Cloning and expression of biologically active *Plantago lanceolata* pollen allergen Pla 11 in the yeast *Pichia pastoris*

Belén CALABOZO*, Araceli DÍAZ-PERALES†, Gabriel SALCEDO†, Domingo BARBER* and Florentino POLO*¹

*Departamento I + D, ALK-Abelló, S.A., c/Miguel Fleta 19, 28037 Madrid, Spain, and †Unidad de Bioquímica, Departamento de Biotecnología, E.T.S. Ingenieros Agrónomos, Ciudad Universitaria, 28040 Madrid, Spain

The glycoprotein Pla 11 is the major allergen from English plantain (*Plantago lanceolata*) pollen, which is a common cause of pollinosis in temperate areas. Three complete cDNAs for Pla 11 isoforms were isolated by PCR using specific 3' and 5' primers. All three Pla 11 cDNAs code for a 25-residue leader peptide and a 131-residue mature protein that contains two polymorphic positions, an N-glycosylation site at position 107 and six cysteine residues involved in three disulphide bridges. The allergen variant Pla 11.0101 was produced in *Pichia pastoris* at a yield of 20 mg per litre of culture as a mixture of non-glycosylated (17 kDa), glycosylated (23 kDa) and dimeric (32–39 kDa) forms. Recombinant Pla 11 (rPla 11) was purified by affinity chromatography with an anti-natural Pla 11 (anti-nPla 11) monoclonal antibody, and its molecular and immunological properties were compared with the natural

allergen by CD spectroscopic analysis, enzymic deglycosylation, lectin-binding assay, immunodetection and ELISA-inhibition assays using sera from plantain-allergic patients. The recombinant allergen is properly folded, as deduced from CD spectra, and the immunodominant allergenic epitopes of the natural allergen are preserved in rPla 11. These results allow us to conclude that *P. pastoris* is a convenient system for the efficient production of biologically active rPla 11, which could have a potential use for clinical purposes. Furthermore, a sequence similarity of Pla 11 to the major allergen from the olive tree pollen, Ole e 1, is revealed in this work, and the allergenic cross-reactivity between both allergens has been studied.

Key words: allergy, English plantain, pollinosis, recombinant allergen, ribwort.

INTRODUCTION

Pollen of *Plantago lanceolata* (English plantain, or ribwort) is known to be a common cause of pollinosis in the temperate zones of North America, Australia and Europe [1–3], where it is widely distributed. Although the clinical importance of this dicotyledon weed has been largely demonstrated in several studies carried out in different countries [1,2,4,5], its role in the aetiology of pollinosis has usually been overlooked in clinical practice. This is probably because plantain pollinates in the same season as other relevant allergenic plants, such as grasses, and monosensitization to plantain pollen is not frequently found among allergic patients. Nevertheless, only a 30 kDa plantain allergen cross-reactive with the grass group 5 allergens has been identified to date, yet this cross-reactivity shows little or no clinical relevance, as suggested by Asero et al. [6]. These authors also reported in the same article that IgE from monosensitized plantain-allergic patients mainly reacted with 17, 19 and 40 kDa allergens, which is in accordance with our previous findings. Thus 13 out of 14 plantain-sensitive individuals had IgE to 17, 20 and 32–36 kDa proteins [7], which could be the same allergens identified by Asero et al. [6]. Furthermore, we found that the 17 and 20 kDa allergens are the unglycosylated and glycosylated forms of the same protein (Pla 11) respectively [8], and that the 32–36 kDa protein is a dimeric form of the same Pla 11 allergen [7].

Moreover, the major complex N-glycan of Pla 11 might be a potential source of cross-reactivity with other glycosylated pollen allergens [8] that could be misleading in terms of false positive

diagnoses of allergy to plantain when using natural extracts. Therefore the generation of a recombinant Pla 11 without the complex N-glycan as part of its structure would be a very useful tool to diagnose patients specifically sensitized to plantain.

Here, we report the cDNA cloning, sequencing and expression in *Pichia pastoris* of recombinant Pla 11. The physicochemical equivalence of the natural and recombinant allergens was studied, as well as the allergenic cross-reactivity with the major allergen of olive (*Olea europaea*) pollen, which has been found in the present study to share sequence similarity to Pla 11.

EXPERIMENTAL

Isolated allergens and sera

Pl. lanceolata and *O. europaea* pollens were obtained from Allergen Pollens (Madrid, Spain). Natural Pla 11 (nPla 11) and natural Ole e 1 were purified from their respective pollen extracts by affinity chromatography with monoclonal antibodies (mAbs), following methods previously described [9,10].

Sera from 14 individuals allergic to plantain pollen were acquired from PlasmaLab International (Everett, WA, U.S.A.). These sera were of radioallergosorbent test ('RAST') classes 3 and 4 to *Pl. lanceolata* (Pharmacia, Uppsala, Sweden). A serum pool was obtained by mixing equal-volume aliquots of the individual sera, and then was stored in aliquots at –80 °C.

Methods other than those reported below were carried out according to standard procedures [11].

Abbreviations used: DSA, *Datura stramonium* agglutinin; GNA, *Galanthus nivalis* agglutinin; IUIS, International Union of Immunological Societies; MAA, *Maackia amurensis* agglutinin; mAb, monoclonal antibody; nPla 11/rPla 11, natural and recombinant Pla 11 respectively; PNA, peanut agglutinin; PNGase F, peptide-N-glycosidase F; RACE, rapid amplification of cDNA ends; SNA, *Sambucus nigra* agglutinin.

¹ To whom correspondence should be addressed (e-mail florentino.polo@es.alk-abello.com).

The nucleotide sequence data reported for Pla 11.0101, Pla 11.0102 and Pla 11.0103 appear in DDBJ, EMBL, GenBank® and GSDB Nucleotide Sequence Databases under the accession numbers AJ313166, AJ313167, and AJ313168 respectively.

Isolation of RNA from pollen

Pl. lanceolata pollen was disrupted by sonication in denaturing solution [4 mol/l guanidinium thiocyanate/25 mMol/l sodium citrate (pH 4.0)/0.5 % (w/v) lauryl sarcosinate/100 mMol/l 2-mercaptoethanol], and RNA was extracted by essentially following the guanidinium thiocyanate method [12].

PCR-amplification of cDNA encoding Pla 11

First-strand cDNA synthesis and amplifications of Pla 11-specific cDNA fragments were carried out by using the 3' RACE (rapid amplification of cDNA ends) and 5' RACE system kits from Gibco BRL (Gaithersburg, MD, U.S.A.), and with specific primers. A Gene ATAQ (Pharmacia) programmable thermal controller was used in PCR amplifications.

Following the synthesis of first-strand cDNA from RNA using Superscript reverse transcriptase and a primer containing an oligo(dT)₁₇ sequence, according to the manufacturer's instructions, first amplification of the 3' region of the Pla 11 cDNA was achieved by PCR using the degenerate oligonucleotide CAYCCNGCNAARTTYCAYGT, deduced from the N-terminal amino acid sequence of natural Pla 11 at positions 5–10 [7] as the sense primer, and the oligo(dT)₁₇ primer as the antisense primer.

Amplification of the 5' region of the Pla 11 cDNA was carried out with the 5' RACE system, using the gene-specific primers CGGAATTCTCACTACAATCGGGCCTGCCGC and CCAATTGCACTTGTGCCCTGCCATGCGTTCGC, the latter being nested internally to the former.

To obtain the complete sequence of Pla 11 cDNA from the N-terminal end of the mature protein, the cDNA was amplified again with the 3' RACE system using the primer ACACAAACATCTCATCCCGC, corresponding to positions 1–7 of the amino acid sequence of natural Pla 11 allergen.

The PCR products were purified and cloned into pGEM T Easy vector (Promega, Madison, WI, U.S.A.), and the constructs were used to transform DH5 α *Escherichia coli* cells. The complete nucleotide sequences of both strands from selected clones were determined using the ABI PRISM Dye Terminator kit and an ABI 377 sequencer (PerkinElmer Biosystems, Warrington, U.K.).

Expression of Pla 11 in *P. pastoris*

The coding region of Pla 11 cDNA was amplified by PCR with *Taq* Gold DNA polymerase (Gibco BRL), a sense primer (CTCGAGAAAAGAGAGACACAAACATCTCATCCCGC; the *Xho*I restriction site is shown underlined) and an anti-sense primer (GGGAATTCCTTAACACCCAGGGGC; the *Eco*RI restriction site is shown underlined). The hybridizing sequences are extended in the synthetic primers with in-frame sequences coding for the pre-pro-sequence of the α -mating factor of *Saccharomyces cerevisiae* present in plasmid pPIC9. The sense primer also included a codon for glutamic acid, which enabled the processing of the signal peptide by the yeast. PCR products were isolated and used directly for ligation into pGEM T Easy vector, as described above. This construct was used to transform DH5 α *E. coli* cells, before sequencing. Plasmid DNA from one clone representative of the Pla 11.0101 variant was isolated and digested with *Xho*I/*Eco*RI restriction enzymes. The DNA fragment was subcloned into pPIC9 vector, rendering pPIC9–Pla 11.0101, which was sequenced to confirm the in-frame arrangement of the leader sequence and insert.

The construct pPIC9–Pla 11.0101 was linearized with *Bgl*II restriction enzyme and used to transform GS115 cells by

electroporation. Transformed cells were incubated on minimal-glucose plates at 30 °C for 4–6 days, and screening was performed by patching the colonies in replica plating on minimal-glucose and minimal-methanol plates.

Selected transformed strains were cultured for 4 days at 30 °C in buffered glycerol complex medium. Cells were then collected by centrifugation, and resuspended in buffered methanol complex medium for induction of recombinant Pla 11 (rPla 11). The production of rPla 11 in the supernatant of the culture medium was analysed at different times by SDS/PAGE and a 'sandwich' ELISA designed for quantification of Pla 11 [9]. The highest expression level of recombinant Pla 11 was reached at day 4 after induction.

Purification of rPla 11

The extracellular medium obtained after the induction was dialysed against PBS and used as the starting material to purify rPla 11. Purification was carried out by affinity chromatography with an anti-Pla 11 monoclonal antibody as previously described for the natural allergen [9], and then dialysed extensively against PBS.

In order to separate the dimeric from the monomeric components that form rPla 11, size-exclusion chromatography was performed. Briefly, affinity-purified rPla 11 was dialysed against water, freeze-dried, dissolved in 100 mM Na₂SO₄/20 mM NaH₂PO₄ buffer, pH 6.8, injected into a size-exclusion HPLC column (Ultraspherogel SEC 2000; 7.5 mm $\times$ 300 mm, 5 μ m particle size; Beckman Instruments, Inc., San Ramon, CA, U.S.A.) and eluted with the same solvent at a constant flow rate of 0.25 ml/min.

Enzymic deglycosylation

rPla 11 monomer was subjected to enzymic deglycosylation using N-acetylneuraminidase II, O-glycosidase DS and peptide:N-glycosidase F (PNGase F; Enzymic Deglycosylation Kit, Bio-Rad, Hercules, CA, U.S.A.) under reducing conditions, following the manufacturer's instructions. Fetuin was used as a positive control.

Characterization of rPla 11 glycan

The structure of the carbohydrate moiety of rPla 11 was identified by lectin-binding assay using the DIG Glycan Differentiation Kit from Roche Molecular Biochemicals (Mannheim, Germany). Briefly, protein samples were separated on SDS/PAGE and electroblotted on to nitrocellulose membranes (Hybond ECL[®]; Amersham Biosciences, Uppsala, Sweden) that were subsequently incubated with digoxigenin-labelled lectins, polyclonal sheep anti-digoxigenin Fab fragments conjugated with alkaline phosphatase and the enzyme substrate 4-Nitro Blue Tetrazolium chloride. The kit included the following digoxigenin-labelled lectins: GNA (*Galanthus nivalis* agglutinin), SNA (*Sambucus nigra* agglutinin), MAA (*Maackia amurensis* agglutinin), PNA (peanut agglutinin) and DSA (*Datura stramonium* agglutinin).

IgE-binding inhibition-ELISA

The binding of specific IgE from a human serum pool to nPla 11, adsorbed on to plastic microtitre wells, was inhibited by addition of increasing amounts of rPla 11.0101 or nOle e 1. Briefly,

nPla 11 (1 µg/ml in PBS; 100 µl/well) was adsorbed on to polystyrene microtitre wells (EIA/RIA plates; Costar, Cambridge, MA, U.S.A.) and, after blocking with 1 % (w/v) BSA in PBS, wells were incubated for 3 h at room temperature with a mixture of 50 µl of a serum pool from *Pl. lanceolata*-allergic patients [diluted 1:2 with PBS containing 0.1 % (w/v) BSA and 0.05 % (v/v) Tween 20] and 50 µl of serial dilutions of rPla 11 or nOle e 1 in the same buffer. After washing with 0.05 % (v/v) Tween 20 in PBS, wells were incubated with rabbit anti-human IgE antibodies conjugated with horseradish peroxidase (Dako, Glostrup, Denmark) for 2 h at room temperature. After washing as before, a solution of *o*-phenylenediamine (Dako) was added to each well, and allowed to react for 30 min at room temperature in the dark. The reaction was stopped by the addition of 2 M HCl, and the absorbance was measured at 490 nm with a reference filter at 650 nm in an automatic microplate reader (E-max; Molecular Devices, Sunnyvale, CA, U.S.A.). A control curve constructed with serial dilutions of nPla 11 was also included. Determinations were carried out in triplicate.

Immunodetection

SDS/PAGE was performed in Tricine/10–20 % (w/v) polyacrylamide gels (Novex, San Diego, CA, U.S.A.). Electrophoresis was performed under reducing conditions. Low-range SDS/PAGE standards (Bio-Rad) were used as molecular-mass markers. Proteins were stained with Coomassie Blue.

Immunodetection of IgE-binding proteins with a *Pl. lanceolata*-sensitive human serum pool was performed as described previously [7].

Amino acid composition analysis

Samples were hydrolysed at 110 °C for 24 h with 6 M HCl in the vapour phase. The hydrolysates were analysed in a Biochrom 20 Analyzer (Pharmacia) using a post-column derivative-formation method with ninhydrin. Norleucine was used as an internal standard.

MS analysis

Samples were mixed (1:1, v/v) with a matrix solution consisting of saturated α -cyano-4-hydroxycinnamic acid in aq. 30 % (v/v) acetonitrile and 0.1 % (v/v) trifluoroacetic acid, and measured on a Biflex III (Bruker-Franzen Analytik, Bremen, Germany) matrix-assisted laser-desorption ionization–time-of-flight mass spectrometer. Mass spectra were recorded in the linear positive mode, accumulating 100 spectra of single laser shots. The equipment was externally calibrated using myoglobin.

CD spectra

CD spectra in the far-UV region were measured in a Jasco 720 dichrograph at a scanning speed of 20 nm/min. A 10^{-5} ΔE /mm sensitivity scale was used, together with time constants of 4 s. Samples, prepared in 10 mM sodium phosphate buffer, pH 7.0, were analysed at 25 °C in 0.1 cm optical-path-length cells.

RESULTS

Cloning and sequencing of cDNA encoding Pla 11

cDNAs encoding a partial sequence of Pla 11 were obtained from RNA of *Pl. lanceolata* pollen by specific amplification

with a degenerate primer deduced from the N-terminal amino acid sequence (positions 5–10) of mature Pla 11. The PCR products showed a size of approx. 700 bp (results not shown). Some differences in sequence between the distinct clones were observed, both in the coding and the non-coding region. Specific primers from sequence stretches with no polymorphic positions were used for 5' RACE in order to elucidate unambiguously the nucleotide sequence corresponding to the N-terminal end of the mature protein and the signal peptide. Three clones were sequenced, and the sequence established for the leader peptide had the same length (75 bases) in all of them. The deduced amino acid sequence for the leader peptide was M⁻²⁵VKLTQVAAILLIGAFFLIASPSIA⁻¹. Only one base substitution in the leader peptide sequence stretch that implies an amino acid change at position -4 was observed in one clone (P → T). No polymorphic positions were found in the sequence stretch of these clones encoding the N-terminal end of the mature protein. Therefore a non-degenerate, 20-nucleotides-long primer, corresponding to positions 1–7, was designed and used in a subsequent 3' RACE amplification. cDNA clones obtained from this 3' RACE ranged in size from 674–704 bp, because of distinct lengths of the polyadenylated tract and some divergence in the 3' untranslated region. On the basis of nucleotide sequence analysis of the coding region, these cDNA clones were classified into three groups, each one encoding a variant of Pla 11, which have been named according to the International Union of Immunological Societies (IUIS) rules for allergen nomenclature. In Figure 1(A), the nucleotide sequence for one clone representative of each group is given. The sequences of these clones showed that all of them coded for a 131-amino-acid residue mature protein (Figure 1B). Nucleotide sequence polymorphism was detected at four positions, although only changes at positions 173 and 244 lead to amino acid substitutions, at positions 58 and 82 respectively, of the amino acid sequence. All of the variants displayed six cysteine residues in the sequence and a potential N-glycosylation site at position 107. The molecular-mass values estimated for the polypeptide backbone of the mature Pla 11 allergen are 14 521, 14 463 and 14 433 for variants 0101 (Asp⁵⁸, Ser⁸²), 0102 (Gly⁵⁸, Ser⁸²), and 0103 (Gly⁵⁸, Gly⁸²) respectively. Comparison of the hydrophilicity profiles deduced from the amino acid sequences showed only a small difference in the vicinity of position 58 between variant 0101 and the two other variants (results not shown).

The sequences of the three variants of Pla 11 were compared with known protein sequences in protein sequence databases. A significant degree of sequence identity was found with allergens related to Ole e 1, i.e. the main allergen from *O. europaea* pollen. Depending on the Pla 11 variant and the 'Ole e 1-like' allergen considered, the sequence identity ranged from 38–42 %, and the sequence similarity, when conservative changes were taken into account, increased to 59–61 %. Figure 2 shows the best alignment between the amino acid sequences of Pla 11.0101 and Ole e 1.

Expression, purification and characterization of rPla 11

The cDNA coding for the mature Pla 11.0101 allergen was expressed in the yeast *P. pastoris*, as described in the Experimental section. Recombinant allergen secreted into the media by GS115 cells transformed with the construct pPIC9-Pla 11.0101 on induction with methanol was detected by SDS/PAGE and by a sandwich ELISA, with an anti-nPla 11 mAb on the solid phase and an anti-*Pl. lanceolata* rabbit serum as the second antibody [9]. The highest expression level was observed at day 4 after induction.

A	Pla 1 1.0101	ACACAAACATCTCATCCCGCAAATTCACGTTGAGGGAGAGGTATACTGCAATGTTTGT	60
	Pla 1 1.0102	-----	
	Pla 1 1.0103	-----	
	Pla 1 1.0101	CACAGCAGAAATTTAATCAATGAACTCAGCGAACGCATGGCAGGGGCACAAGTGCAATTG	120
	Pla 1 1.0102	-----T-----	
	Pla 1 1.0103	-----T-----	
	Pla 1 1.0101	GATTGCAAAGATGATTCTAAAAAAGTCATATACTCTATAGGGGGTGAGACTGATCAAGAT	180
	Pla 1 1.0102	-----G-----	
	Pla 1 1.0103	-----G-----	
	Pla 1 1.0101	GGTGTTTACCGCCTGCCTGTTGTAGGCTATCACGAAGATTGTGAAATCAAACAGTGAAG	240
	Pla 1 1.0102	-----	
	Pla 1 1.0103	-----	
	Pla 1 1.0101	AGCAGCAGGCCCCGATTGTAGTGAAATTCGAAACTTGCAAAGGGAACAATTCAAACCTCG	300
	Pla 1 1.0102	-----	
	Pla 1 1.0103	---G-----A---	
	Pla 1 1.0101	AAAGTGGACCTTTCAAAAACACAACCATCACCGAAAAACACGTATGTCAAGCCACTG	360
	Pla 1 1.0102	-----	
	Pla 1 1.0103	-----	
B	Pla 1 1.0101	AGCTTTCGCGCAAAGACGGATGCCCTGGGTGT TAA aggatgctgcagagggcgattaat	420
	Pla 1 1.0102	-----	
	Pla 1 1.0103	-----	
	Pla 1 1.0101	tgtacaagtccaattttcataataacaagcgttcaatgtgattcctttttctgttttct	480
	Pla 1 1.0102	-----	
	Pla 1 1.0103	-----	
	Pla 1 1.0101	tgttttcttttttgttccccagttttgtagtagtcttcaaagttcaataaggtgtttcc	540
	Pla 1 1.0102	-----a-----	
	Pla 1 1.0103	-----aa-----	
	Pla 1 1.0101	agacctggttgagggttggtttctcaggatactgatgaacctttttattatctatgagta	600
	Pla 1 1.0102	-----	
	Pla 1 1.0103	-----	
	Pla 1 1.0101	gttatacgcaaaagttttgcaagcagtgagaaatcactgtagtttatatttaaatggaat	660
	Pla 1 1.0102	-----	
	Pla 1 1.0103	-----g-at--t-----	
	Pla 1 1.0101	ttgatgttcttac(a) _n	
	Pla 1 1.0102	----- (a) _n	
	Pla 1 1.0103	----- (a) _n	
		10 20 30 40 50 60	
	Pla 1 1.0101	TQTSHPAKFH VEGEVYCNVC HSRNLNELS ERMAGAQQVL DCKDDSKKVI YSIGGETDQD	
	Pla 1 1.0102	-----	
	Pla 1 1.0103	-----	
		70 80 90 100 110 120	
	Pla 1 1.0101	GVYRLPVVGY HEDCEIKLVK SSRPDCSEIP KLAKGTIQTS KVDLSKNTTI TEKTRHVKPL	
	Pla 1 1.0102	-----	
	Pla 1 1.0103	-----G-----	
		131	
	Pla 1 1.0101	SFRAKTDAPG C	
	Pla 1 1.0102	-----	
	Pla 1 1.0103	-----	

Figure 1 Nucleotide and deduced amino acid sequences of Pla 11 variants

(A) Nucleotide sequence of cDNA clones encoding Pla 11 starting from the N-terminal end of the mature protein. Dashes indicate sequence identity among the clones, and blank spaces indicate deletions. The sequence encoding the mature protein is shown in capital letters, and the 3' untranslated region is shown in lowercase letters. The stop codon (TAA) is in bold typeface. The name of each clone according to IUIS nomenclature rules is indicated at the beginning of each sequence. (B) Deduced amino acid sequence from the nucleotide sequence of Pla 11 cDNA clones. Dashes indicate identity with the amino acid residue in the upper line. The potential N-glycosylation site is in a box.

Figure 2 Alignment of amino acid sequences of Pla I 1 and Ole e 1

(A) SDS/PAGE analysis of purified rPl a 1, and its low- (M) and high- (D) molecular-mass components isolated by size-exclusion chromatography; natural Pl a 1 was included as a control. Proteins were stained with Coomassie Brilliant Blue. (B) IgE immunodetection of the same samples with a serum pool from *Pl. lanceolata*-allergic patients. (C) SDS/PAGE analysis of low-molecular-mass components before (M) and after (dM) treatment with PNGase F. Staining was carried out with Coomassie Brilliant Blue. (D) Detection of proteins electroblotted on to nitrocellulose membranes with digoxigenin-labelled *Galanthus nivalis* agglutinin (GNA) as described in the Experimental section. Lanes M and dM are as in (C); CY, carboxypeptidase Y, which was included as a positive control for GNA binding. The positions of molecular-mass markers (in kDa) are shown on the left. Potential N-glycosylation sites are in a box.

The 23 kDa band observed in the electrophoretogram of rPl a 1 could correspond to a glycosylated form of the recombinant allergen, and the discrepancy in the apparent molecular mass with respect to the natural glycosylated form (20 kDa) might arise from the different glycosylated modifications introduced by the yeast or the pollen. Further experiments were carried out to clarify this point. Treatment of a sample of low-molecular-mass components (Figure 3C, lane 'M') with PNGase F caused the conversion of the 23 kDa band into the 17 kDa band, demonstrating that they corresponded to the glycosylated and

Figure 4 MS analysis

Mass spectra recorded for rPlal1 monomer not subjected (M) or subjected (dM) to enzymic deglycosylation with PNGase F, exhibiting the single-charged protein molecular ions. The scale is in arbitrary units.

non-glycosylated monomeric forms of the recombinant allergen respectively (Figure 3C, lane 'dM').

Recently, we reported that the natural Plal1 allergen bears an N-linked glycan of the 'complex' type [8]. In the present study, we have studied the structure of the carbohydrate moiety of the recombinant allergen using lectins with different specificities. Samples of monomeric rPlal1 (M), including the 23 and 17 kDa components, and PNGase F-treated monomeric rPlal1 (dM), then consisting only of the 17 kDa component, were used in these experiments. None of the samples reacted with SNA, MAA, PNA or DSA (results not shown), and only a positive reaction was found for the 23 kDa component of M with GNA (Figure 3D, lane 'M'). GNA reacts with N-glycosidically-linked 'high-mannose'- or 'hybrid'-type carbohydrate chains, and with O-glycosidically-linked mannose of yeast glycoproteins. The fact that this response was abolished after treatment with N-glycosidase F (no positive reaction in Figure 3D, lane 'dM'), and that there is no reaction of M with DSA, which recognizes 'complex' and 'hybrid' N-glycan structures, indicates that the recombinant allergen bears N-glycosidically-linked 'high-mannose'-type carbohydrate chains. All these results were confirmed by MS analysis (Figure 4). Thus, whereas MS spectra of the sample treated with PNGase F (dM) showed a sharp peak corresponding to the molecular ion of the non-glycosylated protein (14665 Da), the untreated sample (M) gave additional overlapped peaks ranging between 16 700 and 18 500 Da, with a small difference in mass values between consecutive apexes that approximately accounts for one mannose residue. This is in agreement with the existence of several forms of rPlal1 with different degrees of glycosylation. The most prominent peaks exhibited molecular masses of 17 169, 17 328, 17 491 and 17 659 Da, which would

Figure 5 CD spectra of natural and recombinant Plal1 in the far-UV region

Samples of the proteins in 10 mM sodium phosphate buffer, pH 7.0, were analysed as described in the Experimental section. Values are expressed as mean residue ellipticities.

correspond to, respectively, $\text{Man}_{13}\text{GlcNAc}_2$, $\text{Man}_{14}\text{GlcNAc}_2$, $\text{Man}_{15}\text{GlcNAc}_2$ and $\text{Man}_{16}\text{GlcNAc}_2$ oligosaccharide moieties linked to the polypeptide chain of rPlal1.

Evidence of the correct folding of rPlal1 was obtained through CD spectroscopic analysis in the far-UV region. The CD spectrum of the recombinant protein was similar both in shape and molar ellipticity values to that of the natural allergen (Figure 5), which indicates the maintenance of the native conformation of the pollen allergen in the recombinant protein expressed in *P. pastoris*.

Immunochemical properties

As mentioned above, rPlal1 is recognized, in the ELISA experiment for Plal1 quantification, by an anti-*Pl. lanceolata* rabbit serum and an mAb raised against the natural allergen that was also used for purification of the recombinant allergen by affinity chromatography. rPlal1 was also recognized, like nPlal1, by a serum pool from plantain-allergic patients in a Western blot analysis (Figure 3B). Furthermore, the IgE-binding capacity of rPlal1 was analysed by means of an inhibition ELISA with the same pool of sera from plantain-allergic patients. The inhibition curve obtained with rPlal1 overlaps with that of the nPlal1 in the middle- and low-inhibition range, reaching up to 80% inhibition of IgE-binding (Figure 6). In the same experiment, an inhibition curve with natural Ole e 1 was also included to test the allergenic significance of the sequence similarity revealed in the present paper. Ole e 1 was able to inhibit the IgE-binding to nPlal1 up to 25%, although an Ole e 1 concentration about 1000-fold higher than that of Plal1 was needed to attain such an inhibition level.

Figure 6 Inhibition ELISA of specific IgE-binding to nPla 1

The binding of specific IgE from a pool of sera from plantain-allergic patients was inhibited by the addition of rPla 1.0101 or natural Ole e 1. An inhibition curve with nPla 1 was used for comparison. Samples were analysed in triplicate. Error bars indicate the standard deviation of the measurements.

DISCUSSION

Through cDNA cloning and PCR amplification, we have isolated clones encoding the major English plantain pollen allergen Pla 1. Three different variants of the allergen have been identified by sequencing these cDNA clones. For all of the variants, the open reading frame encodes a mature Pla 1 polypeptide of 131 amino acids, with six cysteine residues and one potential N-glycosylation site. Two 25-residue signal-peptide variants, differing only at position -4, have been identified for Pla 1. Both sequences follow the '(-3,-1) rule' for cleavage of the signal peptide [13], since small and neutral residues are found at these positions (alanine and serine respectively). Thus the processing of both signal peptides would result in the same mature protein, the N-terminal sequence of which fully coincides with that found for the natural allergen [7].

The sequence polymorphisms of the mature peptide that give rise to different Pla 1 variants were found in sequences of clones from different PCR reactions, which rules out the possibility that these variants arise from PCR-generated errors. The most abundant variant detected among all the clones sequenced throughout this work, Pla 1.0101, was selected to be expressed as a recombinant protein. Since Pla 1 is a glycoprotein with three disulphide bridges in its structure, expression was carried out in *P. pastoris*, a eukaryotic expression system capable of introducing post-translational modifications in order to facilitate proper protein folding. The recombinant protein was efficiently secreted into the culture medium, and easily isolated by mAb-affinity chromatography. rPla 1 consists of a mixture of glycosylated and non-glycosylated forms, and it showed a tendency to aggregate into dimers as the natural protein [7]. N-terminal amino acid sequencing and amino acid composition analysis indicated that the polypeptide chain of rPla 1 was identical with that of the natural allergen, and additional techniques applied for the characterization of the recombinant protein supported physicochemical equivalence with nPla 1, except for the attached carbohydrates. The natural protein bears a major 'complex'-type N-glycan [8], whereas the N-glycosylation introduced by the yeast is of 'high-mannose' type. Thus the apparent molecular mass observed in SDS/PAGE for glycosylated rPla 1 is slightly higher than that of the corresponding glycosylated form of nPla 1 (Figures 3A and 3B). Moreover, glycoproteins expressed in

P. pastoris usually consist of micro-heterogeneous mixtures of isomers with a variable number of mannose residues [14,15], and this seems to occur also with rPla 1. Thus the glycosylated monomeric and dimeric forms produced broad bands upon SDS/PAGE, in contrast with the sharp band observed for the non-glycosylated Pla 1 monomer (Figure 3). Moreover, detection of the monomeric glycosylated form with the lectin GNA revealed a long band-smearing pattern, probably caused by different degrees of glycosylation. Furthermore, MS analysis revealed a wide heterogeneous peak corresponding to the glycosylated monomer (Figure 4). The difference in molecular mass between the highest apexes of this complex peak and the peak corresponding to the non-glycosylated monomer is compatible with the presence of 13–16 mannose residues attached to the core of two GlcNAc moieties of the N-linked oligosaccharide. Nevertheless, the different glycosylation pattern of rPla 1 as compared with the natural allergen does not seem to affect the folding of the protein, as demonstrated by the similarity of CD spectra.

Different lines of evidence suggest that the conformation of antigenic and allergenic epitopes is preserved in the recombinant protein. The immunochemical equivalence of the natural and recombinant allergen was tested with monoclonal and polyclonal antibodies. In fact, the selection of recombinant clones was performed by analysing the culture supernatants by a 'sandwich' ELISA that was originally designed for the quantification of nPla 1 in allergenic extracts [9], in which an anti-Pla 1 mAb on the solid phase and an anti-*Pl. lanceolata* rabbit serum as the second antibody are used. Moreover, the most convincing demonstration that rPla 1 bears the antigenic epitope defined by the mAb is provided by the fact that it could be purified by immunoaffinity chromatography. Regarding the IgE-binding epitopes, at least the immunodominant ones are preserved in rPla 1, since it was able to inhibit to a large extent the binding of specific IgE from human sera to nPla 1. The possibility that the residual non-inhibited IgE binding could be due to the lack of the variants that have not been expressed does not seem to be very plausible, since the hydrophilicity profiles of these variants do not significantly differ from the profile of rPla 1.0101 (results not shown). However, although this possibility cannot be completely ruled out until these variants are produced and tested, a more convincing explanation could be the presence of different carbohydrate moieties in each allergen. It has been demonstrated [8] that specific IgE in sera of some plantain-allergic patients can react with the 'complex' N-glycan of nPla 1, and hence these sera are cross-reactive with unrelated N-glycoproteins having 'complex' type N-glycans. If this explanation were correct, rPla 1 would be a very useful tool for the diagnosis of allergic patients specifically sensitized to *Pl. lanceolata*, avoiding false positive results caused by cross-reactive carbohydrate epitopes.

Another interesting finding in this work is the significant sequence similarity disclosed between Pla 1 and the major allergen from the distantly related olive pollen, Ole e 1. Nevertheless, only a rather limited allergenic cross-reactivity between these allergens has been demonstrated by means of an inhibition ELISA, which could explain why an association between the allergic sensitization to these pollens has not been reported so far. Availability of the Pla 1 sequence will allow future studies on the basis of molecular modelling and comparison of surface-exposed residues to investigate the molecular basis of the limited cross-reactivity between these important pollen allergens.

In conclusion, we report here the complete amino acid sequence of the major allergen from plantain pollen, which has been produced as a recombinant protein in the yeast *P. pastoris*, in a soluble, correctly folded form, with most of its

biochemical and immunochemical features identical with those of its natural counterpart. The results described here may have future applications both for clinical purposes and for basic research, to study the influence of interspecies and isoform substitutions on B- and T-cell responses in allergic subjects.

We thank Linda Puebla for language revision, and Carmen Barrio for expert secretarial assistance. This work was supported by ALK-Abelló.

REFERENCES

- Lewis, W. H. and Imber, W. E. (1975) Allergy epidemiology in the St. Louis Missouri area. IV. Weeds. *Ann. Allergy* **35**, 180–187
- Peat, J. K. and Woolcock, A. J. (1991) Sensitivity to common allergens: relation to respiratory symptoms and bronchial hyper-responsiveness in children from three different climatic areas of Australia. *Clin. Exp. Allergy* **21**, 573–581
- Spiekma, F. T. M., Charpin, H., Noland, N. and Stix, E. (1980) City spore concentration in the European Economic Community (EEC). *Clin. Allergy* **10**, 319–329
- D'Amato, G. and Lobefalo, G. (1989) Allergenic pollens in the southern Mediterranean area. *J. Allergy Clin. Immunol.* **83**, 116–122
- Subiza, J., Jerez, M., Jiménez, J. A., Narganes, M. J., Cabrera, M., Varela, S. and Subiza, E. (1995) Clinical aspects of allergic disease. Allergenic pollen and pollinosis in Madrid. *J. Allergy Clin. Immunol.* **96**, 15–23
- Asero, R., Mistrello, G., Roncarolo, D. and Casarini, M. (2000) Detection of allergens in plantain (*Plantago lanceolata*) pollen. *Allergy* **55**, 1059–1062
- Calabozo, B., Barber, D. and Polo, F. (2001) Purification and characterization of the main allergen of *Plantago lanceolata* pollen, Pla I 1. *Clin. Exp. Allergy* **31**, 322–330
- Calabozo, B., Barber, D. and Polo, F. (2002) Studies on the carbohydrate moiety of Pla I 1 allergen. Identification of a major N-glycan and significance for the immunoglobulin E-binding activity. *Clin. Exp. Allergy* **32**, 1628–1634
- Calabozo, B., Duffort, O., Carpizo, J. A., Barber, D. and Polo, F. (2001) Monoclonal antibodies against the major allergen of *Plantago lanceolata* pollen, Pla I 1. *Allergy* **56**, 429–435
- Lombardero, M., Quirce, S., Duffort, O., Barber, D., Carpizo, J., Chamorro, M. J., Lezaun, A. and Carreira, J. (1992) Monoclonal antibodies against *Olea europaea* major allergen: allergenic activity of affinity-purified allergen and depleted extract and development of a radioimmunoassay for the quantitation of the allergen. *J. Allergy Clin. Immunol.* **89**, 884–894
- Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: a Laboratory Manual*, 2nd edn, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
- Chomczynski, P. and Sacchi, N. (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate–phenol–chloroform extraction. *Anal. Biochem.* **162**, 156–159
- Von Heijne, G. (1984) How signal sequences maintain cleavage specificity. *Mol. Biol.* **173**, 243–251
- Trimble, R. B., Atkinson, P. H., Tschopp, J. F., Townsend, R. R. and Maley, F. (1991) Structure of oligosaccharides on *Saccharomyces SUC2* invertase secreted by the methylotrophic yeast *Pichia pastoris*. *J. Biol. Chem.* **266**, 22807–22817
- Huecas, S., Villalba, M., González, E., Martínez-Ruiz, A. and Rodríguez, R. (1999) Production and detailed characterization of biologically active olive pollen allergen Ole e 1 secreted by the yeast *Pichia pastoris*. *Eur. J. Biochem.* **261**, 539–546

Received 25 September 2002/21 February 2003/accepted 19 March 2003

Published as BJ Immediate Publication 19 March 2003, DOI: 10.1042/BJ20021491