

In silico analyses of structural and allergenicity features of sapodilla (*Manilkara zapota*) acidic thaumatin-like protein in comparison with allergenic plant TLPs[☆]



Hassan G. Ashok Kumar, Yeldur P. Venkatesh^{*}

Department of Biochemistry & Nutrition, CSIR – Central Food Technological Research Institute, Mysore 570020, Karnataka, India

ARTICLE INFO

Article history:

Received 31 May 2013

Received in revised form 14 August 2013

Accepted 18 August 2013

Keywords:

B-cell epitope prediction

IgE epitope prediction

Order Ericales

Sapodilla acidic TLP

Thaumatococcus-like protein

TLP structural features

ABSTRACT

Thaumatococcus-like proteins (TLPs) belong to the pathogenesis-related family (PR-5) of plant defense proteins. TLPs from only 32 plant genera have been identified as pollen or food allergens. IgE epitopes on allergens play a central role in food allergy by initiating cross-linking of specific IgE on basophils/mast cells. A comparative analysis of pollen- and food-allergenic TLPs is lacking. The main objective of this investigation was to study the structural and allergenicity features of sapodilla (*Manilkara zapota*) acidic TLP (TLP 1) by *in silico* methods. The allergenicity prediction of composite sequence of sapodilla TLP 1 (NCBI B3EWX8.1, G5DC91.1) was performed using FARRP, Allermatch and Evaller web tools. A homology model of the protein was generated using banana TLP template (1Z3Q) by HHPRED-MODELLER. B-cell linear epitope prediction was performed using BCPreds and BepiPred. Sapodilla TLP 1 matched significantly with allergenic TLPs from olive, kiwi, bell pepper and banana. IgE epitope prediction as performed using AlgPred indicated the presence of 2 epitopes (epitope 1: residues 36–48; epitope 2: residues 51–63), and a comprehensive analysis of all allergenic TLPs displayed up to 3 additional epitopes on other TLPs. It can be inferred from these analyses that plant allergenic TLPs generally carry 2–3 IgE epitopes. ClustalX alignments of allergenic TLPs indicate that IgE epitopes 1 and 2 are common in food allergenic TLPs, and IgE epitopes 2 and 3 are common in pollen allergenic TLPs; IgE epitope 2 overlaps with a portion of the thaumatococcus family signature. The secondary structural elements of TLPs vary markedly in regions 1 and 2 which harbor all the predicted IgE epitopes in all food and pollen TLPs in either of the region. Further, based on the number of IgE epitopes, food TLPs are grouped into rosid and non-rosid clades. The number and distribution of the predicted IgE epitopes among the allergenic TLPs may explain the specificity of food or pollen allergy as well as the varied degree of cross-reactivity among plant foods and/or pollens.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Plant allergens from both food and pollen sources are currently being studied extensively to understand the nature of clinical and immunological cross-reactivity for recommending the avoidance of certain foods for pollen- and/or food-sensitized subjects (Prescott, 2013). Sensitization process to an allergen in a susceptible individual is yet to be established and the clinical expression of sensitization is extremely variable (Asero and Pravettoni, 2013). Variation in the expression levels of allergens in plant tissues and

presence of isoforms makes it difficult to extrapolate the results of one allergenic case to a generalized group of food allergy. However, attempts are being made all over the world to predict IgE epitopes, which are one of the critical parameters for the development of allergy, by using *in silico* methods (Aluffi-Pentini et al., 2009; Dall'Antonia et al., 2011; Maleki et al., 2011; Chen et al., 2012; Sharma et al., 2013). B-cell epitopes are principally involved in the interaction of food allergens with specific IgE on basophils and mast cells. Among the B-cell epitopes, some epitopes are recognized more often than others, probably indicating the role of sensitization process (Aalberse and Cramer, 2011).

Sapodilla or sapodilla plum (*Manilkara zapota*; known by the names sapota or chiku in India) is a tropical delicious fruit, rich in sugars, and mainly consumed for table purposes (Paull, 2008; Small, 2011). Sapodilla is native to Mexico and Central America, but is now widely cultivated in almost all the tropical countries; it belongs to the family Sapotaceae of the order Ericales. India accounts for 10% of the world fruit production and ranks first in the production of banana and sapodilla (APEDA agri Xchange, 2013). The fruit is

[☆] A portion of this work was presented at the 2012 WAO International Scientific Conference held on 6–9 December 2012 at Hyderabad, India (Abstract No. 3033).

^{*} Corresponding author at: Department of Biochemistry & Nutrition, CSIR – Central Food Technological Research Institute (CFTRI), KRS Road, Mysore 570020, Karnataka, India. Tel.: +91 821 2514876; fax: +91 821 2517233.

E-mail addresses: ashokame84@gmail.com (H.G. Ashok Kumar), venkatyp@yahoo.com, ypv@cftri.res.in (Y.P. Venkatesh).

becoming popular as a specialty fruit in restaurants across Europe and North America. The protein content of sapodilla is very low (0.4–0.7 g per 100 g pulp).

Allergic reactions to sapodilla are rare. Three cases of oral allergy syndrome (OAS) to the ingestion of sapodilla were first described from India, and the causative allergen was identified as a 21 kDa protein in sapodilla pulp (Hegde and Venkatesh, 2002). Subsequently, the major allergen was isolated and identified as a basic thaumatin-like protein (TLP) using sera from 7 sapodilla-allergic subjects (Hegde et al., 2013). Basic TLP (TLP 2) is rich in arginine, glycine and threonine with a pI of ≥ 9.5 , and its N-terminal sequence is ATFDIQNNXY- (UniProt Accn. B3EWE5); however, its gene has not been cloned yet. Recently, a case of OAS to sapodilla and custard apple (*Annona squamosa*) was investigated wherein both the acidic and basic TLPs were identified as the allergens; the isolation and characterization of sapodilla TLP 1 (22 kDa), and its genomic clone have been described (Ashok Kumar et al., 2013). Sapodilla TLP 1 is a single polypeptide of 207 residues stabilized by 8 disulfides and has a pI of ~ 4.3 .

TLPs are plant pathogenesis-related proteins belonging to the PR-5 family produced in response to biotic and environmental stress (Liu et al., 2010). TLPs show high sequence similarity to thaumatin, an intensely sweet-tasting protein obtained from the berries of a West African shrub, *Thaumatococcus daniellii*. During safety evaluation of thaumatin (Talin protein), it was found to be a weak sensitizer in rats upon systemic administration; allergy testing of laboratory workers who had been intermittently exposed to commercial thaumatin for periods up to 7 years showed that 9.3% were positive (Higginbotham et al., 1983). Since thaumatin is available as a natural sweetener for consumption in only few countries, allergic reactions to thaumatin are expected to be uncommon. Despite lack of reports on allergy to thaumatin in humans, several pollen and food TLPs have been identified as inhalant and ingestant allergens, respectively, in susceptible atopic individuals (Breiteneder, 2004).

Only a limited number of TLPs have been identified from plant pollens and foods. Among these, TLPs from kiwi fruit, apple, peach, pear and cherry have been well investigated in molecular allergology. Only recombinant forms of TLPs from apple and cherry have tested positive for IgE-reactivity with allergic sera (Krebitz et al., 2003; Fuchs et al., 2006). In the case of allergenic pollen TLPs, TLP from Japanese cedar pollen (Jun a 3) has been well studied with respect to its structure and IgE-epitopes (Soman et al., 2000). So far, only 6 plant TLPs have been crystallized and their 3D structures elucidated (Narasimhan et al., 2009). However, comparative analyses of the structural features of allergenic TLPs in the context of prediction of IgE epitopes have not been reported so far. The main objective of this study was to investigate whether the number and distribution of IgE epitopes among allergenic food TLPs have a possible bearing on the specificity and cross-reactivity of food and/or pollen allergy. Therefore, we investigated the structural characteristics of sapodilla TLP 1 in comparison with other allergenic TLPs.

2. Materials and methods

2.1. Sapodilla TLP 1

The sapodilla acidic TLP mature sequence as retrieved from NCBI (G5DC91.1; partial) was devoid of the N-terminal 7 residues. The complete mature sequence was obtained by overlapping the experimentally determined N-terminal sequence of 25 residues (UniProt No. B3EWX8; TLP 1b). The composite sequence of sapodilla acidic TLP (Ashok Kumar et al., 2013), shown in Fig. 1, was used for all *in silico* analyses described hereafter, and is referred to as sapodilla TLP 1.

2.2. Allergenicity assessment of sapodilla TLP 1 as per FAO/WHO guidelines

Sapodilla TLP 1 has been shown to be an allergen based on positive *in vitro* and *in vivo* allergy diagnostic tests using a sapodilla-allergic subject (Ashok Kumar et al., 2013). The allergenicity assessment *in silico* (Vaughan et al., 2010) of sapodilla TLP 1 sequence was performed using online tools at the Food Allergy Research and Resource Program (FARRP) allergen protein database (www.allergenonline.org). Here, allergenicity analysis of a protein sequence is performed by 3 different methods, viz., the full-length alignment method, the 80 amino acid alignment method and the 8 amino acid exact match method, all of which were defined by the FAO/WHO guidelines in order to assess the allergenicity of newly introduced proteins in genetically modified (GM) foods.

Allergenicity of sapodilla TLP 1 sequence was also predicted using Allermatch web tool (<http://allermatch.org/>). Assessment is performed using 3 methods, viz., full-length alignment, 80-aa match and exact 6-aa match. Evaller (<http://www.slv.se/en-gb/Group1/Food-Safety/e-Testing-of-protein-allergenicity/>), another allergenicity/cross-reactivity prediction web tool based on peptide selection-query method was also performed. Here, peptides are selectively compared against allergenic and non-allergenic peptides in the database.

2.3. Linear B-cell epitope prediction of sapodilla TLP 1

BepiPred Linear Epitope Prediction tool of the IEDB Database (<http://tools.immuneepitope.org/tools/bcell/iedb.input>) was used to predict the epitope residue scores for sapodilla TLP 1. BepiPred predicts the location of linear B-cell epitopes using a combination of a hidden Markov model and a propensity scale method. B-cell epitopes were also predicted using BCPreds epitope prediction server (<http://ailab.cs.iastate.edu/bcpreds/>). A fixed length epitope prediction of 20-residues and 75% specificity was used for scoring potential B-cell epitopes.

2.4. IgE epitope prediction for sapodilla TLP 1 and allergenic TLPs

IgE epitope prediction for sapodilla TLP 1 was carried out using AlgPred server (Saha and Raghava, 2006; <http://www.imtech.res.in/raghava/algpred/submission.html>). The currently known allergenic food and pollen TLPs having full-length sequences as retrieved from Allergome database (www.allergome.org) is compiled in Table 1; these TLPs were also subjected to IgE epitope prediction. Allergenic TLPs with partial sequences are not listed in Table 1 and these include English mugwort (*Artemisia vulgaris*), birch (*Betula verrucosa*), hazelnut (*Corylus avellana*), Eastern red cedar (*Juniperus virginiana*), garden lettuce (*Lactuca sativa*), American sycamore (*Platanus acerifolia*) and obeche (*Triplochiton scleroxylon*). Theoretical pI and molecular mass were determined using ExPASy's compute pI/MW tool (http://web.expasy.org/compute_pi/). N-Glycosylation sequons were searched using NetNGlyc server (<http://www.cbs.dtu.dk/services/NetNGlyc/>). In order to investigate the distribution pattern of the predicted IgE epitopes among allergenic plant food and pollen TLPs (Table 1), a relationship tree was constructed using their mature sequences retrieved from NCBI, and Phylogeny.fr was employed for this purpose.

2.5. Homology modeling and annotation of sapodilla TLP 1

Homology modeling, also known as comparative modeling, is a method for predicting the three-dimensional structure of a given protein up to its atomic resolution based on its primary sequence; it involves identification of protein structures

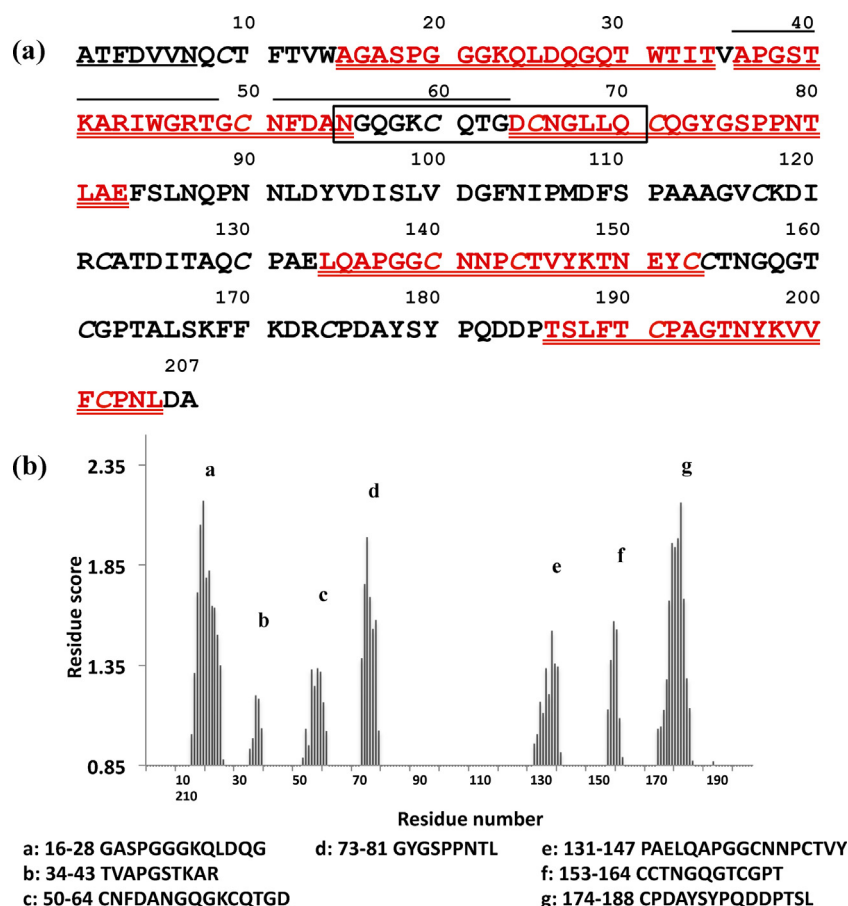


Fig. 1. Panel (a) The composite amino acid sequence of sapodilla TLP 1. The first 7 residues were obtained by N-terminal sequencing (underlined). Cysteines are shown in italics and the thaumatin family signature is boxed. The sequences representing the predicted IgE epitopes 1 and 2 obtained from this study are shown by solid lines above the sequences. The B-cell epitopes predicted by BCPreds are double underlined. Modified from Ashok Kumar et al. (©2013, with permission from the Japanese Society of Allergology). Panel (b) B-cell linear epitope prediction scores for sapodilla TLP 1 as analyzed by BepiPred. Only residue scores above 0.85 (threshold score for this analysis) are displayed. Potential B-cell epitopes are sequentially named 'a' to 'g'. The underlined sequence in epitope c represents a portion of the thaumatin family signature shown in panel (a).

or templates which are likely to resemble the structure of the query sequence, and a structure of the query is built upon it. The servers HHpred (<http://toolkit.tuebingen.mpg.de/hhpred>) and Modeller (<http://toolkit.tuebingen.mpg.de/modeller>) of the Max-Planck Institute for Developmental Biology were used for this purpose. Annotation of the predicted structure was performed using PDBsum (<http://www.ebi.ac.uk/pdbsum/>) which provides the secondary structure and cleft information, and evaluates the stereochemical quality of the modeled protein. The predicted IgE epitopes were mapped on to the modeled 3D structure of sapodilla TLP 1 using Jmol 3D structure visualization software.

2.6. Mapping of the predicted B-cell and IgE epitopes, and structural comparison of TLPs

Based on the distribution pattern obtained from the IgE epitope prediction, representative TLPs from each group were selected and their crystal structures were retrieved from PDB. The predicted B-cell and IgE epitopes were mapped onto the secondary structures of the selected TLPs. Common and variable secondary structural features were identified manually and visualized on their 3D structures using RasMol. PDBsum was used to generate the secondary structures and the homology model of sapodilla TLP 1 was also used alongside for comparison.

3. Results

3.1. Allergenicity assessment of sapodilla TLP 1 according to FAO/WHO guidelines

In the full-length alignment method, an identity match with >50% for the query is considered as a potential cross-reactive allergen with the match. An identity match with >35% for the query is considered as a potential cross-reactive allergen in the 80 amino acid alignment method. However, in the 8 amino acid exact match (8-mer) method, any exact match for the query is considered to identify the protein as a potential cross-reactive allergen. Table 2 displays the results of full-length alignment method showing allergen matches (with identities greater than 50%) to allergenic food and pollen TLPs. Allergenic food TLPs from olive and kiwi fruits showed significant identity (69–78%), while pollen TLP allergens from cedars and cypresses showed identity percentage close to 50. Although there are other allergenic plant TLPs in the database, interestingly, their identity with sapodilla TLP 1 stayed below 50%. Using the 80 amino acid alignment search, total of 128 sequences of 80 amino acid stretch were produced from sapodilla TLP 1, of which twenty eight 80 amino acid stretch matched the corresponding stretches of the deposited allergens (Table 2). By 8-mer method, two hundred 8-mers were found, and 11 allergens contained at least one 8-mer exact match (Table 2). In consistent with the full

Table 1List of putative allergenic TLPs from plant foods and pollens.^a

Food TLPs	pl	Food TLPs	pl	Pollen TLPs	pl
<i>Actinidia chinensis</i> (gold kiwi, 3)		<i>Olea europaea</i> (olive, 1)		<i>Cryptomeria japonica</i> (Japanese cedar, 8)	
AGC39176.1	7.91	E3SU11.1	4.76	BAC15615.1	8.46
AGC39181.1	4.46	<i>Prunus avium</i> (sweet cherry, 1)		BAC15614.1	8.46
AGC39177.1	7.55	P50694.1	4.57	BAC15616.1	4.63
<i>Actinidia deliciosa</i> (green kiwi, 5)		<i>Prunus domestica</i> (plum, prune, 1)		BAD90815.1	4.80
AGC39180.1	4.61	ADM22305.1	9.15	BAD99299.1	4.62
AGC39175.1	7.91	<i>Prunus dulcis</i> (almond, 1)		BAD90813.1	5.47
ABQ42566.1	7.92	ACE80974.1	4.62	BAF51970.1	4.43
AGC39179.1	5.35	<i>Prunus persica</i> (peach, 5)		BAD90814.1	4.54
AGC39178.1	5.34	AEV57470.1	4.27	<i>Cupressus arizonica</i> (Arizona cypress, 1)	
<i>Actinidia eriantha</i> (kiwi, velvet vine, 1)		P83335.1	4.58	CAC05258.1	4.82
AGC39182.1	4.61	AEV57472.1	4.73	<i>Cupressus sempervirens</i> (Mediterranean cypress, 2)	
<i>Brassica oleracea</i> (cabbage, 1)		AEV57471.1	4.61	AAR21074.1	5.32
AAO12209.1	7.97	P83332.1	8.14	AAR21075.1	5.32
<i>Capsicum annuum</i> (bell pepper, 3)		<i>Pyrus pyrifolia</i> (apple pear, 8)		<i>Juniperus ashei</i> (mountain cedar, 1)	
AAK97184.1	7.53	ACN97419.1	4.51	P81295.1	4.79
ACB30367.1	8.48	ACN97417.1	4.51	<i>Juniperus rigida</i> (Japanese juniper, 2)	
AAP86781.1	5.77	AEO36939.1	4.51	AAR21071.1	4.82
<i>Castanea sativa</i> (chestnut, 1)		ACI04550.1	4.57	AAR21072.1	4.82
CAB62167.1	8.46	O80327.1	5.01	<i>Thuja occidentalis</i> (Northern-white cedar, 1)	
<i>Cucumis melo</i> (musk melon, 2)		BAC78212.1	5.29	AAV65287.1	4.76
ADN33946.1	4.86	ACN97418.1	4.63		
ADN33945.1	4.19	ACN97416.1	4.51		
<i>Manilkara zapota</i> (sapodilla, 1)		<i>Triticum aestivum</i> ^b (wheat, 1)			
AEP84104.1	4.52	AAM15877.1	7.91		
<i>Malus domestica</i> (apple, 4)		<i>Vitis vinifera</i> (grape, 5)			
AAX19851.1	4.43	Q9M4G6	4.76		
ABC47923.1	4.43	Q9LLB7	8.46		
AAX19848.1	4.43	O04708	4.76		
JC7201	4.72	F6HUH1	4.76		
<i>Musa acuminata</i> (banana, 2)		A9CSP0	4.76		
1Z3Q	6.33	<i>Thaumatococcus daniellii</i> ^b (miracle fruit, 4)			
AFK29763.1	4.76	P02883.1	8.46		
		BAF44567.1	6.50		
		P02884.1	7.51		
		1RQW_A	8.46		

^a Allergen list was obtained from Allergome database (www.allergome.org) and the sequence data was retrieved from NCBI database; the common name and the number of gene sequences deposited are shown in parentheses following the botanical name. Theoretical pIs were calculated for the mature sequences using ExPASy pI/MW tool. Underlined sequences have putative N-glycosylation sites as analyzed by NetNGlyc server. Only allergenic TLPs having full-length sequences have been listed except for sapodilla acidic TLP (residues 8–207).

^b Inhalant food allergens.

Table 2

Allergen matches for sapodilla acidic TLP in FARRP database.

Allergen (botanical name, Accn. No.)	Percent identity		No.
	FL ^a	80-mer ^a	
<i>Olea europaea</i> , E3SU11.1	77.80	82.50	26
<i>Actinidia deliciosa</i> , AGC39175.1	69.50	76.23	30
<i>Actinidia deliciosa</i> , ABQ42566.1	69.00	74.95	30
<i>Juniperus virginiana</i> , Q9LD79.2	61.40	62.50	01
<i>Juniperus ashei</i> , P81295.1	55.00	63.70	03
<i>Cryptomeria japonica</i> , BAF51970.1	53.50	62.50	02
<i>Juniperus rigida</i> , AAR21071.1, AAR21072.1	54.00	62.50	03
<i>Cupressus sempervirens</i> , AAR21075.1, AAR21074.1	53.00	61.30	03
<i>Cupressus arizonica</i> , CAC05258.1	52.50	60.00	02
<i>Triticum aestivum</i> , P27357.1	<50.00	53.70	00
<i>Prunus dulcis</i> , ACE80974.1	<50.00	50.00	00
<i>Prunus persica</i> , P83335.1	<50.00	48.80	00
<i>Prunus dulcis</i> × <i>Prunus persica</i> , ACE80955.1	<50.00	48.80	00
<i>Prunus avium</i> , P50694.1	<50.00	48.78	00
<i>Prunus persica</i> , AEV57471.1	<50.00	48.20	00
<i>Malus domestica</i> , AAX19851.1, CAT99611.1, AAX19848.1, AFM77001.1, Q9FSG7.1, CAT99612.1	<50.00	47.50	00
<i>Prunus dulcis</i> × <i>Prunus persica</i> , ACE80957.1	<50.00	46.50	00
<i>Prunus dulcis</i> × <i>Prunus persica</i> , ACE80956.1	<50.00	45.80	00
<i>Prunus dulcis</i> × <i>Prunus persica</i> , ACE80959.1	<50.00	45.30	00
<i>Prunus persica</i> , P83332.1	<50.00	45.10	00
<i>Prunus dulcis</i> × <i>Prunus persica</i> , ACE80958.1	<50.00	45.03	00

^a Full-length (FL) alignment method lists allergen matches with >50% identity, 80-mer match method lists allergen matches with >35% identity, and 8-mer match method lists allergens with at least 1 exact match.

Table 3

Allergenicity assessment of sapodilla acidic TLP by Allermatch (full length, 80-aa sliding window and 6-aa Wordmatch).

Allergen ID ^a , Accn. No.	Full length ^b	80-mer sliding window ^c			6-mer Wordmatch ^d	
		(i)	(ii)	(iii)	(i)	(ii)
al.Act.d.2, P81370	69.46/203	76.25	128	100.00	44	21.78
wi.Cap.a.1w, Q9ARG0	66.35/208	73.75	128	100.00	26	12.87
wi.Mus.a.4, Q22322	65.88/170	72.50	122	95.31	12	5.94
al.Jun.a.3, P81295	54.95/202	63.75	128	100.00	9	4.46
wi.Jun.v.3.b, Q9LD79	61.45/083	62.50	36	28.12	5	2.48
wi.Jun.v.3.a, Q9LD79	61.45/083	62.50	36	28.12	5	2.48
wi.Cup.s.3.c, Q69CS2	52.97/202	61.25	128	100.00	8	3.96
wi.Cup.s.3.b, Q69CS3	52.97/202	61.25	128	100.00	8	3.96
wi.Cup.s.3.a, Q69CS2	52.97/202	61.25	128	100.00	8	3.96
wi.Pru.p.2.c, B6CQT3	41.63/221	48.78	122	95.31	4	1.98
al.Pru.av.2, P50694	40.99/222	48.75	119	92.97	4	1.98
al.MaI.d.2, Q9FSG7	40.54/222	47.50	123	96.09	2	0.99
wi.Pru.p.2.b, B6CQT5	40.36/223	46.51	122	95.31	3	1.49
wi.Pru.p.2.a, B6CQT7	39.91/223	45.35	124	96.88	2	0.99

^a Allergen ID. The prefix before the allergen nomenclature indicates the source database (al: both UniProt and WHO-IUIS; wi: WHO-IUIS isoallergen list).^b Results of FASTA alignment of the complete input sequence against UniProt and WHO-IUIS database. The first number refers to percentage identity, and the second number refers to the length of sequence over which the FASTA aligned.^c (i) Percent identical amino acids in the aligned 80-aa sliding window, (ii) the number of hits the input sequence had with this allergen, and (iii) the percentage of windows analyzed for this input sequence hitting this allergen.^d (i) The number of exact 6 aa hits the input sequence had with this allergen, and (ii) the percentage of exact hits the input sequence is found to hit this allergen sequence.

length alignment data, 8 mer exact matches were found to be 26 for olive TLP and 30 for kiwi TLPs compared to allergenic pollen TLPs which showed only 1–3 exact matches.

The results of allergenicity assessment of sapodilla TLP 1 by Allermatch using the full-length, 80-aa sliding window and 6-aa Wordmatch are shown in Table 3. TLPs from kiwi, bell pepper, banana, apple, cherry and peach as well as from pollens (red cedar and Mediterranean cypress) showed >35% identity with sapodilla TLP 1 by the full length analysis. The resulting 80-aa matches were fairly identical with full-length matches indicating the overall similarity of sapodilla TLP 1 with the above-mentioned TLPs. However, the 6-aa Wordmatch analysis showed that only TLPs from kiwi (Act d 2), bell pepper (Cap a 1) and banana (Mus a 4) seem to have higher number of Wordmatches with sapodilla TLP 1.

Evaller assessment of sapodilla TLP 1 showed 7 peptide matches – 5 from kiwi TLP (Act d 2; matched sequences of the query: 71–98, 91–115, 159–179, 164–184 and 167–192), 1 each from mountain cedar TLP (Jun a 3; matched sequence 42–62) and eastern red cedar major pollen allergen (Jun v 3; matched sequence 42–62) with Smith-Waterman scores ranging from 94 to 132. Sapodilla TLP 1 is classified as 'presumably an allergen' based on Evaller assessment.

3.2. Linear B-cell epitope scoring for sapodilla TLP 1

A total of 5 B-cell epitopes, named sequentially as b1 to b5, spanning the entire length of sapodilla TLP 1 were predicted by BCPreds (Fig. 1, panel a). The prediction scores for each residue of sapodilla TLP 1, as obtained by IEDB BepiPred, is shown in Fig. 1, panel b. From this analysis, it is seen that a total of 7 B-cell linear epitopes (sequentially named 'a' to 'g') can be predicted for sapodilla TLP 1. Epitopes 'a', 'd' and 'g' show high scores, whereas epitopes 'b', 'c', 'e' and 'f' show intermediate scores.

3.3. IgE epitope prediction of sapodilla TLP 1 by AlgPred

Upon analysis by AlgPred, two epitopes were predicted on sapodilla TLP 1 (Fig. 1, panel a): IgE epitope '1', stretching from residues 36 to 48 (APGSTKARIWGRT) and IgE epitope '2' stretching from residues 51 to 63 (NFDANGQGKQCTG). A portion of the IgE epitope '2' overlaps with the thaumatin family signature as shown in Fig. 1 (panel a). Further, it was observed from BepiPred analysis that the residues having high scores in epitope 'b' and 'c' are APGST and QGKC, respectively; these stretches are found in IgE epitope

1 and 2, respectively, which indicate their higher accessibility for antibody recognition.

3.4. IgE epitope prediction of allergenic food TLPs

Using AlgPred allergenicity prediction server, IgE epitopes were predicted for thaumatins and each of the plant TLP sequences listed as allergens in Table 1. The prediction of IgE epitope was based on matching the query sequence with the experimentally obtained IgE epitope sequence deposited in the AlgPred database. The clustalX alignments for allergenic food TLPs displaying the predicted IgE epitopes are shown in Fig. S1 (Supplementary data). It is readily apparent from Fig. S1 that there are 4 predicted IgE epitopes matching with the IgE epitopes available in the database: epitope 1 (AAGTASARFWGRT), epitope 2 (TFDASGKGSCQTG), epitope 3 (ADINAVCPSELK) and epitope 4 (VDGGCNSACNVFKT). From this analysis, a few interesting observations can be made, namely, (i) the number of predicted IgE epitopes varies from 0 to 4, (ii) thaumatin sequences contain IgE epitopes 2 and 4, (iii) 27 out of 46 (59%) allergenic food TLPs displayed either one or two IgE epitopes, (iv) 7 TLPs (green kiwi TLPs: AGC39175.1, ABQ42566.1; banana TLP: 1Z3Q; grape TLPs: all except Q9LLB7) displayed all the 4 IgE epitopes and six TLPs displayed 3 IgE epitopes in two combinations (banana TLP: AFK29763.1; kiwi/velvet vine TLP: AGC39182.1; green kiwi TLP: AGC39180.1; gold kiwi TLPs: AGC39176.1, AGC39181.1; grape TLP: Q9LLB7) and (v) six TLPs do not display any putative IgE epitopes (13%).

Most of the food TLPs (29 out of 41) ends in the sequence-FCP, whereas 7 food TLPs including sapodilla TLP 1 have C-terminal extensions ranging from 2 to 8 residues. Another 7 food TLPs show C-terminal extensions ranging from 17 to 29 residues. It is interesting to note that two food TLPs have unusually long C-terminal extensions – 78 residues (musk melon TLP, ADN33946.1) and 83 residues (peach TLP, AEV57472.1).

3.5. IgE epitope prediction of allergenic pollen TLPs

Similarly, the clustalX alignments of allergenic pollen TLPs displaying the predicted IgE epitopes are shown in Fig. S2 (Supplementary data). Most pollen TLPs (12 out of 15) range in length from 199 to 206 residues; however, 3 pollen TLPs are longer (218 to 223 residues). In addition to the 4 predicted IgE epitopes described for food allergenic TLPs, 7 out of 15 pollen allergenic

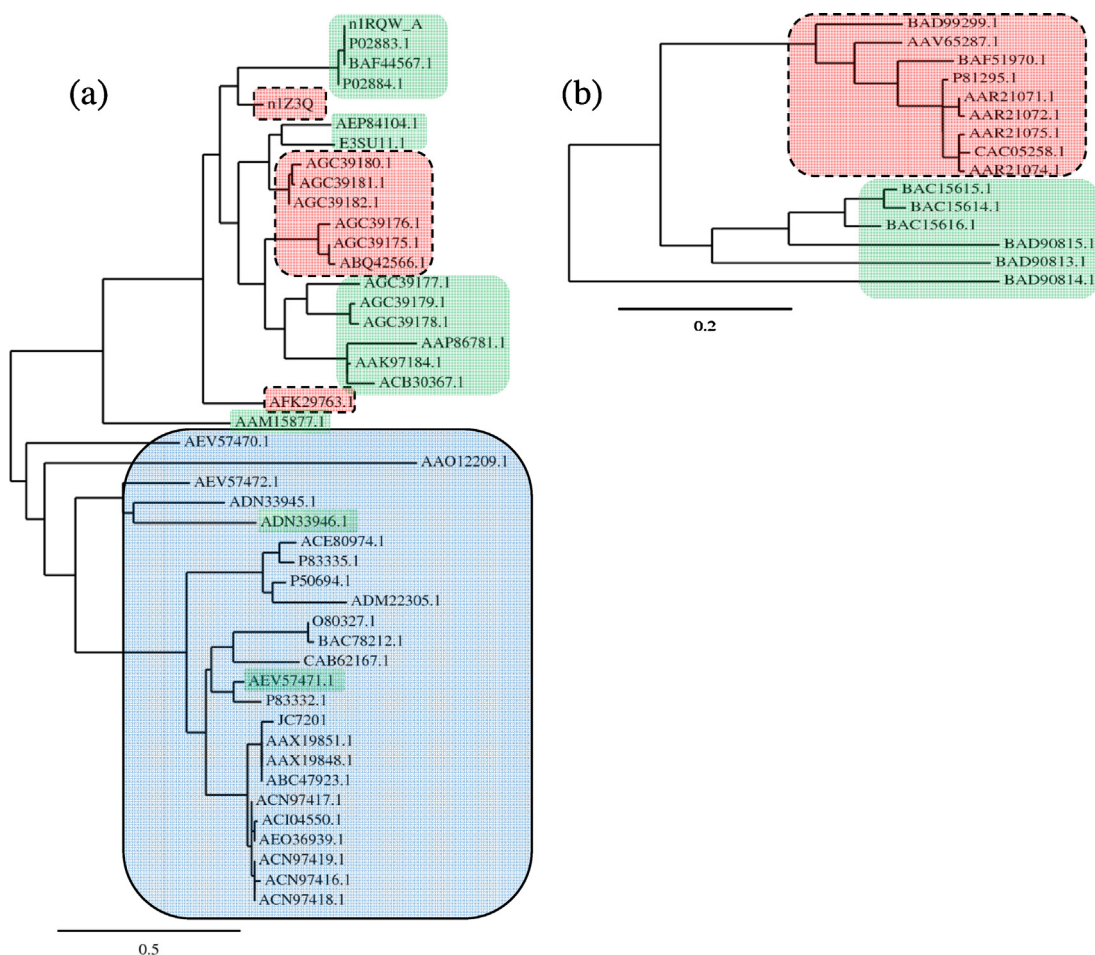


Fig. 2. Distribution of predicted IgE epitopes among allergenic plant TLPs. Food allergenic TLPs are shown in panel (a) and pollen allergenic TLPs in panel (b). The color shadings represent the number of predicted IgE epitopes in the sequence. Blue shade with solid border represents sequences showing either none or one epitope; exceptions are highlighted which displayed 2 epitopes. Red shade with broken border represents sequences having 3 or 4 epitopes in case of food TLPs, and 4 or 5 epitopes in case of pollen TLPs. Sequences harboring only 2 epitopes are shown under a green shade with no border. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

TLPs (47%) showed a match with the 5th IgE epitope (NAYVDNC-PATNYSK) present in the AlgPred database. Interestingly, pollen allergenic TLPs with either 0 or 1 epitope were absent in this analysis. Pollen allergenic TLPs with 2 IgE epitopes were found in only four (27%), whereas those with either 3 or 4 IgE epitopes accounted for only 2 each (13% each).

3.6. Distribution of predicted IgE epitopes among allergenic plant TLPs

The distribution pattern of predicted IgE epitopes of food allergenic TLPs (shown in Fig. S1) is represented phylogenetically in Fig. 2, panel a. It is observed from this representation that the TLPs from the lower half were found to have either 0 or 1 IgE epitope, the exceptions being TLPs from musk melon (ADN33946.1) and peach (AEV57471.1) which show two IgE epitopes. In contrast, the majority of the TLPs from the upper half (sapodilla, olive, bell pepper, kiwi fruit and wheat), as well as thaumatococins, contain two IgE epitopes. However, several TLPs from kiwi fruit contain 3 or 4 IgE epitopes.

Of the 15 allergenic pollen TLPs shown in Fig. 2 (panel b), all belonging to the family Cupressaceae of order Pinales (class Pinopsida), 9 pollen TLPs harbor 4 or 5 predicted IgE epitopes, and form a separate clade (upper half of panel b). Six out of 15 pollen TLPs (lower half of panel b) contain only 2 or 3 predicted

IgE epitopes. Unlike allergenic food TLPs, no allergenic TLPs with 0 or 1 IgE epitope were found.

Further, it can be observed that the lower half of the phylogenetic tree (Fig. 2, panel a) comprises of TLPs from rosaceae (apple, plum, peach, almond, cherry and apple pear), cucurbitaceae (musk melon), fagaceae (chestnut) and brassicaceae (cabbage) all of which belong to the rosids; the rosids and the asterids are the largest clades in the eudicots. The upper half of the tree comprises of TLPs from actinidiaceae (kiwi fruit), solanaceae (bell pepper), oleaceae (olive), musaceae (banana), poaceae (wheat), marantaceae (miracle fruit) and sapotaceae (sapodilla). It should be noted here that banana, wheat and miracle fruit belong to the Commelinids clade, while kiwi fruit, sapodilla, olive and bell pepper belong to the Asterids clade.

3.7. Homology modeling of sapodilla TLP 1

For homology modeling of sapodilla TLP 1, six deposited crystallographic structures of plant TLPs were available on PDB, any of which could be used as a template. However, based on alignment scores, banana TLP showed the highest score with sapodilla TLP 1; therefore, the top scoring crystal structure of banana TLP was used as the template. The generated homology model of sapodilla TLP 1 is displayed in Fig. 3, which shows that it has an overall structural similarity with other plant TLPs (Ghosh and Chakrabarti, 2008;

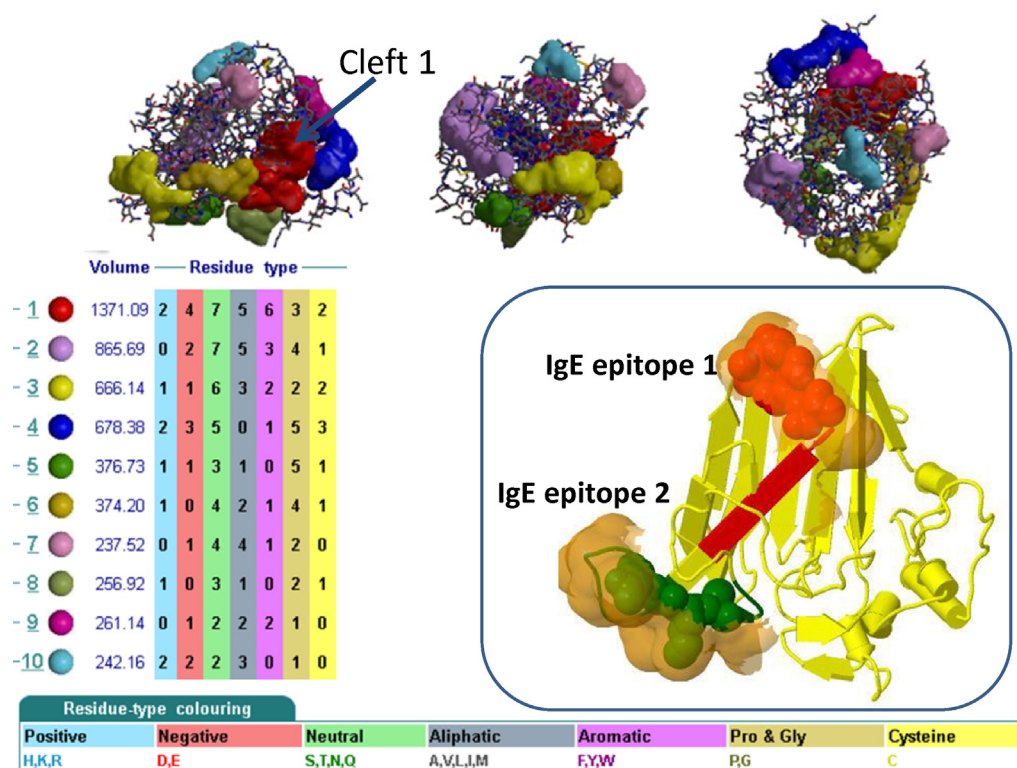


Fig. 3. Predicted clefts on the generated 3D structure of sapodilla TLP 1. Protein structure is shown in 3 different views. Clefts are shown as color-filled spaces; other regions are shown as backbones. Cleft numbering is assigned based on its volume, highest being 1. Types of residues in the clefts are shown in different colors. (Inset) 3D structure of sapodilla TLP 1 showing IgE epitopes predicted by AlgPred; the structure is shown as a cartoon. IgE epitopes 1 and 2 are shown in red and green, respectively. The epitope residues with high prediction scores by IEDB BepiPred are shown as space-filled; the solvent-accessible surface for the IgE epitopes is shown as translucent halos. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article.)

Menu-Bouaouiche et al., 2003) in having a central domain made up of antiparallel β -sheets supported by 2 domains on either side, one of which is made up of only β -sheets and the other made up of both β -sheets and α -helices; the entire structure is strongly stabilized by 8 disulfide bridges. The stereochemical quality of the generated model was evaluated using PROCHECK available in PDBsum; the results indicated that 89.2% residues lie in the most favorable region of the Ramachandran plot (a good quality model generally has >90% residues lying in the most favorable region).

The putative clefts in the modeled sapodilla TLP 1 structure are shown in Fig. 3. Among the 10 clefts observed in the structure, 4 clefts appear to be of significance since they have volumes of >500 Å³, and the clefts with higher volumes tend to possess functional characteristics. The first cleft with volume size of 1371.09 Å³ seems to be the biggest cleft containing 2 cysteines, 4 negative and 2 positive residues; this could very well correspond to the electronegative cleft described in the case of tobacco PR-5d, a vacuolar neutral anti-fungal TLP (Koiwa et al., 1999). An important feature of the first cleft is the presence of high number of neutral (7) and aromatic (6) residues. The two predicted IgE epitopes of sapodilla TLP 1 are mapped on the 3D model – epitope 1 and 2 shown in red and green, respectively (Fig. 3, inset).

3.8. Epitope mapping and structural variation

The predicted B-cell and IgE epitopes were mapped on the secondary structures of some representative TLPs. We chose apple and cherry TLPs belonging to the family rosaceae from the rosid clade (Fig. 2, lower half of panel a); sapodilla TLP 1 and banana TLP were selected from the non-rosid clade (Fig. 2, upper half of panel a). A grape TLP was also included in this analysis, since several cases

of allergy to grape and wine have been reported wherein TLP has been identified as one of the grape allergens (Pastorello et al., 2003; Armentia et al., 2007; Cardinale et al., 2010). Among the grape TLPs listed in Table 1, the crystal structure of one is available (UniProt Accn. O04708); IgE epitope prediction showed that this grape TLP has 4 IgE epitopes (data shown in Fig. 4).

The annotations and epitope mappings on the secondary structures of banana, grape, sapodilla, apple and cherry TLPs are depicted in Fig. 4. It can be observed that 5–7 high-scoring predicted B-cell epitopes are present in these TLPs (indicated as solid boxes) spanning the entire sequence, of which only two seem to be highly conserved – the second and the last one. The predicted IgE epitopes (indicated as broken boxes) after mapping on the secondary structure was observed to lie in two distinct regions. IgE epitope 1 is found in all except cherry TLP, and IgE epitope 2 is found only in banana, grape and sapodilla TLPs. IgE epitope 3 is present only on banana, grape and cherry TLPs, whereas IgE epitope 4 is found only on banana and grape TLPs.

Although the five TLPs have a similar overall structure, they can be grouped into two based on their fine variations in their structure (Fig. 4). The N-terminal, the C-terminal and the central regions of all these TLPs have identical secondary structures; however, variations in the length of β -sheets and β -hairpins are seen prominently in the central and the C-terminal regions. Further, a small helix is present in the central region of apple TLP and in the C-terminal region of grape and sapodilla TLPs.

The region 1 (between the N-terminal and central regions) and region 2 (between the central and C-terminal regions) have more number of variations with respect to the secondary structural elements. It can also be observed that most of the helices appear in

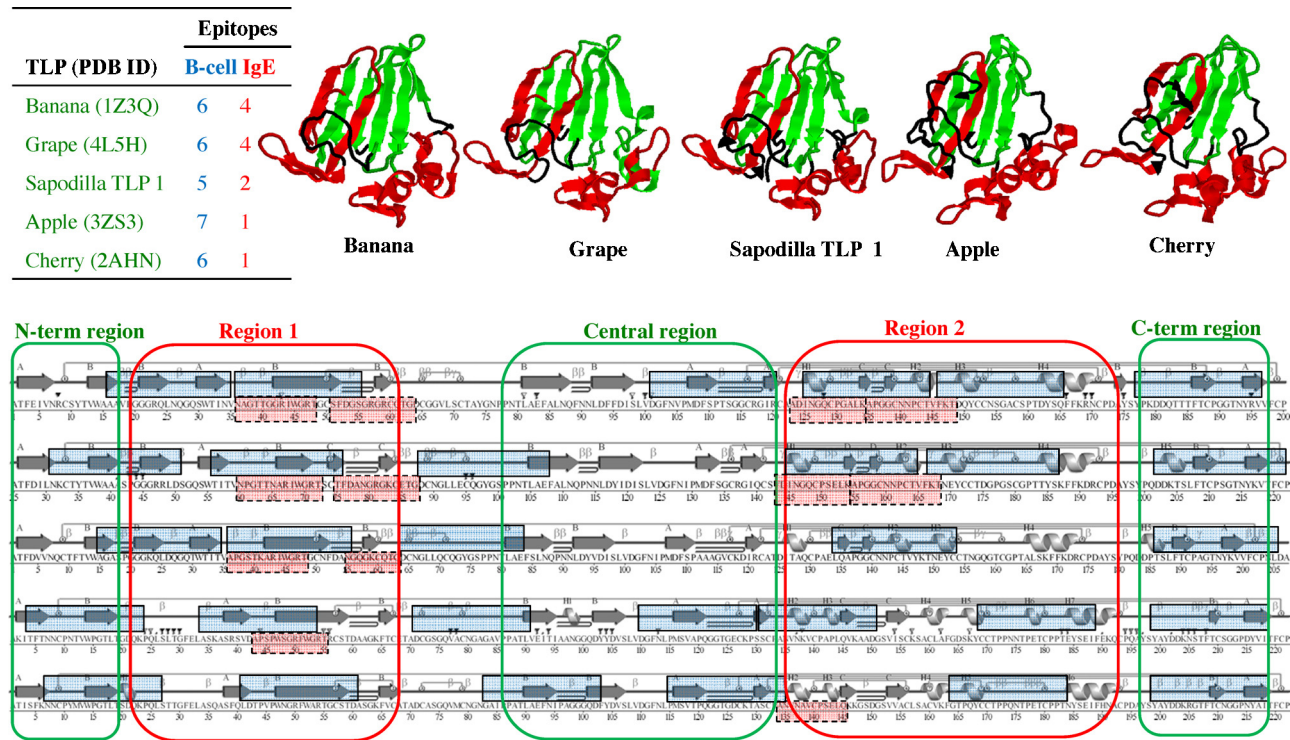


Fig. 4. Structural features and mapping of predicted epitopes on selected allergenic food TLPs. X-ray crystallographic structures of banana, grape, apple and cherry were retrieved from PDB. Structure of sapodilla TLP 1 was modeled based on homology. Secondary structure annotation was performed using PDBsum; the display from top to bottom corresponds to the TLPs as listed in the inset table. B-cell epitopes predicted by BCPreds are shown in solid boxes, and IgE epitopes predicted by AllgPred are shown in broken boxes. 3D structures are represented in cartoons using Rasmol. The secondary structural elements highlighted in green correspond to the N-terminal, central and C-terminal regions; similarly, the secondary structural elements highlighted in red corresponds to regions 1 and 2.

region 2. It is interesting to note that all the predicted IgE epitopes in all TLPs lie in either region 1 or 2.

It is easily discernible from the 3D structures (Fig. 4) that all the TLPs have a similar topology, consisting of a central lectin-like domain of antiparallel β -sheets surrounded by two disulfide-stabilized domains of both β -sheets and α -helices. The variable regions 1 and 2 observed in the secondary structures harboring the predicted IgE epitopes are seen to lie in the side domains. It can be seen from Fig. 4 that the rosoid TLPs (apple and cherry) fold differently than the non-rosoid TLPs (banana, grape and sapodilla).

3.9. Cladogram of different botanical families in the order Ericales

Since sapodilla TLP 1 is evolutionarily related to olive TLP (family oleaceae of order Lamiales and clade asterids) and several kiwi TLPs (family actinidiaceae of order Ericales and clade asterids), it appeared interesting to look at the allergens reported from the plants in the order Ericales. This order also includes families which produce several edible berries (blueberry, cranberry and bilberry) and fruits (persimmon, date plum, star apple and sapote). A phylogenetic tree of the different families in the order Ericales (Smets and Pyck, 2003) was retrieved from Tree of Life web project (2002) in order to assess the relationship of sapotaceae family with other families known to cause inhalant and ingestant type I allergies. Order Ericales consists of 28 families of which only plants from 7 families are reported, till date, to cause either pollen or food allergies (Fig. S3, Supplementary data). Kiwi fruit allergy is the most widely investigated and highest number of allergens has been reported from kiwi. Other food allergies have been reported from Brazil nut, persimmon, blueberry, sapodilla, argan oil and tea.

4. Discussion

Plant proteins known to be allergens have been grouped into different families of which the pathogenesis-related proteins (PRPs) comprise the biggest group named PR-1 to PR-17 (Radauer and Breiteneder, 2007; Radauer et al., 2008). All plant TLPs belong to the PR-5 family and share in common a sequence similarity to thaumatin, a sweet-tasting protein from the fruits of the West African shrub, *Thaumatococcus daniellii*. Although there is high degree of sequence and structural similarity among TLPs and thaumatin, not all TLPs share common functional features, be it allergenicity, β -1,3-glucanase activity, antifungal activity, β -glucan binding activity or kinase activity. The reason for this functional diversity is unknown at present (Liu et al., 2010; Narasimhan et al., 2009). It is known that osmotin (*Nicotiana tabacum*) domain I can be overlapped to the β -barrel domain of adiponectin; recently, a nonapeptide spanning 157–165 of osmotin has been shown to bind to adiponectin receptor on human synovial fibroblasts and causing release of cytokines (Miele et al., 2011).

Many food and pollen TLPs are reported as allergens from various plants (Breiteneder, 2004). Although several TLP genes have been reported in eudicots, there is variation in gene sequences; further, the tissue-specific expression profiles of these TLP genes are not known completely. In view of the total number of plant TLPs known and the >50% identity with sapodilla TLP 1 by the full-length alignment and 80-mer window methods (FARRP and Allermatch), it is readily evident that olive, kiwi, bell pepper, banana, wheat and other rosoid TLPs from plant foods show >50% identity; the high-scoring TLPs belong to the non-rosoid group. The pollen TLPs do show considerable similarity although the number of pollen TLP sequences available for comparison are very few (Radauer and Breiteneder, 2006). The number of 8-mer matches with sapodilla

TLP 1 by FARRP, Allermatch and Evaller was found to be high for kiwi, bell pepper and olive TLPs.

AlgPred epitope analysis shows that the predicted IgE epitopes 1 and 2 are present in almost all allergenic TLPs, IgE epitopes 1 and 2 being the predominant ones in the case of food TLPs; on the contrary, IgE epitopes 1, 2 and 3 occur very frequently in most pollen TLPs. IgE epitope 5, absent in allergenic food TLPs, is found in a majority (47%) of allergenic pollen TLPs. Since only a few pollen TLP sequences have been deposited, a significant comparison cannot be made of the predicted IgE epitopes.

Only a few plant genomes (papaya, *Carica papaya*; poplar, *Populus trichocarpa*; rice, *Oryza sativa*; thale cress, *Arabidopsis thaliana*) have been investigated for the multigenic nature and linkage analysis of TLPs (Liu et al., 2010; Looze et al., 2009; Petre et al., 2011; Gao et al., 2005; Chen et al., 2008); allergy to these plant TLPs have not been reported. Grapevine (*Vitis vinifera*) genome was compiled recently (Jaillon et al., 2007). IgE epitope prediction of the *Vitis* TLPs from its sequenced genome showed that the number of epitopes present varied among the 42 putative TLPs identified *in silico*; 24 out of 42 (57%) have only one predicted IgE epitope (epitope 2). IgE epitope 2 is present in the majority (81%) of the *Vitis* TLPs (data not shown). The grape TLPs (Accn. Q9M4G6, A9CSP0, O04708 and F6HUH1) were found to have 4 predicted IgE epitopes (Fig. S1, inset).

IgE epitope prediction of all known allergenic plant TLPs using AlgPred showed that not all the epitopes are present in all TLPs analyzed. TLPs were found to differ in the number of epitopes ranging from 0 to 5. This was observed not only within allergenic plant TLPs, but also in a single plant genome (grape) as discussed above. It would be worth studying the clinical reactivity of a particular TLP isoforms in relation to its abundance in the respective fruits and fruit-specific TLP gene expression. It can be observed from the epitope predictions that few of the predicted IgE epitopes overlap with the B-cell epitopes.

The score values of each residue of the predicted IgE epitope in sapodilla TLP 1 indicated the critical 4–6 residues which are likely to interact with the IgE antibody. Because even though the epitope stretches from 15 to 20 residues, in a stereochemical interaction between an antibody and antigen, only 4–6 residues of the epitope are involved (Ponomarenko and van Regenmortel, 2009). It is essential to have the structure of the antigen–antibody complex to precisely identify the epitope residues interacting with the hypervariable regions of the specific IgE.

The wiring diagram of the predicted secondary structure of sapodilla TLP 1 indicates that the disulfide bonding patterns and the preponderance of β -sheets and hairpins over α -helices are similar to those observed for thaumatin and other plant TLPs. This seems to correlate very well with the high percentages of glycine (10.6%), threonine (9.7%) and alanine (8.7%) in the deduced composition of sapodilla TLP 1 (Ashok Kumar et al., 2013). The two predicted IgE epitopes (1 and 2) of sapodilla TLP 1 have been mapped on the modeled structure of the protein. It is observed that almost half of the residues in IgE epitopes 1 and 2 are buried inside and non-accessible to the solvent, but the other half is readily accessible and can bind to the specific IgE of a sensitized patient to produce allergic reactions.

TLPs have been studied for their role as an allergen in causing type I allergy (immediate hypersensitivity) to the ingestion of many fruits. Plant food cross-reactive allergens and pollen-food cross-reactive allergens also form a major contributor for the prevalence of inhalant and ingestant allergies (Cartier, 2010; Ramirez and Bahna, 2009; Leone et al., 2006; Palacín et al., 2010, 2012; Oberhuber et al., 2008). The concept of epitope diversity and allergen variability with reference to plant TLPs in a clinical scenario has been documented by several researchers (Bublin et al., 2004; Maddumage et al., 2013; Palacín et al., 2011). It has been reported

that mountain cedar pollen TLP (Jun a 3) contribute not only to the overall allergenicity, but variable levels of Jun a 3 may alter the allergenic potency of pollens produced under different environmental conditions (Midoro-Horiuti et al., 2000). Soman et al. (2000) have identified 4 tryptic peptides from Jun a 3 based on their reactivity with IgE from patients with mountain cedar pollen allergy. These peptides have the following stretches in Jun a 3 sequence: 120–131, 132–144, 152–165 and 169–179; among these the first 3 experimentally determined IgE epitope match exactly with the predicted IgE epitopes 3, 4 and 5 as shown in Fig. S2. Computational analysis of thaumatin-II and some TLPs indicated that residues Thr12, Leu74, Gln133 and Thr161 of thaumatin-II (or equivalent residues of other TLPs) could be used as a target for mutagenesis with the objective of developing TLPs with reduced allergenicity (Baniulis et al., 2008).

5. Conclusions

Based upon bioinformatics analyses, sapodilla TLP 1 closely matches allergenic TLPs from olive, kiwi, banana and bell pepper which are also known to cause food allergy upon ingestion. Further, two IgE epitopes (1 and 2) have been identified for sapodilla TLP 1. ClustalX alignments of all known putative allergenic TLPs indicate that the occurrence of predicted IgE epitopes 1 and 2 are more prevalent in food allergenic TLPs, and IgE epitopes 2 and 3 in pollen allergenic TLPs. This contrasts with the two predicted IgE epitopes (2 and 4) for thaumatin sequences. Overall, it appears that IgE epitope 2 which partially overlaps with the thaumatin family signature appears to be the most conserved. Allergenic TLPs from the rosids are predicted to harbor 0 or 1 IgE epitope. The number and distribution of the predicted IgE epitopes among the allergenic TLPs may be responsible for the specificity of food or pollen allergy as well as the cross-reactivity observed among plant foods and/or pollens. To our knowledge, this is the first comprehensive report of structural and allergenicity characteristics of an allergenic protein from sapodilla fruits.

Conflict of interest

Both the authors declare no conflicts of interest.

Acknowledgment

We thank the Council of Scientific & Industrial Research (CSIR), New Delhi, for financial support (YPV, institutional project; HGA, research fellowship).

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at <http://dx.doi.org/10.1016/j.molimm.2013.08.010>.

References

- Aalberse, R.C., Crameri, R., 2011. IgE-binding epitopes: a reappraisal. *Allergy* 66, 1261–1274.
- Aluffi-Pentini, F., De Fonzo, V., Parisi, V., 2009. An attempt to improve the quantitative epitope prediction by modelling alternative binding modes. *Open Bioinform. J.* 3, 51–58.
- APEDA agri Xchange, India Production – Search Product Result(s), http://agriexchange.apeda.gov.in/India%20Production/Result_SearchProduct.aspx (last accessed 30.05.2013).
- Armentia, A., Pineda, F., Fernández, S., 2007. Wine-induced anaphylaxis and sensitization to hymenoptera venom. *N. Engl. J. Med.* 357, 719–720.
- Asero, R., Pravettoni, V., 2013. Anaphylaxis to plant-foods and pollen allergens in patients with lipid transfer protein syndrome. *Curr. Opin. Allergy Clin. Immunol.* 13, 379–385.

- Ashok Kumar, H.G., Hegde, V.L., Shetty, S.M., Venkatesh, Y.P., 2013. Characterization and gene cloning of an acidic thaumatin-like protein (TLP 1), an allergen from sapodilla fruit (*Manilkara zapota*). *Allergol. Int.* 62, <http://dx.doi.org/10.2332/allergolint.12-OA-0522> (in press).
- Baniulis, D., Liobikas, J., Gelvonauskienė, D., Sasnauskas, A., Vaitiekaitis, G., Stanys, V., 2008. Computational analysis of thaumatin-II allergenicity and prediction of antigenic elements of thaumatin-like family proteins. *Biologija* 54, 202–207.
- Breiteneder, H., 2004. Thaumatin-like proteins – a new family of pollen and fruit allergens. *Allergy* 59, 479–481.
- Bublin, M., Mari, A., Ebner, C., Knulst, A., Scheiner, O., Hoffmann-Sommergruber, K., Breiteneder, H., Radauer, C., 2004. IgE sensitization profiles toward green and gold kiwifruits differ among patients allergic to kiwifruit from 3 European countries. *J. Allergy Clin. Immunol.* 114, 1169–1175.
- Cardinale, F., Berardi, M., Chinellato, I., Damiani, E., Nettis, E., 2010. A child with anaphylaxis to grapes without reaction to grape seed oil. *Allergy* 65, 800–801.
- Cartier, A., 2010. The role of inhalant food allergens in occupational asthma. *Curr. Allergy Asthma Rep.* 10, 349–356.
- Chen, J.C., Chiu, L.L., Lee, K.L., Huang, W.N., Chuang, J.G., Liao, H.K., Chow, L.P., 2012. Identification of critical amino acids in an immunodominant IgE epitope of Pen c 13, a major allergen from *Penicillium citrinum*. *PLoS ONE* 7, e34627, <http://dx.doi.org/10.1371/journal.pone.0034627>.
- Chen, L., Zhang, S., Illa, E., Song, L., Wu, S., Howald, W., Arús, P., van de Weg, E., Chen, K., Gao, Z., 2008. Genomic characterization of putative allergen genes in peach/almond and their synteny with apple. *BMC Genomics* 9, 543, <http://dx.doi.org/10.1186/1471-2164-9-543>.
- Dall'Antonia, F., Gieras, A., Devanaboyina, S.C., Valenta, R., Keller, W., 2011. Prediction of IgE-binding epitopes by means of allergen surface comparison and correlation to cross-reactivity. *J. Allergy Clin. Immunol.* 128, 872–879.e8.
- Fuchs, H.C., Bohle, B., Dall'Antonia, Y., Radauer, C., Hoffmann-Sommergruber, K., Mari, A., Scheiner, O., Keller, W., Breiteneder, H., 2006. Natural and recombinant molecules of the cherry allergen Pru av 2 show diverse structural and B cell characteristics but similar T cell reactivity. *Clin. Exp. Allergy* 36, 359–368.
- Gao, Z.S., Weg, W.E., Schaart, J.G., Arkel, G., Breiteneder, H., Hoffmann-Sommergruber, K., Gilissen, L.J., 2005. Genomic characterization and linkage mapping of the apple allergen genes Mal d 2 (thumatin-like protein) and Mal d 4 (profilin). *Theor. Appl. Genet.* 111, 1087–1097.
- Ghosh, R., Chakrabarti, C., 2008. Crystal structure analysis of NP24-1: a thaumatin-like protein. *Planta* 228, 883–890.
- Hegde, V.L., Ashok Kumar, H.G., Sreenath, K., Hegde, M.L., Venkatesh, Y.P., 2013. Identification and characterization of a basic thaumatin-like protein (TLP 2) as an allergen in sapodilla plum (*Manilkara zapota*). *Mol. Nutr. Food Res.* <http://dx.doi.org/10.1002/mnfr.201300261> (in press).
- Hegde, V.L., Venkatesh, Y.P., 2002. Oral allergy syndrome to sapodilla (*Achras zapota*). *J. Allergy Clin. Immunol.* 110, 533–534.
- Higginbotham, J.D., Snodin, D.J., Eaton, K.K., Daniel, J.W., 1983. Safety evaluation of thaumatin (Talin protein). *Food Chem. Toxicol.* 21, 815–823.
- Jaillon, O., et al., The French-Italian public consortium for grapevine genome characterization, 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature* 449, 463–467.
- Koiwa, H., Kato, H., Nakatsu, T., Oda, J., Yamada, Y., Sato, F., 1999. Crystal structure of tobacco PR-5d protein at 1.8 Å resolution reveals a conserved acidic cleft structure in antifungal thaumatin-like proteins. *J. Mol. Biol.* 286, 1137–1145.
- Krebitz, M., Wagner, B., Ferreira, F., Peterbauer, C., Campillo, N., Witty, M., Kolarich, D., Steinkellner, H., Scheiner, O., Breiteneder, H., 2003. Plant-based heterologous expression of Mal d 2, a thaumatin-like protein and allergen of apple (*Malus domestica*), and its characterization as an antifungal protein. *J. Mol. Biol.* 329, 721–730.
- Leone, P., Menu-Bouaouiche, L., Peumans, W.J., Payan, F., Barre, A., Roussel, A., Van Damme, E.J., Rougé, P., 2006. Resolution of the structure of the allergenic and antifungal banana fruit thaumatin-like protein at 1.7-Å. *Biochimie* 88, 45–52.
- Liu, J.J., Sturrock, R., Ekramoddoullah, A.K., 2010. The superfamily of thaumatin-like proteins: its origin, evolution, and expression towards biological function. *Plant Cell Rep.* 29, 419–436.
- Looze, Y., Boussard, P., Huet, J., Vandenbusche, G., Azarkan, M., Raussens, V., Wintjens, R., 2009. Purification and characterization of a wound-inducible thaumatin-like protein from the latex of *Carica papaya*. *Phytochemistry* 70, 970–978.
- Maddumage, R., Nieuwenhuizen, N.J., Bulley, S.M., Cooney, J.M., Green, S.A., Atkinson, R.G., 2013. Diversity and relative levels of actinidin, kiwifruit, and thaumatin-like allergens in 15 varieties of kiwifruit (*Actinidia*). *J. Agric. Food Chem.* 61, 728–739.
- Maleki, S.J., Teuber, S.S., Cheng, H., Chen, D., Comstock, S.S., Ruan, S., Schein, C.H., 2011. Computationally predicted IgE epitopes of walnut allergens contribute to cross-reactivity with peanuts. *Allergy* 66, 1522–1529.
- Menu-Bouaouiche, L., Vriet, C., Peumans, W.J., Barre, A., Van Damme, E.J., Rougé, P., 2003. A molecular basis for the endo-beta 1,3-glucanase activity of the thaumatin-like proteins from edible fruits. *Biochimie* 85, 123–131.
- Midoro-Horiuti, T., Goldblum, R.M., Kurosky, A., Wood, T.G., Brooks, E.G., 2000. Variable expression of pathogenesis-related protein allergen in mountain cedar (*Juniperus ashei*) pollen. *J. Immunol.* 164, 2188–2192.
- Miele, M., Costantini, S., Colonna, G., 2011. Structural and functional similarities between osmotin from *Nicotiana tabacum* seeds and human adiponectin. *PLoS ONE* 6, e16690, <http://dx.doi.org/10.1371/journal.pone.0016690>.
- Narasimhan, M.L., Bressan, R.A., D'Urzo, M.P., Jenks, M.A., Mengiste, T., 2009. Unexpected turns and twists in structure/function of PR-proteins that connect energy metabolism and immunity. *Adv. Bot. Res.* 51, 439–489.
- Oberhuber, C., Bulley, S.M., Ballmer-Weber, B.K., Bublin, M., Gaier, S., DeWitt, A.M., Briza, P., Hofstetter, G., Lidholm, J., Vieths, S., Hoffmann-Sommergruber, K., 2008. Characterization of Bet v 1-related allergens from kiwifruit relevant for patients with combined kiwifruit and birch pollen allergy. *Mol. Nutr. Food Res.* 52 (Suppl. 2), S230–S240.
- Palacín, A., Quirce, S., Sanchez-Monge, R., Bobolea, I., Diaz-Perales, A., Martin-Muñoz, F., Pascual, C., Salcedo, G., 2011. Sensitization profiles to purified plant food allergens among pediatric patients with allergy to banana. *Pediatr. Allergy Immunol.* 22, 186–195.
- Palacín, A., Rivas, L.A., Gomez-Casado, C., Aguirre, J., Tordesillas, L., Bartra, J., Blanco, C., Carrillo, T., Cuesta-Herranz, J., Bonny, J.A., Flores, E., Garcia-Alvarez-Eire, M.G., Garcia-Nunez, I., Fernandez, F.J., Gamboa, P., Munoz, R., Sanchez-Monge, R., Torres, M., Losada, S.V., Villalba, M., Vega, F., Parro, V., Blanca, M., Salcedo, G., Diaz-Perales, A., 2012. The involvement of thaumatin-like proteins in plant food cross-reactivity: a multicenter study using a specific protein microarray. *PLoS ONE* 7, e44088, <http://dx.doi.org/10.1371/journal.pone.0044088>.
- Palacín, A., Tordesillas, L., Gamboa, P., Sanchez-Monge, R., Cuesta-Herranz, J., Sanz, M.L., Barber, D., Salcedo, G., Diaz-Perales, A., 2010. Characterization of peach thaumatin-like proteins and their identification as major peach allergens. *Clin. Exp. Allergy* 40, 1422–1430.
- Pastorello, E.A., Farioli, L., Pravettoni, V., Ortolani, C., Fortunato, D., Giuffrida, M.G., Perono Garoffo, L., Calamari, A.M., Brenna, O., Conti, A., 2003. Identification of grape and wine allergens as an endochitinase 4, a lipid-transfer protein, and a thaumatin. *J. Allergy Clin. Immunol.* 111, 350–359.
- Paull, R.E., 2008. In: Janick, J., Paull, R.E. (Eds.), *The Encyclopedia of Fruit & Nuts*. CAB International, Wallingford, Oxfordshire, pp. 828–831.
- Petre, B., Major, I., Rouhier, N., Duplessis, S., 2011. Genome-wide analysis of eukaryote thaumatin-like proteins (TLPs) with an emphasis on poplar. *BMC Plant Biol.* 11, 33, <http://dx.doi.org/10.1186/1471-2229-11-33>.
- Ponomarenko, J.V., van Regenmortel, M.H.V., 2009. B-cell epitope prediction. In: Gu, J., Bourne, P.E. (Eds.), *Structural Bioinformatics*, 2nd ed. Wiley-Blackwell, Hoboken, New Jersey, pp. 847–877.
- Prescott, S.L., 2013. Early-life environmental determinants of allergic diseases and the wider pandemic of inflammatory noncommunicable diseases. *J. Allergy Clin. Immunol.* 131, 23–30.
- Radauer, C., Breiteneder, H., 2007. Evolutionary biology of plant food allergens. *J. Allergy Clin. Immunol.* 120, 518–525.
- Radauer, C., Breiteneder, H., 2006. Pollen allergens are restricted to few protein families and show distinct patterns of species distribution. *J. Allergy Clin. Immunol.* 117, 141–147.
- Radauer, C., Bublin, M., Wagner, S., Mari, A., Breiteneder, H., 2008. Allergens are distributed into few protein families and possess a restricted number of biochemical functions. *J. Allergy Clin. Immunol.* 121, 847–852.e7.
- Ramirez Jr., D.A., Bahna, S.L., 2009. Food hypersensitivity by inhalation. *Clin. Mol. Allergy* 7, 4, <http://dx.doi.org/10.1186/1476-7961-7-4>.
- Saha, S., Raghava, G.P.S., 2006. AlgPred: prediction of allergenic proteins and mapping of IgE epitopes. *Nucleic Acids Res.* 34 (Web Server issue), W202–W209.
- Sharma, P., Gaur, S.N., Arora, N., 2013. *In silico* identification of IgE-binding epitopes of osmotin protein. *PLoS ONE* 8, e54755, <http://dx.doi.org/10.1371/journal.pone.0054755>.
- Small, E., 2011. *Top 100 Exotic Food Plants*. CRC Press, Boca Raton, pp. 515–518.
- Smets, E., Pyck, N., 2003. *Ericales* (Rhododendron). In: *Encyclopedia of Life Sciences*. Macmillan Reference Ltd., New York, Section A3730.
- Soman, K.V., Midoro-Horiuti, T., Ferreón, J.C., Goldblum, R.M., Brooks, E.G., Kurosky, A., Braun, W., Schein, C.H., 2000. Homology modeling and characterization of IgE binding epitopes of mountain cedar allergen Jun a 3. *Biophys. J.* 79, 1601–1609.
- Tree of Life Web Project, 2002. *Ericales*, Version 01 January 2002 (temporary), <http://tolweb.org/Ericales/20720/2002.01.01> In The Tree of Life Web Project, <http://tolweb.org/> (last accessed on 30.05.2013).
- Vaughan, K., Greenbaum, J., Kim, Y., Vita, R., Chung, J., Peters, B., Broide, D., Goodman, R., Grey, H., Sette, A., 2010. Towards defining molecular determinants recognized by adaptive immunity in allergic disease: an inventory of the available data. *J. Allergy (Cairo)* 2010, 628026, <http://dx.doi.org/10.1155/2010/628026>.



Characterization and gene cloning of an acidic thaumatin-like protein (TLP 1), an allergen from sapodilla fruit (*Manilkara zapota*)

Journal:	<i>Allergology International</i>
Manuscript ID:	JSA-AI-2012-OA-0522.R3
Manuscript Type:	Original Article
Date Submitted by the Author:	01-May-2013
Complete List of Authors:	Ashok Kumar, Hassan; CSIR-CFTRI, Biochemistry & Nutrition Hegde, Venkatesh; University of South Carolina School of Medicine, Pathology, Microbiology & Immunology Shetty, Santoshkumar; CSIR-CFTRI, Biochemistry & Nutrition Venkatesh, Yeldur; CSIR-CFTRI, Biochemistry & Nutrition
Keyword:	allergen, food allergy, protein structure, skin prick test, allergen-specific IgE
Specialty Area:	Basic Science, Allergy, Biochemistry
Sub Specialty Area:	Allergen, Food Allergy, Immunology

1 **Original article**

2

3 **Characterization and gene cloning of an acidic thaumatin-like protein (TLP 1), an**
4 **allergen from sapodilla fruit (*Manilkara zapota*)[#]**

5 **Hassan G. Ashok Kumar, Venkatesh L. Hegde[†], Santoshkumar M. Shetty[‡], Yeldur P.**
6 **Venkatesh***

7 Dept of Biochemistry and Nutrition, CSIR–Central Food Technological Research Institute (a
8 Council of Scientific and Industrial Research laboratory), Mysore–570020, Karnataka, India

9

10 **Running title:** Acidic thaumatin-like protein is a sapodilla allergen

11 ***Corresponding author**

12 Dr. Yeldur P. Venkatesh, Dept. Biochemistry & Nutrition, CSIR–Central Food Technological
13 Research Institute (CFTRI), KRS Road, Mysore-570020, Karnataka, India

14 Tel: +91-821-2514876; Fax: +91-821-2517233

15 Email: venkatyp@yahoo.com; ypv@cftri.res.in

16

17 [†]Present address: Department of Pathology, Microbiology and Immunology, University of
18 South Carolina School of Medicine, Columbia, South Carolina 29208, USA.

19 [‡]Present address: Bhat Bio-Tech India Pvt. Ltd., Research & Development, No. 11-A, 4th
20 Cross, Veerasandra Industrial Area, Electronics City, Bengaluru–560100, Karnataka, India.

21

22 [#]This work was presented at the 2012 WAO International Scientific Conference held on 6–9
23 December 2012 at Hyderabad, India (Abstract No. 2046).

24 **Number of words: 51~~56~~15**

Abstract

Background: Allergy to sapodilla (*Manilkara zapota*) fruit ingestion is rare. An independent study from our group has identified a basic thaumatin-like protein (TLP 2) as the major allergen. The present study was aimed at identifying and characterizing additional allergens from sapodilla.

Methods: Allergic subjects were identified by case history, skin prick test (SPT) and allergen-specific IgE. Sapodilla extract was fractionated using SP-Sepharose into 3 components (SP1, SP2 and SP3) which were analyzed by native/SDS-PAGE, IgE-immunoblot, isoelectric focusing (IEF) and N-terminal sequencing. The conserved regions of plant TLPs and the N-terminal sequence ~~and the conserved regions of plant TLPs~~ were used to design primers for PCR.

Results: SPT and ELISA confirmed a subject with oral allergy syndrome (OAS) to sapodilla and custard apple. Two proteins (26.9 and 24.5 kDa; reducing conditions) were detected as allergens, of which the latter in SP2 has already been identified as basic TLP (TLP 2). The 26.9 kDa protein present in SP1 was identified as an acidic TLP based on native PAGE, IEF and N-terminal sequencing. Presence of a basic β -1,3-glucanase in SP3 was inferred by zymography. Sequence analysis of the genomic clone of the acidic TLP gene revealed that it is intronless and non-glycosylated. Evolutionary relatedness to olive, grape and kiwi fruit allergenic TLPs were inferred by phylogenetic analysis.

Conclusions: An acidic TLP (TLP 1) was identified as a new allergen in sapodilla. TLP 1 is a single polypeptide (207 residues) belonging to the thaumatin family of the GH64-TLP-SF superfamily. Clinically, sapodilla should be considered in the list of fruits causing OAS.

Keywords: acidic TLP; exotic fruit; food allergen; *Manilkara zapota*; oral allergy syndrome; sapodilla; thaumatin-like protein

Introduction

Sapodilla (*Manilkara zapota* syn. *Achras zapota*), also known as sapota, chikoo and by various other names, yields delicious fruits that are round, ovoid, ellipsoid or conical, flattened at the stem end, 5 to 10 cm wide.^{1,2} The fruit is highly perishable and sensitive to cold storage; hence, the bulk of the produce is used for table purpose. Sapodilla is cultivated in the tropical parts of the world with India being the largest grower. Commonly available varieties in India are the round cricket ball variety and the oval Kalipatti variety. The fruits are rich in sugars (10.2%) and fiber (11.5%); however, the protein content is very low (~0.4–0.7%).

Fruits and vegetables form a major part of the food pyramidal system. The importance of regular consumption of fruits and vegetables in the prevention of cancer and cardiovascular diseases has been widely acknowledged.³ WHO recommends daily intake of approx. 400 g of fresh fruits and vegetables per person. Though uncommon, fruits and vegetables elicit allergic reactions in sensitized individuals; fruit allergy may increase in future due to increased consumption.³⁻⁵ Although several common fruits (apple, banana, citrus, grape, and other Rosaceae fruits) have been studied with respect to their allergenicity, some exotic fruits, consumed commonly in a particular region have not been well studied in terms of Oral Allergy Syndrome (OAS) and cross-reactive fruits.⁶

Generally, fruits cause allergic reactions confined to the oropharyngeal region, and is referred to as OAS or pollen-food allergy.^{7,8} With import of several tropical fruits in many countries, sensitization to many uncommon fruits may occur more commonly. However,

allergy to sapodilla is very rare, and only three cases of OAS have been reported; a 21 kDa allergen has been identified from sapodilla pulp⁹. Further, it has been shown that this allergen from sapodilla pulp is a basic thaumatin-like protein (TLP 2) based on its amino acid composition, isoelectric point and N-terminal sequence (Hegde VL et al. and others, unpublished observations).

During the course of investigation of the basic TLP from sapodilla for further biochemical and immunological studies, we detected another protein with a similar molecular mass which eluted earlier to the basic TLP ~~by from~~ SP-Sepharose cation-exchange chromatography of sapodilla extract. Therefore, it appeared interesting to purify this acidic protein from sapodilla fruit and characterize it in relation to fruit allergy with the objective of identifying additional allergens. The present study describes the characterization of the biochemically purified acidic TLP from sapodilla pulp and its ~~the~~ genomic cloning ~~of the acidic thaumatin-like protein and characterization of the biochemically purified protein from sapodilla pulp~~. Further, a comparative phylogenetic analysis has been carried out on acidic TLPs which have been identified as food allergens till date.

Materials and methods

Materials

Sapodilla (cv. cricket ball) saplings were obtained from Indian Institute of Horticultural Research, (Bangalore, India), and maintained at the nursery of this institute (CFTRI, Mysore, India). Round and oval sapodilla fruits were purchased from local market. InsTAclone™ PCR Cloning Kit was procured from Fermentas International (Burlington, ON, Canada). Primers were synthesized at Eurofins Genomics India Pvt. Ltd., Bangalore, India. Cetyltrimethylammonium bromide (CTAB), 3-

(cyclohexylamino)-1-propanesulfonic acid (CAPS), laminarin (from *Laminaria digitata*), laminarinase (from *Trichoderma* sp.; 100-400 units/g solid), mouse monoclonal anti-human IgE-alkaline phosphatase (ALP) conjugate, Immobilon-P^{SO} PVDF membranes, SP-Sepharose Fast Flow and JumpStart™ AccuTaq™ LA DNA Polymerase Mix were purchased from Sigma-Aldrich Chemical Co., St. Louis, MO. 2,3,5-Triphenyl-tetrazolium chloride (TTC, A.R. grade) was a product of Reanal, Budapest, Hungary. Ultrafiltration membrane (PLGC, 10000 NMWL, 44.5 mm diameter) was purchased from Millipore Corporation, Bedford, MA.

Diagnostic tests for sapodilla allergy

This study was undertaken following approval by the Institutional Ethics Committee. Informed consent was obtained from normal and allergic subjects prior to collection of blood for *in vitro* tests. Diagnostic tests for confirming allergy to sapodilla included case history, skin prick test (SPT), and allergen-specific IgE by ELISA; these were carried out as described previously for diagnosis of allergy to the ingestion of sapodilla,⁹ pomegranate, mango, a chewable drug containing excipient mannitol, and eggplant.¹⁰⁻¹⁴ A total of 300 subjects in the age range 18-60 y were screened at random. A 22-year-old female (~~K.M.~~) always experienced allergy whenever she ate sapodilla fruit or custard apple (sweetsop; *Annona squamosa* of family Annonaceae). She first noticed itching, swelling and redness (mainly confined to the oropharyngeal area) to these fruits at the age of 10, within 15 min of ingestion of sapodilla or custard apple. Since then, she has avoided eating these fruits.

Purification of an acidic thaumatin-like protein

Pulp from round ripe sapodilla fruits (250 g) was homogenized in an equal volume (250 mL) of extraction buffer (20 mM KH₂PO₄, pH 7.4 containing 3 mM EDTA, 0.9% NaCl, and 0.02% NaN₃). The homogenate was clarified by cheese cloth filtration and centrifuged at

7000 $\times$ g at 4 °C for 10 min. The supernatant was dialyzed against the same buffer using 12-14 kDa cut-off dialysis tubing to remove low molecular weight substances. Proteins were precipitated by 75% ammonium sulfate saturation, and the precipitate was resuspended in a minimum volume of the starting buffer (20 mM sodium acetate buffer, pH 4.0) and dialyzed against the starting buffer using 12-14 kDa cut-off tubing. The dialyzed solution was loaded onto a cation-exchange column (SP-Sepharose FF, 1.45 $\times$ 67.5 cm; 10 mL bed volume) which had been pre-equilibrated with the starting buffer. After loading, the column was washed with the starting buffer to remove unbound proteins. The bound proteins were step-eluted as follows: (a) 20 mM sodium acetate buffer, pH 5.0, (b) buffer in step (a) + 0.1 M NaCl, (c) buffer in step (a) + 0.2 M NaCl, and (d) buffer in step (a) + 1 M NaCl. The eluted protein fractions were pooled, desalted and concentrated using a 10 kDa mol. wt. cut-off membrane in an Amicon stirred-cell ultrafiltration system. Protein quantitation was performed by the dye-binding method using BSA as standard.¹⁵

Native PAGE, SDS-PAGE and Electrofocusing

The chromatographic components from SP-Sepharose chromatography of sapodilla extract were analyzed by 12% SDS-PAGE (dimensions: 14 cm width $\times$ 16 cm length for long gel, and 8 cm width $\times$ 6 cm length for mini gel) according to Laemmli¹⁶ under both reducing and non-reducing conditions. PAGE at alkaline and acidic pH under native conditions were carried out using Tris-glycine buffer, pH 9.5 (pH 8.9 system or Davis system) and β -alanine-acetic acid buffer, pH 4.3 (pH 4.3 system or Reisfeld system), respectively.¹⁷ Isoelectric point was evaluated by analytical electrofocusing in an immobilized pH gradient containing ampholyne covering the pH range 3.5–9.5 using a Multiphor II unit (Amersham Biosciences, Uppsala, Sweden). An isoelectric focusing (IEF) calibration kit (Amersham Pharmacia

Biotech) containing 11 proteins of known isoelectric points (pI) was used as reference. The proteins were allowed to focus initially at 500 V for 20 min, and then at 1500 V for 1 h.

Zymography for β -1,3-glucanase activity

β -1,3-Glucanase activity on native polyacrylamide gels was carried out according to Stahmann et al.¹⁸ After the native PAGE run, the gel was washed 3 times with distilled water and equilibrated in 0.5 M sodium acetate buffer, pH 4.5. The gel was then incubated in the same buffer containing laminarin (6.7 mg/mL) at 37 °C for 2 h. After washing 3 times in distilled water, the gel was immersed in a solution containing 0.15 g TTC in 100 mL of 1 M NaOH, and placed in a boiling water bath until red bands on a clear background appear. The gel was then immediately transferred to a solution of 7.5% acetic acid to conserve the clear background.

β -1,3-Glucanase assay

β -1,3-Glucanase activity for the SP-Sepharose chromatographic components was analyzed using laminarin as the substrate as per the protocol described by Sigma Chemical Co., [St. Louis, MO \(L-5272 – Enzyme Assay\)](#). The assay mixture consisting of 100 mM sodium acetate buffer, pH 5 (0.6 mL), 2.5% laminarin in water (0.2 mL), and 5 μ g test protein in water (0.2 mL) was incubated at 37 °C for 60 min. Release of reducing sugars was measured as glucose by dinitrosalicylic acid (DNS) method.¹⁹ Briefly, 0.2 mL of DNS reagent and 0.2 mL of sample (test), or distilled water (blank), were taken in test tubes. The tubes were placed in a water bath at 100 °C for 10 min, and then allowed to cool to room temperature. After the addition of 0.8 mL distilled water to each tube and mixing, the

absorbance of the solutions was measured at 540 nm. Appropriate controls were used (BSA as negative control and laminarinase as positive control), and a calibration curve was prepared using D-glucose as standard in the range 0–200 µg. The protein content of laminarinase was determined by Bradford assay.¹⁵

N-Terminal sequencing

Protein components from SP-Sepharose chromatography were subjected to SDS-PAGE (12%, reducing). Multiple lanes of a sample were run; after electrophoresis, one half of the gel was stained using 0.3 M CuCl₂²⁰ for visualization of protein bands. The bands of interest were located on the gel based on molecular mass, and the corresponding regions in the other half of the gel were excised. The protein band of interest was transferred on to an Immobilon-P^{8Q} PVDF membrane by wet transfer in 10 mM CAPS buffer with 20% methanol overnight. After transfer, the membrane was transiently stained using 0.2% Ponceau S in 10% acetic acid for the visualization of the protein band; after excising the band, it was destained using water. The membrane was then subjected to N-terminal sequencing by Edman degradation on a PROCISE Protein Sequencer, Model 491 (Applied BioSystems, Foster City, CA) at the Protein Sequencing Facility, (Indian Institute of Science, Bangalore, and CSIR-CFTRI, Mysore).

Cloning of an acidic thaumatin-like protein from sapodilla

Genomic DNA from the fresh leaves of sapodilla plant (cv. cricket ball) was prepared according to the method described by Saghai-Maroo et al.²¹ Fresh leaves (1 g) were homogenized using an autoclaved mortar and pestle in 20 mL of extraction buffer (50 mM Tris-HCl buffer, pH 8, containing 0.7 M NaCl, 10 mM EDTA, 1% CTAB, 0.1% 2-

186 mercaptoethanol) and incubated at 60 °C for 1 h with occasional mixing by gentle swirling.
187 Ten milliliters of chloroform-isoamyl alcohol (24:1 v/v) were added and the solution was
188 mixed by inversion to form an emulsion that was centrifuged at 6000 × g at 25 °C for 10 min.
189 The aqueous phase was removed, and two-thirds volume of isopropanol was added and
190 mixed 2-4 times by quick gentle inversions. The precipitated DNA was lifted out with a bent
191 glass rod, transferred to a tube containing 20 mL of 76% ethanol/10 mM ammonium acetate,
192 kept for 20 min, and then dissolved in 1.5 mL of 10 mM ammonium acetate/0.25 mM EDTA.

193 All the primers designed and used for cloning sapodilla acidic TLP are listed in Table
194 S1 (Supplementary information). Two sets of nested degenerate primers were designed. N-
195 Terminal sequence of the 26.9 kDa sapodilla acidic allergen (ATFDV V~~N~~QCT FTVWA
196 GASPG GGKQL-; identified as thaumatin-like protein) was used to design forward primer
197 for set 1 (SAP2: 5' CAGTGT-ACCTTCACCGTCTGGGC 3'). Also, this sequence was used
198 as a query to retrieve gene sequences and their respective protein translations from NCBI
199 Database (Genbank ID: *Nicotiana tabacum* P07052.1, *Olea europaea* ACZ57583.1, *Vitis*
200 *vinifera* AAB61590.1, *Camellia sinensis* ABE01396.1, *Solanum tuberosum* AAU95244.1,
201 *Sambucus nigra* AAK59278.1 and *Nepenthes rafflesiana* ACU31848.1). Multiple sequence
202 alignment was performed using DIALIGN ([http://bibiserv.techfak.uni-](http://bibiserv.techfak.uni-bielefeld.de/dialign/submission.html)
203 [bielefeld.de/dialign/submission.html](http://bibiserv.techfak.uni-bielefeld.de/dialign/submission.html)) for these gene and protein sequences. A highly
204 conserved region was identified from the protein sequences' alignment (named conserved
205 region A representing the sequence -CNNPCTV-) and a degenerate reverse primer was
206 designed (SAP4: 5' TGTAACAACCCC-TGTACCGTG 3'). Primers were analyzed using
207 Integrated DNA Technologies (IDT) oligo analyzer software
208 ([http://www.idtdna.com/analyzer/](http://www.idtdna.com/analyzer/Applications/OligoAnalyzer/)-Applications/OligoAnalyzer/). PCR amplification was

then carried out using JumpStart™ AccuTaq™ LA DNA Polymerase Mix according to the manufacturer's instructions.

PCR product formation was confirmed by 1% agarose gel electrophoresis and then ~~was~~ cloned using InsTAclone™ PCR Cloning Kit and sequenced at Eurofins Genomics India Pvt. Ltd., Bangalore, India. The sequence of the product was used to design forward primer for set 2 (SAP5: 5' CGACATATCTCTGGTTGACGG 3'). Further, the sequence was also used as a query to retrieve sequences from NCBI Database (Genbank ID: *Camellia sinensis* ABE01396.1, *Vitis vinifera* AAB61590.1, *Olea europaea* ACZ57583.1, *Sambucus nigra* AAK59278.1, *Solanum tuberosum* AAU95246.1, *Sambucus nigra* AAK59276.1, *Coffea arabica* ABW76503.1 and *Actinidia deliciosa* ABQ42566.1). Multiple sequence alignment was performed. A highly conserved region was identified at the C-terminus of plant TLPs and a reverse primer was designed (SAP9: 5' GGTTGTCTTCTGCCCTTGAAATTG 3'). Primer analysis and synthesis, PCR amplification and vector sequencing were carried out as described above.

Nucleotide sequence and the deduced protein sequence analyses

The two nucleotide sequences obtained were assembled using DIALIGN, and BLAST search was performed against NCBI database (<http://blast.ncbi.nlm.nih.gov/>). The cloned nucleotide sequence was translated into the protein sequence using ExPASy translate tool (<http://web.expasy.org/translate/>). The protein sequence was then analyzed for conserved domains using NCBI Conserved Domain Database (CDD) search (<http://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi>). The physical and chemical characteristics of the deduced protein sequence were determined using ProtParam tool of ExPASy (<http://web.expasy.org/protparam/>). Phylogenetic relationship between sequences was analyzed using Phylogeny.fr one click mode (<http://www.phylogeny.fr/>). Putative *N*-

glycosylation site prediction was analyzed using the NetNGlyc algorithm available online at <http://www.cbs.dtu.dk/services>.

Results

Identification of a subject having allergy to sapodilla and custard apple

The allergic subject showed a wheal/flare diameter (SPT response) of 6/20 mm for 50% sapodilla extract and 7/25 mm with pseudopods for 50% custard apple extract (Table 1); during the SPT, the allergic subject experienced irritation at the prick points. When allergen-specific IgE was carried out by ELISA using the allergic-subject's serum, there was a 2-fold increase in the absorbance for specific IgE compared to pooled normal sera for both these fruit extracts. These results indicate the confirmation of allergy to both sapodilla and custard apple fruits.

Purification of the allergen from sapodilla extract by chromatography

An extract of sapodilla pulp followed by dialysis at neutral pH resulted in a protein yield of 95 mg from 250 g sapodilla pulp. After ammonium sulfate precipitation, reconstitution and dialysis at acidic pH, approx. 22 mg protein was recovered as a clear solution. The precipitated proteins at pH 4 amounting to ~77% of proteins in sapodilla extract could be resolubilized at neutral pH, and contained mainly high molecular weight proteins as analyzed by SDS-PAGE; further, the precipitated protein pool did not contain any low molecular weight proteins (<30 kD) described hereinafter.

The chromatographic profile of sapodilla protein fractionation is shown in Fig. 1. The protein components eluted using 20 mM sodium acetate buffer, pH 5 is labeled as SP1 (1.7 mg); similarly, the protein components eluted using 20 mM sodium acetate buffer, pH 5 with 0.1 and 0.2 M NaCl are labeled as SP2₁ (1.4 mg) and SP3₁ (1.9 mg), respectively. When

1 M NaCl in 20 mM sodium acetate buffer, pH 5 was used, no protein was detected in the eluted pool.

The SP-Sepharose components were analyzed by ELISA for detection of the allergens using sapodilla-allergic subject's serum. The results are shown in Table 2. In comparison to the values obtained for normal serum, the sapodilla-allergic serum showed increased ELISA values for both SP-1 and SP-2 indicating the presence of allergenic proteins.

Analyses by gel electrophoresis and electrofocusing

Upon analysis by SDS-PAGE under reducing conditions (Fig. 2, panel a), SP1 showed a major band of 26.9 kDa (band 'a'). SP2 showed two distinct bands – one of 24.5 kDa, (band 'b') and another of 28.2 kDa (band 'c'). Only a single protein band of 37.4 kDa (band 'd') is evident in SP3. Under non-reducing conditions (Fig. 2, panel a), the mobility of band 'a' shifted to 18.6 kDa, and that of band 'b' shifted to 19.9 kDa. However, the mobilities of the other two protein bands ('c' and 'd') did not alter much under non-reducing conditions (27.3 and 38.2 kDa, respectively). It should be noted here that in order to increase the resolution of the protein bands, a long SDS-PAGE gel was used; the protein bands 'a' and 'b' normally appear at 20–22 kDa position in the regular mini gel (Fig. 2, panel b). Only the protein bands 'a' in SP1 and 'b' in SP2 were detected in IgE-immunoblotting using sapodilla-allergic serum (Fig. 2, panel c); no bands were detected in SP3 pool.

Upon examining the protein components on native PAGE under basic system (Fig. 3, panel a), only band 'a' moved through the gel indicating its acidic nature, whereas bands 'b', 'c', and 'd' did not enter the gel, indicating their basic nature. On the other hand, when these components were subjected to native PAGE under acidic system (Fig. 3, panel b), band

279 ‘a’ did not enter the gel, whereas bands ‘b’ and ‘c’ resolved as two closely spaced protein
280 bands and band ‘d’ resolved as a single band indicating that all of them are basic proteins.

281 Electrofocusing analysis (Fig. 3, panel c) revealed that SP1 component showed 4
282 proteins of pI values 3.9, 4.1, 4.3 and 5.5 out of which the first three protein bands appear to
283 represent variants of band ‘a’ shown in Fig. 3 (panel a), since band ‘a’ runs as a single intense
284 band in SDS-PAGE gels. The protein band with a pI of 5.5 is likely to be a different co-
285 eluting protein in SP1. On the other hand, the protein bands in SP2 and SP3 components do
286 not resolve in the electrofocusing gel and appear as a single protein band having a pI >9.5
287 (Fig. 3, panel c).

288
289 **N-terminal sequencing and ~~allergen~~-identification of the allergen as a thaumatin-like**
290 **protein**

291 SP1 component was blotted onto PVDF membrane; upon N-terminal sequencing of
292 the major band ‘a’ by Edman degradation, the sequence was found to be ~~resulted as~~
293 ATFDVVNQCTFTVWAGASPGGGKQL- (UniProt Accession No. B3EWX8; TLP 1b).
294 BLAST search against protein sequence database picked up thaumatin-like protein sequences
295 (Supplementary information, Table S2). Also, on another occasion, a separate purification
296 was carried out using commercial sapodilla fruits and the purified SP1 component was
297 sequenced. The 5th residue was found to be Ile, rather than Val. Based on these results, it can
298 be inferred that there are at least two acidic TLP genes encoding the two isoforms of acidic
299 TLP (1a and 1b).

300 The 24.5 kDa protein band ‘b’ present in SP2 component was also sequenced and
301 the N-terminal sequence was found to be ATFDIQNNC- (UniProt Accession No. B3EWE5;

TLP 2). The isolation and characterization of the basic TLP (TLP 2) are described in a separate study (Hegde VL et al., unpublished observations).

N-Terminal sequencing of band 'c' present in SP2 component resulted in the sequence DVSXFA- which did not yield any significant matches upon BLAST search. In the case of sequencing of band 'd' present in SP3 component, no amino acids were released right from the first cycle indicating the presence of a blocked N-terminus.

Analysis of β -1,3-glucanase activity

Table 3 shows the results of the β -1,3-glucanase assay performed on the SP-Sephadex chromatographic components of sapodilla extract. Since both SP1 and SP2 components showed significantly higher enzymic activity than the negative control (BSA), it was inferred that these components exhibit β -1,3-glucanase activity. On the other hand, SP3 shows β -1,3-glucanase activity which is significantly higher than the activity displayed by either SP1 or SP2. Laminarinase, which hydrolyzes both β -1,3 and β -1,4 glycosidic bonds, was used as a positive control in this assay. Since laminarin contains β -1,3 and β -1,6 linkages in the ratio 3:1, it appears that the protein present in SP3 component is likely to be β -1,3-glucanase.

Analysis of SP1 component for β -1,3-glucanase activity by zymography on native basic PAGE demonstrated detectable enzymic activity by acidic TLP shown as band 'a' (Fig. 4, panel a). In addition to the β -1,3-glucanase activity seen for acidic TLP, one can also observe a marked glucanase activity exhibited by a less acidic or a neutral protein (not detectable by Coomassie staining in Fig. 4, panel a) which appears to have co-purified with acidic TLP in SP1 component; this protein is likely to be an acidic or neutral β -1,3-glucanase in sapodilla pulp. Similarly, SP2 component shows detectable β -1,3-glucanase activity by

zymography on native acidic PAGE (Fig. 4, panel b) indicating that the basic TLP (the more intense band 'b' in Coomassie staining) has weak β -1,3-glucanase activity. SP3 (protein band 'd') displays significant β -1,3-glucanase activity (Fig. 4, panel b) indicating that it is most likely a basic β -1,3-glucanase in sapodilla pulp; this seems to parallel the enzymic activity seen in the test tube assay (Table 3).

Sequence analysis of cloned sapodilla acidic TLP

The partial nucleotide sequence of the acidic TLP gene has 603 bp including the stop codon, and has been deposited at NCBI GenBank sequence repository (JN624813.1). BLAST search identified the deduced protein sequence (AEP84104.1) as belonging to TLP. CDD search revealed that the sapodilla acidic allergen has domains which fit into the superfamily glycoside hydrolase family 64 (GH64) and thaumatin-like proteins (TLP). The deduced partial protein sequence (200 residues) lacks the N-terminal 7 residues as well as the signal peptide. The composite sequence of sapodilla acidic TLP based on the N-terminal sequence (residues 1–25) and the deduced sequence (residues 8–207) is shown in Fig. 5. The calculated molecular mass is 21922.4 with a theoretical pI of 4.44. Notably, the protein lacks histidine and contains only one methionine at position 107; the 3 highly conserved tryptophans are present at positions 14, 31 and 45. Sapodilla acidic TLP is rich in glycine (10.6%), threonine (9.7%), alanine (8.7%), cysteine/proline (7.7% each) and aspartic acid (7.3%), and lacks a putative N-glycosylation site (-N-X-S- or -N-X-T-).

The alignment of the complete sequence of sapodilla acidic TLP with the sequences of the top 10 matches from BLAST search and their percent identities are annotated in Fig. 6. The residue numbers given henceforth refers to the residue number of sapodilla acidic TLP. The salient features of this annotation include the high degree of conservation of all the 16 cysteine residues and the 3 tryptophans. However, there is an additional Trp at position 74 in

castor bean TLP, and two additional tryptophans at positions 34 and 74 in soybean and chickpea TLPs. The single Met at position 107 is highly conserved in all the sequences shown except in olive TLP. The thaumatin family signature²² is clearly evident from the sequence stretch of -G-X-[G/F]-X-C-X-T-[G/A]-D-C-X(1,2)-[G/Q]-X(2,3)-C- which represents amino acid residues -GQGKCQTGDCNGLLQC- from 56-71 in the mature sapodilla acidic TLP.

Sequence comparison and phylogenetic analysis of sapodilla acidic TLP with known food-allergenic TLPs

The alignment of the complete sequence of sapodilla acidic TLP with those of allergenic (food and pollen) acidic TLPs along with their percent identities are shown in Fig. 7. Met-107 is conserved in Cry j 3.4, Mal d 2, Pru av 2.0101, Pru p 2.0201 and Pru p 2.0301, whereas in all the other allergenic acidic TLPs it is replaced by Leu. Trp-45 is conserved in all the TLPs listed except Pru p 2.0201 which has Phe. The missing Trp-31 in some acidic TLPs (Cry j 3.5, Mal d 2, Pru av 2.0101, Pru p 2.0201, Pru p 2.0301 and Pyr py 2) appears to be compensated by Trp-40; these 6 TLPs also have a longer sequence which may provide them with additional structural or functional features.

The phylogenetic analysis of sapodilla acidic TLP with known allergenic acidic and basic TLPs from several food sources (olive, kiwi fruit, bell pepper, apple, pear, peach and sweet cherry) is shown in Fig. S1 (Supplementary information). It is readily seen that sapodilla acidic TLP is closely related to the TLPs from olive, grape, kiwi, bell pepper and tobacco; among these, tobacco TLP (osmotin) is an inhalant allergen. Further, it is apparent that sapodilla acidic TLP is evolutionarily related to the TLPs from olive and kiwi fruits, both of which belong to the same order as sapodilla (order Ericales). Interestingly, this order also

includes families which produce several edible berries (blueberry, cranberry and bilberry) and fruits (persimmon, kaki, date plum, star apple and sapote).

Discussion

Allergic reactions to ingestion of sapodilla fruit are rare; only 3 cases of sapodilla allergy displaying oral allergy syndrome has been described earlier.⁹ The major allergen in sapodilla causing OAS has been identified as a basic TLP in an extended study (Hegde VL et al., unpublished observations). An important caveat of their study is the non-detection of an acidic TLP in sapodilla extracts. In the present study, we have described an additional case of sapodilla allergy based on allergy diagnostic tests, and found that an acidic protein also reacted with the serum from the sapodilla-allergic subject's sera. The allergic subject experienced symptoms of oral allergy to the ingestion of sapodilla as well as custard apple, another exotic fruit commonly consumed in India. Several reports of allergy to custard apple have appeared in the literature;²³⁻²⁵ in one study, a 20-25 kDa protein was identified as the allergen by IgE-immunoblotting,²³ which may likely be TLP. An important constraint in the present study is the non-availability of sera from earlier sapodilla-allergic subjects to address the issue of prevalence of sapodilla allergy.

Biochemical purification of the acidic allergen eluting at pH 5 (band 'a') from SP-Sepharose chromatography of sapodilla extract, and N-terminal sequencing up to 12 or 25 residues identified this protein as a TLP by BLAST search. Upon native PAGE using basic system, the TLP in SP1 moves as a single protein band indicating its acidic nature; conversely, this protein cannot be detected in native PAGE using acidic system. By N-terminal sequencing of the band 'a' on two occasions, it was found that the amino acid at position 5 can be either isoleucine or valine ~~or isoleucine~~ indicating that there are at least 2 variants of sapodilla acidic TLP (TLP 1a and 1b, respectively). Further, IEF analysis showed

that ~~the~~-TLP 1 resolves into 3 variants which differed in their pI values by ~0.2 units. Further, both TLP 1 and TLP 2 show a remarkable difference in electrophoretic mobility between reducing and non-reducing conditions of SDS-PAGE. A similar observation has been made for TLP from kiwi fruits.²⁶ This anomalous behavior of the protein indicates that it is highly rich in disulfide bonds which contribute to the stability of the protein. Purified acidic TLP has been obtained in a yield of 6.8 mg per kg sapodilla fruit.

Many TLPs have been identified as allergens in plant foods and pollens in the last decade.²⁷ Allergenic TLPs have been described from kiwi, apple, grape, sweet cherry, peach, bell pepper and almond.^{26,28-33} Palomares et al.³⁴ have shown that occupational asthma can be caused by ~~a~~-TLP from olive fruit. Many of these allergenic TLPs have been cloned and expressed as recombinant proteins for detailed studies on allergenicity in relation to protein structure and allergy diagnosis.

Since very little information is available on the proteins of sapodilla pulp, it appeared important to clone the gene coding for acidic TLP for allergological studies. A partial acidic TLP gene, found to be intronless, has been obtained by genomic cloning which codes for 200 amino acid residues. The cloned genomic DNA lacked the N-terminal heptapeptide as the primer was designed from the 8th residue in order to avoid the high degeneracy. Functional primers could not be designed for up to the 7th residue. Moreover, attempts with various combinations of primers (Supplementary information, Table S1) and temperatures were unsuccessful. Based on the position of the first cysteine occurring at residue 9 in TLPs, it was inferred that the deduced sequence of the partial acidic TLP represented residues 8-207 of the mature sapodilla acidic TLP. Residues 8 to 25 of the deduced sequence matched completely with the corresponding sequence obtained by N-terminal sequencing of biochemically purified acidic TLP from sapodilla pulp. As expected,

the deduced protein sequence contains 16 highly conserved cysteines which are involved in disulfide bonding for providing structural stability in TLPs.³⁵

TLPs occur in all organisms including fungi, animals and plants.³⁵ Although TLPs are known to be expressed by specific stress, non-specific general stress conditions as causative factors inflammatory response have also been considered.³⁶ Because of their pathogen-induced expression, TLPs are classified under group 5 of the pathogenesis-related proteins (PR-5 family of PRPs),³⁷ which includes cysteine-rich proteins sharing a sequence similarity with thaumatin (THN), an intensely sweet-tasting protein from the fruit of a west African shrub *Thaumatococcus daniellii*, commonly known as miracle fruit or katemfe.³⁸ PR-5 proteins have a clearly recognizable structure and are annotated in protein databases as the THN domain which belongs to the thaumatin family (PROSITE PS00316).

Several TLPs have been shown to exhibit antifungal activity,³⁹ while others possess β -1,3-glucanase activity.^{40,41} Some TLPs have also been shown to bind β -1,3-glucans.⁴² Although our results are indicative of a weak β -1,3-glucanase activity by zymography for both acidic and basic TLPs, it cannot be ruled out that the weak enzymic activity is due to the presence of contaminating β -1,3-glucanases in our purified preparations of SP1 and SP2 components. Further work on obtaining highly purified acidic or basic TLPs as well as recombinant sapodilla TLPs should help to resolve the ambiguity of whether sapodilla TLPs possess inherent weak β -1,3-glucanase activity or not.

The major protein in SP3 component (~37 kDa) exhibits strong β -1,3-glucanase activity. Most basic β -1,3-glucanases are approx. 35 kDa in size and an examination of their N-terminus shows that Gln is present in most cases, and Glu in one case, viz., *Nicotiana plumbaginifolia* (Supplementary information, Fig. S2). Pyroglutamic acid formation occurs when the N-terminal residue of a protein is either glutamine or glutamic acid,⁴³ resulting in a

444 blocked N-terminus. In view of the blocked N-terminus of SP3 component and strong
445 enzymic activity, SP3 component has been suggested to be a basic β -1, 3-glucanase.

446 A comparison of the complete sequence of sapodilla acidic TLP (207 residues) with
447 other plant TLPs (BLAST search results) indicated 109 out of 207 residues are identical
448 (52.7%). On the contrary, a similar comparison with food and pollen allergenic acidic TLPs
449 revealed that only 47 out of 207 residues are identical (22.7%). This indicates that allergenic
450 acidic TLPs seem to have more variation than plant TLPs that have not been identified as
451 allergens till date. Phylogenetic analysis of the allergenic TLPs from foods indicated that
452 sapodilla acidic TLP is closely related to the TLPs from olive, grape, kiwi, and bell pepper
453 and forms one cluster; this cluster is distinct from the other cluster containing allergenic
454 acidic TLPs of the Rosaceae family. A recent study on the involvement of TLPs in plant
455 food cross-reactivity using specific protein microarray indicated that TLPs are significant
456 allergens in plant food allergy and should be considered when diagnosing and treating pollen-
457 food allergy.⁴⁴

458 In conclusion, an acidic form of TLP (TLP 1) has been isolated from sapodilla fruit
459 which causes oral allergy syndrome in a sapodilla-allergic individuals. Sapodilla TLP 1 is a
460 single polypeptide occurring as at least two variants and exhibits weak β -1,3-glucanase
461 activity. The intronless gene coding for sapodilla acidic TLP representing 97% of the mature
462 sequence has been cloned. The complete sequence of sapodilla TLP 1 comprises of 207
463 residues stabilized by 8 disulfide bonds and lacks an *N*-glycosylation site. Phylogenetic
464 analysis indicates close relationship with TLPs from olive, grape, kiwi fruit and bell pepper.
465 The identification of an allergic subject experiencing oral allergy to sapodilla and custard
466 apple may indicate co-sensitization to TLP among fruits.

467 Acknowledgments

The assistance rendered by the staff of the Health Centre, [CSIR](#)-CFTRI, Mysore, is gratefully acknowledged. We thank the Council of Scientific & Industrial Research (CSIR), New Delhi, for financial support (Y.P.V., institutional project; H.G.A. and V.L.H., research fellowship). S.M.S. thanks the Department of Biotechnology (DBT), New Delhi for the award of a research fellowship. We gratefully acknowledge the assistance received from Dr. Purnima Tiku (Protein Chemistry & Technology Department, [CSIR](#)-CFTRI, Mysore) for protein analysis by electrofocusing. Thanks are also due to Dr. Ram Rajasekaran (Director, CSIR-CFTRI) and Dr. Asha Martin (Food Safety & Analytical Quality Control Laboratory Dept., CSIR-CFTRI) for protein sequencing of SP2 component (28 kD) and SP3 component (37 kD) in this study.

Conflict of interest: The authors declare no conflicts of interest.

Supplementary Information

Table S1. Primers designed for use in cloning sapodilla acidic TLP gene

Table S2. [Homology of N-terminal sequence \(25 residues\) of sapodilla acidic TLP as analyzed by BLAST](#)~~BLAST search results for the N-terminal sequence (25 residues) of sapodilla acidic TLP (Man z TLP1)~~

Figure S1. Phylogenetic analysis of sapodilla acidic TLP with known allergenic TLPs from foods

Figure S2. Signal peptide and N-terminal sequences of plant basic β -1,3-glucanases (glucan endo-1,3-beta-glucosidase)

491

492

493

494

495

496

497

498

499

500

501

502

503

504

505

506

507

508

For Peer Review

509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530

References

1. Paull RE. *Manilkara zapota*, chiku. In: Janick J, Paull RE, eds. *The Encyclopedia of Fruit & Nuts*. Wallingford, Oxfordshire, U.K.: CAB International, 2008:828-31.

2. Small E. Sapodilla. In: *Top 100 Exotic Food Plants*. Boca Raton: CRC Press, 2011; 515-8.

3. Marzban G, Mansfeld A, Hemmer W, Stoyanova E, Katinger H, da Câmara Machado ML. Fruit cross-reactive allergens: a theme of uprising interest for consumers' health. *Biofactors* 2005;**23**:235-41.

4. Ballmer-Weber BK, Hoffmann-Sommergruber K. Molecular diagnosis of fruit and vegetable allergy. *Curr Opin Allergy Clin Immunol* 2011;**11**:229-35.

5. Salcedo G, Díaz-Perales A, Sánchez-Monge R. Fruit allergy: plant defence proteins as novel potential panallergens. *Clin Exp Allergy* 1999; **29**:1158-60.

6. Knudsen I, Søborg I, Eriksen F, Pilegaard K, Pedersen J. Risk management and risk assessment of novel plant foods: concepts and principles. *Food Chem Toxicol* 2008;**46**:1681-705.

7. Webber CM, England RW. Oral allergy syndrome: a clinical, diagnostic, and therapeutic challenge. *Ann Allergy Asthma Immunol* 2010;**104**:101–8.

8. Katelaris CH. Food allergy and oral allergy or pollen-food syndrome. *Curr Opin Allergy Clin Immunol* 2010;**10**:246–51.

- 531 9. Hegde VL, Venkatesh YP. Oral allergy syndrome to sapodilla (*Achras zapota*). *J Allergy*
532 *Clin Immunol* 2002;**110**:533-4.
- 533 10. Hegde VL, Mahesh PA, Venkatesh YP. Anaphylaxis caused by mannitol in pomegranate
534 (*Punica granatum*). *ACI Int* 2002;**14**:37-9.
- 535 11. Hegde VL, Venkatesh YP. Anaphylaxis following ingestion of mango fruit. *J Investig*
536 *Allergol Clin Immunol* 2007;**17**:341-4.
- 537 12. Hegde VL, Venkatesh YP. Anaphylaxis to excipient mannitol: evidence for an
538 immunoglobulin E-mediated mechanism. *Clin Exp Allergy* 2004;**34**:1602-9.
- 539 13. Harish Babu BN, Mahesh PA, Venkatesh YP. A cross-sectional study on the prevalence
540 of food allergy to eggplant (*Solanum melongena* L.) reveals female predominance. *Clin*
541 *Exp Allergy* 2008;**38**:1795-802.
- 542 14. Harish Babu BN, Venkatesh YP. Clinico-immunological analysis of eggplant (*Solanum*
543 *melongena*) allergy indicates preponderance of allergens in the peel. *WAO Journal*
544 2009;**2**:192-200.
- 545 15. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities
546 of protein utilizing the principle of protein-dye binding. *Anal Biochem* 1976;**72**:248-54.
- 547 16. Laemmli UK. Cleavage of structural proteins during the assembly of the head of
548 bacteriophage T4. *Nature* 1970;**227**:680-5.
- 549 17. Gabriel O. Analytical disc gel electrophoresis. *Methods Enzymol* 1971;**22**:565-78.
- 550 18. Stahmann KP, Schimz KL, Sahm H. Purification and characterization of four
551 extracellular 1,3- β -glucanases of *Botrytis cinerea*. *J Gen Microbiol* 1993;**139**:2833-40.
- 552 19. Gonçalves C, Rodriguez-Jasso RM, Gomes N, Teixeira JA, Belo I. Adaptation of
553 dinitrosalicylic acid method to microtiter plates. *Anal Methods* 2010;**2**:2046-8.

- 554 20. Sullivan ML, Hatfield RD, Thoma SL, Samac DA. Cloning and characterization of red
555 clover polyphenol oxidase cDNAs and expression of active protein in *Escherichia coli*
556 and transgenic alfalfa. *Plant Physiol* 2004;**136**:3234-44.
- 557 21. Saghai-Marooof MA, Soliman KM, Jorgensen RA, Allard RW. Ribosomal DNA spacer-
558 length polymorphisms in barley: Mendelian inheritance, chromosomal location, and
559 population dynamics. *Proc Natl Acad Sci U.S.A.* 1984;**81**:8014-18.
- 560 22. Edens L, Heslinga L, Klok R, Ledeboer AM, Maat J, Toonen MY, Visser C, Verrips CT.
561 Cloning of cDNA encoding the sweet-tasting plant protein thaumatin and its expression in
562 *Escherichia coli*. *Gene* 1982;**18**:1-12.
- 563 23. Gonzalo MA, Moneo I, Ventas P, Polo F, García JM. IgE-mediated hypersensitivity to
564 custard-apple. *Allergy* 1997;**52**:597.
- 565 24. Sánchez-Guerrero IM, Escudero AI, Tortosa JA, Lombardero M. Anaphylaxis to
566 cherimoya. *Allergy* 2000;**55**:976-7.
- 567 25. Sánchez-Morillas L, Moneo I, Sedano E, Iglesias A, Caballero ML, Rodríguez M. Oral
568 allergic syndrome after ingestion of custard apple. *Allergy* 2003;**58**:260-1.
- 569 26. Gavrović-Jankulović M, Ćirković T, Vucković O, Atanasković-Marković M, Petersen A,
570 Gojgić G, Burazer L, Jankov RM. Isolation and biochemical characterization of a
571 thaumatin-like kiwi allergen. *J Allergy Clin Immunol* 2002;**110**:805-10.
- 572 27. Breiteneder H. Thaumatin-like proteins – a new family of pollen and fruit allergens.
573 *Allergy* 2004;**59**:479-81.
- 574 28. Inschlag C, Hoffmann-Sommergruber K, O'Riordain G, Ahorn H, Ebner C, Scheiner O,
575 Breiteneder H. Biochemical characterization of Pru a 2, a 23-kD thaumatin-like protein
576 representing a potential major allergen in cherry (*Prunus avium*). *Int Arch Allergy*
577 *Immunol* 1998;**116**:22-8.

- 578 29. Krebitz M, Wagner B, Ferreira F, Peterbauer C, Campillo N, Witty M, Kolarich D,
579 Steinkellner H, Scheiner O, Breiteneder H. Plant-based heterologous expression of Mal d
580 2, a thaumatin-like protein and allergen of apple (*Malus domestica*), and its
581 characterization as an antifungal protein. *J Mol Biol* 2003;**329**:721-30.
- 582 30. Pastorello EA, Farioli L, Pravettoni V, Ortolani C, Fortunato D, Giuffrida MG, Perono
583 Garoffo L, Calamari AM, Brenna O, Conti A. Identification of grape and wine allergens
584 as an endochitinase 4, a lipid-transfer protein, and a thaumatin. *J Allergy Clin Immunol*
585 2003;**111**:350-9.
- 586 31. Fuchs HC, Bohle B, Dall'Antonia Y, Radauer C, Hoffmann-Sommergruber K, Mari A,
587 Scheiner O, Keller W, Breiteneder H. Natural and recombinant molecules of the cherry
588 allergen Pru av 2 show diverse structural and B cell characteristics but similar T cell
589 reactivity. *Clin Exp Allergy* 2006;**36**:359-68.
- 590 32. Smole U, Bublin M, Radauer C, Ebner C, Breiteneder H. Mal d 2, the thaumatin-like
591 allergen from apple, is highly resistant to gastrointestinal digestion and thermal
592 processing. *Int Arch Allergy Immunol* 2008;**147**:289-98.
- 593 33. Palacín A, Tordesillas L, Gamboa P, Sanchez-Monge R, Cuesta-Herranz J, Sanz ML,
594 Barber D, Salcedo G, Díaz-Perales A. Characterization of peach thaumatin-like proteins
595 and their identification as major peach allergens. *Clin Exp Allergy* 2010;**40**:1422-30.
- 596 34. Palomares O, Alcántara M, Quiralte J, Villalba M, Garzón F, Rodríguez R. Airway
597 disease and thaumatin-like protein in an olive-oil mill worker. *N Engl J Med*
598 2008;**358**:1306-8.
- 599 35. Liu JJ, Sturrock R, Ekramoddoullah AK. The superfamily of thaumatin-like proteins: its
600 origin, evolution, and expression towards biological function. *Plant Cell Rep*
601 2010;**29**:419-36.

36. Narasimhan ML, Bressan RA, D’Urzo MP, Jenks MA, Mengiste T. Unexpected turns and twists in structure/function of PR-proteins that connect energy metabolism and immunity. *Adv Bot Res* 2009;**51**:439-89.

37. Hoffmann-Sommergruber K. Pathogenesis-related (PR)-proteins identified as allergens. *Biochem Soc Trans* 2002;**30**(Pt 6):930-5.

38. Temussi PA. Natural sweet macromolecules: how sweet proteins work. *Cell Mol Life Sci* 2006;**63**:1876-88.

39. Van Loon LC, Rep M, Pieterse CMJ. Significance of inducible defense-related proteins in infected plants. *Annu Rev Phytopathol* 2006;**44**:135-62.

40. Grenier J, Potvin C, Trudel J, Asselin A. Some thaumatin-like proteins hydrolyse polymeric β -1,3-glucans. *Plant J* 1999;**19**:473-80.

41. Menu-Bouaouiche L, Vriet C, Peumans WJ, Barre A, Van Damme EJM, Rougé P. A molecular basis for the endo- β -1,3-glucanase activity of the thaumatin-like proteins from edible fruits. *Biochimie* 2003;**85**:123-31.

42. Trudel J, Grenier J, Potvin C, Asselin A. Several thaumatin-like proteins bind to beta-1,3-glucans. *Plant Physiol* 1998;**118**:1431-8.

43. Liu YD, Goetze AM, Bass RB, Flynn GC. N-terminal glutamate to pyroglutamate conversion *in vivo* for human IgG2 antibodies. *J Biol Chem* 2011;**286**:11211–17.

44. Palacín A, Rivas LA, Gómez-Casado C, Aguirre J, Tordesillas L, Bartra J, Blanco C, Carrillo T, Cuesta-Herranz J, Bonny JA, Flores E, García-Alvarez-Eire MG, García-Núñez I, Fernández FJ, Gamboa P, Muñoz R, Sánchez-Monge R, Torres M, Losada SV, Villalba M, Vega F, Parro V, Blanca M, Salcedo G, Díaz-Perales A. The involvement of thaumatin-like proteins in plant food cross-reactivity: A multicenter study using a specific protein microarray. *PLoS ONE* 2012;**7**(9):e44088. Doi:10.1371/journal.pone.0044088.

627
628
629
630
631
632
633
634
635
636
637
638
639
640
641

642 Legends to Figures

643 **Fig. 1.** Chromatographic profile of sapodilla extract on SP-Sepharose FF. Column size: 1.54
644 × 76.5 cm; protein load: 22 mg of 75% (NH₄)₂SO₄ precipitated sapodilla extract,
645 reconstituted and dialyzed against 20 mM sodium acetate buffer, pH 4; flow rate: 20 mL/h.
646 Component SP1 (1.7 mg): elution with 20 mM sodium acetate buffer, pH 5; component SP2
647 (1.4 mg): elution with 20 mM sodium acetate buffer, pH 5 + 0.1 M NaCl; component SP3
648 (1.9 mg): elution with 20 mM sodium acetate buffer, pH 5 + 0.2 M NaCl. Elution with 1 M
649 NaCl in 20 mM sodium acetate buffer, pH 5 did not elute any protein (data not shown).

650 Inset: 12% SDS-PAGE (reducing, long gel) pattern of sapodilla extract that was used for
651 chromatography.

652 **Fig. 2.** Panel (a): SDS-PAGE (12% long gel) pattern of SP-Sepharose chromatographic
653 components under reducing (R) and non-reducing (NR) conditions. Lane M, molecular
654 weight markers; lanes 1 and 4: SP1 component; lanes 2 and 5: SP2 component; lanes 3 and 6:
655 SP3 component. Protein bands labeled as 'a, b, c and d' in different components are for easy
656 referral throughout the text. Protein load: 5 μ g in all cases. Stain: Coomassie blue. Panel
657 (b): 12% SDS-PAGE (mini gel, Coomassie stain) showing chromatographic pools from Fig.
658 1. Lane numbers and other details are as shown in panel (a). Panel (c): IgE-immunoblot of
659 SP-Sepharose pools with sapodilla-allergic serum (serum dilution, 1:3; secondary antibody
660 conjugate dilution, 1:2000).

661 **Fig. 3.** Native (non-denaturing) PAGE (12%) of SP-Sepharose chromatographic components.
662 Panel (a): native PAGE using basic system (Tris-glycine buffer system which runs at pH 9.5).
663 Panel (b): native PAGE using acidic system (β -alanine–acetic acid buffer system which runs
664 at pH 4.3). Protein bands labeled as 'a, b, c and d' correspond to the identical bands shown in
665 Fig. 2, panel a. Protein load: 5 μ g in all cases; stain: Coomassie blue. Panel (c): isoelectric
666 focusing of SP-Sepharose chromatographic components of sapodilla extract. The pI markers
667 are indicated on the left. The 3 closely spaced bands grouped as 'a' have pI values of 3.9, 4.1
668 and 4.3. The band marked 'x' represents the position where sample strips are placed for IEF
669 analysis.

670 **Fig. 4.** Zymogram of SP-Sepharose chromatographic components of sapodilla extract. β -
671 1,3-Glucanase activity staining of SP1, SP2 and SP3 components was carried out using
672 laminarin as substrate. In each panel, the left half represents Coomassie-stained gel and the
673 right half represents the zymogram. Protein load: 5 μ g in all cases. Panel (a): native PAGE

(12%), basic system. The band labeled 'a' shows acidic TLP having weak β -1,3-glucanase activity and the solid arrow shows an acidic or neutral protein with strong β -1,3-glucanase activity. Panel (b): native PAGE (12%), acidic system. SP2 component has 2 closely spaced bands ('b' and 'c') which are identical to the corresponding bands in Figs. 2 and 3 based on Coomassie staining intensity. SP3 component shows a single protein (band labeled 'd') which displays strong β -1,3-glucanase activity.

Fig. 5. The complete amino acid sequence of sapodilla acidic TLP (TLP 1). Residues 8 to 207 represent the deduced sequence of sapodilla acidic TLP gene cloned from genomic DNA (JN624813.1). The top line shows the nucleotide sequence (italicized numbers) and the bottom line represents the deduced amino acid sequence (non-italicized numbers). The stop codon (TAG) is indicated by an asterisk. The sequence of residues 1 to 7~~is~~ taken from the N-terminal sequence of acidic TLP (SP1 component, band 'a') is underlined up to 25th residue (UniProt Accn. B3EWX8; sapodilla acidic TLP 1b). The 5th position is an Ile in another N-terminal sequence analysis up to 12 residues (UniProt Accn. B3EWS0; sapodilla acidic TLP 1a). Residues 56–71 shown in double underline represent the thaumatin family signature (PROSITE PS00316). The boxed regions of the sequence were used for designing the primers in 2 sets: boxes with solid line (set1); boxes with broken line (set2). The acidic TLP sequence does not contain an *N*-glycosylation sequon. Cysteine residues are shown in italics. Based on analogy with thaumatin (thaumatin-1, UniProt ID P02883; thaumatin-2, P02884), the disulfide bond pairing are as follows: 9–202, 50–60, 65–71, 117–191, 122–174, 130–140, 144–153, 154–161.

Fig. 6. Sequence alignment of sapodilla acidic TLP (Man z TLP1) with other plant TLPs. Alignments were created using T-Coffee multiple sequence alignment program. Symbols: (*), complete identity; (:), strongly similar; (.), weakly similar; no symbol, different. The

region denoting thaumatin family signature is underlined. UniProt ID and the taxonomic name are shown below (the common name, percent identity with Man z TLP1 and theoretical pI of the mature sequence are shown in parentheses): A2T1L9 - *Camellia sinensis* (tea, 80%, 4.59); O04708 - *Vitis vinifera* (grape, 79%, 4.76); E3SU11 - *Olea europaea* (common olive, 78%, 4.76); K4DFX0 - *Solanum lycopersicum* (tomato, 78%, 4.76); B9RC64 - *Ricinus communis* (castor bean, 78%, 4.73); Q5XUG9 - *Solanum tuberosum* (potato, 76%, 4.76); Q5DJS5 - *Nicotiana tabacum* (common tobacco, 75%, 4.94); Q93XD1 - *Sambucus nigra* (European elder, 74%, 4.85); B6ZHC0 - *Glycine max* (soybean, 72%, 6.19); O81926 - *Cicer arietinum* (chickpea, 73%, 5.05). The TLPs with pI <6 is arbitrarily considered as acidic and shown as underlined in the first segment.

Fig. 7. Alignment of the complete sequence of sapodilla acidic TLP (Man z TLP1) with known acidic TLPs currently identified as pollen and food allergens. Alignments were created using T-Coffee multiple sequence alignment. Symbols: (*), complete identity; (:), strongly similar; (.), weakly similar; no symbol, different. The region denoting thaumatin family signature is underlined. Common name, allergome ID, percent identity with Man z TLP1 and theoretical pI of the mature sequence are shown in parentheses following the taxonomic name. Pollens: *Cryptomeria japonica* (Japanese cedar, Cry j 3, 54%, 4.62; Cry j 3.3, 65%, 4.63; Cry j 3.4, 53%, 5.47; Cry j 3.5, 42%, 4.54; Cry j 3.6, 53%, 4.80; Cry j 3.7, 54%, 4.62; Cry j 3.8, 53%, 4.43); *Cupressus sempervirens* (Mediterranean cypress, Cup s 3.0101, 53%, 5.32; Cup s 3.0102, 53%, 5.32); *Cupressus arizonica* (Arizona cypress, Cup a 3, 52%, 4.82); *Juniperus ashei* (mountain cedar, Jun a 3.0101, 55%, 4.79); *Juniperus rigidus* (Japanese juniper, Jun r 3.1, 54%, 4.82; Jun r 3.2, 54%, 4.82). Foods: *Malus domestica* (apple, Mal d 2, 39%, 4.72); *Olea europaea* (common olive, Ole e 13, 78%, 4.76); *Prunus avium* (cherry, Pru av 2.0101, 41%, 4.57); *Prunus persica* (nectarine/peach, Pru p 2.0201, 40%, 4.80; Pru p 2.0301, 42%, 4.58); *Pyrus pyrifolia* (apple pear, Pyr py 2, 43%, 5.01).

Table 1. Results of SPT and ELISA for diagnosis of sapodilla allergy

Sample	Wheal/flare diameter ^a (mm)	ELISA units ^b (A405 nm)	
		Normal serum (pooled)	Allergic serum
Negative control ^c	1/0	n.a.	n.a.
Positive control ^d	5/25	n.a.	n.a.

Sapodilla extract	6/20	0.206	0.408
Custard apple extract	7/25 ^e	0.188	0.359

742

743 ^a SPT is considered positive if the wheal diameter is 3 mm or more

744 ^b mean of triplicate analysis; n.a. = not applicable

745 ^c 50% glycerinated phosphate-buffered saline

746 ^d 1.66 mg/mL histamine.2HCl equivalent to 1 mg/mL histamine base

747 ^e with pseudopods

748

Table 2. Analysis of SP-Sepharose components of sapodilla extract for allergen-specific**IgE by ELISA**

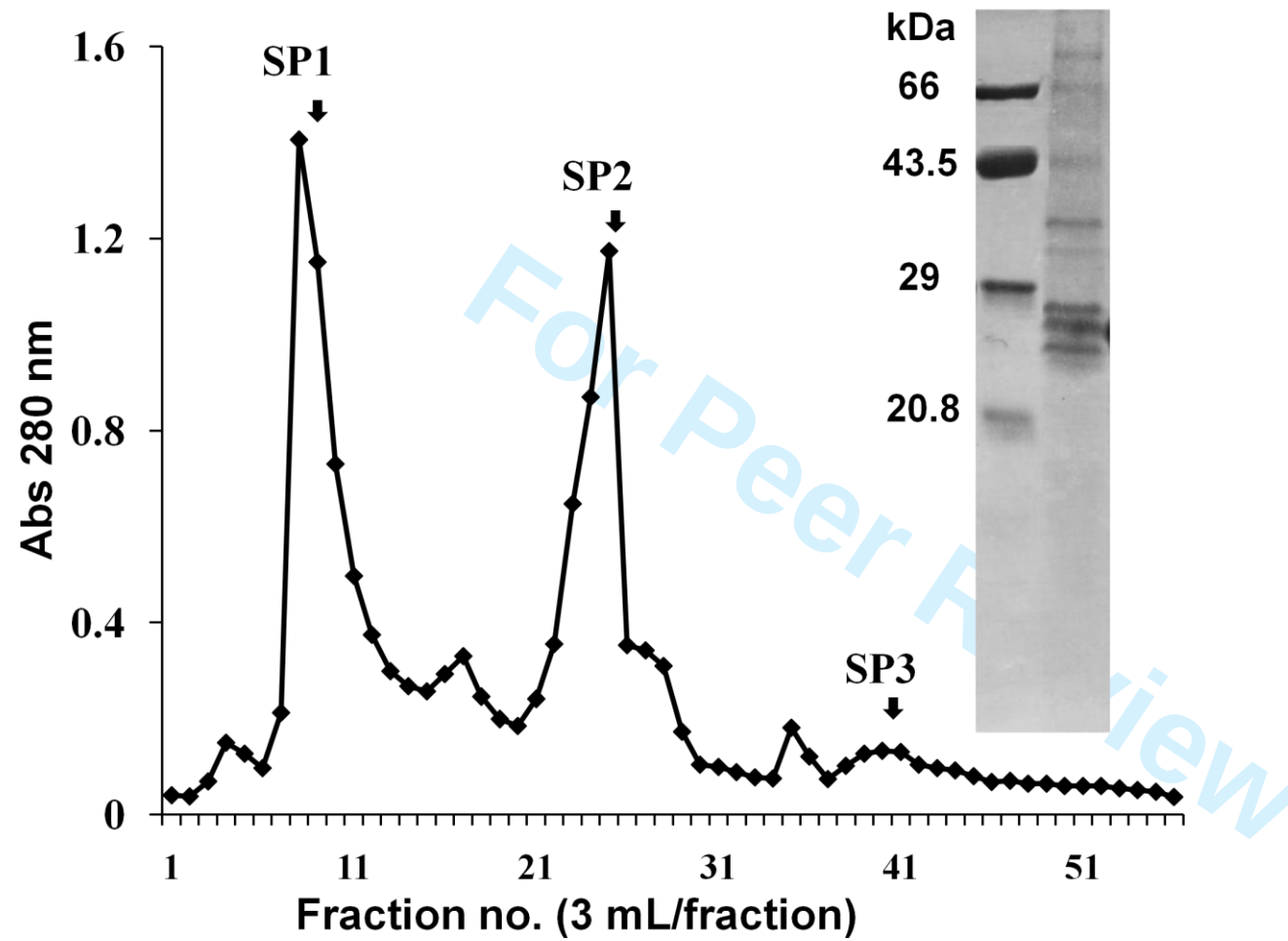
Sample	ELISA units ^a (A405 nm)	
	Normal serum (pooled) ^b	Allergic serum ^b
Sapodilla extract	0.197	0.419
SP-1 component	0.187	0.297
SP-2 component	0.183	0.258
SP-3 component	0.176	0.175

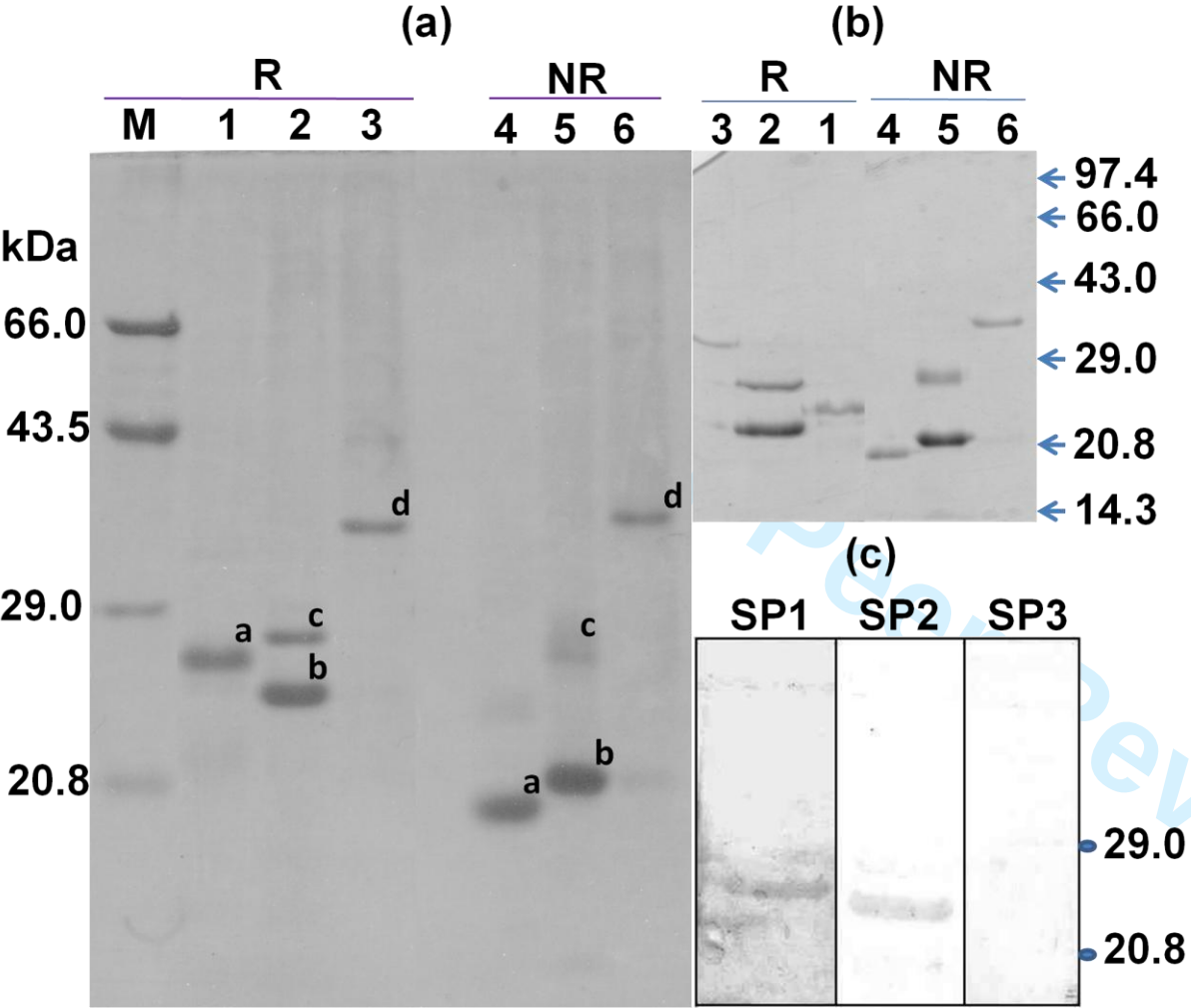
^amean of triplicate analysis^bserum dilution, 1:~~3~~ to 4

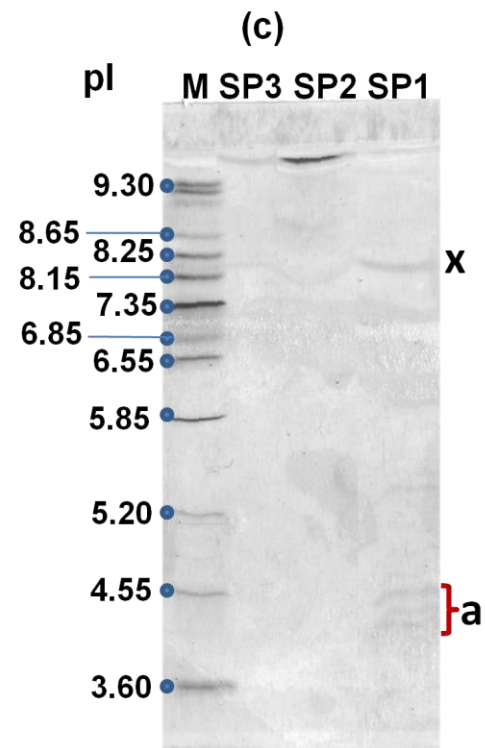
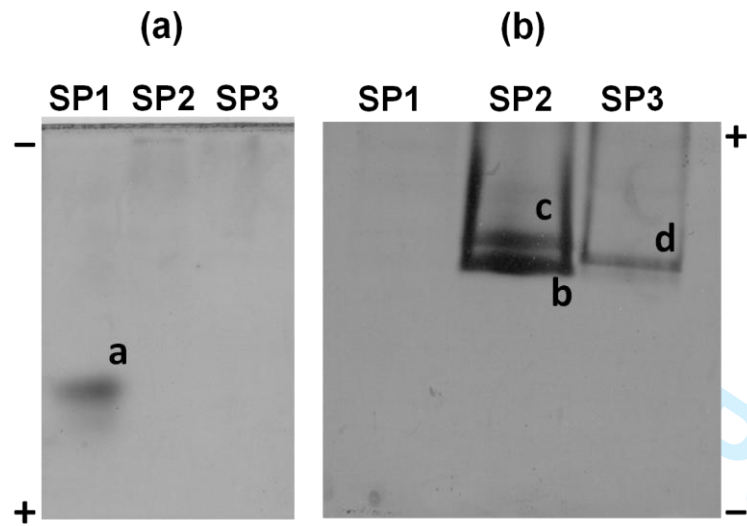
758 **Table 3. β -1,3-Glucanase activity of SP-Sepharose chromatographic components of**
759 **sapodilla extract**

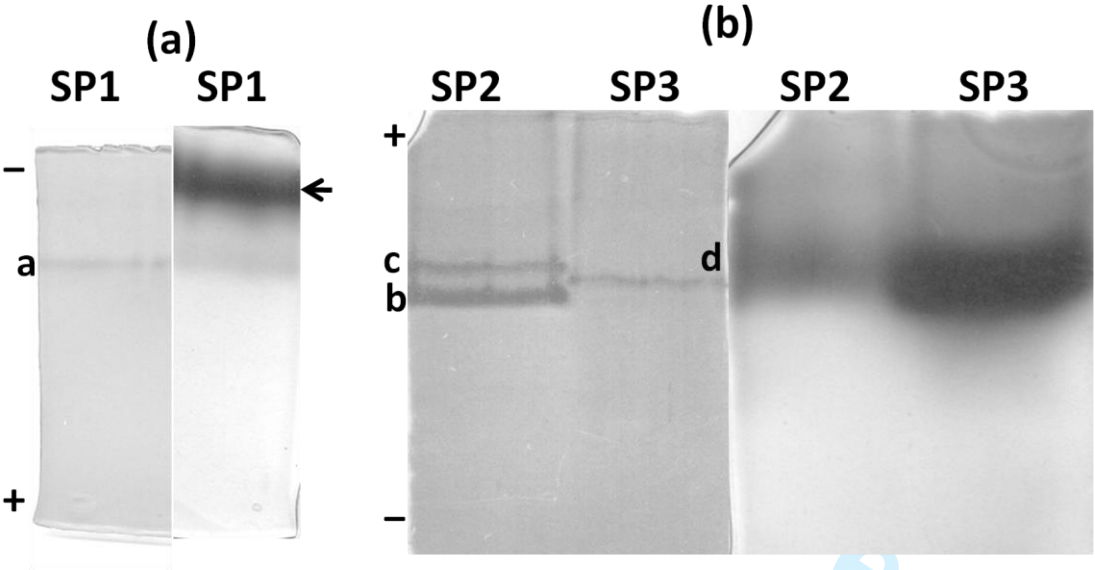
Protein sample	Enzymic activity (Units/mg protein) ^a
SP1	1.6
SP2	1.6
SP3	10.8
Laminarinase	26.0
BSA	0.0

760 ^aOne unit liberates 1 mg of reducing sugar (measured as glucose)
761 from laminarin per min at pH 5.0 at 37 °C.









Peer Review

1 CAGTGTACC TTCACCGTCTGGGCT GGTGCCAGTCCCGGT GGGGGCAAGCAGCTT
 1 A T F D V V N Q C T F T V W A G A S P G G G K Q L
 55 GACCAAGGCCAGACC TGGACCATCACGGTT GCCCCCGGCTCCACC AAGGCTCGCATTTGG GGCCGAACCGGCTGC
 26 D Q G Q T W T I T V A P G S T K A R I W G R T G C
 130 AACTTTGACGCCAAT GGTCAAGGAAAATGC CAGACCGGTGACTGC AATGGGCTTCTACAA TGCCAGGGCTATGGG
 51 N F D A N G Q G K C Q T G D C N G L L Q C Q G Y G
 205 AGCCCCCAACACC CTAGCCGAATTTTCA TTGAACCAACCCAAT AACTTGGACTACGTC GACATATCTCTGGTT
 76 S P P N T L A E F S L N Q P N N L D Y V D I S L V
 280 GACGGGTTCAATATC CCCATGGACTTCAGC CCCGCCGCCCGGC GTATGCAAGGATATC CGGTGCGCCACCGAC
 101 D G F N I P M D F S P A A A G V C K D I R C A T D
 355 ATAACCGCGCAGTGT CCGGCCGAGCTCCAG GCACCGGGTGGCTGC AACAAACCGTGTACT GTTTACAAGACCAAC
 126 I T A Q C P A E L Q A P G G C N N P C T V Y K T N
 430 GAGTATTGTTGTACT AATGGACAGGGGACC TGTGGACCCACCGCT CTGTCCAAATTCTTC AAGGACAGGTGCCCA
 151 E Y C C T N G Q G T C G P T A L S K F F K D R C P
 505 GATGCCTATAGCTAC CCTCAGGATGACCCA ACCAGTTTGTTCACT TGCCCCGCTGGCACC AATTACAAGTTTGTCT
 176 D A Y S Y P Q D D P T S L F T C P A G T N Y K V V
 580 TTCTGCCCTAATCTA GATGCATAG
 201 F C P N L D A *

ManzTLP1	-----ATFDVVNQCTFTVWAGASPGGGKQL	25
A2T1L9	MSFPKSLTILFPLIFT-LLIPFTNAATFAVLNKCTYTVWAAASPGGGMRL	49
O04708	MRFTTTLPILIPLLLS-LLFTSTHAATFDILNKCTYTVWAAASPGGGRRRL	49
E3SU11	MNFSKNLPLLVSLWAI-TFFAYTHAATFDIVNQCTYTVWAAASPGGGRRRL	49
K4DFX0	MHFLKFFPLFVFLYFYGQYYLYVTHAATFDITNRCTYPVWAGASPGGGRRRL	50
B9RC64	MSLVNIPILPTILSL-AL-LFITSTHAATFDVVNRCTYTVWAAASPGGGRRRL	48
Q5XUG9	MHFLKFFPLFVFLYFYGQYYLYVTHAATFDITNRCTYPVWAAASPGGGRRRL	50
Q5DJS5	MNFLKSFPFFAFLYFYGFVAVTHAATFDIVNKCTYTVWAAASPGGGRRRL	50
Q93XD1	MSITKNLPILLSTLSC-LYFTFSHGATFDIVNRCSYTVWAAASPGGGRRRL	49
B6ZHC0	MVYLALCSLLT---LA-LSLATTHAANFEIVNNCPYTVWAAASPGGGRRRL	46
O81926	TRSILTITLCS---LL-FLLTPSQANFEIVNNCPYTVWAAASPGGGRRRL	46
	. : *.* : *.***.***** :*	
ManzTLP1	DQGQTTWITIVAPGSTKARIWGRITGNCNFDANGQKQCQTGDCNGLLQCQGYG	75
A2T1L9	DPGQSWTVNVNPGTTQARIWGRITNCNFDANGNGQCQTGDCNGLLQCQGYG	99
O04708	DSGQSWTITVNPGTTNARIWGRITSCTFDANGRGKCETGDCNGLLECQGYG	99
E3SU11	DQGQSWNINVAPGTTQARIWGRITNCNFDANGRGQCETGDCNGLLECQGYG	99
K4DFX0	DSGQSWNINVNPGTQARIWGRITNCNFDGSGRGKCETGDCNGLLECQGYG	100
B9RC64	DQGQTTWTFDVAQGTNMARIWGRITNCNFDANGQGRCECTGDCNGLVCQGWG	98
Q5XUG9	DSGQSWNINVNPGTQARIWGRITNCNFDGSGRGKCETGDCNGLLECQGYG	100
Q5DJS5	DSGQSWNINVNPGTQARIWGRITNCNFDGSGRGNCETGDCNGLMLECQGYG	100
Q93XD1	DQGQSWNINVNPGTQARIWGRITNCNFDASGRGQCQTGDCNGLLECQGYG	99
B6ZHC0	DRGQSWNINVNPGTAMARIWGRITGNCNFDGSGRGRCQTGDCNGLVCQGWG	96
O81926	DRGQSWNINVNPGTAMARIWGRITGNCNFDGSGRGRCQTGDCNGLVCQGWG	96
	* ** : * . * : *****.*.*.*.*.*.* : * . * * * * *	
ManzTLP1	SPPNTLAEFSLNQPNNDYVDISLVDGFNIPMDFSPPAAAGVCKDIRCATD	125
A2T1L9	KAPNTLAEFALNQPNNDYVDISLVDGFNIPMDFSPTT-AVCKSLRCAAN	148
O04708	SPPNTLAEFALNQPNNDYVDISLVDGFNIPMDFS-----GCRGIQCSVD	144
E3SU11	RPPNTLAEFALNQPNNDYVDISLVDGFNIPLEFSPTT-NVCRRLVCNAP	148
K4DFX0	SPPNTLAEFALNQPNNDYVDISLVDGFNIPMEFSPTT-GGCRNLLCNAP	149
B9RC64	SPPNTLAEFALNQPNNDYVDISLVDGFNIPMDFSPTT-GACRGIRCAAD	147
Q5XUG9	SPPNTLAEFALNQPNNDYVDISLVDGFNIPMEFSPTT-GGCRNLLCNAP	149
Q5DJS5	KAPNTLAEFALNQPN-QDFVDISLVDGFNIPMEFSPTT-GGCRNLLCNAP	148
Q93XD1	SPPNTLAEFALNQPNNDYVDISLVDGFNIPMDFTGIN---CRSIVCSAP	146
B6ZHC0	VPPNTLAEFALNQPNQDFYDIDISLVDGFNIPMDFTGIN---GGCHKISCSAD	145
O81926	VPPNTLAEFALNQPNQDFYDIDISLVDGFNIPMDFTGIN---GGCHKISCSAD	145
	.***** * : . * : *** ***** : * * : * .	
ManzTLP1	ITAQCPAELQAPGGCNPCTVYKTNQYCCCTNGQGTGCGPTALSKFFKDRCP	175
A2T1L9	IVGECPAELQTPGGCNPCTVYKTNQYCCCTNGQGTGCGPTPLSKFFKDRCP	198
O04708	INGQCPSELKAPGGCNPCTVYKTNQYCCCTNGQGTGCGPTPLSKFFKDRCP	194
E3SU11	IVQCPSELKAPGGCNPCTVYKTNQYCCCTNGQGTGCGPTPLSKFFKDRCP	198
K4DFX0	INDQCPNELRAPGGCNPCTVYKTNQYCCCTNGQGTGCGPTPLSKFFKDRCP	199
B9RC64	INGQCPSELKAPGGCNPCTVYKTNQYCCCTNGQGTGCGPTPLSKFFKDRCP	197
Q5XUG9	INDQCPNELRAPGGCNPCTVYKTNQYCCCTNGQGTGCGPTPLSKFFKDRCP	199
Q5DJS5	INEQCPAELKAPGGCNPCTVYKTNQYCCCTNGQGTGCGPTPLSKFFKDRCP	198
Q93XD1	INEQCPAELKAPGGCNPCTVYKTNQYCCCTNGQGTGCGPTPLSKFFKDRCP	196
B6ZHC0	INGQCPSELKAPGGCNPCTVYKTNQYCCCTNGQGTGCGPTPLSKFFKDRCP	195
O81926	INGQCPSELKAPGGCNPCTVYKTNQYCCCTNGQGTGCGPTPLSKFFKDRCP	195
	* : ** * : : ***** * : : *** : * * : * * * * *	
ManzTLP1	DAYSYPQDDPTSLFTCPAGTNYKVVFPCPNL-----DA	207
A2T1L9	DAYSYPQDDPTSLFTCPAGTNYKVVFPCPNL-----DA	226
O04708	DAYSYPQDDPTSLFTCPAGTNYKVVFPCPNL-----DA	222
E3SU11	DAYSYPQDDPTSLFTCPAGTNYKVVFPCPNL-----DA	226
K4DFX0	DAYSYPQDDPTSLFTCPAGTNYKVVFPCPNL-----DA	227
B9RC64	DAYSYPQDDPTSLFTCPAGTNYKVVFPCPNL-----DA	225
Q5XUG9	DAYSYPQDDPTSLFTCPAGTNYKVVFPCPNL-----DA	227
Q5DJS5	DAYSYPQDDPTSLFTCPAGTNYKVVFPCPNL-----DA	226
Q93XD1	TSYSYPQDDPTSLFTCPAGTNYKVVFPCPNL-----DA	224
B6ZHC0	DSYSYPQDDPTSLFTCPAGTNYKVVFPCPNL-----DA	240
O81926	DAYSYPQDDPTSLFTCPAGTNYKVVFPCPNL-----DA	239
	:***** : * ***** : * : * : .***	

ManzTLP1	-----ATFDVVNQCTFTVWAGASP-GGG	22
Cryj3	MATVSDIALLLVAGLVAISLHMQEAGAVKFELKNQCEYTVWAAAGLP-GGG	49
Cryj3.3	RAIGV--WIALVAA-LSVFLHGMVEVRAATFDITNQCPYTVWAAASP-GGG	46
Cryj3.4	MARAILWVLLTVMA-VSLLH-AGVEGVNFDIENQCPYTVWAAAGTPFGGG	48
Cryj3.5	MASLR-LATL--AM-MVL-FGSCRAGATVFTLVNKCSTVWPGTLS-GSG	44
Cryj3.6	GSIPF--WIALIAS-FSVFLQGVKVKAPTFEITNKCPYTVWAAAFP-GGG	46
Cryj3.7	MATVSDIALLLVAGLVAISLHMQEAGAVKFELKNQCEYTVWAAAGLP-GGG	49
Cryj3.8	MAKVSDIALLLVAG-MAISLYIQETGAVKFEDIKNQCGYTVWAAAGLP-GGG	48
Cupa3	-----VKFDIKNQCGYTVWAAAGLP-GGG	22
Cups3.0101	MARVSELALLLVAT-SAISLHMQEAGAVKFEDIKNQCRYTVWAAAGLP-GGG	48
Cups3.0102	MARVSELALLLVAT-LAISLHMQEAGAVKFEDIKNQCGYTVWAAAGLP-GGG	48
Juna3.0101	MARVSELAFLLAAT-LAISLHMQEAGVVKFDIKNQCGYTVWAAAGLP-GGG	48
Junr3.1	MARVSELALLLVAT-LAISLHMQEAGAVKFEDIKNQCGYTVWAAAGLP-GGG	48
Junr3.2	MARVSELALLLVAT-LAISLHMQEAGAVKFDIRNQCGYTVWAAAGLP-GGG	48
Mald2	MMKSQ-VASLLGLT-LAI-LFFSGAHAAKITFTNNCPNTVWPGTLT-GDQ	46
Olee13	MNFSKNLPLL-VSL-WAITTF-AYTHAATFDIVNQCTYTVWAAASP-GGG	46
Pruav2.0101	MMKTL-VVVL-SLS-LTI-LSFGGAHAATISFKNNCPYMWVPGTLT-SDQ	45
Prup2.0201	MMKSQ-AALLGLTT-LAI-LFFSGAHAAKITFTNKCSYTVWPGTLT-GDQ	46
Prup2.0301	MMKTL-VAVL-SLS-LTL-LSFGGAHAATMSFKNNCPYTVWPASFG-NP-	44
Pyrpy2	MKFEA-LIGL--VL-VFL-SEHAGVYSAKFTFTNKCPNTVWPGTLT-GGG	44
	: . *:* **.. .	
ManzTLP1	-----KQLDQGQTTWTITVAPGSTKARIWGRTGCNFDANGQGGKQCTGD	64
Cryj3	-----QQLDQGQTTWPVEVPAGTKGAREFWGRTGCSFDASGRGTCKTGD	91
Cryj3.3	-----QQLDQGQTTWTIQVAAGTTQARIWARTGCSFDGSGRGTCQCTGD	88
Cryj3.4	-----IELKRGQSWRVNVP-GARGRFWGRTGCSFDGNGRGRCNTGD	89
Cryj3.5	SSVLGEGGFTLAPGQSVPLTASS-RWSGRFWGRTDCSFDASGKGSCITGD	93
Cryj3.6	-----KQLAQGQSWSVQPDAGTSTGRIWGRTDCSFDGSGRGTCQSGD	88
Cryj3.7	-----QQLDQGQTTWPVEVPAGTKGAREFWGRTGCSFDASGRGTCKTGD	91
Cryj3.8	-----QQLTQGQTTWTVNLAAGTQSAREFWGRTGCSFDASGKGTCQCTGD	90
Cupa3	-----KEFDQGQTTWTVNLAAGTASAREFWGRTGCTFDASGKGSCRS GD	64
Cups3.0101	-----KRLDQGQTTWTVNLAAGTASAREFWGRTGCTFDASGKGSCRS GD	90
Cups3.0102	-----KRLDQGQTTWTVNLAAGTASAREFWGRTGCTFDASGKGSCRS GD	90
Juna3.0101	-----KRLDQGQTTWTVNLAAGTASAREFWGRTGCTFDASGKGSCQCTGD	90
Junr3.1	-----KRLDQGQTTWTLNLAAGTASAREFWGRTGCTFDASGKGSCKTGD	90
Junr3.2	-----KRLDQGQTTWTLNLAAGTASAREFWGRTGCTFDASGKGSCKTGD	90
Mald2	KPQLSLTGFELASKASRSVDAPS-PWSGRFWGRTRCSTDAAGKFTCETAD	95
Olee13	-----RRLDQGQSWNINVAPGTTQARIWGRTNCNFDANGRGQCETGD	88
Pruav2.0101	KPQLSTTGFE LASQAS FQ L D T P V - P W N G R F W A R T G C S T D A S G K F V C A T A D	94
Prup2.0201	KPQLSLTGFELATGISNSVDAPS-PWSGRFFGRTRCSTDASGKFTCATAD	95
Prup2.0301	--QLSTTGFE LASQAS FQ L D T P V - P W S G R F W A R T R C S T D A S G K F V C E T A D	91
Pyrpy2	GPQLLSTGFELASGASTSLTVQA-PWSGRFWGRSHCSIDSSGKFKCSTGD	93
	: : : .*::*: *. *. *: * :.*	
ManzTLP1	CN-GLLQCQGYGSP-PNTLAEFSLNQPNNLDYVDISLVDGFNIPMDFSPA	112
Cryj3	CN-SQLSCQVSGGV-PTTLAEYTLNGDGNKDFYDVSLVDGFNVPLSINPT	139
Cryj3.3	CN-GMLSCQGYGQV-PATLAEYALNQYMNLD FYD I S L V D G F N V P L S M T P T	136
Cryj3.4	CG-GLLNCQGS GG V - P S T L L E Y A L N Q Y Q N L D F Y D I S L V D G F N L R M T V V L S	137
Cryj3.5	CG-NVLNCAQAGGTPPVSLAEFTL---GDKDFYDVSLVDGYNVPLSIAAV	139
Cryj3.6	CN-GTLNCQGDASA-PVTLVEYTHNPSMNLD FYD I S L L D G F N L P L S I T P T	136
Cryj3.7	CN-SQLSCQVSGGV-PTTLAEYTLNGDGNKDFYDVSLVDGFNVPLSINPT	139
Cryj3.8	CG-GQLSCTVSGAV-PATLAEY T Q --- S D Q D Y Y D V S L V D G F N I P L S I N P T	135
Cupa3	CG-GQLSCTVSGAV-PATLAEY T Q --- S D Q D Y Y D V S L V D G F N I P L A I N P T	109
Cups3.0101	CG-GQLSCTVSGAV-PATLAEY T Q --- S D Q D Y Y D V S L V D G F N I P L A I N P T	135
Cups3.0102	CG-GQLSCTVSGAV-PATLAEY T Q --- S D K D Y Y D V S L V D G F N I P L A I N P T	135
Juna3.0101	CG-GQLSCTVSGAV-PATLAEY T Q --- S D Q D Y Y D V S L V D G F N I P L A I N P T	135
Junr3.1	CG-GQLSCTVSGAV-PATLAEY T Q --- S D Q D Y Y D V S L V D G F N I P L A I N P T	135
Junr3.2	CG-GQLSCTVSGAV-PATLAEY T Q --- S D Q D Y Y D V S L V D G F N I P L A I N P T	135
Mald2	CGSGQVACNGAGAVPPATLVEITIAANGQDYDVSLVDGFNLPM SVAPQ	145
Olee13	CN-GLLECQGYGRP-PNTLAEFALNQPNNLDFVDISNVDGFNIPLEFSPT	136
Pruav2.0101	CASGQVMCNGNGAIPPATLAEFNI PAGGGQDFYDVSLVDGFNLPM SVTPQ	144
Prup2.0201	CGSGQVSCNGNGAIPPATLVEITIAENGQDFYDVSLVDGFNLPM SVAPQ	145
Prup2.0301	CDSGQLMCNGKTGIPPATLAEFTIAAGGGQDFYDVSLVDGFNLPM SVTPQ	141
Pyrpy2	CGSGQISCNGAGASPPASLVELTLATNGGQDFYDVSLVDGFNLPIKLAPR	143
	* . : * * : * * . *: *: * :*: *: : .	
ManzTLP1	A-AGVCKD-IRCATDITAQCPAELQAPG-----GCNNPCTVYKTNEYCC	154
Cryj3	N--SQCFA-PACKADVNAACPAQLKVDG-----GCNSACTVFQTDEYCC	180
Cryj3.3	STDPNCKGRIACLSDINSQCPSDLKVTG-----GCKSACARYNTPEYCC	180
Cryj3.4	N--TNCKR-IACNSDINSKCPGELKVVD-----GCRSACAAFNTPAYCC	178
Cryj3.5	GGTGDCRT-AGCVSDLRTSCPAELSVTSN-GQVIACKSACAAFSTPEYCC	187
Cryj3.6	STNPNCKGIIITCLSDINSQCPNELKVSG-----GCLSACVKYNTDDHCC	180
Cryj3.7	N--SQCFA-PACKADVNAACPAQLKVDG-----GCNSACTVFQTDEYCC	180
Cryj3.8	N--AQCTA-PACKADVNAVCPAELKVDG-----GCKSACAAFQTDQYCC	176
Cupa3	N--TKCTA-PACKADINAVCPSELKVDG-----GCNSACNVLQTDQYCC	150
Cups3.0101	N--TKCTA-PACKADINAVCPSELKVDG-----GCNSACNVLQTDQYCC	176
Cups3.0102	N--TKCTA-PACKADINAVCPSELKVDG-----GCNSACNVLQTDQYCC	176
Juna3.0101	N--AQCTA-PACKADINAVCPSELKVDG-----GCNSACNVFKTDQYCC	176
Junr3.1	N--AQCTA-PACKADINAVCPSELKVEG-----GCNSACNVFQTDQYCC	176
Junr3.2	N--AQCTA-PACKADINAVCPSELKVEG-----GCNSACNVFQTDQYCC	176
Mald2	GGTGECKP-SSCPANVNKVC PAPLQVKAADGSVISCKSACLAFGDSKYCC	194
Olee13	T--NVCRR-LVCNAPIVQQCPSELRTPG-----GCNNPCTVFNTNEYCC	177
Pruav2.0101	GGTGDCKT-ASCPANVNAVCPSELQKKGSDGSVVACLSACVKFGTPQYCC	193
Prup2.0201	GGTGECKA-STCPADINKVCPAELQVKGS DGSV IACKSACLALNQPYCC	194
Prup2.0301	GGTGCTKM-GSCANVNLCVCPSELQKIGSDGSVVACLSACVKFGEPPQYCC	190
Pyrpy2	GGSGDCNS-TSCAANINTVCPAELSDKGS DGSV I GCKSACLALNQPYCC	192
	* * : : ** * . * . * : **	
ManzTLP1	TNGQ---GTCGPTALSKFFKDRCPDAYSYPQDDPTS LFTCPAG-TNYKV V	200
Cryj3	RGTH--VDNCSPSSYSMIFKNQCPQAYSYAKDDSSSTFTCPSGTTDYSIV	228
Cryj3.3	TGAS--ENTCGPTDYSKFFKGQCPQAYSYAKDDATSTFTCPSG-TNYKV V	227
Cryj3.4	TGSF--LDNCPPTGYSQFFKRECPPLAYSYAKDDPTSTFTCPGG-SDYKII	225
Cryj3.5	TGDHGSPTCSPSKYSQVFKSACPTAYSAYYDDATSTFTCS-N-ADYTIT	235
Cryj3.6	RGA----NCNQTLYSRFFKEQCPQAYSYAKDDVTSFTCPSG-TDYKIV	224
Cryj3.7	RGTH--VDNCSPSSYSMIFKNQCPQAYSYAKDDSSSTFTCPSGTTDYSIV	228
Cryj3.8	TGTY--ANSCPATNYSMIFKNQCPQAYSYPKDD-TATFACPSG-TDYSIV	222
Cupa3	RNAY--VNNCPATNYSKIFKNQCPQAYSYAKDD-TATFACASG-TDYSIV	196
Cups3.0101	RNAY--VDNCPATNYSKIFKNQCPQAYSYAKDD-TATFACASG-TDYSIV	222
Cups3.0102	RNAY--VDNCPATNYSKIFKNQCPQAYSYAKDD-TATFACASG-TDYSIV	222
Juna3.0101	RNAY--VDNCPATNYSKIFKNQCPQAYSYAKDD-TATFACASG-TDYSIV	222
Junr3.1	RNAY--VDNCPATNYSKIFKNQCPQAYSYAKDD-TATFACASG-TDYSIV	222
Junr3.2	RNAY--VDNCPATNYSKIFKNQCPQAYSYAKDD-TATFACASG-TDYSIV	222
Mald2	TPPNNTPETCPPTIEYSEIFEKQCPQAYSAYDDKNSTFTCSGG-PDYVIT	243
Olee13	TNGP---GSCGPTPLSRFFKERCPDAYSYPQDDPTS LFTCPAG-TNYRVV	223
Pruav2.0101	TPPQNTPETCPPTNYSEIFHNACPDAYSAYDDKRGFTFTCNGG-PNYAIT	242
Prup2.0201	TPPNDKPETCPPTDYSKLFKTQCPQAYSAYDDKSSTFTCSGR-PDYLIT	243
Prup2.0301	TPPQETKEKCPPTNYSQIFHEQCPDAYSAYFDDNKGLFTCSGG-PNYLIT	239
Pyrpy2	TGAYGTPDTCPPPTDFSKVFNQCPQAYSAYDDKSSTFTCFGG-PNYEIT	241
	. * : * *. * ** ** . *: * : * :	
ManzTLP1	FCPNLD-----A	207
Cryj3	FCP-----	231
Cryj3.3	FCG-----	230
Cryj3.4	FCGNGDSQSNSSTSTSSGSLYAVE	248
Cryj3.5	FCPS-----S	240
Cryj3.6	FCGKSDHIFAL--YLYITINNEE	245
Cryj3.7	FCP-----	231
Cryj3.8	FCP-----	225
Cupa3	FCP-----	199
Cups3.0101	FCP-----	225
Cups3.0102	FCP-----	225
Juna3.0101	FCP-----	225
Junr3.1	FCP-----	225
Junr3.2	FCP-----	225
Mald2	FCP-----	246
Olee13	FCP-----	226
Pruav2.0101	FCP-----	245
Prup2.0201	FCP-----	246
Prup2.0301	FCP-----	242
Pyrpy2	FCP-----	244
	**	

*Allergology International***Original article (Supplementary information)****Characterization and gene cloning of an acidic thaumatin-like protein (TLP 1), an allergen from sapodilla fruit (*Manilkara zapota*)[#]**

Hassan G. Ashok Kumar, Venkatesh L. Hegde[†], Santoshkumar M. Shetty[‡], Yeldur P. Venkatesh*

Dept of Biochemistry and Nutrition, CSIR–Central Food Technological Research Institute (a Council of Scientific and Industrial Research laboratory), Mysore–570020, Karnataka, India

Running title: Acidic thaumatin-like protein is a sapodilla allergen

Table S1. Primers designed for use in cloning sapodilla acidic TLP gene

Primer name	Primer sequence (5'→3')	Region of plant TLP
SET-1		
<i>Forward (from 8th aa residue)</i>		
SAP1	CAGTGTACCTTCACCGTTTGGGC	N-terminus [†]
*SAP2	CAGTGTACCTTCACCGTCTGGGC	N-terminus [†]
<i>Reverse</i>		
SAP3	ACCAACTACCG(AT)GTCGT(GCT)TTCTGTCC	C-terminus
*SAP4	TGTAACAACCCCTGTACCGTG	Conserved region A [‡]
SET-2		
<i>Forward</i>		
*SAP5	CGACATATCTCTGGTTGACGG	Conserved region B [‡]
SAP6	CTGTAACAACCCCTGTACCGTG	Conserved region A [‡]
<i>Reverse</i>		
SAP7	CGTTTTGCCCTTGAAAGTTG	C-terminus
SAP8	GCTATTACCTTCTGCCCTTGA	C-terminus
*SAP9	GGTTGTCTTCTGCCCTTGAAATTG	C-terminus
SET-3		
<i>Forward from 1st aa residue</i>		
SAP10	AYGCCGCCACCTTCGACG	N-terminus [†]
<i>Reverse</i>		
SAP9	GGTTGTCTTCTGCCCTTGAAATTG	C-terminus

All primers were used in different combinations at various temperatures as described in 'Materials and methods'

*indicates the primers which were functional in PCR

[†]Refers to the N-terminal sequence of sapodilla acidic TLP (UniProt Accession No. B3EWX8; TLP 1b)

[‡]Conserved region A represents –CNNPCTV– in the plant TLP sequences

[‡]Conserved region B represents –DISLVD– in the plant TLP sequences

Table S2. Homology of N-terminal sequence (25 residues)^a of sapodilla acidic TLP as analyzed by BLAST

ID	Description	Total score	Query coverage
P07052.1	Thaumatococin-like protein E2 (<i>Nicotiana glauca</i>)	70.2	100%
ACZ57583.1	Allergenic thaumatococin (<i>Olea europaea</i>)	66.0	100%
XP_002509748.1	Osmotin precursor, putative (<i>Ricinus communis</i>)	63.8	100%
P13046.1	Thaumatococin-like protein E22 (<i>Nicotiana glauca</i>)	61.3	100%
BAC15614.1	Thaumatococin-like protein (<i>Cryptomeria japonica</i>)	61.3	100%
AAK59278.1	Thaumatococin-like protein (<i>Sambucus nigra</i>)	58.7	100%
ACU31849.1	Thaumatococin-like protein (<i>Nepenthes singalana</i>)	58.7	100%
ADB97926.1	Thaumatococin-like protein L2 (<i>Pinus monticola</i>)	58.3	100%
XP_002282910.1	PREDICTED: pathogenesis-related protein R major form (<i>Vitis vinifera</i>)	57.5	100%
XP_002874708.1	ATOSM34 (<i>Arabidopsis lyrata</i> subsp. <i>Lyrata</i>)	56.6	100%
1Z3Q	Allergenic and antifungal banana fruit thaumatococin-like protein	55.4	100%
ACV83333.1	Pathogenesis-related protein 5 (<i>Panax ginseng</i>)	54.9	100%
AAB82777.1	Ripening-associated protein (<i>Musa acuminata</i> AAA Group)	54.1	100%

^athe sequence ATFDVVNQCTFTVWAGASPGGGKQL (UniProt Accn. B3EWX8) was used as the query for this analysis

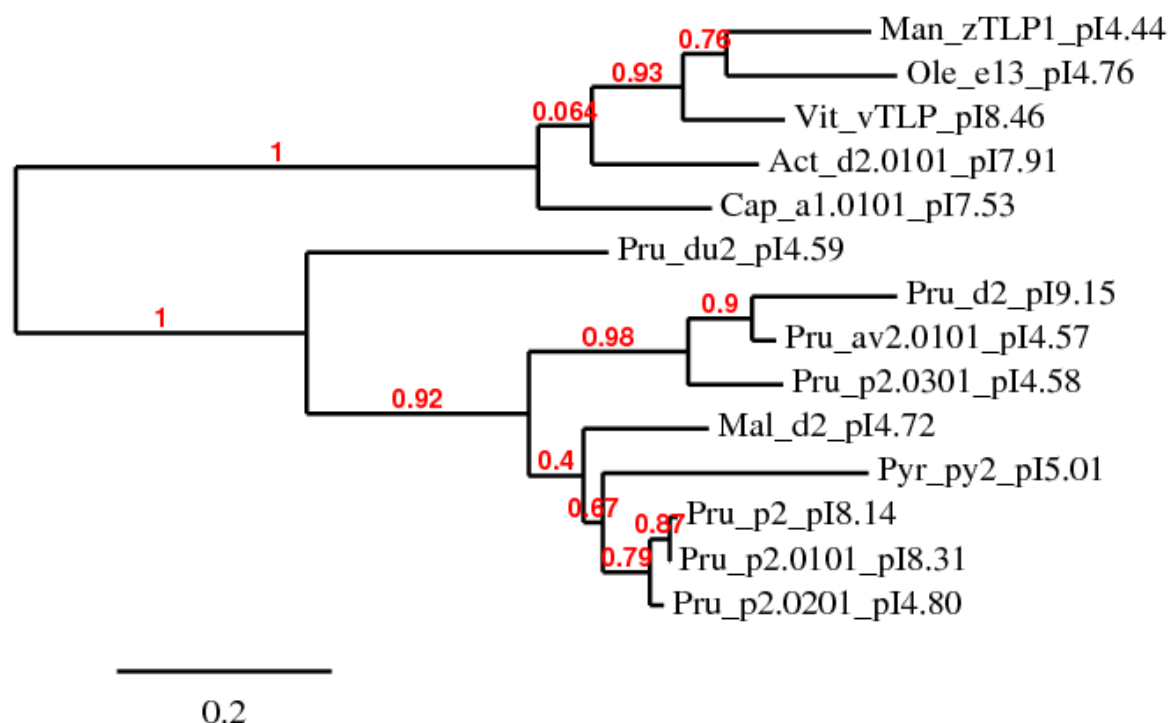
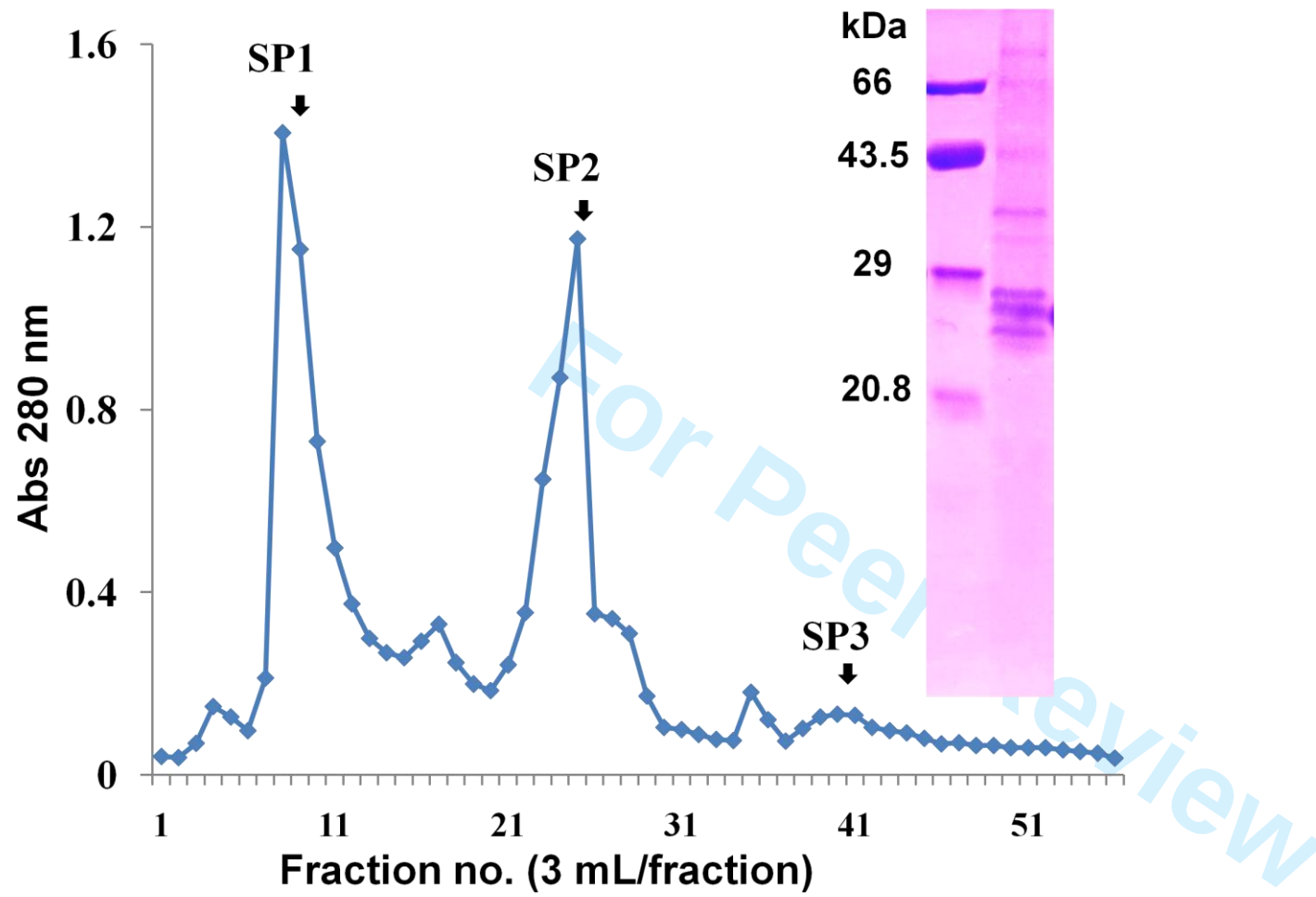
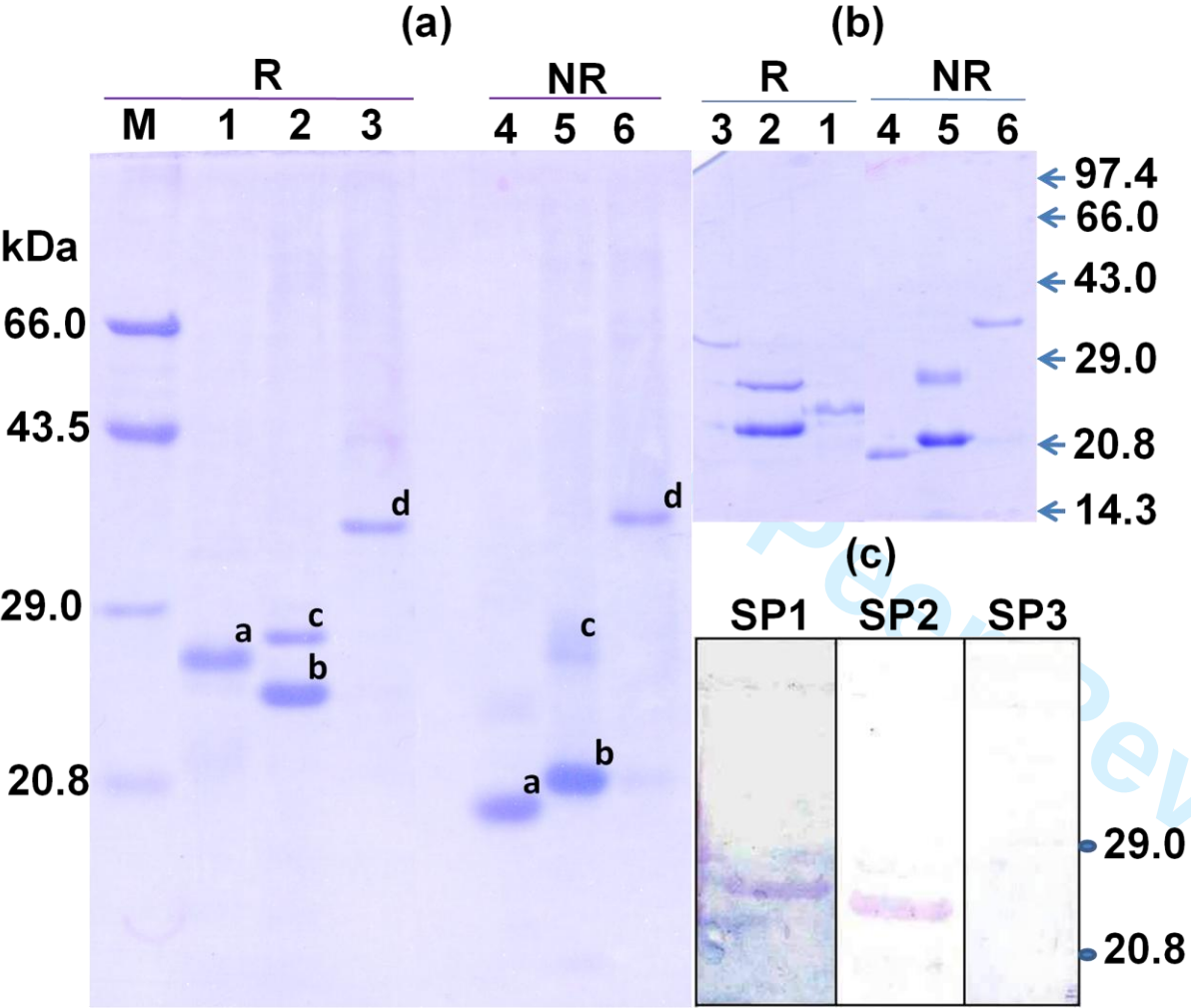
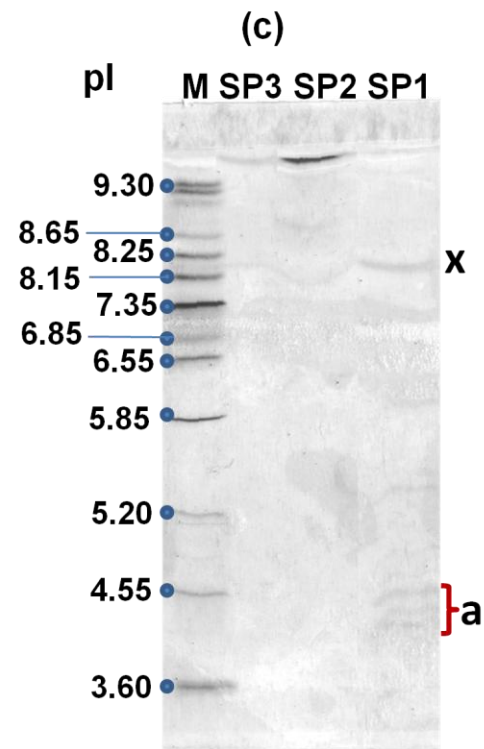
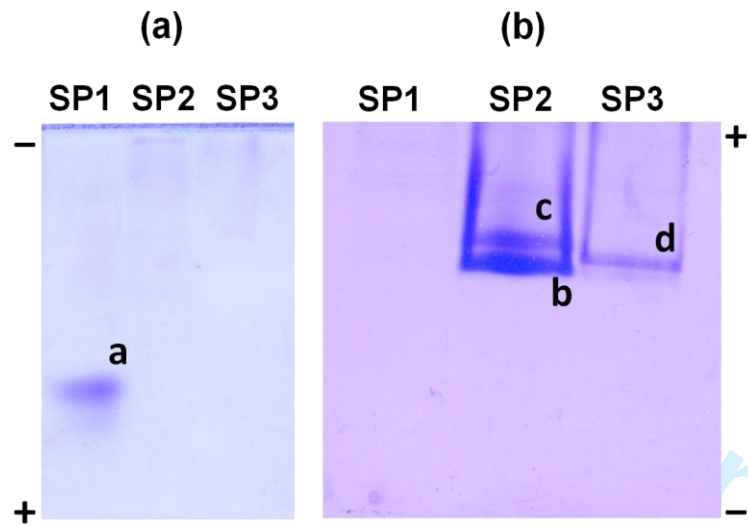
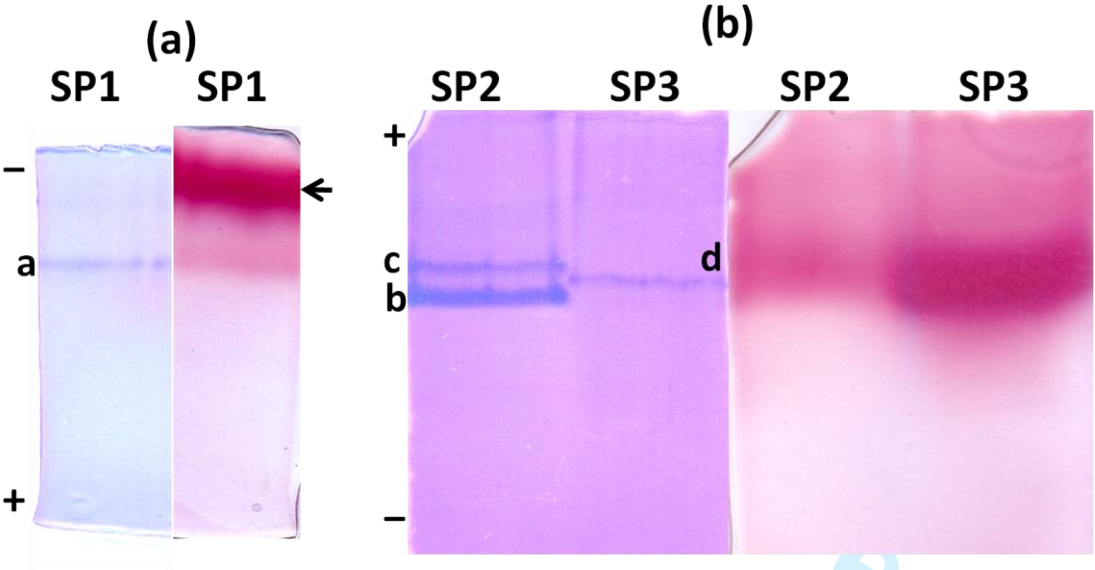


Fig. S1. Phylogenetic analysis of sapodilla acidic TLP with known allergenic TLPs from foods. In this phylogenetic tree, the branch length is proportional to the number of substitutions per site. Allergens were retrieved from Allergome database (www.allergome.org). Each TLP is denoted by its Allergome name and its isoelectric point (pI); the protein sequences used here represent the truncated form of the mature polypeptides as described in ‘**Materials and methods**’. Allergens used for analysis: kiwi (*Actinidia deliciosa*, Act d 2); bell pepper (*Capsicum annuum*, Cap a 1); plum, prune (*Prunus domestica*, Pru d 2); peach (*Prunus persica*, Pru p 2); cherry (*Prunus avium*, Pru av 2); almond (*Prunus dulcis*, Pru du 2); grape (*Vitis vinifera*, Vit v TLP); apple (*Malus domestica*, Mal d 2); olive (*Olea europaea*, Ole e 13); pear (*Pyrus pyrifolia*, Pyr py 2); sapodilla (*Manilkara zapota*, Man z TLP1).









Peer Review

Supplementary data related to

Ashok Kumar, H.G., Venkatesh, Y.P., *In silico* analyses of structural and allergenicity features of sapodilla (*Manilkara zapota*) acidic thaumatin-like protein in comparison with allergenic plant TLPs. Mol. Immunol. (2013), <http://dx.doi.org/10.1016/j.molimm.2013.08.010>

Figure S1. ClustalX alignment of food allergenic TLPs (ingestant and inhalant) displaying predicted IgE epitopes. Allergens were retrieved from Allergome database; sequences were retrieved from NCBI database. The sequence IDs correspond to the description given in Table 1. The predicted IgE epitope stretches in each of the sequences are highlighted in red (epitope 1), green (epitope 2), blue (epitope 3) and pink (epitope 4) in the order of appearance from the N-terminus. The accession numbers of sapodilla acidic TLP and thaumatins are highlighted in cyan. Putative TLPs with accession numbers highlighted in green have no predicted IgE epitopes. Cysteine residues (C) are highlighted in yellow. Symbols: (*), complete identity; (:), strongly similar; (.), weakly similar; no symbol, different. The bar graph shown at the bottom denotes conservancy profile. **Inset: ClustalX alignment of grape TLPs displaying predicted IgE epitopes.** The sequence IDs correspond to the description given in Table 1. The predicted IgE epitopes are highlighted in the same colors as given above.

```

: . * * : * . . * . . . . * : * * * * * : : : . * : * : * :
ACN97419.1 AKITFTTNNCPYTVVPGTTLTGDKPQLSLTGFEELATKASQSVDA P S P - W S G R F W A R T R C S T N A A G K F T C E T A D C S G S Q V A C N G A G A V P P A T L V E I T I A A N - G G Q D F -- Y D V S L V D G F N L P M S V A P Q G G T G -- D C K P S S C P A N V N A A C P A Q L 144
ACN97416.1 AKITFTTNNCPYTVVPGTTLTGDKPQLSLTGFEELATKASQSVDA L S P - W S G R F W A R T R C S T N A A G K F T C E T A D C S G S Q V A C N G A G A V P P A T L V E I T I A A N - G G Q D F -- Y D V S L V D G F N L P M S V A P Q G G T G -- D C K P S S C P A N V N A A C P A Q L 144
ACN97418.1 AKITFTTNNCPYTVVPGTTLTGDKPQLSLTGFEELATKASQSVDA P S P - W S G R F W A R T R C S T N A A G K F T C E T A D C S G S Q V A C N G A G A V P P A T L V E I T I A A N - G G Q D F -- Y D V S L V D G F N L P M S V A P Q G G T G -- D C K P S S C P A N V N A A C P A Q L 144
AEO36939.1 AKITFTTNNCPYTVVPGTTLTGDKPQLSLTGFEELATKASQSVDA P F P P - W S G R F W G R T R C S T N A A G K F T C E T A D C S G S Q V A C N G A G A V P P A T L V E I T I A A N - G G Q D F -- Y D V S L V D G F N L P M S V A P Q G G T G -- D C K P S S C P A N V N A A C P A Q L 144
ACI04550.1 AKITFTTNNCPYTVVPGTTLTGDKPQLSLTGFEELATKASQSVDA P F P P - W S G R F W G R T R C S T N A A G K F T C E T A D C S G S Q V A C N G A G A V P P A T L V E I T I A A N - G G Q D F -- Y D V S L V D G F N L P M S V V P Q G G T G -- D C K P S S C P A N V N A A C P A Q L 144
ACN97417.1 AKITFTTNNCPYTVVPGTTLTGDKPQLSLTGFEELATKASQSVDA P S P - W S G R F W G R T R C S T N A A G K F T C E T A D C S G S Q V A C N G A G A V P P A T L V E I T I A A N - G G Q D F -- Y D V S L V D G F N L P M S V A P Q G G T G -- D C K P S S C P A N V N A A C P A Q L 144
AAX19851.1 AKITFTTNNCPYTVVPGTTLTGDKPQLSLTGFEELATKASQSVDA P S P - W S G R F W G R T R C S T N A A G K F T C E T A D C S G S Q V A C N G A G A V P P A T L V E I T I A A N - G G Q D Y -- Y D V S L V D G F N L P M S V A P Q G G T G -- E C K P S S C P A N V N M A C P A Q L 144
AAX19848.1 AKITFTTNNCPYTVVPGTTLTGDKPQLSLTGFEELATKASQSVDA P S P - W S G R F W G R T R C S T N A A G K F T C E T A D C S G S Q V A C N G A G A V P P A T L V E I T I A A N - G G Q D Y -- Y D V S L V D G F N L P M S V A P Q G G T G -- E C K P S S C P A N V N M A C P A Q L 144
ABC47923.1 AKITFTTNNCPYTVVPGTTLTGDKPQLSLTGFEELATKASQSVDA P S P - W S G R F W G R T R C S T N A A G K F T C E T A D C S G S Q V A C N G A G A V P P A T L V E I T I A A N - G G Q D Y -- Y D V S L V D G F N L P M S V A P Q G G T G -- E C K P S S C P A N V N M A C P A Q L 144
JC7201 AKITFTTNNCPYTVVPGTTLTGDKPQLSLTGFEELATKASRSVDAP S P - W S G R F W G R T R C S T N A A G K F T C E T A D C S G S Q V A C N G A G A V P P A T L V E I T I A A N - G G Q D Y -- Y D V S L V D G F N L P M S V A P Q G G T G -- E C K P S S C P A N V N K V C P A P L 144
AEV57471.1 AKITFTTNNCPYTVVPGTTLTGDKPQLSLTGFEELATKASRSVDAP S P - W S G R F W G R T R C S T N A A G K F T C E T A D C S G S Q V A C N G A G A V P P A T L V E I T I A A N - G G Q D F -- Y D V S L V D G F N L P M A V A P Q G G T G -- E C K A S T C P A D V N A V C P A E L 144
P83332.1 AKITFTTNNCPYTVVPGTTLTGDKPQLSLTGFEELATKASRSVDAP S P - W S G R F W G R T R C S T N A A G K F T C E T A D C S G S Q V A C N G A G A V P P A T L V E I T I A A N - G G Q D F -- Y D V S L V D G F N L P M S V A P Q G G T G -- K C K A S T C P A D I N K V C P A P L 144
CAB62167.1 AKITFTTNNCPYTVVPGTTLTGDKPQLSLTGFEELATKASRSVDAP S P - W S G R F W G R T R C S T N A A G K F T C E T A D C S G S Q V A C N G A G A V P P A T L V E I T I A A N - R G M D F -- Y D V S L V D G Y N L P V S V A T R G G T G -- D C K A T S C R A N V N A V C P A E L 144
P50694.1 ATISFKNNCPYTVVPGTTLTGDKPQLSLTGFEELATKASRSVDAP S P - W S G R F W G R T R C S T N A A G K F T C E T A D C S G S Q V A C N G A G A V P P A T L V E I T I A A N - G G Q D F -- Y D V S L V D G F N L P M S V T P Q G G T G -- D C K T A S C P A N V N A V C P S E L 144
ADM22305.1 AKMSFKNNCPYTVVPGTTLTGDKPQLSLTGFEELATKASRSVDAP S P - W S G R F W G R T R C S T N A A G K F T C E T A D C S G S Q V A C N G A G A V P P A T L V E I T I A A N - G G Q D F -- Y D V S L V D G F N L P M S V N P Q G G T G -- D C K T A S C P K N V N K V C P R E L 144
ACE80974.1 ATISFKNNCPYTVVPGTTLTGDKPQLSLTGFEELATKASRSVDAP S P - W S G R F W G R T R C S T N A A G K F T C E T A D C S G S Q V A C N G A G A V P P A T L V E I T I A A N - G G Q D Y -- Y D V S L V D G F N L R M S V T P Q G G T G -- T C K T G S C A A N V N A V C P S E L 140
P83335.1 ATMSFKNNCPYTVVPGTTLTGDKPQLSLTGFEELATKASRSVDAP S P - W S G R F W G R T R C S T N A A G K F T C E T A D C S G S Q V A C N G A G A V P P A T L V E I T I A A N - G G Q D F -- Y D V S L V D G F N L P M S V T P Q G G T G -- T C K M G S C A A N V N L V C P S E L 141
O80327.1 AKFTFTTNNCPYTVVPGTTLTGDKPQLSLTGFEELATKASRSVDAP S P - W S G R F W G R T R C S T N A A G K F T C E T A D C S G S Q V A C N G A G A V P P A T L V E I T I A A N - G G Q D Y -- Y D V S L V D G F N L P I K L A P R G G S G -- D C N S T S C A A N I N T V C P A E L 144
BAC78212.1 AKFTFTTNNCPYTVVPGTTLTGDKPQLSLTGFEELATKASRSVDAP S P - W S G R F W G R T R C S T N A A G K F T C E T A D C S G S Q V A C N G A G A V P P A T L V E I T I A A N - G G Q D F -- Y D V S L V D G F N L P I K L A P R G G S G -- D C N S T S C A A N I N T V C P A E L 144
ADN33945.1 STFTILNCDYTVVPGVLSGAGTSQSLSTTGFEELATKASRSVDAP S P - W S G R F W G R T R C S T N A A G K F T C E T A D C S G S Q V A C N G A G A V P P A T L V E I T I A A N - G G L D F -- Y D V S L V D G Y N V P M L I T P Q D G T G G G N C T T I G C A T D L N S E C P A E L 146
AEV57472.1 TTFFTNNKCEYTVVPGILTNAAGVPLPTTGFEELATKASRSVDAP S P - W S G R F W G R T R C S T N A A G K F T C E T A D C S G S Q V A C N G A G A V P P A T L V E I T I A A N - G G L D F -- Y D V S L V D G Y N L P M L V V P Q G G S G -- T N C N T T G C V V D L N G L C P S E L 145
AAO12209.1 LFLTIVNHCPTTVVPAIQPNAGHPVLEKGGFALPTNTHRSFFPPSTHWSGRIWGRTC S H Y - N G K F S C V T G D C G - H R L E C N G L G G A T P A S L A Q F D L H H G - H Q D L S S Y G V S L V D G F N V P M T V T P H E G R G -- V C P V V G C R E D L L K T C P A H L 145
ADN33946.1 AKFTFTNNKCEYTVVPGVLSGAGSLKFDTTTGFEELATKASRSVDAP S P - W S G R F W G R T R C S T N A A G K F T C E T A D C S G S Q V A C N G A G A V P P A T L V E I T I A A N - N S P D F -- Y D V S L V D G Y N L A M I V E G T G G T G -- A C G S T G C V T D L N R Q C P L E L 144
AEV57470.1 TVFTLQNRCTYTVVPGTTLTGSGNAAVLGDGGFALAPDSSVQLTAPRG - W S G R F W A R T R C N F D D S G N G K C L T G D C G -- S L K C V G - G G A P P V T L A E F T T G S T P G D K D F -- Y D V S L V D G Y N V A M G L R A N G G T G -- D C Q Y A G C V A D L N G C P P E L 142
AGC39175.1 ATFNIIINNCPYTVVAAAAPV-----GGGKRLDRGQNWIIINPAGSTKGARVWPRTG C N F D G A G R G K C Q T G D C N G L L Q C Q -- A F Q P P N T L A E Y A L N -- Q F N N L D F F D I S L V D G F N V A M E F S P T S G -- G C T R G I K T A D I N G Q C P N E L 134
ABQ42566.1 ATFNIIINNCPYTVVAAAAPV-----GGGKRLDRGQNWIIINPAGSTKGARVWPRTG C N F D G A G R G K C Q T G D C N G L L Q C Q -- A F Q P P N T L A E Y A L N -- Q F N N L D F F D I S L V D G F N V A M E F S P T S G -- G C T R G I K T A D I N G Q C P N E L 134
AGC39176.1 ATFNIIINNCPYTVVAAAAPV-----GGGKRLDRGQNWIIINPAGSTKGARVWPRTG C N F D G A G R G K C Q T G D C N G L L Q C Q -- A F Q P P N T L A E Y A L N -- Q F N N L D F F D I S L V D G F N V A M E F S P T S G -- G C T R G I K T A N I N E Q C P N E L 134
1Z3Q ATFEIVNRCSTYTVVAAAAPV-----GGGRLNQGSWITINVAGTTGGRIWGRTC C S F D G S G R G C Q T G D C G V L S C T -- A Y G N P P N T L A E F A L N -- Q F N N L D F F D I S L V D G F N V P M D F S P T S G -- G C - R G I R C A A D I N G Q C P G A L 133
AFK29763.1 VTFEIVNGCPYTVVAAAAPV-----GGGLQLDQGQTWVLVDVNPPTGGRIWARTG C S F D G S G S G C Q T G D C G V L D C K -- G Y G S P P N T L A E F A V G -- Q F N N -- F D V S L V D G F N V P M E F -----G S C R A I R C A D I N G R C P A E L 126
AGC39181.1 ATFDIVNQCSYTVVAAAAPV-----GGGRLDSDGQSWISITVNPPTIARIWGRTC C N F D A S G R G K C E T G D C N G L L E C Q -- G Y G S P P N T L A E F A L N Q D - Q P T N L D F V D I S L V D G F N I P M D F S P T T G ----V C K S L R C A A A I N E E C P T E L 135
AGC39182.1 ATFDIVNQCSYTVVAAAAPV-----GGGRLDSDGQSWISITVNPPTIARIWGRTC C N F D A S G R G K C E T G D C N G L L E C Q -- G Y G S P P N T L A E F A L N Q D - Q P T N L D F V D I S L V D G F N I P M D F S P T T G ----V C K Q L R C A A A I N E E C P A E L 135
AGC39180.1 ATFDIVNQCSYTVVAAAAPV-----GGGRLDSDGQSWISITVNPPTIARIWGRTC C N F D A S G R G K C E T G D C N G L L E C Q -- G Y G S P P N T L A E F A L N Q D - Q P T N L D F V D I S L V D G F N I P M D F S P T T G ----V C K Q L R C A A P I N E E C P T E L 135
E3SU11.1 ATFDIVNQCSYTVVAAAAPV-----GGGRLDSDGQSWININVAPGTTIARIWGRTC C N F D A S G R G K C E T G D C N G L L E C Q -- G Y G R P P N T L A E F A L N -- Q F N N L D F V D I S N V D G F N I P L E F S P T T N ----V C R R L V C N A P I V Q C C P S E L 133
AEP84104.1 ATFDVVNQCSTYTVVAGASP-----GGGKQLDQCGQTWTITVAPGTTIARIWGRTC C N F D A S G R G K C E T G D C N G L L Q C Q -- G Y G S P P N T L A E F S L N -- Q F N N L D Y V D I S L V D G F N I P M D F S P A A G ----V C R D I R C A D I T A Q C P A E L 134
AGC39179.1 ANFEIRNCPYTVVAAAAPV-----GGGRLDPYQTWVLVDVAGTKMARIWGRTC C N F D A G W G R C E T G D C G - S I D C R S G G W G V P P N T L A E Y A L N -- Q F G N L D F Y D I S L V D G F N I P M E F T P T Y R Q S - D K C R N I L C V A D I N G Q C P G P L 137
AGC39178.1 ANFEIRNCPYTVVAAAAPV-----GGGRLDPYQTWVLVDVAGTKMARIWGRTC C N F D A G W G R C E T G D C G - S I D C R S G G W G V P P N T L A E Y A L N -- Q F G N L D F Y D I S L V D G F N I P M E F T P T Y R Q S - D K C R N I L C V A D I N G Q C P G P L 137
AGC39177.1 ATFEIRNCPYTVVAAAAPV-----GGGRLNLSKQWNLNVAGTKMARIWGRTC C N F D G S G R G H C E T G D C G V L E C K -- G W G V P P N T L A E Y A L N -- Q F G N L D F Y D I S L V D G F N I P M D F S P T R S P S - G K C R P I S C T A N I I G Q C P N P L 136
AAK97184.1 ATFEVRNCPYTVVAASTPV-----GGGRLDRGQTWTINAPGTTAMARIWGRTC C N F D G S G R G S C Q T G D C G V L Q C T -- G W G K P P N T L A E Y A L N -- Q F N N L D F W D I S L V D G F N I P M T F A P T N P S G - G K C H A I Q C T A N I N G E C P G S L 137
ACB30367.1 ATFEVRNCPYTVVAASTPI-----GGGRLNLRGQTWVINA PRGTAMARIWGRTC C N F N A A G R G S C Q T G D C G V L Q C T -- G W G K P P N T L A E Y A L D -- Q F S N L D F W D I S L V D G F N I P M T F A P T K P S G - G K C H A I H C T A N I N G E C P R A L 137
AAP86781.1 GVFEVHNNCPYTVVAAATPV-----GGGKLLERQSWFWFAPGTTAMARIWGRTC C N F D G A G R G W C Q T G D C G V L E C K -- G W G K P P N T L A E Y A L N -- Q F S N L D F W D I S V I D G F N I P M S F G P T K P G P - G K C H I Q C V A N I N G E C P G S L 137
P02883.1 ATFEIVNRCSTYTVVAAAASKGD--AALDAGGRQLNSGESWTINVEPGTNGGKIWARTD C Y F D D S G S G I C K T G D C G G L L R C K -- R F G R P P T L A E F S L N -- Q Y G K - D Y I D I S N I K G F N V P M F S P T T R ----G C R G V R C A A D I V G Q C P A K L 138
BAF44567.1 ATFEIVNRCSTYTVVAAAASKGD--AALDAGGRQLNSGESWTINVEPGTNGGKIWARTD C Y F D D S G S G I C K T G D C G G L L R C K -- R F G R P P T L A E F S L N -- Q Y G K - D Y I D I S N I K G F N V P M F S P T T R ----G C R G V R C A A D I V G Q C P A K L 138
1RQW A ATFEIVNRCSTYTVVAAAASKGD--AALDAGGRQLNSGESWTINVEPGTNGGKIWARTD C Y F D D S G S G I C K T G D C G G L L R C K -- R F G R P P T L A E F S L N -- Q Y G K - D Y I D I S N I K G F N V P M F S P T T R ----G C R G V R C A A D I V G Q C P A K L 138
P02884.1 ATFEIVNRCSTYTVVAAAASKGD--AALDAGGRQLNSGESWTINVEPGTNGGKIWARTD C Y F D D S G S G I C K T G D C G G L L Q C K -- R F G R P P T L A E F S L N -- Q Y G K - D Y I D I S N I K G F N V P M F S P T T R ----G C R G V R C A A D I V G Q C P A K L 138
AAM15877.1 ATITVNRCSYTVVPGALP-----GGGVRLDPGQSWALNMPAGTAGARVWPRTG C T F D G S G R G R C I T G D C G G T L A C R -- V S G Q Q P T L A E Y T L G -- Q G N K D F D L S V I D G F N V P M F E P V G G ----S C R A A R C A T D I T K E C L K E L 133
1.....10.....20.....30.....40.....50.....60.....70.....80.....90.....100.....110.....120.....130.....140.....150

```



Epitope 1

Epitope 2

Epitope 3

ACN97419.1	QVKAADGTV--IACKSACLA	AFGDSKYCCTPPNNTPETCP	PTEYSQFFEQQCPQAYSAYDDKNSTFT	SGG-PDYVITFCP	-----	222									
ACN97416.1	QVKAADGTV--IACKSACLA	AFGDSKYCCTPPNNTPETCP	PTEYSQFFEQQCPQAYSAYDDKNSTFT	SGG-PDYVITFCP	-----	222									
ACN97418.1	QVKAADGTV--IACKSACLA	AFGDSKYCCTPPNNTPETCP	PTEYSQFFEQQCPQAYSAYDDKNSTFT	SGG-PDYVITFCP	-----	222									
AE036939.1	QVKAADGTV--IACKSACLA	AFGDSKYCCTPPNNTPETCP	PTEYSQFFEQQCPQAYSAYDDKNSTFT	SGG-PDYVITFCP	-----	222									
ACI04550.1	QVKAADGTV--IACKSACLA	AFGDSKYCCTPPNNTPATC	PTEYSQFFEQQCPQAYSAYDDKTSTFT	SGG-PDYVITFCP	-----	222									
ACN97417.1	QVKAADGTV--IACKSACLA	AFGDSKYCCTPPNNTPETCP	PTEYSQFFEQQCPQAYSAYDDKNSTFT	SGG-PDYVITFCP	-----	222									
AAX19851.1	QVKAADGSV--ISCKSACLA	AFGDSKYCCTPPNDTPETCP	PTEYSEIFEKQCPQAYSAYDDKNSTFT	SGG-PDYVITFCP	-----	222									
AAX19848.1	QVKAADGSV--ISCKSACLA	AFGDSKYCCTPPNDTPETCP	PTEYSEIFEKQCPQAYSAYDDKNSTFT	SGG-PDYVITFCP	-----	222									
ABC47923.1	QVKAADGSV--ISCKSACLA	AFGDSKYCCTPPNDMPETCP	PTEYSEIFEKQCPQAYSAYDDKNSTFT	SGG-PDYVITFCP	-----	222									
JC7201	QVKAADGSV--ISCKSACLA	AFGDSKYCCTPPNNTPETCP	PTEYSEIFEKQCPQAYSAYDDKNSTFT	SGG-PDYVITFCP	-----	222									
AEV57471.1	QVKGSDGSV--IACKSACLA	LNQPOYCCTPPNDKPETCP	PTEYSKIFKQCPQAYSAYDDKSSTFT	SAR-PDYLITFCP	-----	222									
P83332.1	QVKGSDGSV--IACKSACLA	FNQPKYCCTPPNDKPETCP	PDYSKLFKTCQAYSAYDDKSSTFT	SGR-PAYLITFCP	-----	222									
CAB62167.1	QVKGSDASV--LACKSACLA	FNQPOYCTGAFDARTCP	ATKYSRIFFKQCPQAYSAYDDSTSTFT	SGA-PDYVITFCP	-----	222									
P50694.1	QVKGSDGSV--VACLSACLA	VKFGTPQYCCTPPQNTPETCP	PPTNYSEIFHNACPDAYSAYDDKRGFTFT	CNGG-PNYAITFCP	-----	222									
ADM22305.1	QVKGANGKV--VACLSACLA	VAFQPKPYCCTPPQNTPETCP	PPTNYSQIFSHACPKAYSAYDDKKGFTFT	CKGG-PNYLITFCP	-----	222									
ACE80974.1	QKIGSDGSV--VACLSACLA	VKFNQPOYCCTPPQETKEKCP	PPTNYSQIFHDQCPDAYSAYDDNKGLFT	SGG-PNYLITFCP	-----	218									
P83335.1	QKIGSDGSV--VACLSACLA	VKFGEPQYCCTPPQETKEKCP	PPTNYSQIFHEQCPDAYSAYDDNKGLFT	SGG-PNYLITFCP	-----	219									
O80327.1	SDKSGSDGSV--IGCKSACLA	LNQPOYCCTGAYGTPDTC	PTDFSKVFNKQCPQAYSAYDDKSSTFT	CFG-PNYEITFCP	-----	222									
BAC78212.1	SDKSGSDGSV--IGCKSACLA	LNQPOYCCTGAYGTPDTC	PTDFSKVFNKQCPQAYSAYDDKSSTFT	CFG-PNYEITFCP	-----	222									
ADN33945.1	RVVLSDDGERSV	VACKSACEAFGDAQYCCSG	GEYGNPNTCRSSYSELFKSACPDAYSAYDDGSSTFT	CT-A-ANYLITFCPSLSTR	-----	230									
AEV57472.1	KVTSADGRE--GVACKSACEA	FQPOYCCSGAYGTPDTC	PKSSYQIFKNACPHAYSAYDDRTSTFT	CA-A-ADYITITFCPSPSTSKKASQ	QNTQPOPTNTTTGTGTPTSSTSSSPEPINNTMEYEGASYDQSEASSSTRTTSHAIAGLV	292									
AAO12209.1	QLRSHGHV--VACKSGCEA	FRTDDELCCRNHYNSPQT	CRASSHSLFFKHACPSTMTFAHDSPLMHD	CSSP-RELKVFICH	-----	223									
ADN33946.1	RAEKG-----GACRSACE	KFGTPEYCCSGAYGSPATC	KPSKYSEIFKSACPRSYSAYDDATSTFT	CS-G-ADYTVTFCPSSPSLKSTDS	SPRTTTTTTGTAGVVAQMTPLPDSSWMADMAIGGSTRTYSTDLNFLFLVIFGSSSLVF	286									
AEV57470.1	QVTDSGNVV--ACRSACAA	FNAPFCCTGDHATPQTC	SPTQYSVMFKTACPTAYSAYDDASST	CTCA-G-SDYLITFCPTGSS	-----	222									
AGC39175.1	RAPGG-----CNPCTVFKT	DQYCCN--SGN--CGLT	NFSKFFKDRCPDAYSYPKDDQTSTFT	CAG-TNYKVVF	CP-----	201									
ABQ42566.1	RAPGG-----CNPCTVFKT	DQYCCN--SGN--CGLT	NFSKFFKDRCPDAYSYPKDDQTSTFT	CAG-TNYKVVF	CP-----	201									
AGC39176.1	RAPGG-----CNPCTVFKT	DQYCCN--SGN--CGPT	DYSRFFKTRCPDAYSYPKDDQTSTFT	CPGG-TNYKVVF	CP-----	201									
1Z3Q	KAPGG-----CNPCTVFKT	DQYCCN--SGA--CSPT	DYSQFFKRNCPDAYSYPKDDQTSTFT	CPGG-TNYRVVF	CP-----	200									
AFK29763.1	RTPGG-----CNPCTVFKT	DEYCCCT--SGS--CGPT	TYSKFFKSNC	PDAYSYPKDDATSTFT	CPGN-GDYKVIFCPSVSSLPADGYPTRFGRIAVSFVIPKASGT	222									
AGC39181.1	RAPGG-----CNPCTVFKT	NEYCCT--DGP	GSCPEPTTYSKFFKDRCPDAYSYPQDDPTSLFT	CPGG-TNYRVVF	CP-----	204									
AGC39182.1	RAPGG-----CNPCTVFKT	NEYCCT--DGP	GSCPEPTTYSKFFKDRCPDAYSYPQDDPTSLFT	CPGG-TNYKVVF	CP-----	204									
AGC39180.1	RAPGG-----CNPCTVFKT	NEYCCT--DGQ	GTGCPPTTFSKFFKDRCPDAYSYPQDDPTSLFT	CPGG-TNYKVVF	CP-----	204									
E3SU11.1	RTPGG-----CNPCTVFN	TNEYCCT--NGP	GSCGPTPLSRFFKERC	PDAYSYPQDDPTSLFT	CAG-TNYRVVF	CP-----	202								
AEF84104.1	QAPGG-----CNPCTVYK	TNEYCCT--NGQ	GTGPTALSKFFKDRCPDAYSYPQDDPTSLFT	CAG-TNYKVVF	CPNLDA-----	207									
AGC39179.1	RAPHG-----CNPCTVFKT	QEYCCCT--QGP--CGPT	GYSRFFKDRCPDAYSYPQDDPTSTFT	CPGG-TNYRVVF	CPWGSPLH-----LKMYSANDTSLE	221									
AGC39178.1	RAPHG-----CNPCTVFKT	QYCCCT--QGP--CGPT	DYSRFFKDRCPDAYSYPQDDPTSTFT	CPGG-TNYRVVF	CPWGSPLH-----LKMYSANDTSLE	221									
AGC39177.1	KTAGG-----CNPCTVFKT	QEYCCCT--QGP--CGPT	NYSRFFKERC	PDAYSYPQDDPTSTFT	CPGG-SNYRVTF	CPNGSPH-----FPLEMYGESDVE	220								
AAK97184.1	RVPGG-----CNPCTTFGG	QYCCCT--QGP--CGPT	ELSKFFKRC	PDAYSYPQDDATSTFT	CPSGSTNYRVVF	CPNGVTG-----PNFPLEMPGS-DGVAK	225								
ACB30367.1	KVPGG-----CNPCTTFGG	QYCCCT--QGP--CGPT	ELSKFFKRC	PNAYSYPQDDPTSTFT	CPGGSTNYRVVF	CPNGVAD-----PNFPLEMPTSTDEVAK	226								
AAP86781.1	RVPGG-----CNPCTTFGG	QYCCCT--QGP--CGPT	DLRFFKQRC	PDAYSYPQDDPTSTFT	QSWTQDYKVMFCPYGSTHNETSS	FFLELPTSTHEVA	228								
P02883.1	KAPGGG-----CNDACTVFQT	SEYCCCT--TGK--CGP	TEYSRFFKRLCPDAFSYVLDKPT	-TVT	CPGS-SNYRVTF	CPPTA-----	207								
BAF44567.1	KAPGGG-----CNDACTVFQT	SEYCCCT--TGK--CGP	TEYSRFFKRLCPDAFSYVLDKPT	-TVT	CPGS-SNYRVTF	CPPTALELEDE-----	213								
1RQW A	KAPGGG-----CNDACTVFQT	SEYCCCT--TGK--CGP	TEYSRFFKRLCPDAFSYVLDKPT	-TVT	CPGS-SNYRVTF	CPPTA-----	207								
P02884.1	KAPGGG-----CNDACTVFQT	SEYCCCT--TGK--CGP	TEYSRFFKRLCPDAFSYVLDKPT	-TVT	CPGS-SNYRVTF	CPPTALELEDE-----	213								
AAM15877.1	QVPGG-----CASACGK	FGGDTYCCRG--QFEHN	CPPTNYSKFFKGC	PDAYSIAKDDQTSTFT	CAG-TNYQIVLC	CP-----	203								
	160.....	170.....	180.....	190.....	200.....	210.....	220.....	230.....	240.....	250.....	260.....	270.....	280.....	290.....	300



ACN97419.1 -----
ACN97416.1 -----
ACN97418.1 -----
AEO36939.1 -----
ACI04550.1 -----
ACN97417.1 -----
AAX19851.1 -----
AAX19848.1 -----
ABC47923.1 -----
JC7201 -----
AEV57471.1 -----
P83332.1 -----
CAB62167.1 -----
P50694.1 -----
ADM22305.1 -----
ACE80974.1 -----
P83335.1 -----
O80327.1 -----
BAC78212.1 -----
ADN33945.1 -----
AEV57472.1 TVALAIQLLSQLFF 306
AAO12209.1 -----
ADN33946.1 FSSLNLFSS----- 295
AEV57470.1 -----
AGC39175.1 -----
ABQ42566.1 -----
AGC39176.1 -----
1Z3Q -----
AFK29763.1 -----
AGC39181.1 -----
AGC39182.1 -----
AGC39180.1 -----
E3SU11.1 -----
AEP84104.1 -----
AGC39179.1 -----
AGC39178.1 -----
AGC39177.1 -----
AAK97184.1 -----
ACB30367.1 -----
AAP86781.1 -----
P02883.1 -----
BAF44567.1 -----
LRQW A -----
P02884.1 -----
AAM15877.1 -----
.....310.....

Q9M4G6 ATFDILNKCTYTVWAAASPGGRRRLDSGQSWTITVNP~~GT~~TNARIWGR~~T~~SC~~TF~~DANGRGK~~C~~ETGDCNGLLECQGYGSP~~P~~NT 80
A9CSP0 ATFDILNKCTYTVWAAASPGGRRRLDSGQSWTITVNP~~GT~~TNARIWGR~~T~~SC~~TF~~DANGRGK~~C~~ETGDCNGLLECQGYGSP~~P~~NT 80
O04708 ATFDILNKCTYTVWAAASPGGRRRLDSGQSWTITVNP~~GT~~TNARIWGR~~T~~SC~~TF~~DANGRGK~~C~~ETGDCNGLLECQGYGSP~~P~~NT 80
F6HUH1 ATFDILNKCTYTVWAAASPGGRRRLDSGQSWTITVNP~~GT~~TNARIWGR~~T~~SC~~TF~~DANGRGK~~C~~ETGDCNGLLECQGYGSP~~P~~NT 80
Q9LLB7 ATSTILNKCTYTVWAAASPGRRRLDSGQSWTITV~~K~~P~~GT~~TNARIWGR~~T~~SC~~TF~~DANGRGK~~C~~ETGDCNGLLNCQGYGSP~~P~~NT 80

Q9M4G6 LAEFALNQPNNDYIDISLVDGFNIPMDFS----GCRGIQCSADINGQCPSELKAPGGC~~NNPCTVF~~KTNEYCCTDGP~~G~~SC 156
A9CSP0 LAEFALNQPNNDYIDISLVDGFNIPMDFS----GCRGIQCSADINGQCPSELKAPGGC~~NNPCTVF~~KTNEYCCTDGP~~G~~SC 156
O04708 LAEFALNQPNNDYIDISLVDGFNIPMDFS----GCRGIQCSVDINGQCPSELKAPGGC~~NNPCTVF~~KTNEYCCTDGP~~G~~SC 156
F6HUH1 LAEFALNQPNNDYIDISLVDGFNIPMDFS----GCRGIQCSVDINGQCPSELKAPGGC~~NNPCTVF~~KTNEYCCTDGP~~G~~SC 156
Q9LLB7 LAEFALNQPNNDYIDISLVDGFNIPMGFSPTTSACRGIQCSANINGQCPSELKAPDGC~~NNPCTVF~~KTNNYCCTNGP~~G~~SC 160

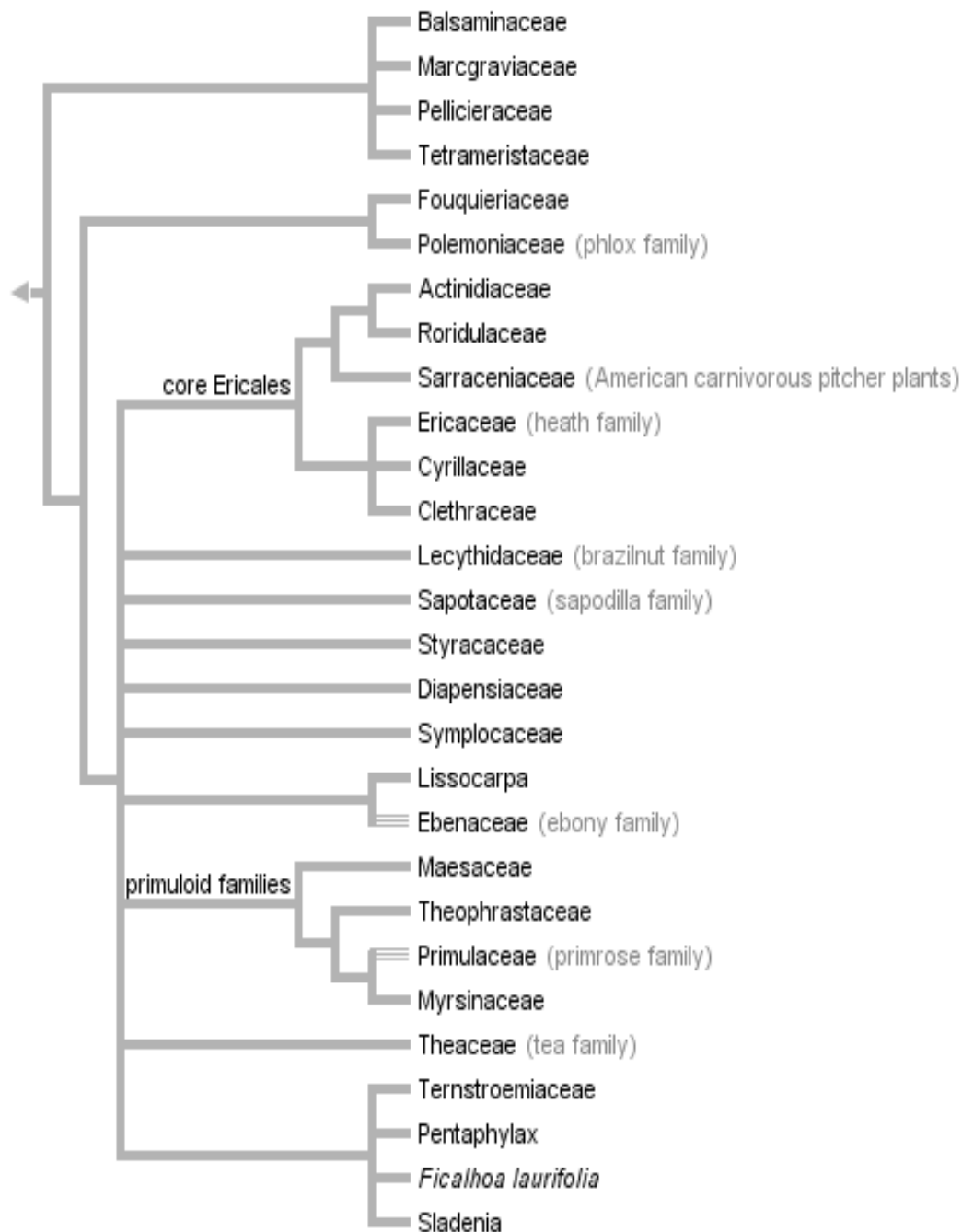
Q9M4G6 GPTTYSKFFKDRCPDAYSY~~PQ~~DDKTS~~LF~~CTSGTNYKV~~T~~FCP 198
A9CSP0 GPTTYSKFFKDRCPDAYSY~~PQ~~DDKTS~~LF~~CTSGTNYKV~~T~~FCP 198
O04708 GPTTYSKFFKDRCPDAYSY~~PQ~~DDKTS~~LF~~CTCPSGTNYKV~~T~~FCP 198
F6HUH1 GPTTYSKFFKDRCPDAYSY~~PQ~~DDKTS~~LF~~CTCPSGTNYKV~~T~~FCP 198
Q9LLB7 GPTTYSKFFKDRCPDAYSY~~PQ~~DDKTS~~LF~~CTCPSGVNYKV~~T~~FCP 202

Inset: ClustalX alignment of grape TLPs displaying predicted IgE epitopes

Supplementary data related to

Ashok Kumar, H.G., Venkatesh, Y.P., *In silico* analyses of structural and allergenicity features of sapodilla (*Manilkara zapota*) acidic thaumatin-like protein in comparison with allergenic plant TLPs. Mol. Immunol. (2013), <http://dx.doi.org/10.1016/j.molimm.2013.08.010>

Figure S3. Cladogram of botanical families in the angiosperm order Ericales displaying the currently known allergens. The cladogram (Tree of Life Web Project, 2002) is shown in the left half. Pollen or food allergies have been reported from only 7 families (listed in the right half along with allergen names where known). Pollen allergens are indicated by an asterisk and inhalant or contact food allergens are indicated by a dagger. Allergens have been identified from *Actinidia arguta* (hardy kiwi), *A. chinensis* (gold kiwi), *A. deliciosa* (green kiwi) and *A. eriantha* (kiwi, velvet vine), *Diospyros kaki* (persimmon, kaki, or sharon fruit), *Vaccinium myrtillus* (bilberry or mirtillo), *Calluna vulgaris* (heather), *Rhododendron spp*, *Camellia japonica* (Japanese camellia), *Camellia sinensis* (tea), *Manilkara zapota* (sapodilla or sapodilla plum) and *Bertholletia excelsa* (Brazil nut). Allergies have been reported for *Argania spinosa* (Argan), *Madhuca indica* (mahua or moa tree), *Palaquium gutta* (gutta percha) and *Cyclamen persicum* (cyclamen or Persian violet); of these, the first 3 belong to Sapotaceae. LTP, lipid transfer protein; TLP, thaumatin-like protein.



Actinidiaceae

Actinidin (Act a 1, Act c 1, Act d 1, Act e 1)

TLP (Act c 2, Act d 2, Act e 2)

Unknown (Act d 3)

Cystatin (Act c 4, Act d 4)

Kiwellin (Act a 5, Act c 5, Act d 5, Act e 5)

Pectin methylesterase inhibitor (Act d 6)

Pectin methylesterase (Act d 7)

PR-10 protein (Act c 8, Act d 8)

Profilin (Act d 9)

LTP (Act c 10, Act d 10)

Bet v 1-like protein (Act d 11)

Chitinase (Act c, Act d)

Ebenaceae

Bet v 1-like protein (Dio k 1)

Profilin (Dio k 4)

Isoflavone reductase (Dio k IFR)

Ericaceae

LTP (Vac m 3)

Unknown (Cal vu)*

W202 (Rho s)*

Theaceae

W213 (Cam j)*

F222 (Cam s)

Lecythidaceae

2S albumin (Ber e 1)

11S globulin (Ber e 2)

Sapotaceae

TLP (Man z TLP 1 and 2)

Unknown (Arg sp[†], Mad i*, Pal g)

Myrsinaceae

Unknown (Cyc p)*

Supplementary data related to

Ashok Kumar, H.G., Venkatesh, Y.P., *In silico* analyses of structural and allergenicity features of sapodilla (*Manilkara zapota*) acidic thaumatin-like protein in comparison with allergenic plant TLPs. Mol. Immunol. (2013), <http://dx.doi.org/10.1016/j.molimm.2013.08.010>

Figure S2. ClustalX alignment of pollen allergenic TLPs (inhalant) displaying predicted IgE epitopes. Allergens were retrieved from Allergome database; sequences were retrieved from NCBI database. The sequence IDs correspond to the description given in Table 1. The predicted IgE epitope stretches in each of the sequences is highlighted in red (epitope 1), green (epitope 2), blue (epitope 3), pink (epitope 4) and brown (epitope 5) in the order of appearance from the N-terminus. Cysteine residues (C) are highlighted in yellow. Symbols: (*), complete identity; (:), strongly similar; (.), weakly similar; no symbol, different. The bar graph shown at the bottom denotes conservancy profile.

```

      * : ** *****      ** : ** :      * * * * * * * * * * * * * * * * : * : * : * : * : * :
BAC15615.1 -ATFDITNQCPYTVWAAASP-----GGGRQLAKGQTWTIQVPAGTTGGRVWARTGCSFDRSSGRGTCQTGDCNGMLSCQGYGQVP-ATLAEYALNQYMNLDFYDISLVDGFNVPISMTPSTNPNCKGRITCLSDINSKCPSELKVNG 140
BAC15614.1 -ATFDITNQCPYTVWAAASP-----GGGRQLAKGQTWTIQVAAGTTGGRVWARTGCSFDGSSGRGTCQTGDCNGMLSCQGYGQVP-ATLAEYGLNKFQNLDFYDISLVDGFNVPLSMTPSTNPNCKGRITCLSHINSMCPAELKVNG 140
BAC15616.1 -ATFDITNQCPYTVWAAASP-----GGGQQLDQGQTWTIQVAAGTTQARIWARTGCSFDGSSGRGTCQTGDCNGMLSCQGYGQVP-ATLAEYALNQYMNLDFYDISLVDGFNVPLSMTPSTDPNCKGRIACLSDINSQCPSDLKVTG 140
BAD90815.1 -PTFEITNKCPYTVWAAAFP-----GGGQQLAQGQSWSVQPDAGTSTGRIWGRDCSFDGSSGRGTCQSGDCNGTLNCGGDASAP-VTLVEYTHNPSMNLDFYDISLDGFLPLSITPTSTNPNCKGIITCLSDINSQCPNELKVSG 140
BAD90813.1 -VNFDIENQCPYTVWAAAGTFP-----GGGIELKRGQSWRVNVPGG-ARGRFWGRTGCSFDGNGRGRCNTGDCGGLLNCQSGSGVP-STLLEYALNQYQNLDFYDISLVDGFNLRMTVVLSNTN--CK-RIACNSDINSKCPGELKVVD 137
AAR21074.1 -VKFDIKNQCGYTVWAAAGLP-----GGGKRLDQGQTWTIVNLAAGTASARFWGRTGCTFDASGKGSCRSGDCGGQLSCTVSGAVP-ATLAEYTS---DKDYYDVSLVDGFNIPLAINPTNTKCTAP---ACKADINAVCPSELKVDG 134
AAR21075.1 AVKFDIKNQCRYTVWAAAGLP-----GGGKRLDQGQTWTIVNLAAGTASARFWGRTGCTFDASGKGSCRSGDCGGQLSCTVSGAVP-ATLAEYTS---DQDYYDVSLVDGFNIPLAINPTNTKCTAP---ACKADINAVCPSELKVDG 135
CAC05258.1 -VKFDIKNQCGYTVWAAAGLP-----GGGKEFDQGQTWTIVNLAAGTASARFWGRTGCTFDASGKGSCRSGDCGGQLSCTVSGAVP-ATLAEYTS---DQDYYDVSLVDGFNIPLAINPTNTKCTAP---ACKADINAVCPSELKVDG 134
AAR21071.1 -VKFDIKNQCGYTVWAAAGLP-----GGGKRLDQGQTWTIVNLAAGTASARFWGRTGCTFDASGKGSCRSGDCGGQLSCTVSGAVP-ATLAEYTS---DQDYYDVSLVDGFNIPLAINPTNAQCTAP---ACKADINAVCPSELKVEG 134
AAR21072.1 -VKFDIRNQCGYTVWAAAGLP-----GGGKRLDQGQTWTIVNLAAGTASARFWGRTGCTFDASGKGSCRSGDCGGQLSCTVSGAVP-ATLAEYTS---DQDYYDVSLVDGFNIPLAINPTNAQCTAP---ACKADINAVCPSELKVEG 134
P81295.1 -VKFDIKNQCGYTVWAAAGLP-----GGGKRLDQGQTWTIVNLAAGTASARFWGRTGCTFDASGKGSCRSGDCGGQLSCTVSGAVP-ATLAEYTS---DQDYYDVSLVDGFNIPLAINPTNAQCTAP---ACKADINAVCPSELKVDG 134
BAF51970.1 -VKFDIKNQCGYTVWAAAGLP-----GGGQQLTQGQTWTIVNLAAGTQSARFWGRTGCSFDASGKGTCQTGDCGGQLSCTVSGAVP-ATLAEYTS---DQDYYDVSLVDGFNIPLSINPTNAQCTAP---ACKADVNAVCPAELKVDG 134
AAV65287.1 -VKFNIRNQCGYTVWAAAGLP-----GGGKQLDQGQTSVDVAAGTKGARFWGRTGCSFDGSSGRGSCQTGDCGGQLSCTLSGAVP-ATLAEYSLNGNDNKDFYDVSLVDGFNVPLSINPTNTGCTAP---ACKADVNAACPAELKVNS 137
BAD99299.1 AVKFELKNQCEYTVWAAAGLP-----GGGQQLDQGQTWPVEVPAGTKGARFWGRTGCSFDASGKGTCQTGDCNSQLSCQVSGGVP-TTLAEYTLNGDGNKDFYDVSLVDGFNVPLSINPTNSQCFAP---ACKADVNAACPAQLKVDG 138
BAD90814.1 -TVFTLVNKCYSYTVPGTSLSGSGSSVLGEGGFTLAPGQSVPLTASSR-WSGRFWGRDCSFDASGKGSITGDCGNVLNCAQAGGTPPYSLAEFTLG---DKDFYDVSLVDGYNVPLSTAAVGGTGDCR-TAGCVSDLRTSCPAELSVTS 144
1.....10.....20.....30.....40.....50.....60.....70.....80.....90.....100.....110.....120.....130.....140.....150

```



Epitope 1

Epitope 2

Epitope 3

```

      * ***      * : **      * : ** ** * * * * * * * * * * * * * * * * : * : * :
BAC15615.1 ----GCKSACARYNTAQYCTGAS--ANNCGPTNYSKFFKGQCPQAYSYAKDDATSTFTCPSG-TNYKVVFCG----- 206
BAC15614.1 ----GCKSACARYNTAQYCTGAS--ANNCGPTNYSKFFKGQCPQAYSYAKDDATSTFTCPSG-TNYKVVFCG----- 206
BAC15616.1 ----GCKSACARYNTPEYCTGAS--ENTCGPTDYSKFFKGQCPQAYSYAKDDATSTFTCPSG-TNYKVVFCG----- 206
BAD90815.1 ----GCLSACVKYNTDDHCCRG--NCNQTLYSRFFKEQCPQAYSYAKDDVTSSAFTCPSG-TDYKIVFCGKSDHIFALYLYIT--INNEE 221
BAD90813.1 ----GCRSACAAFNTPAYCTGSGF--LDNCPPTGYSQFFKRECPLAYSYAKDDPTSTFTCPGG-SDYKIIFCGNGDSQSNSTSTSSGSLYAVE 223
AAR21074.1 ----GNSACNVLQTQDYCCRNAY--VDNCPATNYSKIFKNQCPQAYSYAKDDT-ATFACASG-TDYSIVFCP----- 199
AAR21075.1 ----GNSACNVLQTQDYCCRNAY--VDNCPATNYSKIFKNQCPQAYSYAKDDT-ATFACASG-TDYSIVFCP----- 200
CAC05258.1 ----GNSACNVLQTQDYCCRNAY--VDNCPATNYSKIFKNQCPQAYSYAKDDT-ATFACASG-TDYSIVFCP----- 199
AAR21071.1 ----GNSACNVFQTQDYCCRNAY--VDNCPATNYSKIFKNQCPQAYSYAKDDT-ATFACASG-TDYSIVFCP----- 199
AAR21072.1 ----GNSACNVFQTQDYCCRNAY--VDNCPATNYSKIFKNQCPQAYSYAKDDT-ATFACASG-TDYSIVFCP----- 199
P81295.1 ----GNSACNVFQTQDYCCRNAY--VDNCPATNYSKIFKNQCPQAYSYAKDDT-ATFACASG-TDYSIVFCP----- 199
BAF51970.1 ----GCKSACAAFQTDQYCTGTY--ANSCPATNYSMIFKNQCPQAYSYPKDDT-ATFACPSG-TDYSIVFCP----- 199
AAV65287.1 ----GNSACNVFQTQDYCCRGAN--VDNCPATNYSMIFKNQCPQAYSYAKDDTSTFTCPSGTTDYSIVFCP----- 204
BAD99299.1 ----GNSACNVFQTDEYCCRGTH--VDNCPSSYSMIFKNQCPQAYSYAKDDSSSTFTCPSGTTDYSIVFCP----- 205
BAD90814.1 NGQVIACKSACAAFSTPEYCTGDHGSPQTCSPSKYSQVFKSACPATAYSAYDDATSTFTCSNA--DYTITFCPS----- 218
.....160.....170.....180.....190.....200.....210.....220.....230.....240....

```

Epitope 4

Epitope 5