolved in homogenization buffer and then dialyzed in order to remove the salt. Further, the proteins in each fraction were analyzed on SDS-PAGE. The isolation and purification were done by size exclusion chromatography and BCA concentration assay was used for determining the concentration of purified protein.

2.2 | Protein sequence analysis

Segregated proteins from the selected fraction were transferred on polyvinylidene difluoride (PVDF) membrane. Amino acid sequencing

was done by Edman degradation method on PPSQ-33 A protein sequencer (Shimadzu). N-terminal sequence thus obtained was analyzed for sequence based homology using BLAST database search.¹⁵

2.3 | Crystallization

For crystallization trials, the protein sample was prepared in 50 mM Tris-Cl pH 8.0 and concentration was adjusted to 10 mg mL⁻¹. Crystallization was done using conventional hanging-drop vapor diffusion method at room temperature. Initial screening was done by using various commercially available Hampton PEG/Ion screen, PEG/Ion 2 screen, Crystal screen, Crystal screen 2 and Index screen set up by a mosquito crystallization robot. Drops were composed of equal volume (250 nL) of protein and reservoir solution equilibrated against 100 μ L of reservoir solution. Few hits were obtained in Index (condition 93) and PEG/Ion screen (condition 30). Diffraction quality crystals were obtained after optimization of conditions in a drop size of 2 + 2 μ L of protein and reservoir solution equilibrated against 500 μ L reservoir solution containing 0.05 M zinc acetate dihydrate with 20% w/v polyethylene glycol 3 350 Da.

2.4 | Data collection and statistics

Rigaku FR-E + superbright dual X-ray source (Rigaku) at wavelength of 1.5418 Å having RAXIS IV++ was used as X-ray source. The data were collected at 120 K with a crystal-to-detector distance of 120 mm. Crystals were cryoprotected using 30% w/v glycerol in mother liquor. Data were collected to 2 Å resolution and integrated and scaled with *iMosflm*.¹⁶ For *ab initio* structure solution two data sets were collected at BM14, European Synchrotron Radiation Facility (ESRF), Grenoble, France from the same crystals belonging to space group P6₁ at 0.953 Å and 1.771 Å respectively. These dataset are collected at 2.5 and 1.87 Å resolution. Data processing was done using *iMosflm*.

2.5 | Structure determination

For extracting the initial phases high redundancy sulfur-anomalous data was required, so 2.5-Å resolution data were collected at 3 different kappa angles. This approach resulted in 65 fold redundancy. Anomalous merged data were used by PHENIX software suite¹⁷ to calculate phases. AutoSol was used for the purpose of preparing preliminary model. Further, AutoBuild assisted in improving the phases and model building. Iterative cycles of model building was done on COOT¹⁸ and refinement was carried out by phenix.refine program of PHENIX software. eLBOW¹⁹ was used to generate the topology files for ligand refinement. The final quality was checked by PROCHECK.²⁰ Simulated annealing Ligand omit maps were prepared by CNS.²¹ Figures were prepared using PyMOL.²²

2.6 | Accession numbers

The accession numbers for the SM80.2 structure coordinate is 5TVI.

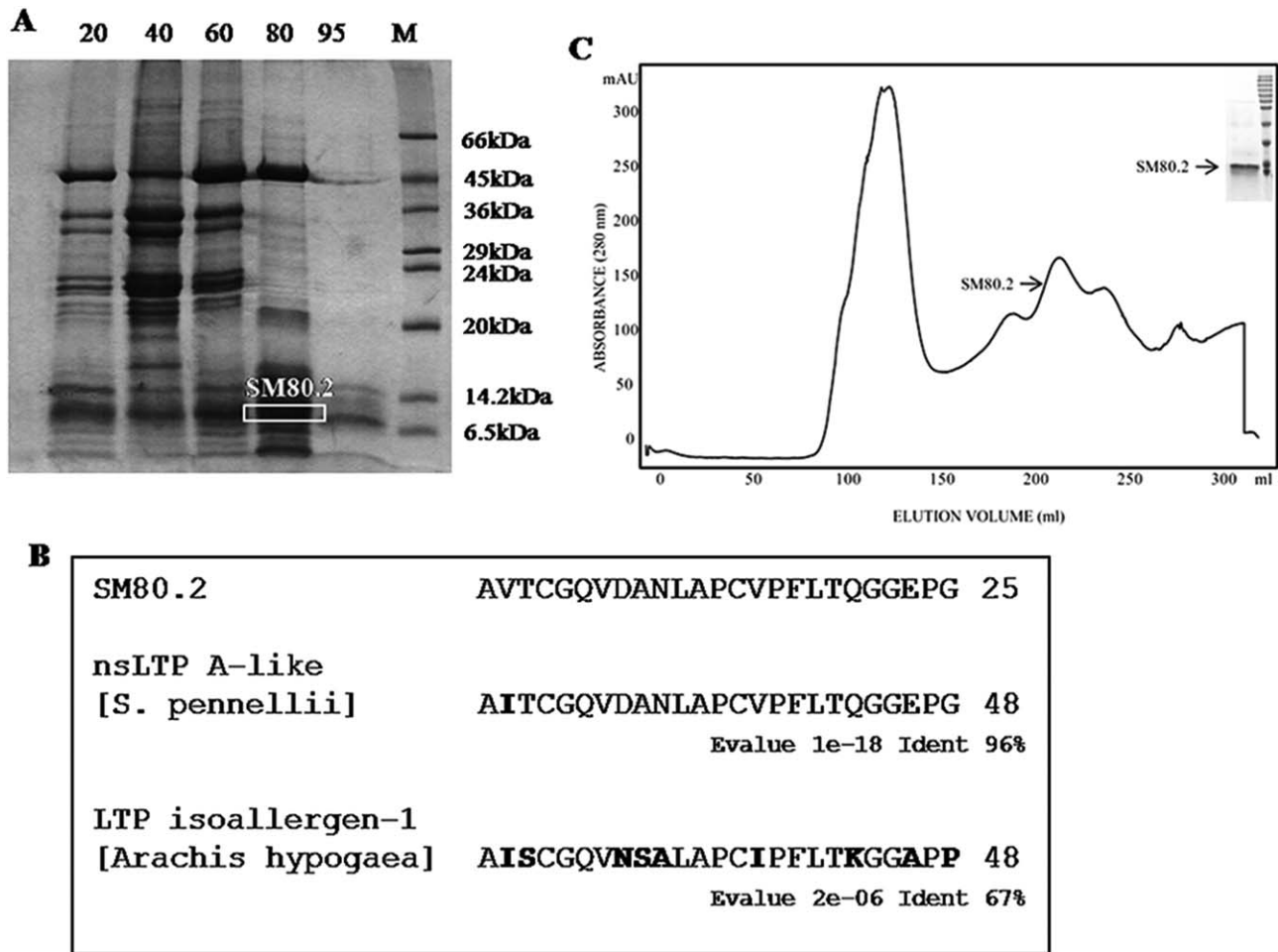


FIGURE 1 Protein identification and purification from seed proteins of *Solanum melongena*. (A) Lanes 1–5 corresponds to 0%–20%, 20%–40%, 40%–60%, 60%–80%, and 80%–95% ammonium sulfate fractions, and lane 6 of 15% SDS PAGE corresponds to the low molecular weight protein markers (sigma) respectively. Targeted band is labelled as SM80.2. (B) N-terminal sequence alignment of SM80.2 with the homologous sequences identified by blast search. In bold are the different residues as compared to query sequence. (C) Purification profile of 80% protein fraction on self-packed sephacryl-200 column. Peak 2 corresponds to the purified protein and labeled as SM80.2. In inset is 15% SDS PAGE of the purified protein along with molecular weight markers. Lane 1 corresponds to SM80.2 and lane two shows molecular weight markers

3 | RESULTS

Proteins associated with lipids are observed to initiate sensitization by regulating immune system.² Many proteins including nonspecific lipid transfer proteins are thought to possess similar potential. Therefore, comparative structural analysis of seed protein associated with lipid transfer from *S. melongena* fruit, which is commonly consumed due to its nutritional value, was addressed. This will also help in understanding the physiological functions of proteins and lipids from edible seeds.

3.1 | Protein purification and crystallization

Protein fractions obtained from ammonium sulfate fractionation were electrophoresed on 15% SDS PAGE and analyzed for proteins with significant molecular weight differences. Apart from proteins characterized earlier,¹⁴ we identified N-terminal sequence of a low molecular weight protein from 80% ammonium sulfate fraction by Edman degradation method. This protein band showed >90% homology with A-like

nonspecific lipid-transfer protein from Solanaceae family on BLAST database search (XP_015084690.1, XP_006347954.1, XP_016544907.1, XP_004233683.1). This protein also showed very high homology with lipid transfer protein isoallergen 1 precursor of *Arachis hypogaea*.²³ Partially purified 80% protein fraction was subjected to further purification using sephacryl-200 column. Peak 2 of chromatographic profile corresponds to the purified protein and was named SM80.2 (Figure 1). This eluted fraction was concentrated to 10 mg mL⁻¹ by 3.0 kDa molecular weight cutoff Amicon (Millipore) for further studies.

SM80.2 protein crystals diffracted to 2.0 Å resolution at home source and 1.87 Å resolution at BM14 European Synchrotron Radiation Facility, Grenoble, France. The cuboidal shaped crystals belonged to space group P6₁, with unit cell parameters $a = b = 88.8$ Å and $c = 49.0$ Å. Data were processed by *iMosflm* software.¹⁶ Data collection and processing statistics are summed up for high resolution data in Table 1. Matthews coefficient suggested the presence of two monomers per asymmetric unit.²⁴

TABLE 1 Data collection and S-SAD phasing statistics for SM80.2 in P6₁ space group

	Dataset1 SAD 0 Kappa	Dataset2 SAD 40 Kappa	Dataset3 SAD 80 Kappa	Merged SAD 0 + 40 + 80Kappa	Dataset4 Native
Space group	P6 ₁	P6 ₁	P6 ₁	P6 ₁	P6 ₁
Unit-cell parameters					
<i>a</i> (Å)	87.55	87.62	87.63	87.60	87.48
<i>b</i> (Å)	87.55	87.62	87.63	87.60	87.48
<i>c</i> (Å)	49.69	49.69	49.66	49.68	49.67
Wavelength (Å)	1.7712	1.7712	1.7712	1.7712	0.95372
Oscillation	1	1	1	1	1
No of images	360	360	360	1 080	300
Total reflections	165 493 (23 318)	166 378 (23 491)	166 465 (23 515)	498 339 (69 416)	169 020 (10 936)
Unique reflections	7 645 (1 093)	7 639 (1 090)	7 637 (1 092)	7 669 (1 095)	18 069 (1 171)
Resolution range (Å)	41–2.50	41–2.50	41–2.50	41–2.50	32–1.87
Completeness (%)	100 (100)	99.7 (99.2)	99.8 (99.5)	100 (100)	99.9 (100)
Redundancy	21.6 (21.3)	21.8 (21.6)	21.8 (21.5)	65.0 (63.4)	9.4 (9.3)
Mean <i>I</i> / <i>σ</i> (<i>I</i>)	34.3 (16.0)	37.9 (16.1)	32.8 (11.9)	53.5 (22.7)	20.9 (5.2)
<i>R</i> _{merge} ^a	0.068 (0.204)	0.060 (0.211)	0.069 (0.298)	0.078 (0.252)	0.061 (0.433)
<i>R</i> _{meas} ^b (all <i>I</i> + & <i>I</i> −)	0.072 (0.213)	0.064 (0.220)	0.074 (0.311)	0.080 (0.256)	0.067 (0.484)
<i>R</i> _{pim} ^c (all <i>I</i> + & <i>I</i> −)	0.015 (0.046)	0.014 (0.047)	0.016 (0.066)	0.010 (0.032)	0.022 (0.158)
<i>CC</i> _{1/2}	0.999 (0.994)	1.000 (0.994)	0.999 (0.986)	1.000 (0.997)	0.999 (0.939)
Mosacity	0.82	0.77	0.71	0.77	0.66

^a $R_{\text{merge}} = \sum_{\text{hkl}} \sum_i |I_i(\text{hkl}) - \langle I(\text{hkl}) \rangle| / \sum_{\text{hkl}} \sum_i I_i(\text{hkl})$ where $I_i(\text{hkl})$ is the i^{th} measurement of the intensity of reflection hkl and $\langle I(\text{hkl}) \rangle$ is the mean intensity of reflection hkl.

^b $R_{\text{meas}}(\text{redundancy-independent } R_{\text{merge}}) = \sum_{\text{hkl}} [N_{\text{hkl}} / (N_{\text{hkl}} - 1)]^{1/2} \sum_i |I_i(\text{hkl}) - \langle I(\text{hkl}) \rangle| / \sum_{\text{hkl}} \sum_i I_i(\text{hkl})$.

^c $R_{\text{pim}}(\text{precision-indicating } R_{\text{merge}}) = \sum_{\text{hkl}} [1 / (N_{\text{hkl}} - 1)]^{1/2} \sum_i |I_i(\text{hkl}) - \langle I(\text{hkl}) \rangle| / \sum_{\text{hkl}} \sum_i I_i(\text{hkl})$.

Data processed using *iMosflm*.

"Values in parentheses are for the highest resolution shell."

3.2 | Structure determination

Although a few crystal structures of nsLTPs were known, we attempted *ab initio* Sulfur-SAD phasing method²⁵ for structure determination (Table 1) to ensure unbiased electron density map. Preliminary model was prepared by PHENIX AutoSol suite followed by iterative model building in AutoBuild.¹⁷ AutoSol program identified 17 sites with correlation coefficient for consensus sequence of 53%. Final preliminary model consists of 150 residues and overall model map correlation of 63. This initial model had *R*_{work} and *R*_{free} of 38.32 and 41.81%, respectively with no side chains. The complete SM80.2 structure was refined to *R*_{work} of 18.36% and *R*_{free} of 22.54% with 98% of the residues in the favored region of the Ramachandran plot. Final refinement statistics are listed in Table 2.

3.3 | Overall structure

SM80.2 consists of 92 residues and exists as dimer in the crystal structure (Figure 2A). The monomer can be divided into three parts that is, four helices (H1-H4), three short loops and a long C-terminal having four disulfide bonds. This long C-terminal contains a 3₁₀ helix and is

linked to N-terminal by a long helix H3 through a pair of disulfide bonds. Another pair of disulfide bonds link helices H1 and H4. In all, disulfide bonds are present between cysteine 4–51, 14–28, 29–74,

TABLE 2 Refinement statistics of SM80.2

Resolution range (Å)	32.83–1.87
Final <i>R</i> _{work} (%)	18.36
Final <i>R</i> _{free} (%)	22.54
R.M.S. deviations	
Bonds (Å)	0.010
Angles (°)	1.380
Average <i>B</i> factors (Å ²)	
Protein	32.20
Ligand	45.70
Water	39.10
Ramachandran plot	
Most favoured (%)	98.88
Allowed (%)	1.12
Outlier (%)	0.00

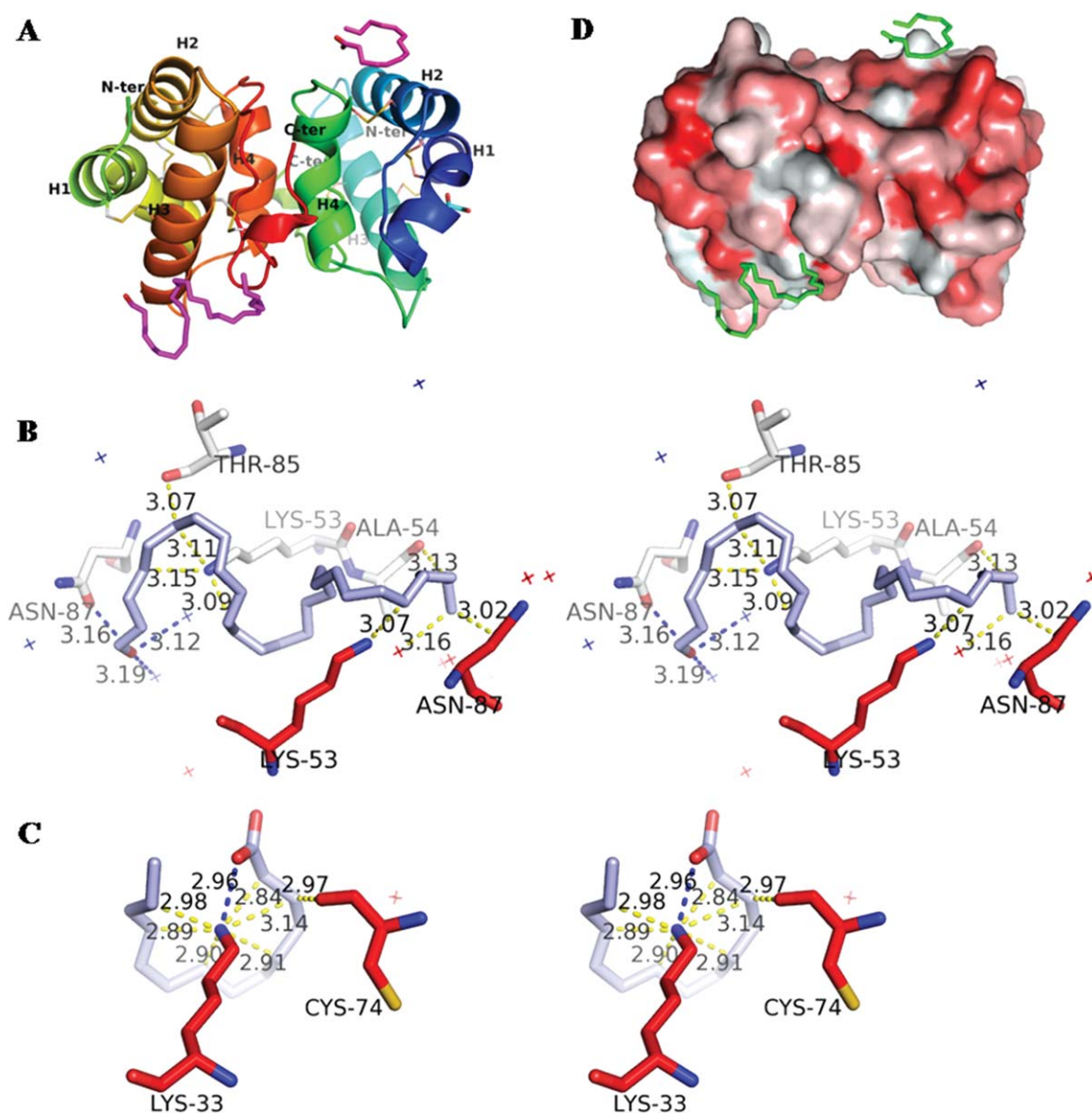


FIGURE 2 Overall structure of SM80.2. (A) Cartoon representation of SM80.2 dimer. The helices of each monomer have been differently colored and marked as 1, 2, 3, and 4, from N to C terminus. Ligands, octacosanol (magenta), myristic acid (magenta) and glycerol (cyan) are shown in sticks. Ligand protein interactions were analyzed between (B) Octacosanol, (C) Myristic acid. Residues interacting by hydrophobic or nonbonded interactions and hydrogen bond were represented as lines in yellow and blue, respectively. Red and white are the residues belonging to SM80.2 structure and symmetry mates, respectively. Octacosanol and myristic acid are shown in blue sticks. Water molecules are represented in asterisk. Figures are prepared in PyMOL. (D) Representation of hydrophobic patches on surface of SM80.2. The intensity of red color is based on hydrophobicity. Hydrophilic residues are lighter/white, whereas hydrophobic residues are darker/red. This coloring is done in PyMOL using the color_h.py python script. Bound lipids are shown in green sticks [Color figure can be viewed at wileyonlinelibrary.com]

49–88. In SM80.2, these four helices are also involved in formation of a hydrophobic tunnel (Figure 2A). Dimer is stabilized by 82 nonbonded interactions and 9 hydrogen bonds. A total of 61 water molecules were identified in SM80.2 structure.

Apart from the protein chains and water molecules, extra electron density for two long continuous stretches and a small entity was seen in the electron density map. Because molecules from the crystallization buffer and cryo-protectant are known to bind to macromolecules,²⁶ cryo-protectant and crystallization components were considered while interpretation of these extra electron densities. In this context,

possibility of the small entity being glycerol was considered since it was present in cryo-protectant. Further refinement by modeling glycerol in the small entity of electron density, with favorable chemical environment and no negative density indicated that it is indeed glycerol (Figure 2A).

It was evident that the other two continuous elongated electron densities could correspond to long hydrocarbon chains. They could either be PEGs as crystallization buffer contained PEG or could be lipids as the protein is a member of the nsLTP family. While we had carried out the delipidification process, the inherent lipid binding features

TABLE 3 Tabular representation of refinement statistics of the protein without ligand, with polyethylene glycol as compared to the protein with lipids

	Values without polyethylene glycol and lipids	Values with polyethylene glycol (2PE and 1PE)	Reported values with lipids (octacosanol and myristic acid)
R_{work}	19.51	18.72	18.36
R_{free}	23.49	22.73	22.54

could be relevant in case of a lipid transfer protein unlike any other soluble protein. Therefore, we had attempted modeling both polyethylene glycols and lipids in the extra electron density. To model these two electron density regions, the expected numbers of atoms were calculated based on size and shape of electron density. We tried to model different hydrocarbons including lipids and polyethylene glycols. For polyethylene glycol, both pentaethylene glycol or 1PE ($\text{C}_{10}\text{H}_{22}\text{O}_6$) and nonaethylene glycol or 2PE ($\text{C}_{18}\text{H}_{38}\text{O}_{10}$) were modeled. To identify biologically relevant lipids of corresponding lengths, Lipid Maps Structure Database (LMSD) was also searched. Initially after exploring different lipids, myristic acid (MYR) for 14-carbon lipid, and octacosanol (O8N) for 28-carbon lipid, were further considered as they showed better fit. The crystallographic refinement parameters (R_{work} and R_{free}) showed less improvement when 1PE and 2PE were modeled as compared to the O8N and MYR in the extra electron density (Table 3). Also, the real space correlation coefficient (RSCC) values for O8N and MYR (0.84 and 0.75, respectively) were significantly better compared to those for 2PE and 1PE (0.73 and 0.61, respectively), indicating that the lipid models fit better. Thus, based on the refinement statistics, octacosanol and myristic acid, were built during the final refinement (Figure 2A). Of course, the possibility of presence of a longer lipid with highly flexible hydrocarbon chain could also not be entirely ignored. However, identification of lipids using mass spectrometry gave ambiguous results.

Based on the local environment, the modeled lipids octacosanol and myristic acid could be stabilized by hydrogen bonds and hydrophobic contacts. Octacosanol, if present, could form hydrogen bonds with neighboring symmetry related water molecules and Asn87 whereas myristic acid will form hydrogen bond with Lys33. These lipids, otherwise will be stabilized by hydrophobic contacts, with Cys74 in case of myristic acid and symmetry related Thr85, Lys53, Ala54, main chain Lys53, Asn87, a water molecule with octacosanol (Figure 2B,C). The average B-factor for refined octacosanol and myristic acid are 42.75 and 50 respectively. These interactions are based on the analysis done using PDBSum²⁷ and COOT. Stability of these lipids was further evaluated by analyzing the hydrophobic region of the protein. It was observed that indeed these regions could allow the stability of lipids on the surface of the protein (Figure 2D). To check whether these ligands are actually present and are not due to refinement artifact, simulated annealing (SA) ligand omit Fo-Fc maps were prepared (Supporting Information Figure S1).

3.4 | Structural comparisons with related proteins

Structure-based comparative analysis was performed toward structure-function analysis for elucidating the possible physiological role of this

class of proteins along with its association with allergy. A protein data bank search led to identification of several proteins structurally similar to SM80.2 from different sources, which were subjected to comparative analysis. In protein data bank, we identified comparable X-ray structures of nsLTP's from maize, barley, peach, rice, hazelnut and wheat^{11,28-36} (Table 4, Supporting Information Figure S2). All the monomers of these structures superimposed well with rmsd <1 Å. The rmsd calculations were done using PyMOL.²² The major variations are mostly seen along two regions that are involved in formation of hydrophobic tunnel. These are H1-H2 loop and the C-terminal region including H3-H4 loop. These deviations are affecting the hydrophobic cavity size.

The structural comparison suggests that size of the cavity varies from apo- to liganded nsLTP structures. Apo-structure showed comparably smaller cavity size than the liganded structure. Out of four apo-nsLTP structures, SM80.2 showed the smallest cavity size followed by rice 1RZL, maize 1MZL and hazelnut 4XUW, respectively. The cavity size comparison was based on the position of the farthest residues that is, position 16 and position 79 of SM80.2 and comparable positions in other nsLTPs. The biggest cavity size is observed in the case of wheat 1BWO that holds two molecule of 22-carbon fatty acid, LPC ([1-myristoyl-glycerol-3yl] phosphoncholine), the largest fatty acid bound so far in the cavity. Table 4 further shows features of all known homologous structures compared with the SM80.2. Lipid mostly binds in this tunnel with head group toward surface and tail buried inside. Range of lipids binds in the cavity from C-7 to C-22 that show plasticity of the cavity as compared to apo-structures. Lipid binding is generally observed in two modes, inside or partially outside in reference to tunnel. It was mentioned that such binding is a result of collapsing of C-terminal into the hydrophobic cavity. Thus, if lipids are fortuitously bound on the surface of SM80.2, it effectively could be an intermediate state before the lipid gets into the cavity.

Structure-based sequence alignment of all the above structures was carried out. Clustal Omega multiple sequence alignment³⁷ identified conserved residues among all. It is seen that Val6, Thr40, Arg44, Leu51, Pro71, Pro78 and aromatic residue at position 79 (Aro79) are consensus residues (residue numbering is as per rice nsLTP sequence). However, in case of SM80.2, Thr40, Leu51, Pro78, and Aro79 are not conserved. The consequences of absence of consensus residue in SM80.2 were checked in crystal structure to understand differences in lipid binding in hydrophobic pocket. In the hydrophobic cavity, at non polar end presence of Tyr59 instead of corresponding residue Ile (in rice, hazelnut, wheat, barley) or Val (in peach, maize) is oriented in such a way that it closes the entry. However, at polar end which anchor carboxylic acid, orientation of Arg45 and Asn79 are closing the end.

TABLE 4 Comparison of all known crystal structures of nonspecific lipid transfer protein with SM80.2

S.no	Plant name	PDB id	Ligand	Lipid Size	RMSD (Å)
1	Barley	1MID	Lauroyl	C ₂₀ H ₄₃ N O ₇ P	0.738
		3GSH	(12E)10-oxooctadec12-enoic acid	C ₁₈ H ₃₂ O ₃	0.844
2	Maize	1FK0	Capric acid	C ₁₀ H ₂₀ O ₂	0.838
		1FK1	Lauric acid	C ₁₂ H ₂₄ O ₂	0.857
		1FK2	Myristic acid	C ₁₄ H ₂₈ O ₂	0.922
		1FK3	Palmitoleic acid	C ₁₆ H ₃₀ O ₂	0.840
		1FK4	Stearic acid	C ₁₈ H ₃₆ O ₂	0.813
		1FK5	Oleic acid	C ₁₈ H ₃₄ O ₂	0.847
		1FK6	α-linolenic acid	C ₁₈ H ₃₀ O ₂	0.869
		1FK7	Ricinoleic acid	C ₁₈ H ₃₄ O ₃	0.852
		1MZL	Unliganded state	-----	0.845
		1MZM	Palmitate	C ₁₆ H ₃₂ O ₂	0.858
3	Wheat	1TUK	Lysophosphatidylglycerol	C ₂₂ H ₄₄ O ₉ P	1.408
		1BWO	Phosphonyl choline	C ₂₂ H ₄₇ N O ₇ P	0.594
4	Peach	2ALG	Lauric acid	C ₁₂ H ₂₄ O ₂	0.642
		2B5S	Lauric acid	C ₁₂ H ₂₄ O ₂	0.727
5	Rice	1RZL	Unliganded state	-----	0.888
		1UVA	Myristic acid	C ₁₄ H ₂₈ O ₂	0.590
		1UVB	Palmitoleic acid	C ₁₆ H ₃₀ O ₂	0.716
		1UVC	Stearic acid	C ₁₈ H ₃₆ O ₂	0.701
6	Hazelnut	4XUW	Unliganded state	-----	1.057
7	Brinjal (SM80.2)	5TVI	Octacosanol; Myristic acid	C ₂₈ H ₅₈ O; C ₁₄ H ₂₈ O ₂	

Presence of Asn79 and Val80 instead of consensus residues Pro78 and Aro79 (as in rice nsLTP) are resulting in reducing the area of hydrophobic pocket.

In case of SM80.2, the hydrophobic pocket is formed by 21 residues that include Val2, Val7, Leu11, Phe17, Leu18, Pro24, Val32, Leu35, Asn36, Ala39, Ile52, Ala55, Leu62, Leu70, Pro71, Ile76, Leu78, Asn79, Val80, Ile82, and Ile86. To identify the reason for compact closing in SM80.2 all the comparable nsLTP structures were superimposed (Figure 3). This plasticity of the binding pocket is governed by Tyr79 in all nsLTPs. This key residue Tyr79 of the hydrophobic pocket, known to play a crucial role in opening and closing of hydrophobic tunnel, is absent in case of SM80.2. In this case, Val80 is present instead of aromatic residue Tyr79 (Figure 3). Val80 occupies the equivalent position in space as observed for Tyr79 in rice nsLTP apo-structure (1RZL). Surprisingly, Asn79, residue before Val80, is further involved in reducing the pocket space. This position is occupied by proline in all other nsLTP structures. Superimposition of all the apo-nsLTPs suggested that indeed the cavity in SM80.2 had smallest size, mainly because of Asn79. It is observed that Asn79 of SM80.2 is involved in hydrogen bonding with Asn36 of helix-2 which results in reduced size of the cavity in SM80.2 (Figure 3). In other nsLTPs, Try79 is mostly stabilized by

forming hydrogen bond with a water molecule or the polar heads of the lipid in the liganded state. We found that owing to the structural similarity of SM80.2 with Peach Pru P3, we could map the identified allergy specific epitopes that can result in cross-reactivity.³² Accordingly, three potential epitopic regions were identified in SM80.2. Interestingly, epitopes mapped overlap with one of the probable lipid binding regions (Figure 4).

4 | DISCUSSION

We have determined the structure of non-specific lipid transfer protein (nsLTP), SM80.2 from *S. melongena* ab initio. SM80.2 is the first structure from A-like class of nsLTPs. A few other structures from this family are deciphered by X-ray crystallography or NMR spectroscopy. Although the known structures show common structural fold, there exist significant differences in the size of hydrophobic pocket to accommodate different lipids. However, by comparing the structures of SM80.2 with other nsLTPs, evident structural difference was mainly due to sequence variation. These deviations were seen mainly in loops H1-H2 and H3-H4, which affect the hydrophobic cavity.

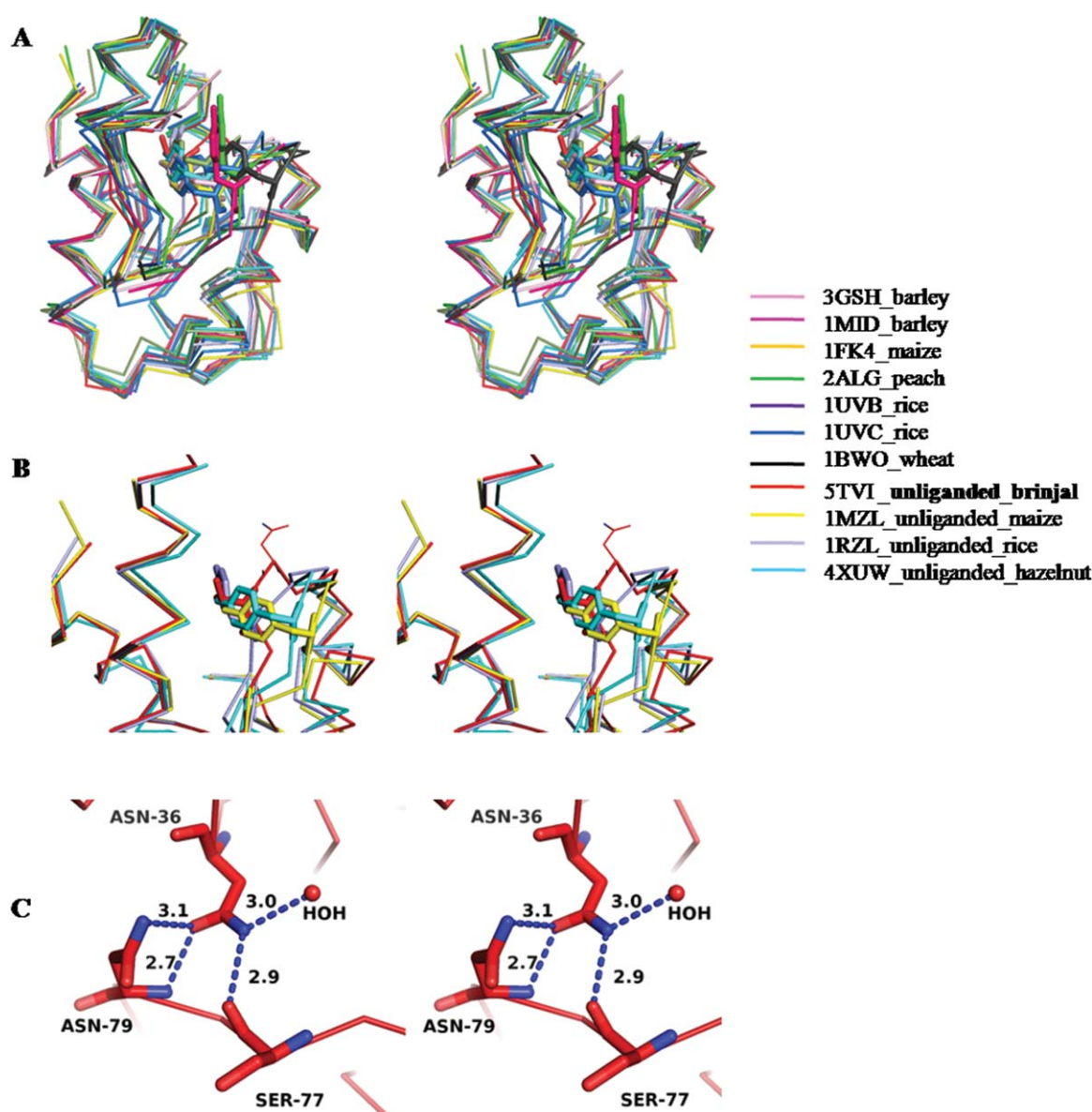


FIGURE 3 Superimposition of SM80.2 with all homologous PDBs. (A) Stereo view of C-alpha superimposition showing the variation of tyrosine in all liganded and unliganded structures with respect of valine (Val80) of SM80.2. (B) Stereo view of C-alpha superimposition showing zoomed in on the important position corresponding to Val80 and Asn79 in SM80.2 suggesting flexibility in loop for unliganded or apo-structures. Comparable tyrosine is also shown in stick. (C) Stereo view of C-alpha superimposition showing hydrogen bond between Asn79 and Asn36 of SM80.2

SM80.2 exhibits smallest cavity size identified so far in nsLTPs. Comparison of the cavity in SM80.2 with that of other known nsLTPs suggested that the size difference in cavity is governed mainly by few critical residues. In maize nsLTP, Arg46, and Tyr81 were identified as key residues for the arrangement of lipids in the hydrophobic cavity.^{29,30} In rice nsLTP unliganded structure, Tyr79 ensures total breakdown of lipid binding environment by occupying the central space in the cavity whereas this residue reorients in the liganded structure such that extra space is created for lipid binding.^{11,33} Similar degree of plasticity is observed in barley and peach nsLTPs despite existence of disulfide bonds.^{31,32} Molecular dynamics experiment also suggests that the tyrosine residue rotates inside or outside in liganded or unliganded

state resulting in plasticity of C-terminal region to facilitate ligand binding.³² Despite being a critical residue, as identified in all other nsLTPs, this tyrosine is not conserved in SM80.2. Absence of this key residue in SM80.2 structure suggests that the harmony of liganded and unliganded state could disrupt. However, for proper equilibrium between entry and exit of different lipids in this class of nsLTP, nature had ensured another mechanism mediated by hydrogen bonding.

The key residue in SM80.2, Asn79 is responsible for even tight closure as compared to tyrosine mediated closure of binding pocket due to formation of hydrogen bond. The presence of these hydrogen bonds between Asn79 and Asn36 are playing key role in reducing the cavity size and probably restricting the entry of lipids in the cavity. This

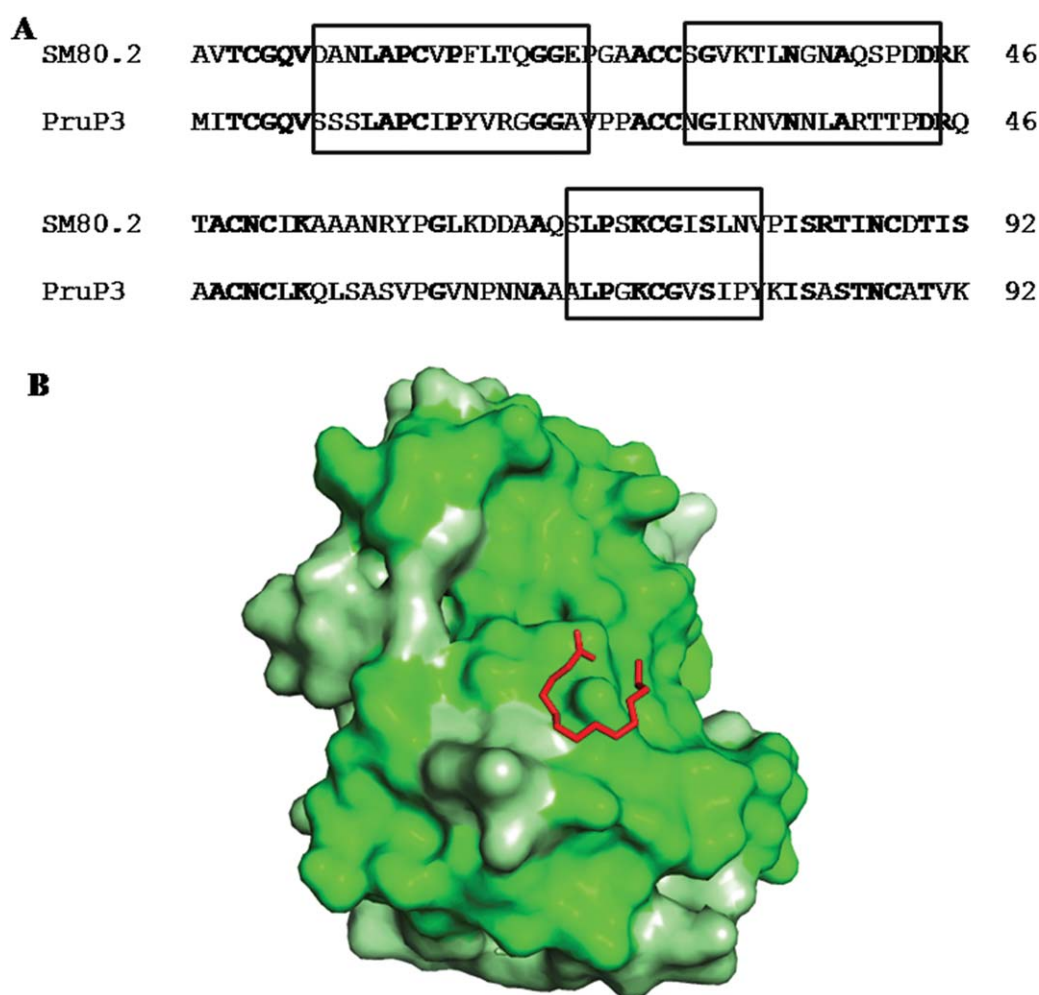


FIGURE 4 Identification and depiction of epitopic regions and lipid binding regions in SM80.2. (A) Sequence alignment of SM80.2 with peach PruP3 sequence. Residues in box highlight the epitopic region in SM80.2. Three boxes represent the epitopes mapped for peach PruP3 and the corresponding in SM80.2. (B) Surface representation of SM80.2 showing myristic acid position with respect to the surface epitope. Protein surface is shown in pale green, epitopes in green, and lipid in red stick [Color figure can be viewed at wileyonlinelibrary.com]

hydrogen bond mediated closure of binding pocket is observed for the first time in nsLTPs as the residue important for formation of this bond is absent in all the known nsLTP structure. This suggested that the absence of consensus residues are playing critical role in relatively tight closure of binding pocket in SM80.2 structure. However, closure of binding pocket via hydrogen bond could possibly be favored in other A-like nsLTP due to Asn/Asp residue at the comparable position (Supporting Information Figure S3). This structure thus is a novel structure with the smallest pocket size so far identified.

The presence of two elongated blobs of continuous electron density observed in the case of SM80.2 structure is curious. While we could not obtain unambiguous interpretation in the mass spectrometric data, it was attractive to model biologically relevant lipids, octacosanol and myristic acid in this electron density, the chemical environment being conducive for such a consideration. SM80.2, being a member of nsLTP, could have modeled lipids at non-canonical position due to closure of binding pocket in contrast to the canonical hydrophobic cavity in which they bind in other nsLTP structures. If this is actually the case, then it is very interesting to state that SM80.2 probably shows an

intermediate state, perhaps favored by residues which keep the lipid binding pocket relatively tight.

Several recent reports indicate that nsLTPs from different species can contribute in plant food allergies. Amino acid sequence identity and immunological cross-reactivity among different nsLTPs strongly suggest that the identical or nearly identical epitopic regions can participate in allergic responses. In case of SM80.2, we found that one of the regions that show long extra density (modeled as lipid) and the epitopes overlap together. This association might be acting as a shield to the epitope and thus could be a plausible cause of increase in stability of the epitopic fragment leading to increase in the gastrointestinal stability of proteins. This suggests that intermediate state, seen in SM80.2 might be causing allergy. This could be one of the ways that promotes allergenic sensitization (Figure 4).

To conclude, we have isolated, purified, crystallized and solved the structure of first A-like non-specific lipid transfer protein, SM80.2 from eggplant by ab initio S-SAD method. SM80.2 shows structural similarity in overall structure with known nsLTP structure but variation in sequence, role and hydrophobic pocket. Structural comparison gave

insights about hydrogen bond mediated closure of binding pocket critically defined by certain key residues.

ACKNOWLEDGMENTS

The authors thank the EMBL staff Dr. Hassan Belrhali and Dr. Babu A. Manjasetty for providing support on the beamline, EMBL-DBT for providing access to the BM14 beamline at the ESRF; Dr. Nirpendra Singh and Mr. Zaid K. Madni for help and discussion. They acknowledge financial support from the Departments of Science & Technology and Biotechnology, Govt of India.

AUTHOR CONTRIBUTIONS

A.J. and D.M.S. conceived and designed the experiments, A.J. performed the experiments, A.J. and D.M.S. analyzed the data, A.J. and D.M.S. wrote the manuscript. Manuscript is reviewed, discussed and approved by all the authors. The authors have declared no conflicts of interest.

REFERENCES

- [1] Cianferoni A, Spergel JM. Food allergy: review, classification and diagnosis. *Allergol Int*. 2009;58(4):457–466.
- [2] Bublin M, Eiwegger T, Breiteneder H. Do lipids influence the allergic sensitization process?. *J Allergy Clin Immunol*. 2014;134(3):521–529.
- [3] Pramod SN, Venkatesh YP. Allergy to eggplant (*Solanum melongena*) caused by a putative secondary metabolite. *J Invest Allergol Clin Immunol*. 2008;18(1):59–62.
- [4] Harish Babu BN, Mahesh PA, Venkatesh YP. A cross-sectional study on the prevalence of food allergy to eggplant (*Solanum melongena* L.) reveals female predominance. *Clin Exp Allergy*. 2008;38(11):1795–1802.
- [5] Hirano K, Hino S, Oshima K, et al. Evaluation of allergenic potential for rice seed protein components utilizing a rice proteome database and an allergen database in combination with IgE-binding of recombinant proteins. *Biosci Biotechnol Biochem*. 2016;80(3):564–573.
- [6] Mendes C, Costa J, Vicente AA, Oliveira MBP, Mafra I. Cashew nut allergy: clinical relevance and allergen characterisation. *Clin Rev Allergy Immunol*. 2016;1–22. doi:10.1007/s12016-016-8580-5
- [7] Sicherer SH, Sampson HA. Peanut allergy: emerging concepts and approaches for an apparent epidemic. *J Allergy Clin Immunol*. 2007;120(3):491–503.
- [8] Rayes H, Abid Raza A, Williams A, Matthews S, Arshad SH. Specific IgE to recombinant protein (Ber e 1) for the diagnosis of Brazil nut allergy. *Clin Exp Allergy*. 2016;46(4):654–656.
- [9] Staswick P. Storage proteins of vegetative plant tissues. *Annu Rev Plant Biol*. 1994;45(1):303–322.
- [10] Pramod SN, Venkatesh YP. Allergy to eggplant (*Solanum melongena*). *J Allergy Clin Immunol*. 2004;113(1):171–173.
- [11] Lee JY, Min K, Cha H, Shin DH, Hwang KY, Suh SW. Rice non-specific lipid transfer protein: the 1.6 Å crystal structure in the unliganded state reveals a small hydrophobic cavity. *J Mol Biol*. 1998;276(2):437–448.
- [12] Breiteneder H, Radauer C. A classification of plant food allergens. *J Allergy Clin Immunol*. 2004;113(5):821–830.
- [13] Wijesinha-Bettoni R, Alexeev Y, Johnson P, et al. The structural characteristics of nonspecific lipid transfer proteins explain their resistance to gastroduodenal proteolysis. *Biochemistry*. 2010;49(10):2130–2139.
- [14] Jain A, Kumar A, Salunke DM. Crystal structure of the vicilin from *Solanum melongena* reveals existence of different anionic ligands in structurally similar pockets. *Sci Rep*. 2016;6:23600.
- [15] Altschul SF, Madden TL, Schaffer AA, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res*. 1997;25(17):3389–3402.
- [16] Battye TG, Kontogiannis L, Johnson O, Powell HR, Leslie AG. iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr D Biol Crystallogr*. 2011;67(Part 4):271–281.
- [17] Adams PD, Afonine PV, Bunkóczi G, et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr*. 2010;66(2):213–221.
- [18] Emsley P, Cowtan K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr*. 2004;60(12):2126–2132.
- [19] Schüttelkopf AW, Van Aalten DM. PRODRG: a tool for high-throughput crystallography of protein–ligand complexes. *Acta Crystallogr D Biol Crystallogr*. 2004;60(8):1355–1363.
- [20] Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Crystallogr*. 1993;26(2):283–291.
- [21] Brünger AT, Adams PD, Clore GM, et al. Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr D Biol Crystallogr*. 1998;54(5):905–921.
- [22] DeLano W. The PyMOL molecular graphics system. Palo Alto, CA, USA: DeLano Scientific LLC; 2008; Available at: www.pymol.org.
- [23] Krause S, Reese G, Randow S, et al. Lipid transfer protein (Ara h 9) as a new peanut allergen relevant for a Mediterranean allergic population. *J Allergy Clin Immunol*. 2009;124(4):771–778. e775.
- [24] Kantardjieff KA, Rupp B. Matthews coefficient probabilities: improved estimates for unit cell contents of proteins, DNA, and protein–nucleic acid complex crystals. *Protein Sci*. 2003;12(9):1865–1871.
- [25] Dauter Z, Dauter M, Dodson E. Jolly sad. *Acta Crystallogr D Biol Crystallogr*. 2002;58(3):494–506.
- [26] Dym O, Song W, Felder C, et al. The impact of crystallization conditions on structure-based drug design: a case study on the methylene blue/acetylcholinesterase complex. *Protein Sci*. 2016;25(6):1096–1114.
- [27] Laskowski RA. PDBsum: summaries and analyses of PDB structures. *Nucleic Acids Res*. 2001;29(1):221–222.
- [28] Gomar J, Sodano P, Sy D, et al. Comparison of solution and crystal structures of maize nonspecific lipid transfer protein: a model for a potential in vivo lipid carrier protein. *Proteins*. 1998;31(2):160–171.
- [29] Han GW, Lee JY, Song HK, et al. Structural basis of non-specific lipid binding in maize lipid-transfer protein complexes revealed by high-resolution X-ray crystallography. *J Mol Biol*. 2001;308(2):263–278.
- [30] Shin DH, Lee JY, Hwang KY, Kim KK, Suh SW. High-resolution crystal structure of the non-specific lipid-transfer protein from maize seedlings. *Structure*. 1995;3(2):189–199.
- [31] Bakan B, Hamberg M, Larue V, Prange T, Marion D, Lascombe MB. The crystal structure of oxylipin-conjugated barley LTP1 highlights the unique plasticity of the hydrophobic cavity of these plant lipid-binding proteins. *Biochem Biophys Res Commun*. 2009;390(3):780–785.

- [32] Pasquato N, Berni R, Folli C, et al. Crystal structure of peach Pru p 3, the prototypic member of the family of plant non-specific lipid transfer protein pan-allergens. *J Mol Biol.* 2006;356(3):684–694.
- [33] Cheng HC, Cheng PT, Peng P, Lyu PC, Sun YJ. Lipid binding in rice nonspecific lipid transfer protein-1 complexes from *Oryza sativa*. *Protein Sci.* 2004;13(9):2304–2315.
- [34] Charvolin D, Douliez JP, Marion D, Cohen-Addad C, Pebay-Peyroula E. The crystal structure of a wheat nonspecific lipid transfer protein (ns-LTP1) complexed with two molecules of phospholipid at 2.1 Å resolution. *Eur J Biochem.* 1999;264(2):562–568.
- [35] Hoh F, Pons JL, Gautier MF, de Lamotte F, Dumas C. Structure of a liganded type 2 non-specific lipid-transfer protein from wheat and the molecular basis of lipid binding. *Acta Crystallogr D Biol Crystallogr.* 2005;61(Part 4):397–406.
- [36] Offermann LR, Bublin M, Perdue ML, et al. Structural and functional characterization of the hazelnut allergen Cor a 8. *J Agric Food Chem.* 2015;63(41):9150–9158.
- [37] Sievers F, Higgins DG. *Clustal Omega, Accurate Alignment of Very Large Numbers of Sequences. Multiple Sequence Alignment Methods.*: Springer; Methods Mol Biol. 2014;1079:105–116.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.

How to cite this article: Jain A, Salunke DM. Crystal structure of nonspecific lipid transfer protein from *Solanum melongena*. *Proteins.* 2017;00:000–000. <https://doi.org/10.1002/prot.25335>

Crystal structure of non-specific lipid transfer protein from *Solanum melongena*

SUPPLEMENTAL FIGURES

Figure S1: Simulated annealing ligand omit FoFc maps. A) myristic acid. B) octacosanol.
The SA ligand omit FoFc maps are prepared at 3.5sigma in COOT.

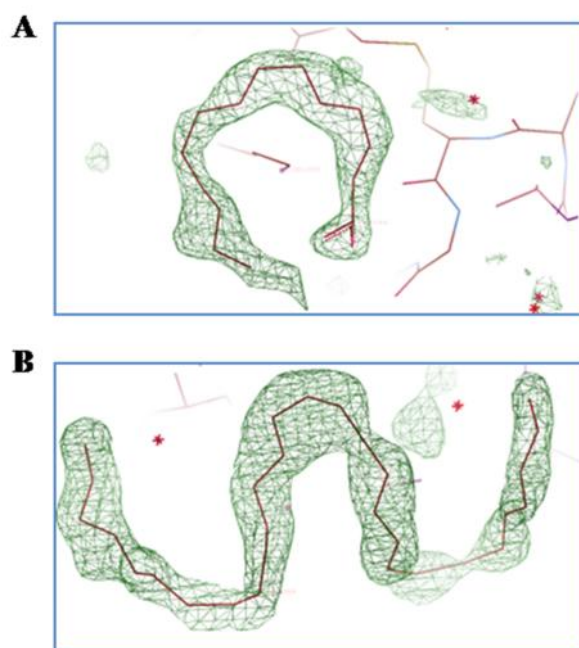


Figure S2: Comparison of available non-specific lipid transfer proteins. Stereo view of C-alpha superimposition of homologous nsLTP structures (green) from maize, peach, barley, rice, hazelnut and wheat with, SM80.2 monomer chains (magenta) suggesting a well fit of monomer with RMSD less than 1 Å. SM80.2 ligands are shown in magenta sticks whereas in cyan is the ligand binding in all the comparable structures at the prescribed hydrophobic pocket.

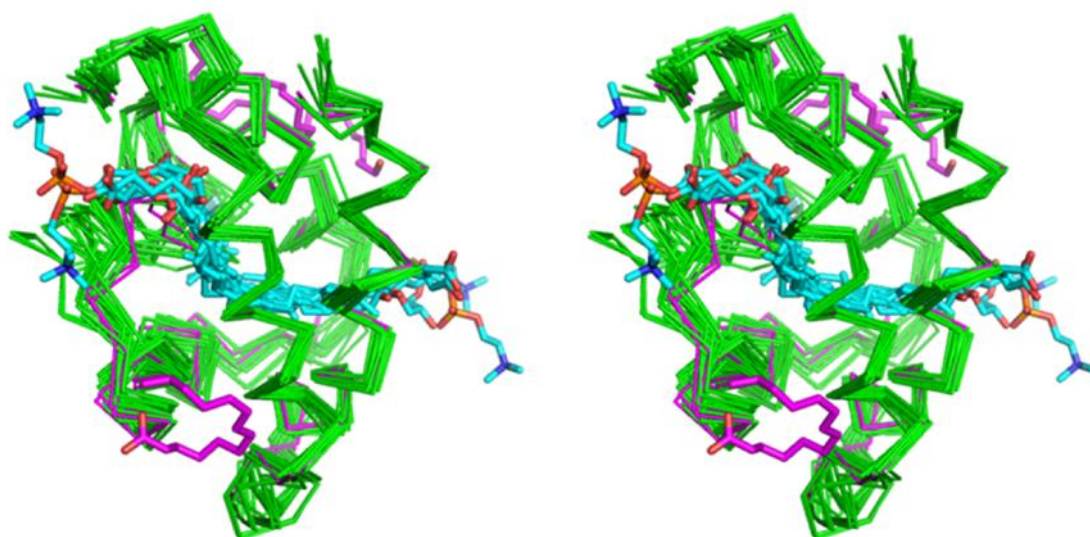


Figure S3: Multiple sequence alignment of SM80.1 from solanum melongena with homologous A-like nsLTPs from different species. In yellow are the conserved residues and in red are the residues corresponding to Asn79 that can form similar H-bond resulting in closure of binding pocket.

S.melongena	AVTCGQVDANLAPCVPF LTQG-GEPGAACCSGVKTLNGNAQSPDDRKTACNCKIAAANRY
C.annuum	AVTCGQVDANLAPCVPF LTQG-GEPGAACCSGVKTLNSLAKSTDERRTTCNCAKAAANRY
S.pennellii	AITCGQVDANLAPCVPF LTQG-GEPGAACCSGVRTLNGNTQSSDDRRTACNCVKAAANRY
S.tuberosum	AITCGQVDANLAPCVSFLTQG-GEPGAACCSGVKTLNGNAQSSDDRKTACNCVKAANRY
N.tomentosiformis	ALTGQVDASLAPCVPY LTQG-GEPGTACCNGVKDLKVMASQSIADRRRTACNCVKAAANRY
N.tabacum	AITCGQVDAALAPCVPY LTQG-GEPGAACCSGVKNLKGMAATTDRRRTACNCVKAAANRY
N.sylvestris	AITCGQVDAALAPCVPY LTQG-GEPGAACCSGVKNLKGMAATTDRRRTACNCVKAAANRY
S.lycopersicum	SINCGEVGGALAPCVPY LTQG-GDPSVSCCDGVKKVETPTQQRQVACECMKSAANRY
G.hirsutum	AVTCQVDDALAACVPY LTNG-GVPSTPCCNGVGELRNAEATTADRQAACNCVKQAAARI
J.curcus	AVDCGQVNAALAPCIPFLTGADTAPSAACCGGVKNLKVLAPTTADRRAACDCVKAASSHY
S.indicum	AVTCGQVDAALVPCISY LTGHVDSPPSAACCGGVKAVKGMAETTADKRACCSCVKAAANRY

S.melongena	PNLKDDAAQSLPSKCGISLNVPIISRTINCDTIS-
C.annuum	PNLKDDAAQSLPSKCGVTLNVPIISRTINCDTIS-
S.pennellii	PNLKDDAAQSLPSKCGISLTVPIISRTVNCDTIS-
S.tuberosum	PNLKDDAAQSLPSKCGISVNVPIISRTVNCDTIS-
N.tomentosiformis	ANLKDDAAQALPSKCGVTMDIPVSRTINCDM---
N.tabacum	TNLKDDAAQALASKCGVAMDVPVSKTINCDTISY
N.sylvestris	TNLKDDAAQALASKCGVAMDVPVSKTINCDTISY
S.lycopersicum	PNVKPDAAASNLPSRCGLTPIPIISPTINCKSVP-
G.hirsutum	PTIKEDNAATLPAKCHVQVDFPIISKNTNCEDIH-
J.curcus	PNINEDSASSLPKKCGVEVNIPIISKNINCQNIN-
S.indicum	ADLKDDAAQSLPTKCGVQMDVPISRSVDCDKIN-