

The amino acid sequence of *Ole e I*, the major allergen from olive tree (*Olea europaea*) pollen

Mayte VILLALBA¹, Eva BATANERO¹, Carlos LÓPEZ-OTÍN², Luis M. SÁNCHEZ², Rafael I. MONSALVE¹, Manuel A. GONZÁLEZ DE LA PEÑA¹, Carlos LAHOZ³ and Rosalía RODRÍGUEZ¹

¹ Departamento de Bioquímica y Biología Molecular, Facultad de Ciencias, Universidad Complutense, Madrid, Spain

² Departamento de Biología Funcional, Facultad de Medicina, Universidad de Oviedo, Spain

³ Departamento de Inmunología, Fundación Jiménez Díaz, Madrid, Spain

(Received May 18, 1993) – EJB 930722/2

The complete primary structure of the major allergen from *Olea europaea* (olive tree) pollen, *Ole e I* (IUIS nomenclature), has been determined. The amino acid sequence was established by automated Edman degradation of the reduced and alkylated molecule as well as of selected fragments obtained by proteolytic digestions. *Ole e I* contains a single polypeptide chain of 145 amino acid residues with a calculated molecular mass of 16331 Da. No free sulfhydryl groups have been detected in the native protein. The molecule contains a putative glycosylation site. A high degree of microheterogeneity has been observed, mainly centered in the first 33% of the molecule. Comparison of *Ole e I* sequence with protein sequence databases showed no similarity with other known allergens. However, it has a 36% and 38% sequence identity with the putative polypeptide structures, deduced, respectively, from nucleotide sequences of genes isolated from tomato anthers and corn pollen, which have been suggested to be involved in the growing of the pollen tube. Therefore, the olive tree allergen may be a constitutive protein of the pollen involved in reproductive functions.

Nowadays, about 15–20% of the human population suffer from some sort of allergy. The exposure by inhalation, ingestion, injection or contact with numerous antigens from diverse sources leads to this hypersensitive reaction in those individuals genetically predisposed. The most common sources of allergens include pollen from different weeds, trees and grasses, as well as molds, animal dander, insect venoms, house dust mites and foods. Moreover, each source usually contains multiple allergens responsible for the hypersensitivity against a particular species. Pollen allergies are dependent on environmental aspects such as geographical and seasonal conditions. Among seasonal allergies, olive-tree(*Olea europaea*)-pollen allergy is a major health problem for humans throughout the mediterranean area [1–3]. At least two major allergens of different size (20.0 kDa and 8.0 kDa) are present in an aqueous extract of olive tree pollen [4]. The allergenic protein of 20.0 kDa, *Ole e I* (according to

IUIS nomenclature, [5]), has been demonstrated as the most important olive tree allergen on the basis of skin reactivity, radioallergosorbent-test inhibition and histamine-release tests [6–8]. This protein was isolated and purified, and some structural features were determined. It is an abundant acidic protein, representing about 1% of the total dry mass of the mature pollen and more than 10% of the total protein extracted in saline media [9]. Although isoforms have been detected, a common 27-amino-acid NH₂-terminal sequence has been determined for highly purified preparations of this allergen [9].

Accurate diagnosis and therapy of Type I allergic diseases require highly purified and well-characterized allergen preparations. Moreover, amino acid sequence analysis of an allergen provides useful information for further assignment of immunological and allergenic roles to specific residues and polypeptide segments. In this study, we report the complete amino acid sequence and other biochemical features of *Ole e I*, the major allergen from *Olea europaea* pollen. Its similarity with other proteins, specific to pollen and anthers, suggests for this allergen a potential role in reproductive functions of the pollen.

EXPERIMENTAL PROCEDURES

Isolation of *Ole e I* allergen

Native *Ole e I* allergen was purified from olive tree (*Olea europaea*) pollen (Allergen AB) as described [9]. Briefly, the

Correspondence to R. Rodríguez, Departamento de Bioquímica y Biología Molecular, Facultad de Química, Universidad Complutense, E-28040 Madrid, Spain

Abbreviations. Pth-Xaa, phenylthiohydantoin-amino acid; RP, reverse-phase; S-V8, Glu-specific staphylococcal V8 protease.

Enzymes. Trypsin (3.4.21.4); chymotrypsin (3.4.21.1); Glu-specific staphylococcal V8 protease (endoproteinase Glu-C) (3.4.21.19); carboxypeptidase A (3.4.17.1); horseradish peroxidase (1.11.1.7).

Note. The novel amino acid sequence data published here have been deposited with the EMBL and Swissprot sequence data bank and are available under accession number P19963.

dried ethyl-ether-defatted olive tree pollen was dissolved in 0.2 M ammonium bicarbonate, pH 8.0, loaded onto a Sephadex G-75 gel-filtration column and the allergenic fractions were lyophilized and chromatographed by reverse-phase (RP)-HPLC using a μ Bondapak C-18 column. The protein was eluted with a linear gradient (from 0 to 60% in 60 min) of acetonitrile containing 0.1% (by vol.) trifluoroacetic acid.

Reduction and carboxyamidomethylation of the allergen was performed as described in [10] by using Iodo[1- ^{14} C]acetamide (0.2 μ Ci/ μ l, Amersham). The alkylated protein was desalted on a Sephadex G-25 column, equilibrated in 0.1 M acetic acid.

Electrophoresis, immunoblotting and carbohydrate detection

Analytical gel electrophoresis was performed according to Laemmli [11], in 15% (mass/vol.) acrylamide gels. Proteins were visualized by Coomassie blue staining. Protein bands in SDS/PAGE were transferred onto ProBlott membranes (Applied Biosystems) by using a LKB Novablot 2117 electrophoretic transfer unit (LKB) in 3-(cyclohexylamino)-1-propanesulfonic acid, pH 11.0, containing 10% (by vol.) methanol, for 2 h at 150 mA, as described in [12]. For immunostaining, a pool of hypersensitive olive tree pollen human sera (1:5 dilution in 10 mM sodium phosphate, pH 7.2, containing 140 mM NaCl, 2.7 mM KCl, and 0.1% bovine serum albumin) and horseradish-peroxidase-labelled goat anti-human IgE (Nordic Immunology) diluted 1:1000 in the same buffer, containing 0.005% Tween 20, were used.

Carbohydrate detection of transferred protein was performed essentially as described Hsi et al. [13] by using a biotinylated-concanavalin-A solution and the ABC reagent (avidin, biotinylated-horseradish peroxidase), diluted 1:400 in 50 mM Tris/HCl, pH 7.5, containing 0.5 M NaCl and 0.1% poly(vinyl pyrrolidone). The staining was developed with 0.05% diaminobenzidine/HCl, 0.03% of 30% H_2O_2 in the Tris/NaCl reaction buffer.

Amino acid analysis

Proteins and peptides (1–5 nmol) were hydrolyzed at 105°C for different lengths of time (20 h for peptides, and 24, 48, 72 and 120 h for proteins) with 0.2 ml 5.7 M HCl containing 0.1% (mass/vol.) phenol and 20 nmol/ml 2-aminohexanoic acid as internal standard, in sealed tubes under vacuum [10]. Hydrolyzed samples were analyzed on a Beckman system 6300 amino acid analyzer with an analog interface module of System Gold. The cysteine content was determined after hydrolysis as carboxymethyl-cysteine from the alkylated protein and peptides. Tryptophan was determined spectrophotometrically [14]. Protein and peptide concentrations were measured by amino acid analysis. Free –SH groups were estimated by reaction with 5,5'-dithiobis(2-nitrobenzoic) acid, in 0.1 M Tris, pH 7.5, containing 1 mM EDTA and 2% (mass/vol.) SDS. The molar absorption coefficient at 412 nm was employed for calculations [15].

Spectroscopic studies

Circular dichroism (CD) spectra were obtained on a Jobin Ivon Mark III dichrograph, fitted with a 250-W xenon lamp [16]. The spectra in the far-ultraviolet region were recorded at 0.2 nm/s scanning speed. The sample was analyzed in 0.05-cm optical-path cells and in 20 mM ammonium bicar-

bonate, pH 8.0, at 25°C. The protein concentration was in the 0.1–0.5 mg/ml range in the far-ultraviolet CD spectrum and 1–2 mg/ml in the near-ultraviolet. Mean residue mass ellipticities were calculated based on 111 as the average molecular mass/residue, obtained from the amino acid composition, and expressed in terms of θ (degree $\times$ cm ^{2} $\times$ dmol $^{-1}$). Secondary structure estimations were performed by computer fit and according to the ellipticity values of Chang et al. [17].

Fluorescence emission spectra have been obtained on an Aminco SLM 6000 spectrofluorimeter. Spectra have been recorded for excitation at 275 nm and 295 nm wavelength, in 0.4 cm optical-path cells and in 20 mM ammonium bicarbonate, pH 8.0 and 25°C, at 110 μ g/ml protein concentration. The slit widths for both excitation and emission beams were 4 nm. The emission spectra for excitation at 275 nm and 295 nm were normalized by considering that the emission of tyrosine is negligible above 380 nm.

Proteolytic treatments

Carboxyamidomethylated protein digestions (20–50 nmol) with sequence grade tosylphenylalanine-chloromethane-treated trypsin and chymotrypsin were carried out essentially as described [10]. The alkylated protein was dissolved in 0.2 M ammonium bicarbonate, pH 8.0, and incubated at 37°C alternatively with trypsin (1:100, by mass, enzyme/substrate ratio) for 2 h, and with chymotrypsin (1:75, by mass, enzyme/substrate ratio) for 90 min. No insoluble material was detected at the end of these incubations.

Glu-specific staphylococcal V8 protease (S-V8) digestion was carried out at an enzyme/substrate ratio of 1:20, in 0.1 M ammonium bicarbonate, pH 8.0, containing 2 mM EDTA, at 37°C for 5 h. Although under these conditions the enzyme should selectively cleave Glu-C bonds [18], unexpected cleavages were observed in high yield. Shorter treatment with this enzyme (90 min) was also carried out in order to obtain larger peptides with a more specific hydrolysis.

Carboxypeptidase A digestion was performed with the alkylated allergen (5 nmol in 210 μ l) in 0.2 M *N*-methylmorpholine, pH 8.2, at an enzyme/substrate ratio of 1:10 (by mass) at 37°C for 10 h. After this reaction, the protein was precipitated with 50% sulfosalicylate and 1 M NaOH. The clear supernatant was analyzed for released amino acids as described above.

Peptide purification

The lyophilized samples obtained from each digestion were dissolved in 0.1% trifluoroacetic acid. The peptides were purified by RP-HPLC on a Nucleosil C-18 column (7 μ m, 250 mm $\times$ 4 mm), on a Beckman system with acetonitrile gradients in 0.1% (by vol.) trifluoroacetic acid specifically designed for each sample. Many peptides required a further purification step which was carried out by RP-HPLC on an Ultrasphere-ODS C-18 column and a gradient of acetonitrile in 10 mM ammonium acetate, pH 6.5.

Automatic sequence analysis

NH $_2$ -terminal Edman degradations were performed automatically by using an Applied Biosystems model 477A sequencer. The resulting phenylthiohydantoin (Pth)-Xaa derivatives were identified by using a 120 A on-line Pth-Xaa analyzer and the standard Applied Biosystems program [19].

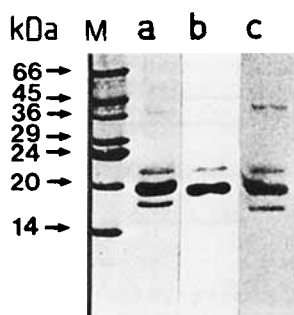


Fig. 1. SDS/PAGE analysis of *Ole e I*. (a) Coomassie Blue staining, (b) biotinylated-concanavalin-A staining and (c) immunostaining with olive-pollen-hypersensitive human serum. 5 μ g protein were applied in lanes (a) and (b), and 2 μ g in lane (c). M, molecular-mass protein markers (BioRad).

A non-redundant SwissProt + PIR + GenPept + GPup-date database was used for searching sequence similarities by using the Altschul et al. [20] algorithm. Alignment of multiple sequences was obtained according to the Needleman and Wunsch algorithm, by using the PAPA programs of Doolittle and Feng [21].

Secondary-structure predictions, based on the determined amino acid sequence, have been performed according to Garnier et al. [22] and Chou and Fasman [23]. Hydrophilicity prediction profiles were performed by using the reference values of Hopp and Woods [24], Kyte and Doolittle [25], and Fraga [26]. Antigenicity predictive analysis was achieved by using the Welling parameters [27].

RESULTS

Isolation and characterization of *Ole e I*

The olive tree pollen allergen, *Ole e I*, has been purified as previously described [9]. The protein exhibits the characteristic SDS/PAGE pattern [4, 9] with two protein bands of 20.0 kDa and 18.5 kDa (Fig. 1a). Two weak bands at 22.0 kDa and 40.0 kDa also appear in the gel. These represent less than 5% of the total protein and were described by Wheeler et al. [7] as components of the olive pollen allergen. All these bands are recognized by the IgE antibodies of allergic human sera when they are transferred onto ProBlott membranes and immunostained (Fig. 1c). When the transferred proteins are treated with biotinylated concanavalin A for detecting carbohydrates, only the three bands with the highest molecular mass are stained, indicating a glycosylated structure (Fig. 1b). Free -SH groups were not detected in the allergen upon titration of native *Ole e I* with 5,5'-dithiobisnitrobenzoate, suggesting that the six cysteine residues determined by amino acid analysis should be linked by disulphide bridges. Further proteolytic treatments and sequence analyses have been performed for reduced and carboxyamidomethylated *Ole e I* (7800–8400 cpm/nmol protein). Edman degradation of both native and alkylated protein renders a unique NH₂-terminal sequence with a microheterogeneity at the 3rd position. The yield of the first residue accounts for the overall amount of the sample sequenced. Therefore, the two bands observed in the electrophoretic pattern of the allergen would correspond to glycosylated (the 20.0 kDa band) and non-glycosylated (the band at 18.5 kDa) forms.

Amino acid sequence determination

Three different proteases, trypsin, chymotrypsin and S-V8 have been used to digest *Ole e I*. Fig. S1 shows the HPLC profile obtained for the separation of the resulting tryptic peptides. All of the major peaks and most of the minor peaks obtained in a significant yield (more than 5%) from the tryptic and chymotryptic treatments were subjected to automated sequencing. The complete primary structure of the olive tree pollen allergen is given in Fig. 2. Tryptic digestion allows the identification of more than 95% of the whole sequence of *Ole e I*, and most of the overlapping were resolved through the chymotryptic peptides. However, a first S-V8 proteolytic treatment was necessary to overlap positions 22–26, 40–41, 45–46, 59–61 and 130–131. A shorter treatment with this enzyme was required for the overlapping of the remaining positions 115–116. However, many pairs of peptides, such as T4/T10a, T6a/T10b, C6/C12, C1/C19, C11/C20, S2/S10 and S3a/S6 were subjected to a further purification step in order to obtain samples suitable for Edman degradation.

More than 60 peptides were automatically sequenced, although only the most significant are presented in Fig. 2. Thus, each region was determined at least twice, and frequently up to four times. From these 60 peptides, only C6 was unable to be sequenced, probably due to a cyclized glutamine at the NH₂-terminal position. Chymotryptic digestion gave rise to the cleavage of the peptide bonds at positions Arg115–Thr116 and Lys130–Cys131, this was attributed to a contaminant tryptic activity in the chymotryptic preparation used. It is also notable that S-V8 protease cleaved unexpectedly the peptide bond between Gly139 and Met140, with a yield greater than 95%. However, carboxypeptidase A digestion of the alkylated allergen confirmed that peptides T14, C19 and S10 correspond to the COOH-terminal sequence of the protein. Only two residues were obtained after 10 h of treatment, Met and Asn (0.9 nmol/nmol protein and 1.0 nmol/nmol protein, respectively).

The protein shows a main putative *N*-glycosylation site at position 111, in the consensus sequence Asn-Xaa-Thr. A gap appeared at this position in the automated sequencing of the corresponding peptides (T10, T10b, C14), and the amino acid compositions of these peptides showed one extra Asx residue in comparison with the determined sequence. Therefore, Asn has been assigned at this position. T10a contained Asp at the same position of the sequence. Thus, the presence of Asn and Asp in position 111 of *Ole e I* is correlated with the existence of the two variants of the allergen, glycosylated and non-glycosylated, respectively. The protein could exhibit a minor *N*-glycosylation site at position 48 because of the sequence heterogeneity in that region. However, the Edman degradation of the corresponding peptides did not corroborate this possibility.

A high degree of microheterogeneity was detected in *Ole e I*. The isolated allergen was found to show alternate amino acids at thirteen positions. The major sequence, considering the yield of the peptides, is shown in the main continuous line of Fig. 2, and the alternate residues below it. Eight substitutions are of conservative character, and five are non-conservative. Some of these changes were simultaneously present in the same peptide.

Spectroscopic analysis

The far-ultraviolet CD spectrum of native *Ole e I*, in sodium phosphate, pH 7.2, is given in Fig. 3. The secondary

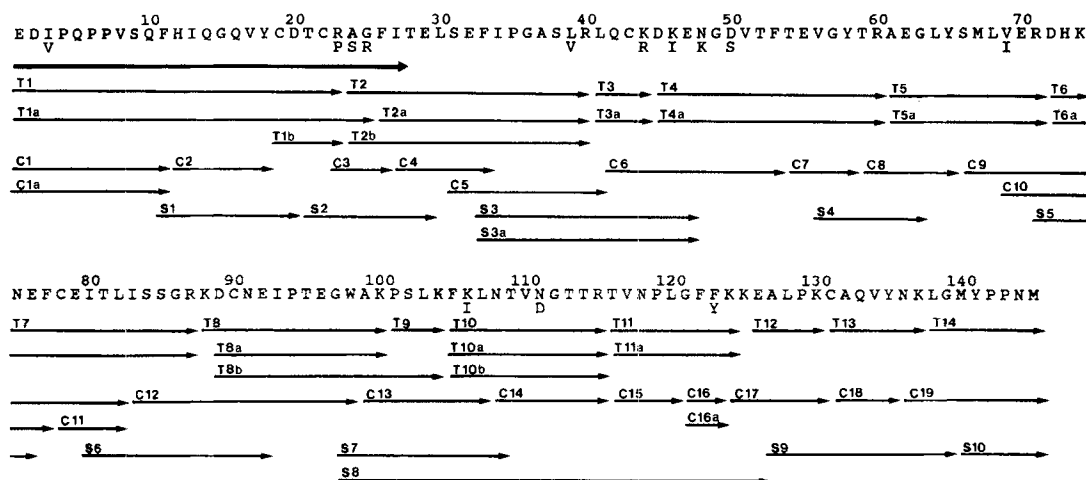


Fig. 2. Primary structure of *Ole e I*. The predominant sequence is shown in the main continuous line, and the corresponding heterogeneities below it. Peptides obtained by cleavage with trypsin (T), chymotrypsin (C), and *Staphylococcus aureus* V8 protease (S) were employed for overlapping as indicated. Peptides are numbered according to their positions from the NH₂-terminus of the protein. A thick arrow indicates Edman degradation data achieved from the overall alkylated allergen.

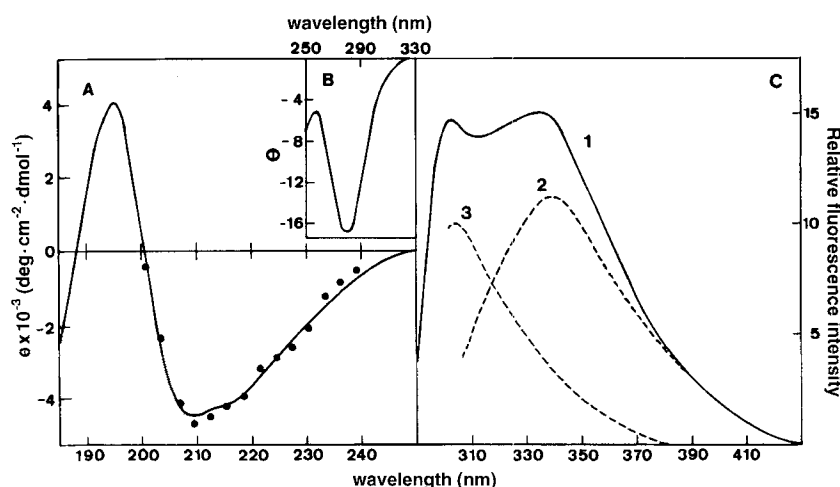


Fig. 3. Spectroscopic analysis of *Ole e I*. Circular dichroism (CD) spectra of olive tree pollen allergen in the far-ultraviolet (A) and in the near-ultraviolet (B) regions. CD values are expressed as θ , mean residue mass ellipticity, in units of degree $\times$ cm² $\times$ dmol⁻¹. (●) The CD theoretical values for the secondary structure determined according to the reference parameters of Chang et al. [17]. (C) Fluorescence emission spectra of *Ole e I*: for excitation at (1) 275 nm and (2) 295 nm after normalization (see Experimental Procedures); (3) calculated difference spectrum.

structure estimated from this spectrum is composed of 21.3% α helix, 21.0% β structure, 26.6% β turn and 31.0% aperiodic conformation and the spectrum fits well to the calculated theoretical values (Fig. 3). The near-ultraviolet CD spectrum (250–340 nm) only displays a negative band centered at 287 nm. Based on the absorbance properties of the protein ultraviolet chromophores, this last band should correspond to the single tryptophan residue of the molecule. The fluorescence emission spectrum of the allergen shows two maxima centered at 300–305 nm and 340 nm which are related to tyrosine and tryptophan, respectively. In fact, excitation at 295 nm produces an emission spectrum exhibiting a single band centered at 340 nm which corresponds to tryptophan emission since the absorbance of tyrosine at 295 is negligible. Both emission spectra have been normalized by considering that there is no tyrosine emission above 380 nm. The calculated difference spectrum (excitation at 275 nm minus

normalized excitation at 295 nm) shows a single band centered at 300–305 nm which thus corresponds to the tyrosine emission of the protein (Fig. 3).

DISCUSSION

Ole e I, the major allergen from olive tree (*O. europaea*) pollen, shows two main variants. These forms differ from each other in the glycosylation state, as it was previously suggested [4, 8, 9]. The high-molecular-mass (20.0 kDa) variant is a glycosylated polypeptide and the low-molecular-mass (18.5 kDa) variant is a non-glycosylated form. Both forms are recognized by the sera of hypersensitive individuals and exhibit a single NH₂-terminal amino acid sequence.

The complete amino acid sequence of *Ole e I* has been determined (Fig. 2). The sequencing strategy involved trypt-

Table 1. Amino acid composition of *Ole e I*. Sequence values were taken from the most abundant form of *Ole e I*. Cysteine was determined as the carboxymethylated derivative.

Amino acid	No. of residues determined from	
	amino acid analysis	sequence
Cys	5.6	6
Asx	14.3	14
Thr	10.4	11
Ser	7.9	7
Glx	18.2	18
Pro	10.6	10
Gly	11.4	11
Ala	5.9	6
Val	7.9	8
Met	2.3	3
Ile	8.3	7
Leu	9.8	11
Tyr	4.7	5
Phe	7.3	8
His	2.5	2
Lys	9.2	11
Arg	6.9	6
Trp	1 ^a	1
Total		145

^a Determined spectroscopically.

tic, chymotryptic and S-V8 protease digestions, and each protein region was determined through at least two independent sequencing experiments. The olive pollen allergen is a polypeptide chain of 145 amino acid residues and contains a putative *N*-glycosylation site at position 111. The six cysteine residues of *Ole e I* seem to be linked by disulphide bridges, since no free -SH groups are detected in the protein. All these residues were susceptible to alkylation after reductive treatment. The molecular mass of the allergen, calculated on the basis of the preponderant amino acid sequence, is 16331 Da and the amino acid composition agrees well with the value obtained through the standard acid hydrolysis and automated analysis (Table 1).

Microheterogeneity has been observed in thirteen positions of the sequence (Fig. 2). This polymorphism would be in agreement with the previously obtained RP-HPLC profiles of the pure protein [9], and the heterogeneity detected for the pI of the allergen [4]. Moreover, the small differences observed in Table 1 can be explained by this microheterogeneity. The existence of multiple isoallergenic forms has been demonstrated for other pollen allergens such as those from species of the genera *Lolium* and *Ambrosia* [28–30].

The hydrophilicity analysis of the sequence of *Ole e I* was performed on the basis of the determined sequence, by using the Hopp and Woods scale (Fig. S2). The profile obtained was almost identical to those resulting from the Kyte and Doolittle, and Fraga parameters. Three regions (positions 44–50, 70–79 and 85–92) of marked hydrophilic character appeared. The first two account for clusters of charged residues. These segments would be exposed in the tridimensional structure of the allergen, and therefore potentially involved in antigenic determinants. In fact, antigenicity prediction using Welling parameters shows the existence of four main regions which may contribute to B-cell epitopes, two of which are coincident with the two charged clusters (Fig. S2). Immuno-

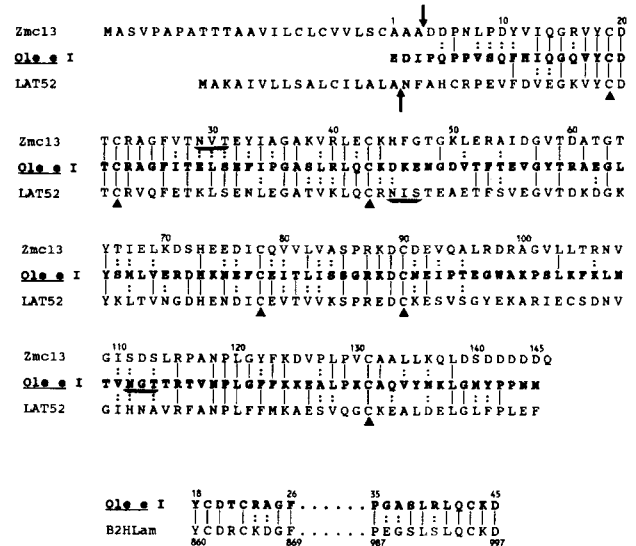


Fig. 4. Amino acid sequence similarity between *Ole e I* and other proteins. Above is shown the similarity of *Ole e I* with the putative proteins encoded by the genes *LAT52* from tomato anthers [32] and *Zmc13* from corn pollen [33]. Common Cys are indicated (▲), and putative sites for *N*-glycosylation are underlined. The suggested processing sites for the polypeptide precursors of *LAT52* and *Zmc13* deduced proteins are marked by arrows. Below is shown amino acid sequence similarity between polypeptide segments of *Ole e I* and B2 subunit of human laminin (B2HLam). Numbering of B2HLam is given according to the unprocessed chain [35]. In all cases, aligned amino acid identities are denoted by (|), and favourable substitutions, according to [20] are denoted by (:).

logical cross-reactions have been observed between olive tree pollen allergens and other pollen allergens such as those from privet (*Ligustrum vulgare*), ryegrass (*Lolium perenne*) and Bermuda grass (*Cynodon dactylon*) [31], but information about the molecular characteristics of the antigenic determinants of these allergens is scarce.

CD analysis showed a relatively high content of periodically ordered secondary structure. The calculated percentages for the main elements of this type of structure are similar to those obtained throughout the predictive analysis on the basis of the determined amino acid sequence: 22% α helix, 21% β bend, 17% β turn and 40% aperiodic conformation (Fig. S2). This theoretical analysis also shows that the glycosylation site is located in a β turn conformation, which is a favourable structure for this attachment. Moreover, the high tyrosine contribution to the fluorescence emission of the protein would indicate a low structural quenching over the five tyrosine residues of the molecule. The fluorescence emission of tryptophan in aqueous model systems is centered at about 345 nm. Therefore, the single tryptophan residue in *Ole e I* is not highly buried in the tridimensional structure since its emission maximum appears at 340 nm. This would be in agreement with the positive hydrophilicity values for the sequence around this residue (Fig. S2).

The primary structure of *Ole e I* did not show any similarity with other known allergens contained in protein and translated nucleotide sequence databases. However, it exhibits substantial amino acid identity with the deduced amino acid sequences of the genes *LAT52* from tomato [32] and *Zmc13* from corn [33] in the open reading frame nucleotide segments. 36% and 38% identity was found, respectively, for the tomato and corn deduced proteins (Fig. 4). The similarity

risers to 56% and 64%, respectively, when considering the conservative changes, according to the scale used by Altschul et al. [20]. It is notable that the three sequences exhibit a putative *N*-glycosylation site, although in different positions. This coincidence could be related with a potential role for the sugar moiety in the common function of these proteins, but the existence of the oligosaccharide group is to be demonstrated in the corn and tomato pollen proteins. Moreover, the Cys residues appeared on conserved positions in the three sequences, which could suggest a similar tridimensional structure for these proteins. The existence of NH₂-terminal signal sequences has been suggested for the putative proteins encoded by these genes [32, 33] but the processing sites are unknown. We suggest a processing site for the corresponding precursors of tomato and corn pollen proteins, respectively, between positions 27–28 and 17–18 from the Met NH₂-terminal residue (Fig. 4), based in both the NH₂-terminal end of *Ole e I* and the existence of the motif Ala-Xaa-Ala, the most frequent sequence preceding the signal-peptidase cleavage site [34].

In addition, only sporadic, short-segment similarities were observed with other proteins in the homology searches performed. However, the close similarity found between the polypeptide segments at positions 18–26 and 35–45 of *Ole e I* and 860–868 and 987–997, respectively, of B2 subunit of the human laminin gene product is striking (Fig. 4). These segments are located in the cell-binding domain of human laminin [35], an extracellular protein involved in multiple protein-protein and protein-cell interactions. Moreover, both sequences are very conserved in the three pollen proteins herein considered, *Ole e I*, and the *LAT52* and *Zmc13* gene products (Fig. 4).

A biological function has been suggested for only a few allergens: *Der p I* and *Der p II* as thiol proteases [36], *Bet v I* as a plant-pathogenesis-related protein [37], and *Bet v II* as a profilin [38]. Thus, it is interesting to speculate about a potential role of *Ole e I* in the germination of the pollen. This hypothesis is based on the observed sequence similarity with the translation product of the tissue-specific gene *LAT52* from tomato, for which a function in such an event has been suggested [32]. If *Ole e I*, as well as the mentioned proteins from tomato and corn pollens, were involved in binding interactions or recognition events during pollen germination, the sequence regions common to that of laminin could play a role in such events. Further analysis will be necessary to demonstrate this.

Many allergens have been characterized as glycoproteins [39–45], but the importance of the glycosyl moiety in the allergenicity is to be defined. *Ole e I* could be a good model for such studies. Moreover, the knowledge of its primary structure should facilitate detailed antigenic and allergenic study as well as the assignment of functions to specific domains of this protein.

This work was supported by grant PB89/087 from the *Dirección General de Investigación Científica y Técnica* (M. E. C. Spain). We thank Dr J. G. Gavilanes for helpful comments and critical reading of the manuscript.

REFERENCES

- Bousquet, J., Guerin, B., Hewitt, B., Lim, S. & Michel, F. B. (1985) *Clin. Allergy* 15, 439–448.
- Tamir, R., Pick, A. I., Topilsky, M. & Kivity, S. (1991) *Clin. Exp. Allergy* 21, 329–332.
- Wheeler, A. W. (1992) *Clin. Exp. Allergy* 22, 1052–1057.
- Lauzurica, P., Maruri, N., Galocha, B., González, J., Díaz, R., Palomino, P., Hernández, D., García, R. & Lahoz, C. (1988) *Mol. Immunol.* 25, 337–344.
- Marsh, D. G., Goodfriend, L., King, T. K., Lowenstein, H. & Platts-Mills, T. A. E. (1988) *Clin. Allergy* 18, 201–209.
- Lauzurica, P., Gurbindo, C., Maruri, N., Galocha, B., Díaz, R., González, J., García, R. & Lahoz, C. (1988) *Mol. Immunol.* 25, 329–335.
- Wheeler, A. W., Hickman, B. E. & Fox, B. (1990) *Mol. Immunol.* 27, 631–636.
- Lombardero, M., Quirce, S., Duffort, O., Barber, D., Carpizo, J., Chamorro, M. J., Lezaun, A. & Carreira, J. (1992) *J. Allergy Clin. Immunol.* 89, 884–894.
- Villaba, M., López-Otín, C., Martín-Orozco, E., Monsalve, R. I., Palomino, P., Lahoz, C. & Rodríguez, R. (1990) *Biochem. Biophys. Res. Commun.* 172, 523–528.
- Menéndez-Arias, L., Moneo, I., Domínguez, J. & Rodríguez, R. (1988) *Eur. J. Biochem.* 177, 159–166.
- Laemmli, U. K. (1970) *Nature* 227, 680–685.
- Monsalve, R. I., González de la Peña, M. A., Menéndez-Arias, L., López-Otín, C., Villalba, M. & Rodríguez, R. (1993) *Biochem. J.* 293, 625–632.
- Hsi, K.-L., Chen, L., Hawke, D. H., Zieske, L. R. & Yuang, P.-M. (1991) *Anal. Biochem.* 198, 238–245.
- Beaven, G. H. & Holiday, E. R. (1952) *Adv. Protein Chem.* 7, 319–386.
- Ellman, G. L. (1959) *Arch. Biochem. Biophys.* 82, 70–77.
- Monsalve, R. I., Villalba, M., López-Otín, C. & Rodríguez, R. (1991) *Biochim. Biophys. Acta* 1078, 265–272.
- Chang, C. T., Wu, C.-S. C. and Yang, J. T. (1978) *Anal. Biochem.* 91, 13–31.
- Drapeau, G. R. (1977) *Methods Enzymol.* 47, 189–191.
- Hewick, R. M., Hunkapiller, M. W., Hood, L. E. & Dreyer, W. (1981) *J. Biol. Chem.* 256, 7990–7997.
- Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. (1990) *J. Mol. Biol.* 215, 403–410.
- Doolittle, R. F. & Feng, D.-F. (1990) *Methods Enzymol.* 183, 659–669.
- Garnier, J., Osguthorpe, D. J. & Robson, B. (1978) *J. Mol. Biol.* 120, 97–120.
- Chou, P. Y. & Fasman, G. D. (1978) *Annu. Rev. Biochem.* 47, 251–276.
- Hopp, T. P. & Woods, K. R. (1981) *Proc. Natl Acad. Sci. USA* 78, 3824–3828.
- Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* 157, 105–132.
- Fraga, S. (1982) *Can. J. Chem.* 60, 2606–2610.
- Welling, G. W., Weijer, W. J., van der Zee, R. & Welling-Wester, S. (1985) *FEBS Lett.* 188, 215–218.
- Johnson, P. & Marsh, D. G. (1965) *Nature* 206, 935–937.
- Rafnar, T., Griffith, I. J., Kuo, M., Bond, J. F., Rogers, B. L. & Klapper, D. G. (1991) *J. Biol. Chem.* 266, 1229–1236.
- Ansari, A. A., Shenbagamurthi, P. & Marsh, D. G. (1989) *J. Biol. Chem.* 264, 11181–11185.
- Baldo, B. A., Panzani, R. C., Bass, D. & Zerboni, R. (1992) *Mol. Immunol.* 29, 1209–1218.
- Twiss, D., Wing, R., Yamaguchi, J. & MacCormick, S. (1989) *Mol. & Gen. Genet.* 217, 240–245.
- Hanson, D. D., Hamilton, D. A., Travis, J. L., Bashe, D. M., & Mascarenhas, J. P. (1989) *Plant Cell* 1, 173–179.
- Carné, T. & Scheele, G. (1985) *Cell biology of the secretory process* (Cantin, M., ed.) p. 73, S. Karger AG, Basel.
- Sasaki, M. & Yamada, Y. (1987) *J. Biol. Chem.* 262, 17111–17117.
- Chua, K. Y., Stewart, G. A., Thomas, W. R., Simpson, R. J., Dilworth, R. J., Plozza, T. M. & Turner, K. J. (1988) *J. Exp. Med.* 167, 175–182.
- Breiteneder, H., Pettenburger, K., Bito, A., Valenta, R., Kraft, D., Rumpold, H., Scheiner, O. & Breitenbach, M. (1989) *EMBO J.* 8, 1935–1938.
- Valenta, R., Duchêne, M., Pettenburger, K., Sillaber, C., Valent, P., Bettelheim, P., Breitenbach, M., Rumpold, H., Kraft, D. & Scheiner, O. (1991) *Science* 253, 557–560.

39. Klapper, D. G., Goodfriend, L. & Capra, J. D. (1980) *Biochemistry* 19, 5729–5734.
40. Tovey, E. R., Chapman, M. D. & Platts-Mills, T. A. E. (1981) *Nature* 289, 592–593.
41. Ford, S. A. & Baldo, B. A. (1986) *Int. Arch. Allergy Appl. Immunol.* 81, 193–203.
42. Heymann, P. W., Chapman, M. D. & Platts-Mills, T. A. E. (1986) *J. Immunol.* 137, 2841–2847.
43. Weber, A., Schröder, H., Thalberg, K. & März, L. (1987) *Allergy (CPH)* 42, 464–470.
44. Polo, F., Ayuso, R. & Carreira, J. (1991) *Mol. Immunol.* 28, 169–175.
45. Sánchez-Monge, R., Gómez, L., Barber, D., López-Otín, C., Armentía, A. & Salcedo, G. (1992) *Biochem. J.* 281, 401–405.

Supplementary material to:

The amino acid sequence of *Ole e I*, the major allergen from olive tree (*Olea europaea*) pollen

Mayte VILLALBA, Eva BATANERO, Carlos LÓPEZ-OTÍN, Luis M. SÁNCHEZ, Rafael I. MONSALVE, Manuel A. GONZÁLEZ DE LA PEÑA, Carlos LAHOZ, and Rosalía RODRÍGUEZ

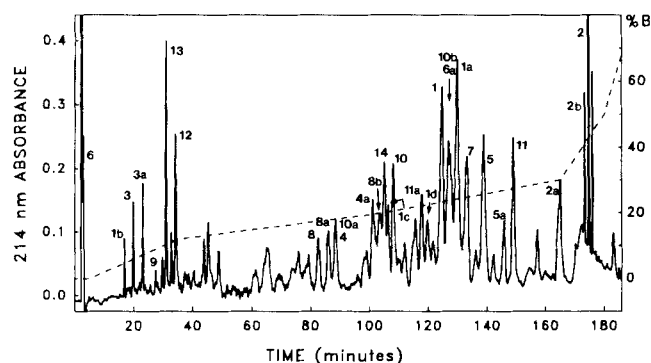


Fig.S1. HPLC profile of tryptic digestion of *Ole e I*. Peptides were eluted with a gradient (---) of acetonitrile (%B) in 0.1% trifluoroacetic acid. HPLC effluent (1.0 ml/min flow rate) was continuously monitored at both 214 nm and 280 nm. Fractions of 0.7 ml were collected. Aliquots of 1.0% or 2.5% of each collected fraction were taken for radioactivity counting. Numbering of the peptides was according to their positions from the NH₂-terminus of the sequence (Fig. 2).

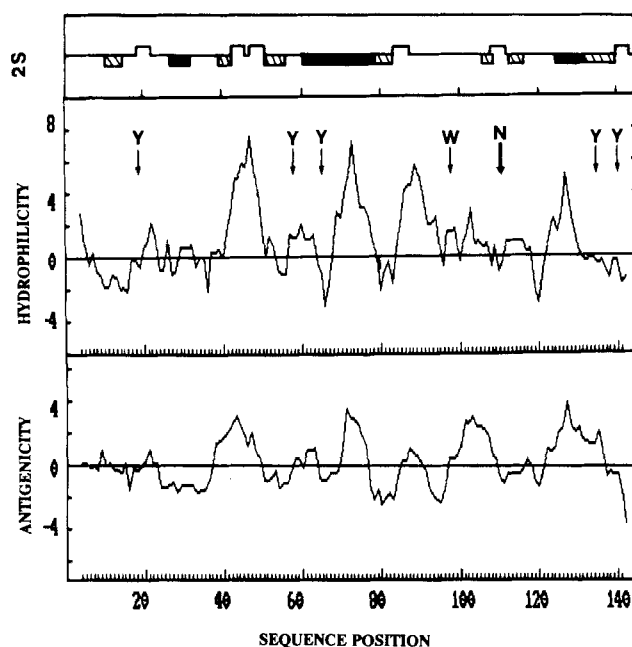


Fig.S2. Secondary structure, hydrophilicity and antigenicity predictions for *Ole e I*. Theoretical secondary structure (2S) was obtained according to Garnier et al. [22] and Chou and Fasman [23]; the upper line shows β turns; the black boxes indicate α -helix segments; the shaded boxes are β bends; and the lower line shows aperiodic conformation. The predictive profiles were obtained by using the Hopp and Woods scale for hydrophilicity [24], and Welling parameters for antigenicity [27]. A span window of seven residues was used. The arrows indicate the sequence positions for W (Trp), Y (Tyr), and N (putative glycosylated-Asn) residues.