

How reliable is the structural prediction of IgE-binding epitopes of allergens? The case study of plant lipid transfer proteins

Jean-Philippe Borges^a, Annick Barre^a, Raphaël Culerrier^a, Nicolas Archimbaud^a,
Alain Didier^b, Pierre Rougé^{a,*}

^a UMR-CNRS 5546, Pôle de Biotechnologie végétale, 24 Chemin de Borde Rouge, 31326 Castanet-Tolosan, France

^b Clinique des Voies Respiratoires, Centre Hospitalier Universitaire Larrey, 24 Chemin de Pouvoirville,
31059 Toulouse Cedex 9, France

Received 27 February 2006; accepted 15 September 2006

Available online 11 October 2006

Abstract

The linear IgE-binding epitopes of non-specific lipid transfer proteins (nsLTP) from plants were predicted using a combination of predictive tools including (1) the hydropathic profiles based on different scales of hydrophilicity, flexibility and exposure to the solvent, (2) the hydrophobic cluster analysis plots, (3) the occurrence of charged residues in the predicted amino acid sequence stretches and, (4) the exposition of the predicted linear IgE-binding epitopes checked on the three-dimensional models built for the nsLTP. A reliable prediction was obtained for nsLTP as compared with the previously characterized IgE-binding epitopes of various proteins. A consensual IgE-binding epitope occurring in other plant nsLTP and responsible for some IgE-binding cross-reactivity among fruit nsLTP has been identified and characterized. Despite some discrepancies, a fairly good prediction resulted in applying our combination of predictive methods to longer nsLTP or plant profilins.

© 2006 Elsevier Masson SAS. All rights reserved.

Keywords: Allergy; Lipid transfer protein; Pru p 3; Mal d 3; IgE-binding epitope; IgE-binding cross-reactivity

1. Introduction

Until now only a few plant dietary allergens have been characterized. However, the increasing amount of information

obtained from both the genome and transcriptome sequencing programs of different plants and the proteomic approaches that are developing for the identification of allergens in fruits and other plant products are still waiting to be assessed in term of allergenicity potency. The suspected allergenicity of plant proteins has to be characterized at the molecular level by the identification of their IgE-binding epitopes that account for their reactivity towards cell-bound IgE. Epitope mapping based on the ELISA inhibition of IgE-binding to a protein allergen by a series or a library of overlapping synthetic oligopeptides covering its entire amino acid sequence (e.g. PEPSCAN) is time consuming and could be greatly accelerated by using the only amino acid sequence stretches that have been accurately predicted to act as linear IgE-binding epitopes. The search for simple predictive methods allowing a reliable identification of the linear IgE-binding epitopes along the amino acid sequences of allergens, yet remains a goal that deserves much attention with respect to the

Abbreviations: Ab, polyclonal antibodies; Ara h 1, vicilin allergen from peanut (*Arachis hypogaea*); Bet v 2, profilin from *Betula verucosa* (birch); HCA, hydrophobic cluster analysis; Hel t 2, profilin from *Helianthus tuberosus* (Jerusalem artichoke); LTP, lipid transfer protein, Mal d 3, *Malus domestica* (apple) allergen 3; nsLTP, non-specific lipid transfer protein; PBS, phosphate buffered saline (pH 7.5); PBSTB, 10 mM phosphate buffered saline (pH 7.5) containing 0.2% Tween-20 (v/v) and 3% BSA (w/v); Par j 1, lipid transfer protein from *Parietaria judaica* (pellitory-of-the-wall); Pru ar 3, *Prunus armeniaca* (apricot) allergen 3; Pru av 3, *Prunus avium* (cherry) allergen 3; Pru d 3, *Prunus domestica* (plum) allergen 3; Pru p 3, *Prunus persica* (peach) allergen 3; Pyr c 3, *Pyrus communis* (pear) allergen 3; TBS, 20 mM Tris-buffered saline (pH 7.5); Zea m 14, *Zea mays* (maize) allergen 14.

* Corresponding author. Tel.: +33 5 62 19 35 58; fax: +33 5 62 19 35 02.

E-mail address: rouge@scsv.ups-tlse.fr (P. Rougé).

characterization of the increasing number of allergens identified so far. Hydropathic profiles based on different propensity scales of both hydrophilicity, flexibility and exposition to the solvent [1–3] are usually used to delineate the linear IgE-binding epitopes or B-cell epitopes that correspond to the most hydrophilic/flexible/accessible regions occurring along the amino acid sequence of protein allergens. These predicted regions often coincide with turns/loops [4] or well-exposed regions protruding out of the molecular surface of the proteins [5]. However, many predicted IgE-binding regions poorly match the experimentally characterized IgE-binding epitopes, as reviewed recently [6]. Obviously, the single use of hydropathic profiles based on a very few properties of the amino acid residues are a too crude predictive tool to reliably predict the location of the IgE-binding regions of the molecular surface of the proteins. Additional structural information [7,8] dealing with the orientation (rotamers, bulky or tiny character, flexibility of the side chain), the properties (net charge) of the surface exposed amino acid residues and their location on the protein surface have to be carefully considered for a more accurate prediction of the amino acid sequence stretches susceptible to act as IgE-binding epitopes. In this respect, all of the linear IgE-binding epitopes that have been characterized so far along the amino acid sequence of plant allergens contain either electropositive, or electronegative residues, or both (Table 1).

Here we report on the use of a combination of different predictive methods including hydropathic profiles, hydrophobic cluster analysis (HCA), molecular modelling and protein surface analysis for the prediction of the linear IgE-binding epitopes of the plant lipid transfer protein allergens from fruits and other plant organs. These non-specific lipid transfer proteins (nsLTP), which belong to the PR 14 family of pathogenesis-related proteins (PR-proteins) [9,10], have been recognized as major food allergens responsible for both the sensitization and allergenic responses of susceptible individuals [11–13]. A conserved consensus IgE-binding epitope common to nsLTP has been identified and characterized on the molecular surface of the allergens.

2. Materials and methods

2.1. Molecular modelling and IgE-binding epitope prediction

Multiple amino acid sequence alignments were carried out with CLUSTAL-X [14] and displayed with ESPript [15].

The HCA (Hydrophobic Cluster Analysis) [16] was performed to delineate the conserved secondary structural features (stretches of α -helix) along the amino acid sequence of Pru p 3 and other fruit LTP by comparison with the maize (*Zea mays*) LTP [17] used as a model. Once the assigned secondary structures have been delineated along the HCA plot of the template protein, the corresponding secondary structure stretches are precisely delineated along the HCA plot of the target protein according to their cluster shape and size similarities. HCA plots were generated using the HCA server (<http://smi.snv.jussieu.fr/hca/hca-form.html>).

Molecular modelling of Pru p 3 was carried out on a Silicon Graphics O2 R10000 workstation, using the programs InsightII, Homology and Discover3 (Accelrys, San Diego, CA, USA). The atomic coordinates of the maize LTP Zea m 14 [17] (RCSB Protein Data Bank code 1MZL) were used to build the three-dimensional model of the allergen. The percentages of both identity ($\sim 35\%$) and homology calculated with the structural homology matrix of Risler et al. [18] ($\sim 75\%$) Zea m 14 shares with Pru p 3 (Fig. 1) allowed us to build rather accurate three-dimensional model using the X-ray coordinates of the maize LTP as a template. Steric conflicts were corrected during the model-building procedure using the rotamer library [19] and the search algorithm implemented in the Homology program [20] to maintain proper side-chain orientation. The geometry of loop regions was corrected using the refine option of TurboFrodo [21]. An energy minimization of the final models was carried out by 50 cycles of steepest descent using Discover3. The program TurboFrodo was run to draw the Ramachandran plot and to perform the superposition of the model with the template protein. PROCHECK [22] was used to assess the geometric

Table 1
List of the IgE-binding epitopes mapped along the amino acid sequences of plant allergens containing either electropositive residues (K/R), or electronegative residues (D/E), or both (K/R + D/E) (epitopes taken from the SDAP database available at http://fermi.utmb.edu/SDAP/sdap_1st.html)

Allergen	No. of IgE-binding epitopes identified	No. of IgE-binding epitopes containing K/R	No. of IgE-binding epitopes containing D/E	No. of IgE-binding epitopes containing K/R + D/E
Ara h1 (<i>Arachis hypogaea</i>)	21	1	2	18
Ara h2 (<i>A. hypogaea</i>)	10	1	1	8
Ara h 3 (<i>A. hypogaea</i>)	4		2	2
Hev b 1 (<i>Hevea brasiliensis</i>)	8	2	2	4
Hev b 3 (<i>H. brasiliensis</i>)	11	4		7
Hev b 5 (<i>H. brasiliensis</i>)	11		3	8
Jug r 1 (<i>Juglans regia</i>)	1			1
Jun a 1 (<i>Juniperus ashei</i>)	4	3		1
Jun a 3 (<i>J. ashei</i>)	5	1		4
Par j 1 (<i>Parietaria judaica</i>)	5	1		4
Par j 2 (<i>P. judaica</i>)	8	3		5
Pru p 3 (<i>Prunus persica</i>)	3	2		1
Total	91	18	10	63

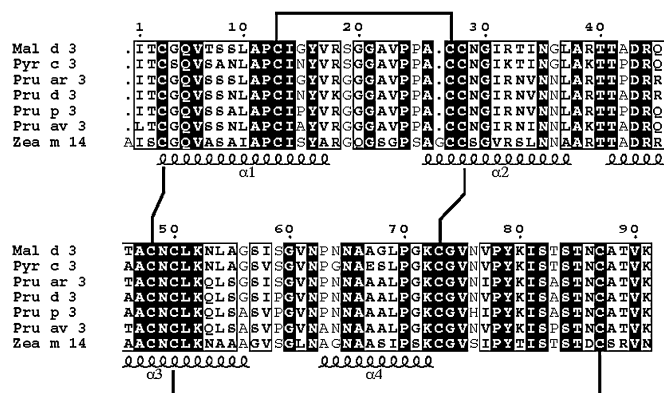


Fig. 1. Multiple alignment of fruit LTP from apple (Mal d 3), pear (Pyr c 3), apricot (Pru ar 3), plum (Pru d 3), peach (Pru p 3) and cherry (Pru av 3) with LTP of maize seedlings (Zea m 14). Identical and homologous residues are black boxed and open boxed, respectively. The four disulfide bridges linking the conserved Cys residues are indicated. Alignment performed with CLUSTAL-X [14] and displayed with ESPript [15].

quality of the three-dimensional models. All the residues of Pru p 3 were correctly assigned on the best allowed regions of the Ramachandran plot with the exception of residues Pro26 and Ala40 which are located in the generously allowed regions of the plot (result not shown). Using ANOLEA [23] to evaluate the models, only a few residues (2, 4 and 6 over 91 for the cherry, apricot and apple LTP, respectively versus 6 over 93 for the maize LTP (1MZL code) used as a template) exhibited energy over the threshold value. Cartoons were drawn with PyMOL (W.L. DeLano (<http://pymol.sourceforge.net>)). nsLTP from other fruit, e.g. from cherry (Pru av 3), apple (Mal d 3) and pear (Pyr c 3), were similarly modelled from the atomic coordinates of Zea m 14.

Electrostatic potentials were calculated and displayed with GRASP using the parse3 parameters [24]. The solvent probe radius used for molecular surfaces was 1.4 Å and a standard 2.0 Å Stern layer was used to exclude ions from the molecular surface [25]. The inner and outer dielectric constants applied to the protein and the solvent were respectively fixed to 4.0 and 80.0 and the calculations were performed keeping a salt concentration of 0.145 M.

Linear B-cell epitopes were predicted from the hydrophobic profiles as being the most hydrophilic, flexible and surface exposed regions. Different scales of hydrophilicity [26], flexibility [27], exposition to the solvent [28] and antigenicity [29], were used to build the hydrophobic profiles with the MacVector (Kodak) software. The most hydrophilic/flexible/accessible amino acid sequence stretches were finally retained as putative IgE-binding epitopes providing they contain either acidic (Asp, Glu) or basic (Arg, Lys). Other predicted hydrophilic/flexible/accessible amino acid sequence stretches devoid of charged residues were discarded.

The surface occupied by the previously characterized [30] or predicted sequential B-cell epitopes along the amino acid sequence of Pru p 3 was calculated and displayed on the molecular surface of the proteins with PyMOL. The overall conformation of the sequential B-cell epitopes on the molecular surface was displayed with PyMOL.

2.2. Isolation of nsLTP

nsLTP were purified from the pulp of mature peach and apple fruits. Usually, 200 g of pulp previously frozen in liquid nitrogen were finely ground in mortar and pestle and suspended in 1 L of 20 mM Tris-buffered saline (TBS) containing 2 mM EDTA and 3% (w/v) insoluble PVP (Polyclar AT, Sigma) (pH 7.5). The slurry was stirred overnight at 4 °C and then centrifuged 20 min at 16,000 × g and 20 min at 30,000 × g. The clear supernatant was poured in SpectraPor dialysis tubing (mol. wt, cut-off 3500 Da) and dialyzed against distilled water for 72 h at 4 °C. After centrifugation at 30,000 × g for 20 min at 4 °C, the supernatant was lyophilized and stored at −80 °C until used. Lyophilized protein samples were dissolved in 50 mM phosphate buffer (pH 7.5) and loaded onto a 5 ml HiTrap SP FF column (Amersham) previously equilibrated with 50 mM phosphate buffer (pH 7.5). The retained LTP were eluted with 50 mM phosphate buffer, 50 mM NaCl (pH 7.5) and checked for purity. Purified Pru p 3 and Mal d 3 were dialysed against bi-distilled water, lyophilized and stored at −80 °C until used.

2.3. Analytical methods

Nonapeptides corresponding to the predicted epitope #2 of Pru p 3 (39RTTPDRQAA 47) and Mal d 3 (39RTTADRQTA47) were synthesized using Fmoc chemistry and coupled to thyroglobulin used as a carrier protein with glutaraldehyde.

The purity of the nsLTP preparations was checked by SDS–PAGE in 15% polyacrylamide gels using Tris–taurine as trailing ion [31] and staining with silver nitrate [32]. Coomassie blue stained bands were digested with trypsin in the gel and mass mapped by MALDI-TOF analysis as previously described [32]. The software Protein Prospector was used for the identification of the protein using the NCBI non-redundant database. For MALDI-TOF-MS analysis, nsLTP samples of 0.75 µl of a 0.5 mg ml^{−1} protein solution in 50 mM phosphate buffer, 50 mM NaCl (pH 7.5) were co-crystallized on the MALDI plate with 0.75 µl of 0.6 mM of 3,5-dimethoxy-4-hydroxy cinnamic acid (sinapinic acid) in 50% (w/v) of azido nitrile used as a matrix solution. Desorption and ionization of crystallized samples were carried out on a Voyager-DE™ STR (Perspective Biosystems, USA) mass spectrometer in positive linear mode using an accelerating voltage of 25 kV, a grid voltage of 93% and an extraction delay time of 800 ns. Acquisition mass was performed between 3000 and 25,000 Da using a mixture of three proteins of known mass molecular weight (insulin: 5734 Da; thioredoxin: 11,674 Da; apomyoglobin: 16,952 Da) for the internal calibration.

The protein content of the nsLTP samples was estimated using the bicinchonic acid kit reagent (Pierce) [33] with bovine serum albumin as a standard.

2.4. Immunochemical methods

Polyclonal antibodies against peach and apple nsLTP were raised in rabbits in the presence of Freund's complete adjuvant (Calbiochem). The allergens (100 µg) freshly dissolved in

saline (400 μ l) and emulsified with adjuvant (400 μ l) were repeatedly injected subcutaneously (200 μ l) at 10-day intervals during a period of 4 months. Every 3 weeks, blood was collected and checked for polyclonal antibodies by ELISA. After blood clotting at room temperature and at 4 °C for 4 h and 12 h, respectively, the immune sera were collected by centrifugation, filtered through a 0.2 μ m membrane and stored at –80 °C until used.

Polyclonal antibodies against synthetic epitope #2 of Pru p 3 (39RTTPDRQA47) and Mal d 3 (39RTTADRQTA47) coupled to thyroglobulin were raised in rabbits following a similar immunization schedule. A series of 12 intradermal injections of 250 μ g of conjugates emulsified in Freund's complete adjuvant were performed over a 6-month period before bleeding and collection of the immune sera. Anti-peptide sera were filtered through a 0.2 μ m membrane and stored at –80 °C until used.

IgE-containing sera from patients allergic to fruit nsLTP were used in Western blots and ELISA measurements as probes for Pru p 3 and Mal d 3. Following 1D SDS–PAGE, proteins were transferred onto a Protran nitrocellulose 0.2 μ m membrane (Schleicher and Schuell) at 20 V for 45 min using a 48 mM Tris/39 mM glycine/20% (v/v) methanol mixture. After an overnight incubation in 10 mM PBS (pH 7.4) containing 0.2% (v/v) Tween and 5% (v/v) skimmed milk, the membrane was soaked in the patient IgE-containing sera diluted 1:10 in the same buffer and incubated for 2 h in a moist chamber. After three washings of 10 min each with the same buffer, the membrane was soaked in rabbit HRP-labelled anti-human IgE diluted 1/5000 in the buffer and incubated for 1 h under gentle stirring. Following three washings of 10 min each with buffer, the immunolabelled spots were detected using the ECL Plus detection (Amersham) after 3 min exposure in cassette. All the handling were carried out at room temperature.

ELISA measurements were performed on standard 96-well microtitration plates (Limbro) using whether polyclonal antibodies (Ab) against Mal d 3 or Pru p 3 or polyclonal antibodies raised against synthetic epitopes. Routinely, the wells were loaded with 50 μ l of 10 mM PBS (pH 7.5) containing 40 ng of Pru p 3/Mal d 3. The plates were incubated overnight at 4 °C and washed 3 times with PBS (pH 7.5). 100 μ l of PBS (pH 7.5) containing 0.2% Tween 20 (v/v) and 3% BSA (w/v) (PBSTB) were added to each well and the plates were incubated at room temperature for 2 h. 50 μ l of tenfold-diluted Ab in PBSTB (pH 7.5) were added and the plates were incubated for 1 h 30 min at room temperature under constant stirring. After 3 washes with PBSTB, 50 μ l of either 1/2000 (v/v) diluted sheep anti-rabbit IgG coupled to alkaline phosphatase (Dako) or 1/300 (v/v) diluted mouse monoclonal anti-human IgE coupled to alkaline phosphatase (Sigma) were added and, after incubation at room temperature for 45 min, 3 washes were performed with PBSTB. 100 μ l of alkaline phosphatase substrate (Sigma) were added and after incubation for 45 min in the dark, 50 μ l of 3 M NaOH were added to stop the colour reaction. The absorbance at 405 nm was recorded on a Titertek Multiscan spectrophotometer. Each value is the

mean of three separate experiments. Appropriate blank and substrate controls were performed under the same conditions.

Increasing concentrations of synthetic peptides ranging from 20 up to 300 μ g ml^{–1} were pre-incubated for 1 h with the anti-epitope Ab or an IgE-containing serum from an allergic patient to inhibit the Pru p 3/Mal d 3–Ab interaction measured by ELISA.

3. Results

nsLTP from edible fruit, e.g. peach, apricot or plum, constitute a very homogeneous group of closely related small proteins that share very similar amino acid sequences (Fig. 1). The overall percentages of identity and similarity (calculated using the structural homology matrix of Risler et al. [19]) among their amino acid sequences amount 70% and 90%, respectively. Eight conserved Cys residues forming four disulfide bridges play an important role in stabilizing the whole molecule. Accordingly, LTP exhibit an extreme stability to heat denaturation or proteolytic attack [34] and thus can escape the thermal denaturation process involved in food processing [35].

Consistent with the high percentage of amino acid sequence identity and the HCA plot similarity (result not shown), a homology modelling approach was performed to build an accurate three-dimensional model of the fruit nsLTP. As an example, the three-dimensional model of Pru p 3 built from the X-ray coordinates of Zea m 14 exhibits the classical plant nsLTP fold [36] which consists of four α -helices α 1, α 2, α 3 and α 4, connected by loops (Fig. 2A). The extended C-terminus of the polypeptide chain has no apparent secondary structure. The accuracy of our model building approach was checked by superposing the three-dimensional model to the recently X-ray solved structure of Pru p 3 (PDB code 2ALG). The superposition of the α -carbons of both 91 residue structures yielded a rmsd of 1.31 Å. Moreover, all the α -helices were correctly delineated but the main discrepancy concerns the shape of the extended loop connecting α 1 to α 2 helices (Fig. 3). Very similar models were accurately built for other fruit nsLTP including Mal d 3 from apple, Pru av 3 from cherry and Pyr c 3 from pear.

Four linear IgE-binding epitopes containing charged residues were predicted to occur along the amino acid sequence of Pru p 3 using both hydropathic profiles based on different scales of hydrophilicity, flexibility and exposition to the solvent and HCA plots. They correspond to the amino acid sequence stretches 18RGGG21 (epitope #1), 29NGIRNVNN36 (epitope #2), 39RTTPDRQA46 (epitope #2') and 72KCGVHIPYKI81 (epitope #3), respectively (Fig. 2C). These predicted linear epitopes correspond to exposed regions of the molecule containing well exposed electropositively charged residues (Fig. 2D,E). Most of these charged residues occur in the surface exposed areas possessing the higher electropositive potentials (Fig. 2F). These sequence stretches nicely match (underlined residues) the three linear epitope sequences 11APCIPYVRGGGAVPP25 (epitope #1), 30GIRNVNNLAR~~TTTPDR~~Q45 (epitope #2) and 70PGKCGVHIPYK80 (epitope #3) previously recognized

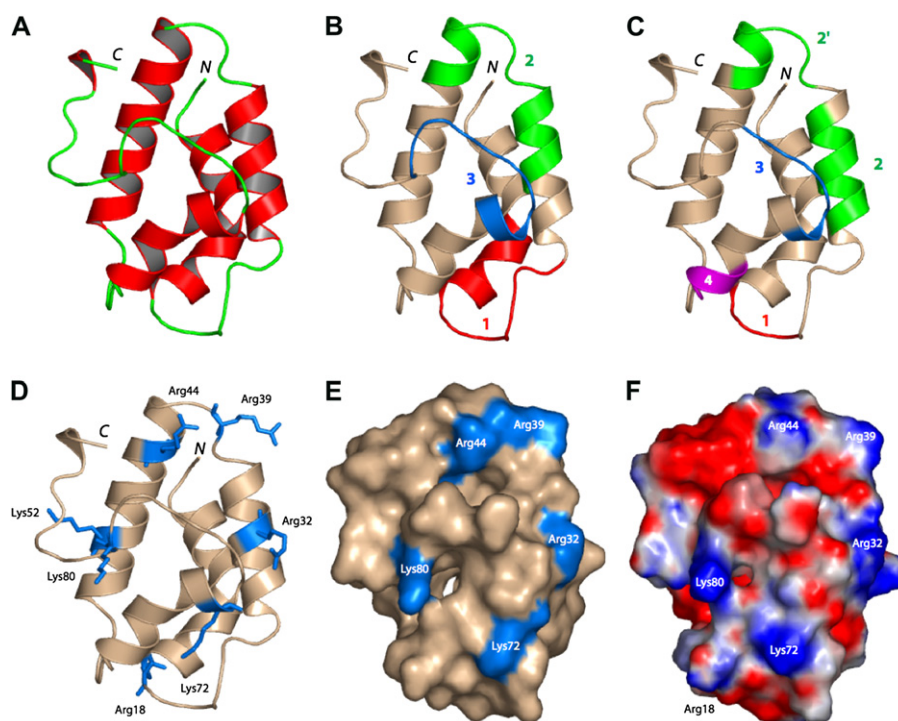


Fig. 2. (A) Ribbon diagram of the modelled Pru p 3 showing the three α -helices α_1 , α_2 , α_3 and α_4 (red) connected by turns and loops (green). *N* and *C* correspond to the N- and C-termini of the polypeptide chain, respectively. (B) Location of the characterized IgE-binding epitopes 1 (red), 2 (green) and 3 (blue) on the ribbon diagram of Pru p 3. (C) Location of the predicted IgE-binding epitopes 1 (red), 2–2' (green), 3 (blue) and 4 (magenta) on the ribbon diagram of Pru p 3. Epitope 4, which does not contain charged residues, was omitted. (D) Exposition of the electropositively charged residues Lys and Arg on the ribbon diagram of Pru p 3. (E) Exposition of electropositively charged residues on the molecular surface of Pru p 3. (F) Mapping of the electrostatic potentials on the molecular surface of Pru p 3. Electropositive and electronegative areas are coloured blue and red, respectively. All the models are similarly oriented. All the cartoons were drawn with PyMOL.

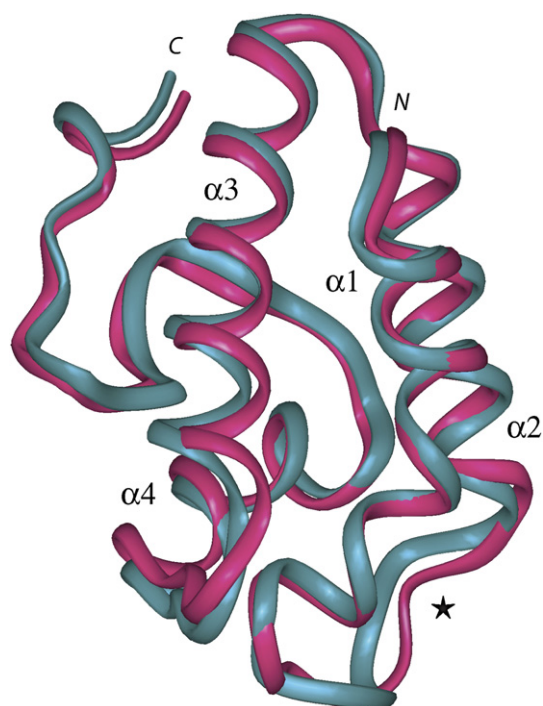


Fig. 3. Ribbon diagram showing the superimposition of the α -carbon tracings from the X-ray solved Pru p 3 (pink ribbon) and the homology modelled Pru p 3 (blue-green ribbon). *N* and *C* correspond to the N- and C-termini of the polypeptide chain, respectively. The star (★) indicates the extended loop connecting α_1 and α_2 helices were the most important discrepancy between the structure and the model occurs. Molecules are similarly oriented as in Fig. 2.

from the IgE-mapping of Pru p 3 using overlapping synthesized oligopeptides covering the entire amino acid sequence [32] (Fig. 2B). In fact, another amino acid sequence stretch, 61VNPNN65, previously predicted as an IgE-binding epitope according to its hydrophilicity, flexibility and accessibility, and referred as epitope #4 (see Fig. 2C), was further discarded since it does not contain any charged residue. Apparently, this sequence stretch does not correspond to an experimentally defined epitope on the peach LTP.

The four putative IgE-binding epitopes predicted on Pru p 3, which correspond to rather conserved amino acid sequence stretches, were also predicted to occur in other fruit nsLTP. Besides fruit nsLTP, the predicted epitope #2' of Pru p 3, which roughly overlaps the loop connecting α_2 to α_3 helices and the beginning of helix α_3 , was also predicted as a putative B-cell epitope in all other plant nsLTP (Fig. 4). This conserved amino acid sequence stretch exhibits a very similar overall conformation in other modelled nsLTP from Rosaceae fruits including Mal d 3 from apple, Pru ar 3 from apricot or Pru av 3 from cherry (Fig. 5). In this respect, the rmsd measured for pairwise superimposed epitopes, e.g. epitope of Pru p 3 vs. epitope of Mal d 3, gave values ≤ 0.1 Å. The calculated accessible surface area gave very similar values of e.g. 1177 Å² for Mal d 3, 1190 Å² for Pru ar 3 and 1136 Å² for Pru p 3. These values indicate, indeed, that the predicted consensus epitope exhibits a similar local structure and occupy a comparable area at the surface of the allergens. In addition, charged

Mal d 3	---ITCGQVSSSLAPFCIGYVRGG-AVPPACCGIRTINGLARTTADRQTACNC-LKLAGSI-SGVNPNNAAGLPKCGVNVFYKISTSTNCATVK-
Pyr c 3	---ITCGQVSSSLAPFCIGYVRGG-AVPPACCGIRTINGLARTTADRQTACNC-LKLAGSV-SGVNPNNAAGLPKCGVNVFYKISTSTNCATVK-
Pru p 3	---ITCGQVSSSLAPFCIGYVRGG-AVPPACCGIRTINGLARTTADRQTACNC-LKLAGSV-SGVNPNNAAGLPKCGVNVFYKISTSTNCATVK-
almond	---ITCGQVSSSLAPFCIGYVRGG-AVPPACCGIRTINGLARTTADRQTACNC-LKLAGSV-SGVNPNNAAGLPKCGVNVFYKISTSTNCATVK-
Pru av 3	---ITCGQVSSSLAPFCIGYVRGG-AVPPACCGIRTINGLARTTADRQTACNC-LKLAGSV-SGVNPNNAAGLPKCGVNVFYKISTSTNCATVK-
Pru ar 3	---ITCGQVSSSLAPFCIGYVRGG-AVPPACCGIRTINGLARTTADRQTACNC-LKLAGSV-SGVNPNNAAGLPKCGVNVFYKISTSTNCATVK-
Pru d 3	---ITCGQVSSSLAPFCIGYVRGG-AVPPACCGIRTINGLARTTADRQTACNC-LKLAGSV-SGVNPNNAAGLPKCGVNVFYKISTSTNCATVK-
cabbage	NRALTCTGVNSNVAPCIGYITGG-TLPGACCTGVSKLNSMARTTADRQTACNC-LKLAGSV-SGVNPNNAAGLPKCGVNVFYKISTSTNCATVK-
rape	---ALSCGTVSGYVAPCIGYLAQAPALPRACCGVTSLSNLMARTTADRQTACNC-LKLAGSV-SGVNPNNAAGLPKCGVNVFYKISTSTNCATVK-
arabidopsis	---ALSCGTVSGYVAPCIGYLAQAPALPRACCGVTSLSNLMARTTADRQTACNC-LKLAGSV-SGVNPNNAAGLPKCGVNVFYKISTSTNCATVK-
hazel nut	---SLTCPQIKGNLTFCVLYLKNKG-VLPSCCGKGVRAVNDASRTTSDRQSACNC-LKDTAKGI-AGLNPNLAAGLPKCGVNVFYKISTSTNCATVK-
soya	---MTCCGVQGNLAQICGLQKGG-VVPPSCCTGVKILNSRTTADRQTACNC-LKLAGSV-SGVNPNNAAGLPKCGVNVFYKISTSTNCATVK-
chickpea	---AITCGRVDTALAPCLGYLQGG-GPSAQCCGGVRLNSAARTTADRQTACNC-LKLAGSV-SGVNPNNAAGLPKCGVNVFYKISTSTNCATVK-
barley	---LNCQVDSKMKPCLTYVQG-GPGSPGCCNGVRDLHNQAGSSGDRQTVCNC-LKGIARGI-HNLLNNAASIPSKCNVNVFYKISTSTNCATVK-
wheat	---IDCGHVDLSVRPCLSYVQG-GPGSPGCCNGVRLNHNQARSQSDRQSACNC-LKGIARGI-HNLLNNAASIPSKCNVNVFYKISTSTNCATVK-
maize	---AISCQVSAALAPCISYARGQSSGVSAGCCSGVSLNHAARTTADRQTACNC-LKLAGSV-SGVNPNNAAGLPKCGVNVFYKISTSTNCATVK-
sorghum	---AISCQVSAALAPCISYARGQSSGVSAGCCSGVSLNHAARTTADRQTACNC-LKLAGSV-SGVNPNNAAGLPKCGVNVFYKISTSTNCATVK-
rice	---AISCQVSAALAPCISYARGQSSGVSAGCCSGVSLNHAARTTADRQTACNC-LKLAGSV-SGVNPNNAAGLPKCGVNVFYKISTSTNCATVK-
millet	---AISCQVSAALAPCISYARGQSSGVSAGCCSGVSLNHAARTTADRQTACNC-LKLAGSV-SGVNPNNAAGLPKCGVNVFYKISTSTNCATVK-

Fig. 4. Multiple alignment of plant LTP from different organs (mainly seeds or seedlings). The more or less extended predicted consensual IgE-binding epitope (corresponding to epitope 2' of Pru p 3) is boxed grey. Except for fruit LTP allergens, other plant LTP are labelled by the common names of the corresponding plants.

residues are similarly distributed within this region as shown from the mapping of the electrostatic potentials on the molecular surfaces. Other X-ray or RMN solved nsLTP also exhibit a very conserved overall conformation in this region (result not shown). This both sequentially and conformationally

conserved region of nsLTP most probably represents a consensus B-cell epitope responsible for IgE-binding cross-reactions in sensitized individuals. The consensual character of this epitopic region was further confirmed by ELISA experiments showing that rabbit antibodies raised against the Pru p 3 epitope cross-reacted with Mal d 3. Antibodies raised against the Mal d 3 epitope similarly cross-reacted with Pru p 3. In addition, the synthetic peptides from Pru p 3 (39RTTPDRQAA) and Mal d 3 (39RTTADRQTA46) both inhibited the Ab-nsLTP interactions (Fig. 6A). Inhibition experiments were similarly performed with IgE-containing sera from allergic patients (Fig. 6B).

4. Discussion

Taking into account the occurrence of electropositively (Arg, Lys) and electronegatively (Asp, Glu) charged residues

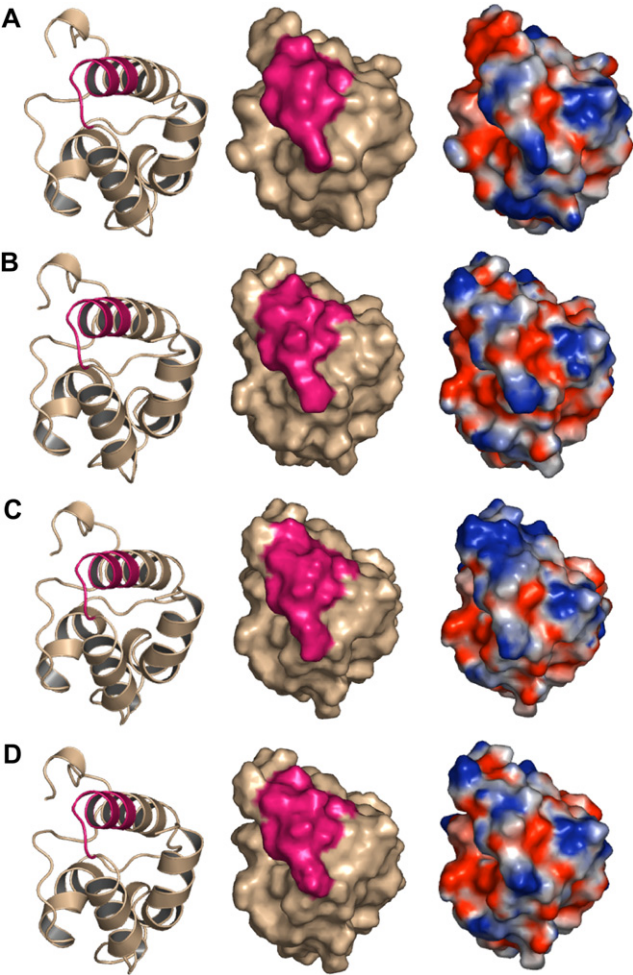


Fig. 5. Ribbon diagram (left), overall conformation of the consensus IgE-binding epitope #3 (middle), and mapping of the electrostatic potentials (left) on the molecular surface of the three-dimensional models of Pru p 3 (A), Pru d 3 (B), Pru ar 3 (C) and Pru av 3 (D). The consensus epitope is coloured pink, and positive and negative electrostatic potentials are coloured blue and red, respectively.

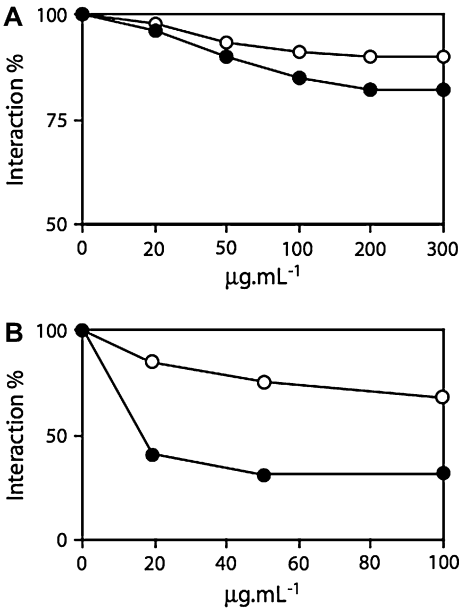


Fig. 6. Inhibition by increasing concentrations (in $\mu\text{g ml}^{-1}$) of the synthetic epitope 39RTTADRQTA47 from Mal d 3 of the interaction of Pru p 3 (○) and Mal d 3 (●) with polyclonal rabbit antibodies raised against the synthetic epitope (A) and an IgE-containing serum from an allergic patient (B). Each value is the mean of two separate ELISA measurements.

in the regions that have been predicted to be the more hydrophilic, flexible and exposed to the solvent, both the hydropathic profiles and HCA plots allowed a rather accurate prediction of the linear IgE-binding epitopes on the surface of fruit nsLTP. Obviously, the aim of our predictive approach is not to correctly predict all the IgE-binding epitopes but to locate with a fairly good accuracy some of these epitopes on the surface of the allergens. When available, the three-dimensional structure of a closely related protein is of paramount importance to build a three-dimensional model that allows great improvement in the prediction of the exposed charged residues on the molecular surface of the allergens. This is the case for the fruit nsLTP that have been easily modelled from the X-ray coordinates of the maize LTP (RCSB PDB code 1MZL), thus allowing a more

accurate localization of the exposed electronegative and electropositive residues protruding out of the protein surface and the mapping of the electrostatic potentials on the protein surface.

Residues Arg39, Thr40, Arg44, Lys80 and Lys91 of Pru p 3 have been previously predicted as potential IgE recognition sites according to both their surface exposition and electrostatic properties [34]. Except for Lys91, all other residues are included in our predicted linear B-cell epitopes. In agreement with the involvement of residues Arg39, Thr40 and Arg44 in a major IgE-binding determinant of Pru p 3, a substantial 5-fold decrease of the IgE binding activity was reported for the triple mutant Arg39Ala/Thr40Ala/Arg44Ala of recombinant Pru p 3. These residues are located in the short

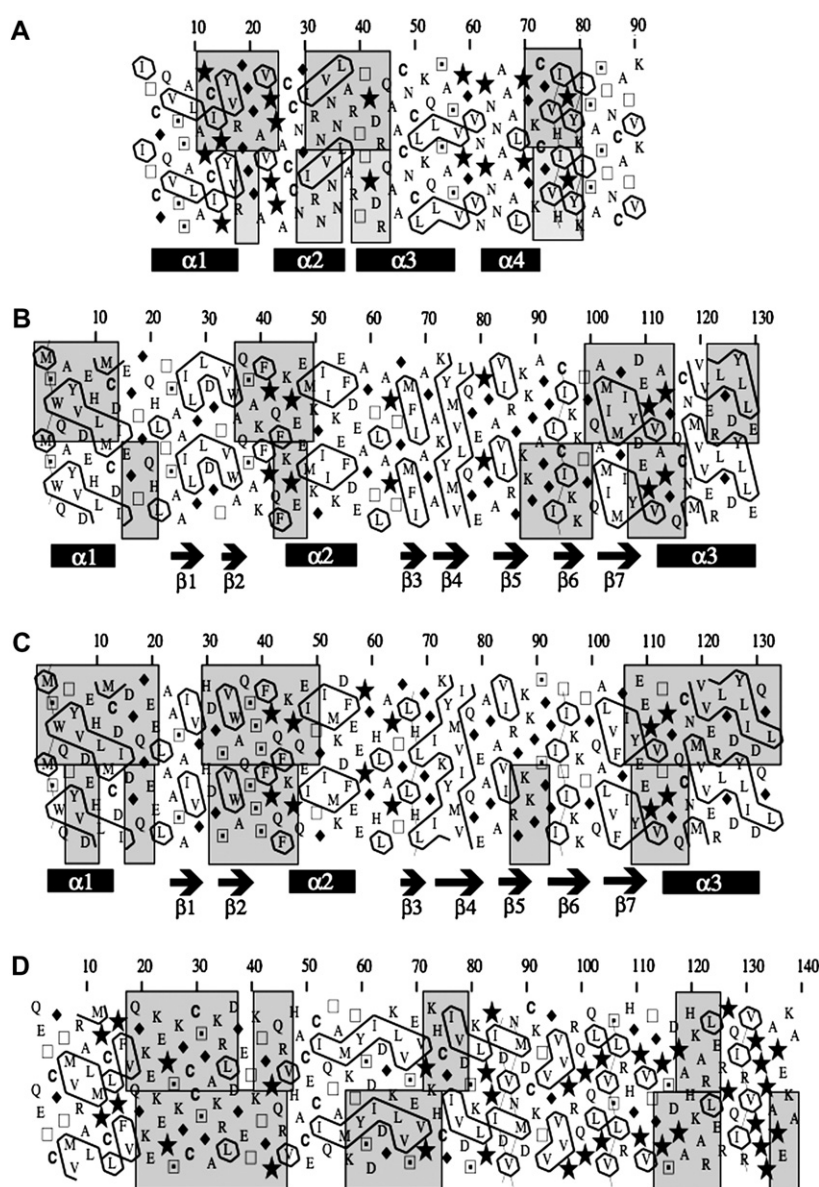


Fig. 7. HCA plots of Pru p 3 (peach LTP) (A), Hel t 2 (profilin from *Helianthus tuberosus*) (B), Bet v 2 (profilin from *Betula verucosa*) (C) and Par j 1 (long LTP from *Parietaria judaica*) (D), showing the delineation of experimentally characterized (upper grey boxes) or predicted (lower grey boxes) linear IgE-binding epitopes. When available, the X-ray- or NMR-assigned β -sheet (β) and α -helix (α) secondary structural features are indicated. Clusters of hydrophobic residues (I, L, M, V, F, Y, W) are circled in black and special symbols are used for Gly (◆), Ser (◻), Thr (⊞), Pro (★) and Cys (C).

loop (Arg39, Thr40) connecting the $\alpha 2$ and $\alpha 3$ helices and at the beginning of helix $\alpha 3$ (Arg 44). They are fully conserved in other fruit LTP (see Fig. 1) but also occur in most of the plant nsLTP of different origins. Interestingly, the well-characterized linear B-cell epitope #3 that contains the three Arg, Thr and Arg residues is systematically predicted as a putative IgE-binding region in all other nsLTP. It could thus correspond to a consensus epitope perhaps responsible for the IgE-binding cross-reactivity commonly observed among the plant nsLTP [13,37]. In this respect, our predictive approach can be of valuable help for the identification of consensual IgE-binding epitopes in conserved families of protein allergens.

Although our predictive approach allowed to rather accurately predict the location of linear IgE-binding epitopes along the amino acid sequence of plant nsLTP, some discrepancies were observed when applied to other plant allergens including other longer nsLTP (Par j 1 from *Parietaria judaica*) or profilins (Hel t 2 from *Helianthus tuberosus* and Bet v 2 from *Betula verucosa*) (Fig. 7). Obviously, most of the experimentally characterized IgE-binding epitopes of Bet v 2 [38] and Par j 1 [39] were correctly predicted whereas the prediction is less reliable for Hel t 2 [40], even though a seven-residue epitopic region predicted near position 90 of Bet v 2 sequence received no experimental evidence. Although the profilins Hel t 2 and Bet v 2 share very similar three-dimensional conformations as checked from their HCA plots, both their experimentally characterized or predicted linear IgE-binding epitopes are rather different. A few amino acid changes occurring along the amino acid sequences of the allergens are apparently responsible for these discrepancies.

When applied to allergens of longer amino acid sequences, e.g. to the major allergen Ara h 1 of peanut (*Arachis hypogaea*), our predictive approach failed to correctly predict the nine linear IgE-binding epitopes previously characterized along the N-terminus of the protein [41]. Almost the entire N-terminal end of the allergen was thus incorrectly predicted as a single epitope. Due to the widespread distribution of hydrophilic and charged residues along this sequence stretch, our combination of predictive tools was unable to discriminate among the nine discrete IgE-binding epitopes that have been previously shown to occur in this protein region. In addition, no structural information is available on this N-terminal region that prevents to check the exposition of the predicted regions on the protein surface. More consistent results were obtained with the C-terminal region of known structure where most of the remaining characterized B-epitopes were correctly predicted [42]. However, a remark has to be made on the accuracy of delineating linear IgE-binding epitopes from IgE-mapping experiments since over the 23 characterized epitopes of Ara h 1 a few of them, e.g. epitopes #17 (RRYTARLKEG) and #18 (ELHLLGFGIN), are poorly exposed on the surface of the modelled monomer (result not shown). Most probably, a very few residues of such amino acid sequence stretches are sufficiently exposed to actually participate in IgE-binding epitopes while the remainder are too buried to be accessible to the IgE antibodies. In this respect, checking the exposure and conformation of the IgE-binding epitopes on the molecular

surface of the allergens appears as an indispensable step to confirm the epitopic character of an amino acid sequence stretch. In addition, surface analysis of the allergens is of great value to recognize more complex conformational epitopes resulting from the coalescence of several discrete linear IgE-binding epitopes on the molecular surface.

Acknowledgements

The financial support of CNRS is gratefully acknowledged.

References

- [1] S. Fraga, Theoretical prediction of protein antigenic determinants from amino acid sequences, *Can. J. Chem.* 60 (1982) 2606–2610.
- [2] B.A. Jameson, H. Wolf, The antigenic index: a novel algorithm for predicting antigenic determinants, *Comput. Appl. Biosci.* 4 (1988) 181–186.
- [3] J.L. Pellequer, E. Westhof, M.H. Van Regenmortel, Predicting location of continuous epitopes in proteins from their primary structures, *Methods. Enzymol.* 203 (1991) 176–201.
- [4] J.L. Pellequer, E. Westhof, M.H. Van Regenmortel, Correlation between the location of antigenic sites and the prediction of turns in proteins, *Immunol. Lett.* 36 (1993) 83–99.
- [5] J.M. Thornton, M.S. Edwards, W.R. Taylor, D.J. Barlow, Location of continuous antigenic determinants in the protruding regions of proteins, *EMBO J* 5 (1986) 409–413.
- [6] M.J. Blythe, D.R. Flower, Benchmarking B cell epitope prediction: underperformance of existing methods, *Protein Sci.* 14 (2005) 246–248.
- [7] M.H. Van Regenmortel, J.L. Pellequer, Predicting antigenic determinants in proteins: looking for unidimensional solutions to a three-dimensional problem, *Pept. Res.* 7 (1994) 224–228.
- [8] D.R. Flower, Towards *in silico* prediction of immunogenic epitopes, *Trends Immunol* 24 (2003) 667–674.
- [9] T. Midoro-Horiuti, E.G. Brooks, R.M. Goldblum, Pathogenesis-related proteins of plants as allergens, *Ann. Allergy Asthma Immunol.* 87 (2001) 261–271.
- [10] H. Breiteneder, C. Radauer, A classification of plant food allergens, *J. Allergy Clin. Immunol.* 113 (2004) 821–830.
- [11] R. Sanchez-Monge, M. Lombardero, F.J. Garcia-Selles, D. Barber, G. Salcedo, Lipid-transfer proteins are relevant allergens in fruit allergy, *J. Allergy Clin. Immunol.* 103 (1999) 514–519.
- [12] E.A. Pastorello, C. Pompei, V. Pravettoni, O. Brenna, L. Farioli, C. Trambaioli, A. Conti, Lipid transfer proteins and 2S albumins as allergens, *Allergy* 56 (2001) 45–47.
- [13] R. van Ree, Clinical importance of non-specific lipid transfer proteins as food allergens, *Biochem. Soc. Trans.* 30 (2002) 910–913.
- [14] J.D. Thompson, T.J. Gibson, F. Plewniak, F. Jeanmougin, D.G. Higgins, The CLUSTAL-X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tool, *Nucleic Acids Res.* 15 (1997) 4876–4882.
- [15] P. Gouet, E. Courcelle, D.I. Stuart, F. Metoz, ESPript: analysis of multiple sequence alignments in PostScript, *Bioinformatics* 15 (1999) 305–308.
- [16] C. Gaboriaud, V. Bissery, T. Benchetrit, J.P. Mornon, Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences, *FEBS Lett.* 224 (1987) 149–155.
- [17] D.H. Shin, J.Y. Lee, K.Y. Hwang, K.K. Kim, S.W. Suh, High-resolution crystal structure of the non-specific lipid-transfer protein from maize seedlings, *Structure* 3 (1995) 189–199.
- [18] J.-L. Risler, M.O. Delorme, H. Delacroix, A. Henaut, Amino acid substitutions in structurally related proteins. A pattern recognition approach. Determination of a new efficient scoring matrix, *J. Mol. Biol.* 204 (1998) 1019–1029.

- [19] J.W. Ponder, F.M. Richards, Tertiary templates for proteins. Use of packing criteria in the enumeration of allowed sequences for different structural classes, *J. Mol. Biol.* 193 (1987) 775–791.
- [20] M.T. Mas, K.C. Smith, D.L. Yarmush, K. Aisaka, R.M. Fine, Modeling the anti-CEA antibody combining site by homology and conformational search, *Proteins Struct. Func. Genet.* 14 (1992) 483–498.
- [21] Roussel, A., Cambillau, C., in: Turbo-Frodo, Silicon Graphics Geometry Partners Directory, Silicon Graphics, Mountain View, CA.
- [22] R.A. Laskowski, M.W. MacArthur, D.S. Moss, J.M. Thornton, PROCHECK: a program to check the stereochemistry of protein structures, *J. Appl. Cryst.* 26 (1993) 283–291.
- [23] F. Melo, E. Feytmans, Assessing protein structures with a non-local atomic interaction energy, *J. Mol. Biol.* 277 (1998) 1141–1152.
- [24] A. Nicholls, K.A. Sharp, B. Honig, Protein folding and association: Insights from the interfacial and thermodynamic properties of hydrocarbons, *Proteins Struct. Func. Genet.* 11 (1991) 281–296.
- [25] M.K. Gilson, B.H. Honing, Calculation of electrostatic potential in an enzyme active site, *Nature* 330 (1987) 84–86.
- [26] J. Kyte, R.F. Doolittle, A simple method for displaying the hydropathic character of a protein, *J. Mol. Biol.* 157 (1982) 105–132.
- [27] P.A. Karplus, G.E. Schulz, Prediction of chain flexibility in proteins, *Naturwissenschaften* 72 (1985) 212–213.
- [28] B. Lee, F.M. Richards, The interpretation of protein structures: estimation of static accessibility, *J. Mol. Biol.* 55 (1971) 379–400.
- [29] T.P. Hopp, K.R. Woods, Prediction of protein antigenic determinants from amino acid sequences, *Proc. Natl. Acad. Sci. USA* 78 (1981) 3824–3828.
- [30] G. Garcia-Casado, L.F. Pacios, A. Diaz-Perales, R. Sanchez-Monge, M. Lombardero, F.J. Garcia-Selles, F. Polo, D. Barber, G. Salcedo, Identification of IgE-binding epitopes of the major peach allergen Pru p 3, *J. Allergy Clin. Immunol.* 112 (2003) 599–605.
- [31] C. Tastet, P. Lescuyer, H. Diemer, S. Luche, A. van Dorsselaer, T. Rabilloud, A versatile electrophoresis system for the analysis of high- and low-molecular-weight proteins, *Electrophoresis* 24 (2003) 1787–1794.
- [32] A. Shevchenko, Wilm, M. Wilm, O. Vorm, O.N. Jensen, A.V. Podtelejnikov, G. Neubauer, A. Shevchenko, P. Mortensen, M. Mann, A strategy for identifying gel-separated proteins in sequence databases by MS alone, *Biochem. Soc. Trans.* 24 (1996) 893–896.
- [33] P.K. Smith, R.I. Krohn, G.T. Hermanson, A.K. Mallia, F.H. Gartner, M.D. Provenzano, E.K. Fujimoto, N.M. Goeke, B.J. Olson, D.C. Klenk, Measurement of protein using bicinchoninic acid, *Anal. Biochem* 150 (1987) 76–85.
- [34] R. Asero, G. Mistrello, D. Roncarolo, S.C. de Vries, M.F. Gautier, C.L. Ciurana, E. Berbeek, T. Mohammadi, V. Knul-Brettlova, J.H. Akkerdaas, I. Bulder, R.C. Aalberse, R. van Ree, Lipid transfer protein: a pan-allergen in plant-derived foods that is highly resistant to pepsin digestion, *Int. Arch. Allergy Immunol.* 124 (2001) 67–69.
- [35] S. Scheurer, I. Lauer, K. Foetisch, M.S. Moncin, M. Retzek, C. Hartz, E. Enrique, J. Lidholm, A. Cistero-Bahima, S. Vieths, Strong allergenicity of Pru av 3, the lipid transfer protein from cherry, is related to high stability against thermal processing and digestion, *J. Allergy Clin. Immunol.* 114 (2004) 900–907.
- [36] E. Gincel, J.P. Simorre, A. Caille, D. Marion, M. Ptak, F. Vovelle, Three-dimensional structure in solution of a wheat lipid-transfer protein from multidimensional 1H-NMR data. A new folding for lipid carriers, *Eur. J. Biochem* 226 (1994) 413–422.
- [37] E.A. Pastorello, F.P. D'Ambrosio, V. Pravettoni, L. Farioli, G. Giuffrida, M. Monza, R. Ansaloni, D. Fortunato, E. Scibola, F. Rivolta, C. Incorvaia, A. Bengtsson, A. Conti, C. Ortolani, Evidence for a lipid transfer protein as the major allergen of apricot, *J. Allergy Clin. Immunol.* 105 (2000) 371–377.
- [38] J.A. Asturias, N. Gómez-Bayón, J.L. Eseverri, A. Martínez, Par j 1 and Par j 2, the major allergens from *Parietaria judaica* pollen, have similar immunoglobulin E epitopes, *Clin. Exp. Allergy* 33 (2003) 518–524.
- [39] J.A. Asturias, N. Gómez-Bayón, M.C. Arilla, L. Sánchez-Pulido, A. Valencia, A. Martínez, Molecular and structural analysis of the panallergen profilin B cell epitopes defined by monoclonal antibodies, *Int. Immunol.* 14 (2002) 993–1001.
- [40] A.A. Fedorov, T. Ball, N.M. Mahoney, R. Valenta, S.C. Almo, The molecular basis for allergen cross-reactivity: crystal structure and IgE-epitope mapping of birch pollen profilin, *Structure* 5 (1997) 33–45.
- [41] D.S. Shin, C.M. Compadre, S.J. Maleki, R.A. Kopper, H. Sampson, S.K. Huang, A.W. Burks, G.A. Bannon, Biochemical and structural analysis of the IgE binding sites on Ara h1, an abundant and highly allergenic peanut protein, *J. Biol. Chem.* 273 (1998) 13753–13759.
- [42] A. Barre, J.-P. Borges, P. Rougé, Molecular modelling of the major peanut allergen Ara h 1 and other homotrimeric allergens of the cupin superfamily: a structural basis for their IgE-binding cross-reactivity, *Biochimie* 87 (2005) 499–506.