

Identification of the cysteine protease Amb a 11 as a novel major allergen from short ragweed

Julien Bouley, PhD,* Rachel Groeme, MSc,* Maxime Le Mignon, PhD, Karine Jain, MSc, Henri Chabre, PhD, Véronique Bordas-Le Floch, PhD, Marie-Noëlle Couret, PhD, Laetitia Bussi eres, MSc, Aur elie Lautrette, PhD, Marie Naveau, PhD, V eronique Baron-Bodo, PhD, Vincent Lombardi, PhD, Laurent Mascarell, PhD, Thierry Batard, PhD, Emmanuel Nony, PhD, and Philippe Moingeon, PhD *Antony, France*

Background: Allergy to pollen from short ragweed (*Ambrosia artemisiifolia*) is a serious and expanding health problem in the United States and in Europe.

Objective: We sought to investigate the presence of undescribed allergens in ragweed pollen.

Methods: Ragweed pollen proteins were submitted to high-resolution gel electrophoresis and tested for IgE reactivity by using sera from 92 American or European donors with ragweed allergy. Pollen transcriptome sequencing, mass spectrometry (MS), and recombinant DNA technologies were applied to characterize new IgE-binding proteins.

Results: High-resolution IgE immunoblotting experiments revealed that 50 (54%) of 92 patients with ragweed allergy were sensitized to a 37-kDa allergen distinct from Amb a 1. The full-length cDNA sequence for this molecule was obtained by means of PCR cloning after MS sequencing of the protein combined with ragweed pollen RNA sequencing. The purified allergen, termed Amb a 11, was fully characterized by MS and confirmed to react with IgEs from 66% of patients. This molecule is a 262-amino-acid thiol protease of the papain family expressed as a combination of isoforms and glycoforms after proteolytic removal of N- and C-terminal propeptides from a proform. Three-dimensional modeling revealed a high structural homology with known cysteine proteases, including the mite Der p 1 allergen. The protease activity of Amb a 11, as well as its capacity to activate basophils from patients with ragweed allergy, were confirmed. The production of a nonglycosylated recombinant form of Amb a 11 in *Escherichia coli* established that glycosylation is not required for IgE binding.

Conclusion: We identified the cysteine protease Amb a 11 as a new major allergen from ragweed pollen. Given the similar physicochemical properties shared by the 2 major allergens, we hypothesize that part of the allergenic activity previously ascribed to Amb a 1 is rather borne by Amb a 11. (J Allergy Clin Immunol 2015;■■■:■■■-■■■.)

Key words: Allergen, Amb a 11, *Ambrosia artemisiifolia*, cysteine protease, mass spectrometry, short ragweed

Abbreviations used

AIT:	Allergen immunotherapy
2D:	Two-dimensional
IUIS:	International Union of Immunological Societies
MS:	Mass spectrometry
MS/MS:	Tandem mass spectrometry
MW:	Molecular weight
nanoLC:	Nanoflow liquid chromatography
pAb:	Polyclonal antibody
PDB:	Protein Data Bank
RACE:	Rapid amplification of cDNA ends

Pollen allergens from short ragweed (*Ambrosia artemisiifolia* var. *elator*) cause severe type I respiratory allergies.¹ Short ragweed pollen exposure has been consistently and significantly expanding in North America and Europe,² possibly as a consequence of climate change.³ As of today, allergen immunotherapy (AIT) with ragweed pollen extract is the only treatment option available to patients to address the underlying mechanisms of allergic inflammation. Specifically, AIT reduces both rhinoconjunctivitis symptoms and use of symptomatic medications as a consequence of a reorientation of inappropriate ragweed pollen-specific T_H2 responses.⁴⁻⁹ Importantly, to be efficacious, allergen extracts used for AIT should contain all major/relevant allergens causing allergic inflammation in patients.

To date, up to 10 distinct allergens have been identified in *A. artemisiifolia* pollen, according to the existing record of the International Union of Immunological Societies (IUIS) allergen nomenclature subcommittee.¹⁰ Among those, the pectate lyase Amb a 1 is currently considered the most allergenic molecule, with greater than 90% of patients with ragweed allergy sensitized to it.¹¹⁻¹³ Ragweed pollen also contains minor allergens, such as Amb a 3 and Amb a 7 (plastocyanins),¹⁴ Amb a 4 (homolog to the major mugwort allergen Art v 1),¹⁵ Amb a 5 (an approximately 5-kDa allergen),¹⁶ Amb a 6 (a nonspecific lipid transfer protein), Amb a 8 (profilin), and Amb a 9 and Amb a 10 (calcium-binding proteins) panallergens.¹⁷ Importantly, however, there is clear evidence that some patients with ragweed allergy exhibit a strong IgE reactivity against various ragweed pollen proteins distinct from known allergens (unpublished results).^{11,12}

In this context we investigated the presence of unidentified ragweed allergens, combining high-resolution gel electrophoresis and mass spectrometric (MS) analysis of pollen proteins exhibiting IgE reactivity. We report the identification and characterization of a new major allergen belonging to the cysteine protease family, which is now recorded as Amb a 11.0101 in the IUIS official nomenclature.

From Research & Pharmaceutical Development, Stallergenes.

*These authors contributed equally to this work.

Disclosure of potential conflict of interest: J. Bouley declares no relevant conflicts of interest. The rest of the authors are employed by Stallergenes.

Received for publication March 28, 2014; revised February 16, 2015; accepted for publication March 2, 2015.

Corresponding author: Philippe Moingeon, PhD, Research and Development, Stallergenes, 6 rue Alexis de Tocqueville, 92183 Antony Cedex, France. E-mail: pmoingeon@stallergenes.com.

0091-6749/\$36.00

  2015 American Academy of Allergy, Asthma & Immunology

<http://dx.doi.org/10.1016/j.jaci.2015.03.001>

METHODS

Patient population

Sera from both European and American allergic donors were selected for proteomics experiments. European patients (n = 69) included in a phase I clinical study (ClinicalTrials.gov identifier NCT01224834) after obtaining informed consent were selected based on a history of symptoms associated with seasonal ragweed pollen allergy, a positive skin prick test response to ragweed pollen, and specific IgE levels to ragweed pollen of greater than 0.70 kU/L, as measured by means of ImmunoCAP (Thermo Scientific, Uppsala, Sweden). Plasma samples from American patients given a diagnosis of ragweed allergy (n = 23, IgE levels to ragweed pollen >1 kU/L) were purchased from PlasmaLab International (Everett, Wash).

Identification of proteins reactive to human IgE

Total proteins were extracted from ragweed pollen (Greer Laboratories, Lenoir, NC) by using ammonium bicarbonate, fractionated by means of 1-dimensional (n = 23) or 2-dimensional (2D; n = 69) PAGE and then blotted onto a nitrocellulose membrane before probing with individual sera from donors with ragweed allergy for IgE reactivity or with a polyclonal antibody (pAb), as described in the [Methods](#) section in this article's Online Repository at www.jacionline.org. Protein spots were excised from 2D gels by using an EXquest spot cutter (Bio-Rad Laboratories, Marnes-la-Coquette, France), processed by means of tryptic in-gel digestion, and analyzed by using MS, as described in the [Methods](#) section in this article's Online Repository. A combination of 2D PAGE, immunoblotting, and Edman sequencing was used to determine the N-terminal amino acid sequence of Amb a 11.¹⁸ The conjugated N-terminal peptide of Amb a 11 (GSAPGSIDTDPNKDF) was used to immunize rabbits and produce anti-Amb a 11 pAb (AGRO-BIO, La-Ferté-Saint-Aubin, France). Basophil activation tests were performed, as described in the [Methods](#) section in this article's Online Repository.

Ragweed pollen RNA sequencing and Amb a 11 cDNA cloning

Deep mRNA sequencing was performed by Beckman Coulter Genomics (Grenoble, France) with a 454 sequencer using titanium chemistry. Up to 43,149 entries were obtained after *de novo* assembly and open reading frame translation with Newbler software (Roche Diagnostics, Meylan, France).

Total RNA was used as a template for 5' and 3' rapid amplification of cDNA ends (RACE) PCR experiments using the SMARTer cDNA synthesis kit (Clontech, Saint-Germain-en-Laye, France), as described in the [Methods](#) section in this article's Online Repository.

Purification and analysis of nAmb a 11 and rAmb a 11

The mature form of nAmb a 11 was purified by means of off-gel electrophoresis (Agilent Technologies, Les-Usis, France), followed by reversed-phase HPLC separation with a high-recovery macroporous reversed-phase C₁₈ column (Agilent Technologies, Santa Clara, Calif), as described in the [Methods](#) section in this article's Online Repository. Purified nAmb a 11.0101 was fully characterized by means of nanoflow liquid chromatography (nanoLC)-MS/MS. Full-length Amb a 11.0101 was produced as a recombinant molecule in *Escherichia coli* and tested for enzymatic activity, as described in the [Methods](#) section in this article's Online Repository.

Bioinformatic analysis and Amb a 11 3-dimensional structure modeling

The following sequences were obtained by running BLAST and PSI-BLAST against the nrNCBI and UniProt databases with the Amb a 11 amino acid sequence as a query: the plant-derived food allergens Act d 1 (ACTN_ACTDE), Ana c 2 (BROM1_ANACO), the cysteine protease-homolog from *Carica papaya* (PAPA1_CARPA, Papain), and the mite allergens Der f 1 (PEPT1_DERFA), Der p 1 (PEPT1_DERPT), and Der p 1

homolog from *Aleuroglyphus ovatus* (A7UNT9_9ACAR, grain mite group 1 allergen). Alignments were made with CLUSTALW and ESPript¹⁹ software. Sequences that were not annotated as "allergen" were removed. Geno3D²⁰ was used as an automatic Web server for protein molecular modeling. The 3-dimensional crystal structure of vignain (Protein Data Bank [PDB] ID 1S4V) determined by using x-ray diffraction²¹ was used as a template structure. The PyMOL software (Schrödinger, Mannheim, Germany) was used to display the 3-dimensional structure of Amb a 11.

RESULTS

Evidence for a new major ragweed pollen allergen

Proteins from an aqueous extract were separated by means of 2D gel electrophoresis, stained with a fluorescent dye (CyDye), and immunoblotted by using individual sera from 69 donors with ragweed allergy (European, n = 46; American, n = 23) to identify new ragweed pollen allergens ([Fig 1, A-D](#)). All IgE-reactive protein spots were retrieved from the gels, trypsin digested, and submitted to tandem mass spectrometry (MS/MS). Among 19 spots analyzed, 14 could be assigned to known ragweed allergens (ie, Amb a 1, Amb a 5, and Amb a 8; see [Table E1](#) in this article's Online Repository at www.jacionline.org). The amino acid sequence of the first 15 N-terminal residues of spot number 14 was identified as GSAPGSIDTDPNKDF, without any homology to known proteins. As expected, spots corresponding to the major allergen Amb a 1 (spots 2-10, 12, 13, and 16 in [Fig 1, B](#)) were confirmed to react to serum IgE from a majority of allergic patients, with 59 (86%) of 69 sera reacting to at least 1 Amb a 1 isoallergen. The 2D immunoblot analysis confirmed that the most prevalent allergens in patient sensitization include Amb a 1.01 and 1.03 isoforms (spots 2, 3, and 6-9; [Fig 1, B and C](#)). Protein spots reactive to serum IgE from approximately 15% of subjects were classified as minor ragweed allergens (ie, spots 1, 11, and 18, as well as spots 17 and 19 corresponding to Amb a 8 and Amb a 5, respectively). It is noteworthy that spots 14 and 15 were reactive to serum IgE from 37 (54%) of 69 donors ([Fig 1, C and D](#)). The same tryptic peptide sequences were obtained by using MS/MS for both spots 14 and 15, as well as for 2 more acidic spots, establishing that they represent a unique unknown protein. On the basis of an extensive characterization illustrated below, the latter was officially recorded as ragweed pollen allergen Amb a 11.0101 by the IUIS allergen nomenclature subcommittee. Collectively, this seroepidemiologic analysis documents 4 prominent patterns of IgE sensitization among patients, involving either Amb a 1 alone ([Fig 1, B](#)), multiple allergens (data not shown), predominantly Amb a 1 and Amb a 11 ([Fig 1, C](#)) or mostly Amb a 11 ([Fig 1, D](#)), representing around 40%, 15%, 30%, and 15% of patients, respectively.

In our 1-dimensional immunoblotting assays with sera from 23 additional European donors with ragweed allergy, Amb a 1 and Amb a 11 resolved as distinct bands, with evidence for a strong reactivity with patient's IgE for both allergens ([Fig 1, E, lane 2](#)). By using a specific rabbit pAb directed against Amb a 11, the latter was essentially observed as a mature form with an apparent molecular weight (MW) of approximately 37 kDa on SDS-PAGE and a minor band at approximately 52 kDa, likely corresponding to a proform ([Fig 1, E, lane 1](#)). As determined by using SDS-PAGE, the observed MW of Amb a 11 is larger than expected based on the amino acid sequence shown below, likely because of a glycosylated N-terminal domain, rigid (proline rich) C-terminus, or both.

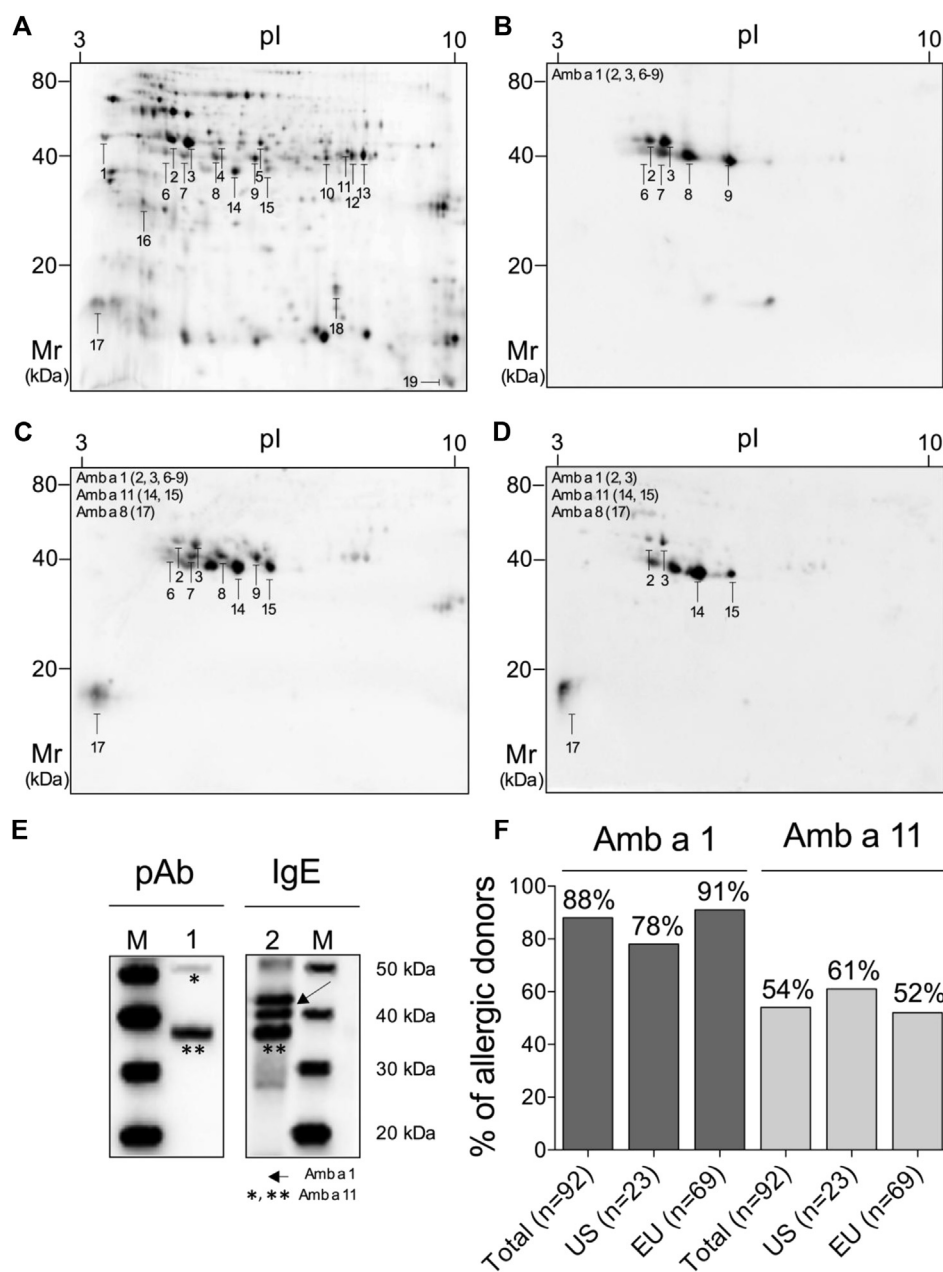


FIG 1. Detection of IgE-reactive proteins in short ragweed pollen extract. **A-D**, 2D gels stained with fluorescent dye (Fig 1, A) or probed with individual patient's serum IgE (Fig 1, B-D). Numbers refer to IgE-reactive spots (see Table E1). *pl*, Isoelectric point. **E**, One-dimensional gel probed with anti-Amb a 11 pAb (lane 1) or serum IgE (lane 2). *M*, MW markers. **F**, Percentage of allergic donors with IgE reactivity to Amb a 1 and Amb a 11 in Western blot analyses.

Prevalence values for IgE reactivity to Amb a 1 and Amb a 11 were determined based on percentages of patients scored as positive among all patients with ragweed allergy tested (Fig 1, F). For counting IgE-reactive proteins, only those with at least 1 clearly visible protein spot or band were considered. Altogether, among the 92 patients with ragweed allergy included in our study, 88% and 54% of donors exhibited IgE reactivity to Amb a 1 and Amb a 11, respectively, with no significant differences between American and European patients. It is noteworthy that such percentages of reactivity are likely underestimated because some IgE epitopes could have been disrupted during PAGE sample preparation.

To complement these analyses, total mRNA was isolated from ragweed pollen, and the transcriptome was analyzed by means of high-throughput sequencing. By using high-throughput sequencing of the ragweed pollen transcriptome and MS (see Table E1), we further identified the amino acid sequence of Amb a 11 (spots 14 and 15), which confirms the results of the RACE-PCRs presented below. Also, 2 new undescribed Amb a 1 isoallergens (spots 1 and 11) were identified, with 72% amino acid sequence identity with the isoallergen Amb a 1.01 for the former and 59% similarity with Amb a 1.05 for the latter. The amino acid sequence of spot 18 was highly similar to Amb a 3.²²

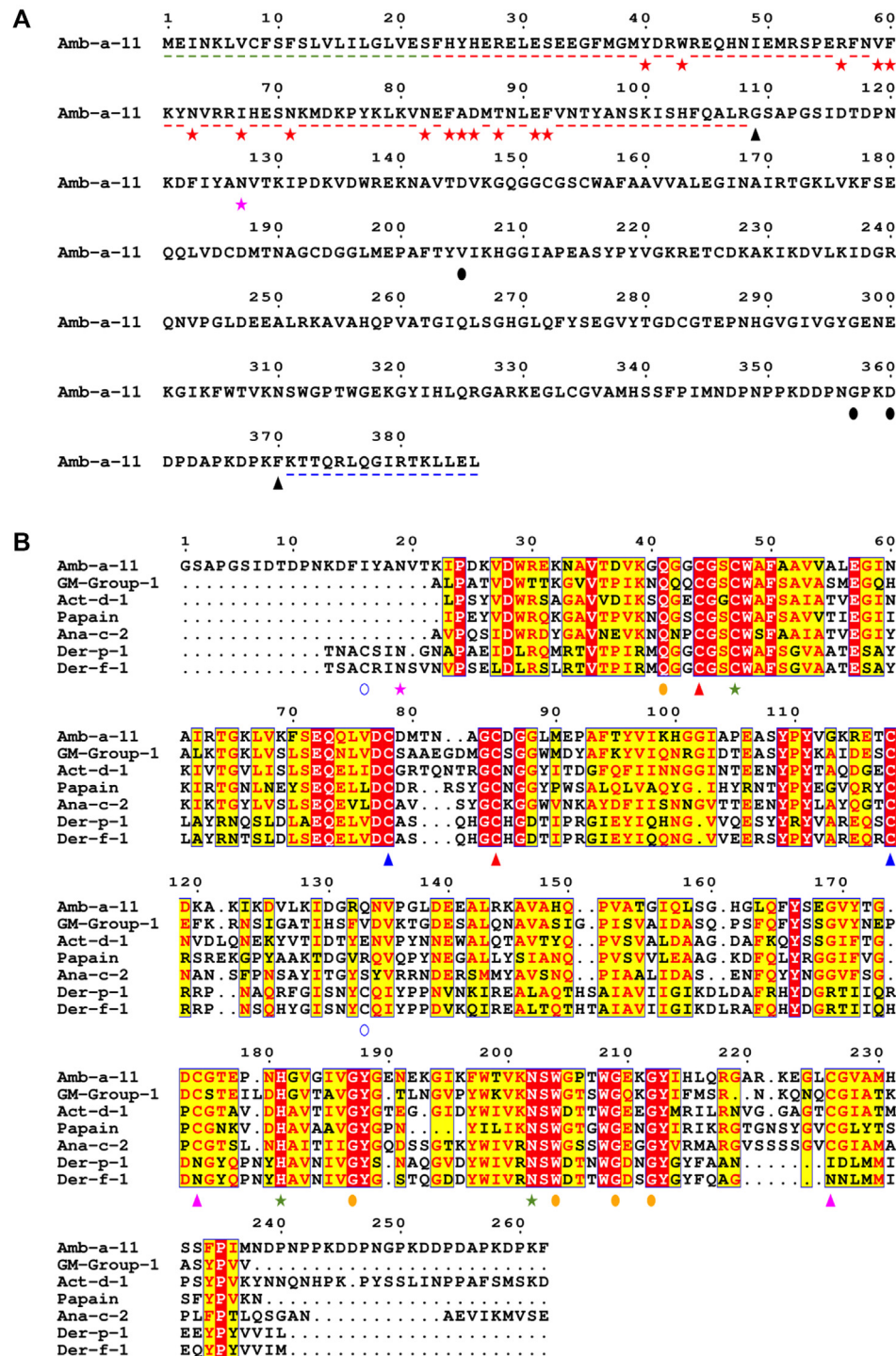


FIG 2. Primary structure of Amb a 11 and alignment with homologous allergens. **A**, Pre/pro-sequence. Green, red, and blue dashes indicate N-terminal signal peptide and N-terminal and C-terminal propeptides, respectively. Black circles, Amino acid variants V1205 G/A/V357 and D/E360; black triangles, first and last residues of the mature form; magenta stars, N-glycosylation site (N₁₂₇); red stars, often highly conserved residues in the protease inhibitor I29 domain. **B**, Alignment of the mature form with homologous allergens (see the [Methods](#) section). N-glycosylation site (magenta stars; N₁₃), active site (green stars; C₄₇, H₁₈₁, and N₂₀₂), peptidase C1A domain (orange circles; Q₄₁, G₁₈₇, W₂₀₄, G₂₀₉, and G₂₁₂), and cysteine (triangles/open circles) residues are highlighted. Dots represent gaps. Residues colored in red or yellow are identical or partially conserved.

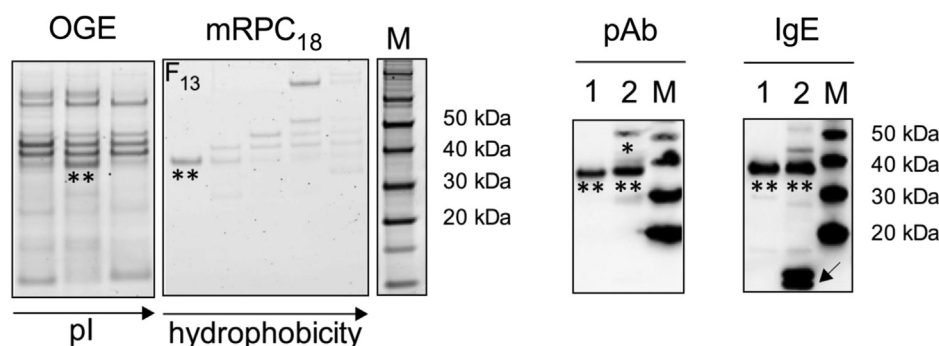


FIG 3. Purification of natural pollen allergen Amb a 11. *Left panel*, The mature form was purified by means of off-gel electrophoresis (OGE) and macroporous reversed-phase HPLC. *Right panel*, One-dimensional gel (F₁₃ fraction, *lanes 1*, and ragweed pollen extract, *lanes 2*) probed with anti-Amb a 11 pAb or serum IgE from a patient sensitized to Amb a 11 (* and **) and Amb a 8 (arrow). M, MW markers; mRPC, macroporous reversed-phase chromatography; pl, isoelectric point.

Cloning of Amb a 11 cDNA

Given its high frequency of IgE sensitization in patients with ragweed allergy, using PCR, we proceeded to clone the gene encoding Amb a 11 with oligonucleotides designed from Amb a 11 peptides sequenced with MS/MS. RACE-PCRs allowed us to identify a full-length coding sequence of 1161 bp (GenBank KF528831) with an open reading frame coding for a 386-amino-acid polypeptide (Uniprot V5LUO1_AMBAR, Fig 2, A). Several mature protein sequences were identified, reflecting the presence of at least 6 isoallergens (see Fig E1 in this article's Online Repository at www.jacionline.org). The prediction results obtained with the Signal P-4.1 algorithm showed that the protein has a signal peptide (1-22) with a possible cleavage site between S₂₂ and F₂₃. A bioinformatic analysis combined with the MS experiments presented below revealed a protease inhibitor I29 domain (40-92, InterPro ID IPR013201) within an N-terminal propeptide (23-108), which is in agreement with the N-terminal amino acid sequence previously determined by means of Edman sequencing, an N-glycosylation site (N₁₂₇), and a 16-amino-acid C-terminal propeptide (371-386; Fig 2, A). The I29 domain is typical of thiol peptidase precursors that belong to the peptidase subfamily C1A. Removal of this proregion by means of proteolytic cleavage results in enzyme activation. Multiple alignments of the mature amino acid sequence with other known proteins confirmed that Amb a 11 shares many of the typical characteristics common to cysteine proteases (Fig 2, B). Specifically, its active site contains both cysteine (C₄₇), histidine (H₁₈₁), and asparagine (N₂₀₂) residues, as well as amino acid residues Q₄₁, G₁₈₇, W₂₀₄, G₂₀₉, and G₂₁₂, together forming a conserved peptidase C1A domain (InterPro ID IPR000668). It is noteworthy that such thiol proteases have been described as major allergens in both plant food (eg, kiwi fruit [Act d 1], pineapple [Ana c 2], papaya [papain]) and mites (Der f 1, Der p 1, and grain mite group 1). Overall, Amb a 11 presents a 40% to 42% amino acid sequence identity (57% to 62% similarity) with plant food allergens and 28% to 31% amino acid sequence identity (45% similarity) with mite allergens.

Characterization of Amb a 11 isoforms and glycoforms

Amb a 11, purified from a ragweed pollen extract by means of off-gel electrophoresis and reversed-phase chromatography,

resolves by means of SDS-PAGE as a single band at approximately 37 kDa (Fig 3). Matrix-assisted laser desorption ionization time-of-flight MS analysis of LysC-peptides from Amb a 11 confirmed 77% of the deduced Amb a 11 protein sequence (see Fig E2, A, and Table E2 in this article's Online Repository at www.jacionline.org). LysC peptides were then submitted to nanoLC-MS/MS, and the amino acid sequences were compiled with the ones identified after in-gel digestion of Amb a 11 protein spots (see Fig E2, B and C, and Tables E3 and E4 in this article's Online Repository at www.jacionline.org). With this approach, the Amb a 11 protein was fully sequenced. Amino acid substitutions were confirmed at the protein level in agreement with the nucleotide sequences obtained by using RACE (ie, V/I₉₇, G/A₂₄₉, and D/E₂₅₂, with numbering corresponding to the mature Amb a 11; Fig 2, A).

Information on the C-terminus of the protein could not be obtained, suggesting that the sequence KTTQRLQGIRTKLLEL (ie, produced after cleavage after residue F₂₆₂) might correspond to a C-terminal propeptide (Fig 2, A), as previously observed for other thiol proteases belonging to the papain family.²³ Moreover, one site of glycan attachment involving residue N₁₉ was evidenced by MS/MS, as well as the associated oligosaccharide structure (eg, MMXF³; for review, see Fig 4, A).²⁴ Full details of the carbohydrate moieties attached to the N₁₉ site are given in Table E3 and Fig E3 in this article's Online Repository at www.jacionline.org. These complex type N-glycans are different from the single monosaccharide (ie, N-acetylglucosamine) identified by means of MS on nDer f 1 or Der p 1 molecules from mite extracts (our data not shown and published structure of natural Der f 1).²⁵

The determination of the intact protein mass confirmed that Amb a 11 is expressed as a combination of isoforms and glycoforms (Fig 4, B, and Fig 5). More than 20 different masses were observed from the deconvoluted mass spectrum, with excellent agreement between theoretic and experimental masses.

A His-tagged unglycosylated recombinant pro-Amb a 11 (rAmb a 11) was produced in *Escherichia coli*, purified, and used to perform immunoblots with sera from patients with ragweed allergy to investigate the role of potential carbohydrate cross-reactive determinants in IgE binding to Amb a 11 (Fig 6 and see Fig E3, A). The determination of the mass of the intact protein by means of matrix-assisted laser desorption ionization MS revealed a 42.9-kDa proform, as expected from the amino

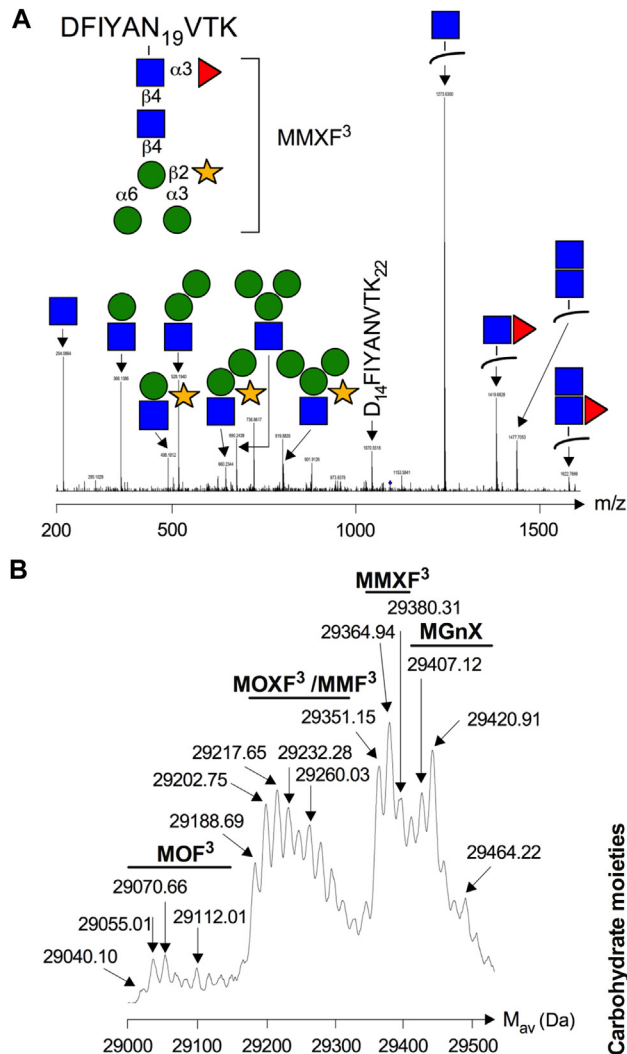


FIG 4. Analysis of Amb a 11 by using MS. **A**, Fragmentation of peptide D₁₄FIYANVTk₂₂ carrying the N-glycan MMXF³ at position N₁₉. Sugar oxonium ions and glycopeptides fragments are indicated. Blue squares, N-acetylglucosamine; green circles, mannose; red triangles, fucose; and yellow stars, xylose. **B**, NanoLC-MS analysis of Amb a 11. The observed molecular masses are reported in Fig 5.

acid sequence (Fig E3, B) and in agreement with the observed MW of unglycosylated rAmb a 11 by using SDS-PAGE (approximately 45 kDa; Fig 6, A). MS/MS analysis confirmed the absence of N-glycosylation at position N₁₉ (see Fig E3, C). The analysis of the disulfide bond pattern indicated that rAmb a 11 contains 7 cysteine residues engaged in 3 disulfide bridges (see Fig E3, B), which is in agreement with structural homologies between Amb a 11 and cysteine proteases of clan CA (Fig 2, B). rAmb a 11 is recognized by pAbs raised against the N-terminal part of Amb a 11. The IgE reactivity observed with rAmb a 11 (Fig 6 and see Fig E3, A) confirmed that N-glycosylation is not essential in this regard, thus establishing the latter as a *bona fide* allergen. The prevalence of IgE reactivity assessed with rAmb a 11 in dot blot experiments (Fig 6, C) was comparable (ie, 51%) with that measured by using PAGE immunoblotting experiments (Fig 1, F). It is noteworthy that a higher prevalence (66%) was obtained when using purified nAmb a 11 under

nondenaturing dot blotting conditions. Such differences observed in the reactivity of nAmb a 11 and rAmb a 11 can be explained by the presence of multiple proteoforms within nAmb a 11 or by the association of N-linked carbohydrates with nAmb a 11 but not rAmb a 11. Other potential causes include a limited access to IgE epitopes hidden by the N-terminal propeptide, as reported for Der f 1 proforms,²⁶ or an incomplete folding of rAmb a 11. In this regard circular dichroism spectra provided evidence for secondary structures that were similar but not identical for nAmb a 11 and rAmb a 11 (see Fig E4 in this article's Online Repository at www.jacionline.org).

Functional characterization of Amb a 11

Amb a 1 and Amb a 11 were purified from short ragweed pollen and tested for their capacity to activate blood basophils from patients with ragweed allergy based on induction of the activation marker CD203c to assess their relative allergenic activity. Fig 7, A, shows the mean results obtained in blood samples from 5 patients. A dose-dependent activation of basophils was observed with the natural ragweed pollen extract, as well as purified nAmb a 1 and nAmb a 11. At the highest dose tested (10 µg/mL), Amb a 11 upregulated CD203c to levels comparable with those observed after cross-linking with an anti-IgE antibody used as a positive control in those experiments.

The cysteine protease activity was assessed with both proforms and matured forms of rAmb a 11 expressed in *E coli* (Fig 7, B, left panel). The enzymatic activity was confirmed for both the zymogen and mature forms, although as expected, it was substantially greater for the latter. This activity was fully abolished by the E-64-specific cysteine protease inhibitor (Fig 7, B, right panel).

Modeling of Amb a 11 3-dimensional structure in comparison with the mite allergen Der p 1

The previous data suggest that both an N-terminal and a C-terminal propeptide, consisting of 86- and 16-amino-acid residues, respectively, are cleaved off during Amb a 11 maturation (Fig 2, A). Furthermore, alignment of the primary structure of Amb a 11 with other homologous protease allergens (Fig 2, B) confirms the presence of identical catalytic triads, conserved amino acid residues within a peptidase C1 domain, and, except for Der p 1 and Der f 1, a conserved position of cysteine residues. Based on those observations, it is concluded that Amb a 11 belongs to the peptidase family C1, subfamily C1A (papain family, clan CA).

A BLAST search against the PDB identified the vignain cysteine endoprotease from *Ricinus communis* (PDB ID 1S4V²¹) as the closest known molecule to Amb a 11 in terms of amino acid sequence (with 58% identity and 70% similarity). Mature Amb a 11 was thus modeled with the known vignain 3-dimensional crystal structure as a template (see Fig E5 in this article's Online Repository at www.jacionline.org). The overall fold of Amb a 11 is presented in Fig 8, A, in comparison with the structure of nDer p 1 (Fig 8, B).²⁷ As a papain-like thiol protease, Amb a 11 is expected to fold into 2 distinct globular domains divided by the active-site cleft, where substrates can bind. The active site comprises the highly conserved catalytic triad, including a cysteine, a histidine, and an asparagine residue (C₄₇, H₁₈₁, and N₂₀₂, respectively). The left domain is mostly

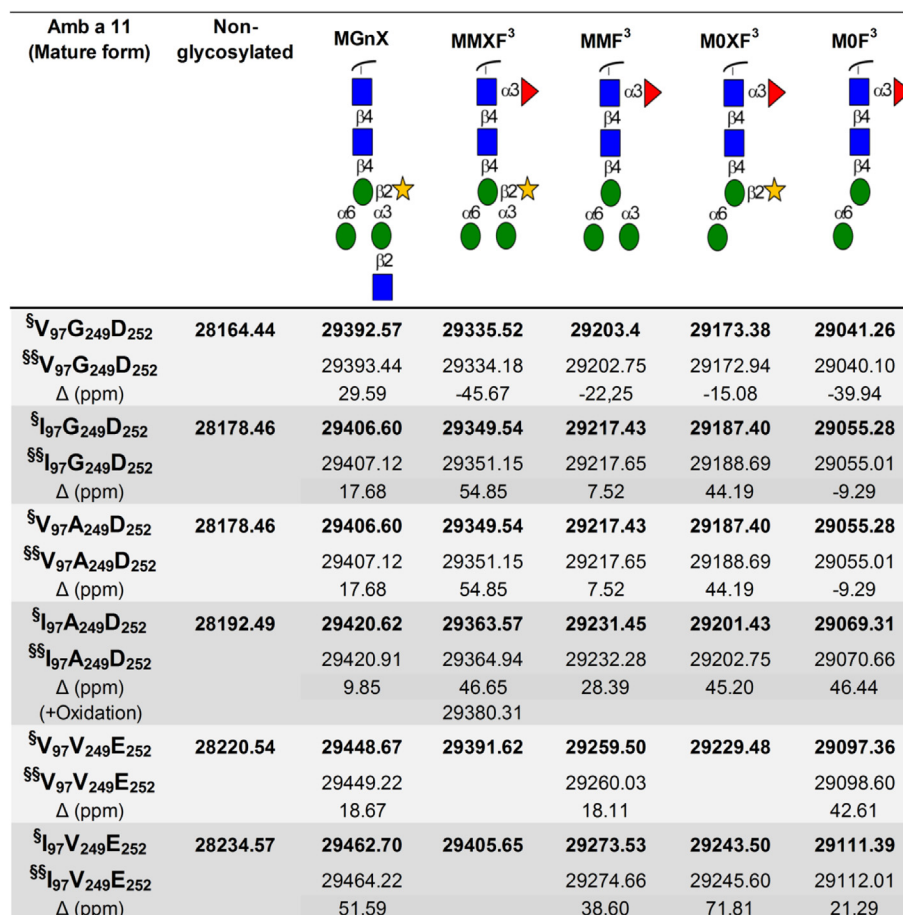


FIG 5. Isoforms and glycoforms of Amb a 11. Calculated (S) and experimental (SS) average molecular masses (considering 7 oxidized cysteine residues) were determined by using nanoLC-MS. The residue numbering corresponds to the mature protein. The oxidized form is +15.99 Da. Blue squares, N-acetylglucosamine; green circles, mannose; red triangles, fucose; yellow stars, xylose.

helical ($\alpha 1$, $\alpha 2$, $\alpha 3$, $\alpha 4$, and $\alpha 5$) but comprises several random coil segments as well. The right domain encompasses large twisted antiparallel β -sheets and contains 2 additional α -helices ($\alpha 6$ and $\alpha 7$). The Amb a 11 fold is further stabilized by 3 disulfide bonds connecting the loop preceding $\alpha 1$ to the $\alpha 2$ - $\alpha 3$ loop (C₄₄₋₈₅), the C-terminus of $\alpha 2$ to the N-terminus of $\alpha 5$ (C₇₈₋₁₁₈), and 2 random coil segments (C₁₇₅₋₂₂₆, with numbering according to the mature sequence), respectively. It is noteworthy that the similarity between the folds of mature Amb a 11 and Der p 1 (PDB-ID 3RVW) is remarkable, with the 3-dimensional model of Amb a 11 and the structure of nDer p 1 superposing with root mean square deviation of 1.29 Å over 138 C α atoms (Fig 8, C). The Amb a 11 disulfide bonds C₄₄₋₈₅ and C₇₈₋₁₁₈ are conserved between the 2 molecules, whereas disulfide C₁₇₅₋₂₂₆ is replaced by C₄₋₁₁₇ within the Der p 1 molecule. Collectively, these observations are in agreement with the observation that Amb a 11 is catalytically active.

DISCUSSION

Short ragweed is known to cause severe type I respiratory allergies. At a molecular level, up to 10 allergens (ie, Amb a 1 to Amb a 10) have been identified in short ragweed pollen. Among those, Amb a 1 (comprising 12 isoallergens)¹⁰ is considered the

only major allergen. The observation that some patients with ragweed allergy have IgE that react with unknown molecules^{11,12} prompted our search for undescribed allergens in short ragweed pollen.

We implemented an approach based on high-resolution PAGE combined with MS characterization and high-throughput pollen transcriptome sequencing. This study confirmed that both the Amb 1.01 and 1.03 isoallergens exhibit the most frequent IgE reactivity,²⁸ as detected by up to 88% of tested sera from patients with ragweed allergy. We also identified a new major IgE-binding protein, termed Amb a 11, which had not been previously reported, likely because it shares nearly the same MW and isoelectric point with Amb a 1. Herein we clearly demonstrate IgE recognition of Amb a 11 by sera from 50 (54%) of 92 American or European patients with ragweed allergy. Using 2-dimensional Western blotting, we estimate that 45% of those patients exhibit a similar or even stronger IgE reactivity with Amb a 11 when compared with Amb a 1 (Fig 1, C and D), suggesting that the former contributes to allergic symptoms in a large number of patients. Importantly, because of the potential partial denaturation of Amb a 11 during sample preparation for PAGE, we observed a higher IgE reactivity (with 66% of patients' sera) when using purified nAmb a 11 under

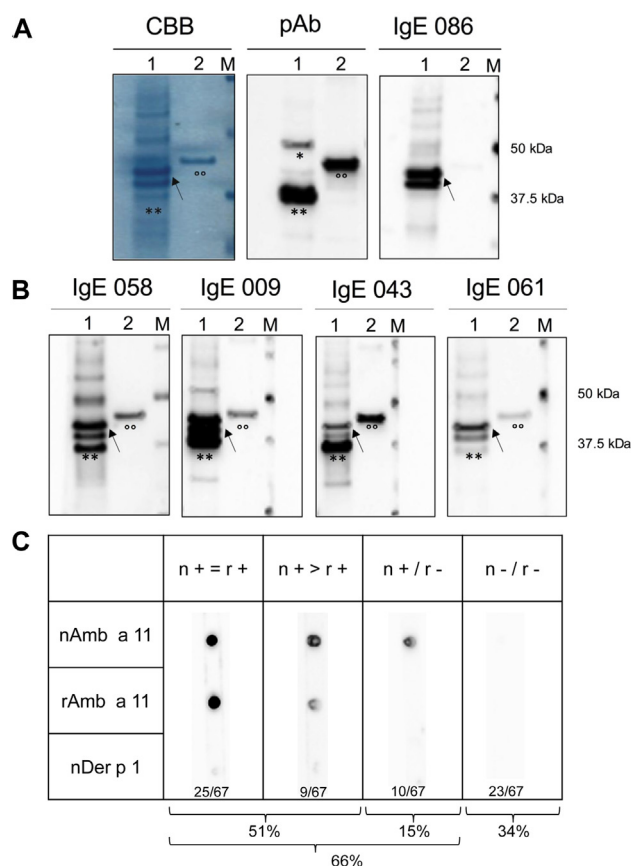


FIG 6. IgE reactivity with nAmb a 11 and rAmb a 11 molecules. **A**, **37-kDa mature nAmb a 11 (lane 1) and nonglycosylated rAmb a 11 (~45-kDa proform, lane 2) resolved on a 1-dimensional gel (Coomassie stained [CBB]) and probed with anti-Amb a 11 pAb or serum IgE 086 (negative control) reacting only to nAmb a 11 (arrow). **B**, Amb a 11 IgE reactivity with sera from 4 representative donors (see also Fig E3). Lanes 1, Ragweed pollen; lanes 2, rAmb a 11. M, MW markers. **C**, Dot blot analysis of patients' IgE reactivity (n = 67) with 100 ng of either purified nAmb a 11, rAmb a 11, or nDer p 1 (negative control). n + or n - and r + or r -, Positive or negative reactivity to nAmb a 11 and rAmb a 11, respectively.

nondenaturing dot blot experiments. Amb a 11 was confirmed to induce activation of basophils from patients with ragweed allergy, although to a lower extent than Amb a 1. In the absence of a precise assay to quantify Amb a 11, the relative abundance of this allergen in the ragweed pollen extract versus Amb a 1 is still unclear. Our first estimate is that Amb a 11 amounts correspond to approximately 1% of total proteins, as determined by using 2-dimensional gel electrophoresis. In comparison, Amb a 1 is reported to account for 6% to 15% of total proteins from a ragweed pollen extract.^{10,29}

Extensive MS characterization confirmed the 262-amino-acid sequence of Amb a 11, including the presence of N- and C-terminal propeptides that could be required for correct folding, resistance to proteolysis, or delivery of proteins to cellular compartments. MS analysis also established that Amb a 11 is N-glycosylated on the N₁₉ residue with complex carbohydrate moieties commonly found in plant allergens.²⁴ Because carbohydrate cross-reactive determinants are known to bind IgE,³⁰ we confirmed that nonglycosylated rAmb a 11 expressed as a proform in *E coli* can also react to IgE from allergic subjects.

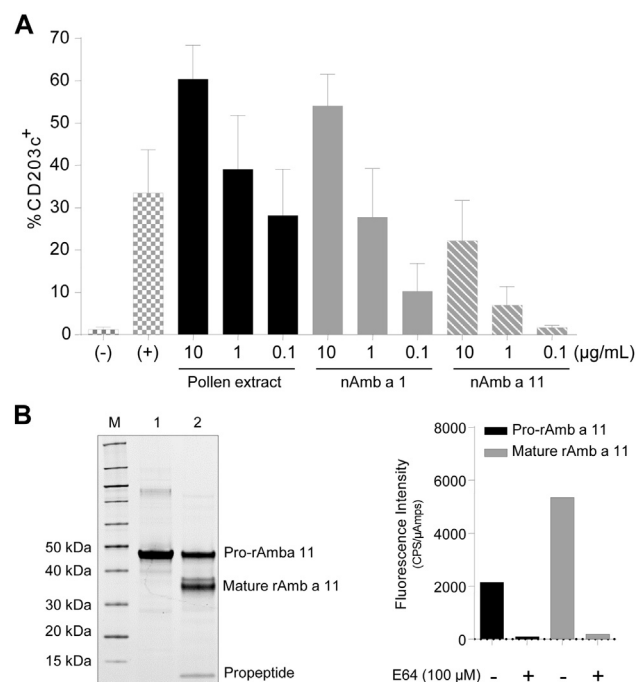


FIG 7. Functional characterization of Amb a 11. **A**, Allergen-induced basophil activation (percentage of CD203c^{bright}CRTh2⁺ basophils) in blood samples from 5 patients. Negative (-) and positive (+) controls were used. **B**, Proforms (lane 1) and mature rAmb a 11 (lane 2) were resolved on a one-dimensional gel, and protease activity was monitored with (+) or without (-) cysteine protease inhibitor. M, MW markers.

After sequence alignment with homologous known proteins, Amb a 11 was categorized as a member of the clan CA cysteine proteases, a family of proteases that includes numerous major allergens from either plants (eg, Act d 1 and Ana c 2) or mites (eg, Der f 1 and Der p 1). Among those allergens, Act d 1 is the closest plant homolog to Amb a 11, with 42% sequence identity. Of interest, Le et al³¹ have reported that the plant-derived Act d 1 is frequently recognized by kiwi-monosensitized patients and associated with severe kiwifruit allergy. Furthermore, 3-dimensional structure modeling confirmed that Amb a 11 is highly similar to the mite allergen Der p 1, one of the most potent known allergens. The broad evidence that cysteine proteases can alter epithelial barriers and further cause direct immune cell dysregulation argues for a significant contribution of Amb a 11 to ragweed pollen allergy.³²⁻³⁶

Cysteine protease allergens share a highly conserved active site and likely have a similar mode of action, although the substrate specificity of each enzyme might vary. Alignment of the Amb a 11 amino acid sequence with other cysteine proteases and molecular modeling reveals a high conservation of the disulfide pattern and regions involved in the catalytic site of those enzymes, strongly suggesting that Amb a 11 is functionally active. Using recombinant forms of the molecule, we confirmed that mature Amb a 11 and, to a lower extent, the proform of the molecule exhibits a cysteine protease activity that can be totally inhibited by the specific E64 inhibitor. The latter is in agreement with the cysteine protease activity observed in ragweed pollen extracts.³⁷ It is noteworthy that members of the papain group of proteases preferentially cleave peptide bonds with Arg and Leu in the P1 and P2 positions, respectively, suggesting that activation of

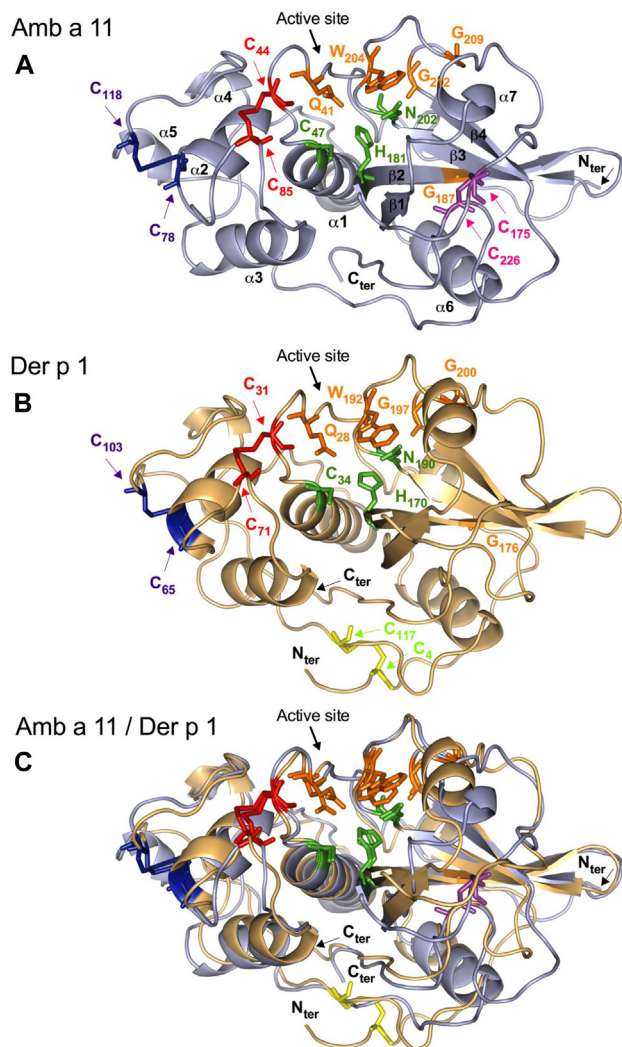


FIG 8. Model of Amb a 11 3-dimensional structure. **A** and **B**, Ribbon of Amb a 11 (model; Fig 8, A) and Der p 1 (PDB-ID 3RVW; Fig 8, B). Active site, peptidase C1A domain, and cysteine residues, as well as the termini and secondary structure elements of Amb a 11, are labeled. Residues are numbered according to the mature protein. **C**, Three-dimensional superposition of Amb a 11 and Der p 1.

Amb a 11 can take place by means of intramolecular proteolysis of the N-terminal region (ie, $L_{P2}R_{P1}G_{109}S_{P1}A_{P2}$). The presence of proline residues N-terminally located from the C-terminal cleavage site (ie, F_{262}) could also be critical in the activation process.

In conclusion, we identified Amb a 11 as a new major allergen from *A. artemisiifolia* pollen, exhibiting both a confirmed thiol protease activity and a capacity to activate the patient's basophils. Consequently, we conclude that Amb a 11 should be considered a critical component of allergen extracts used for the diagnosis and treatment of ragweed pollen allergies. The potency of Amb a 11 relative to Amb a 1 remains to be further addressed in future studies.

We thank Dr Sébastien Brier for helpful discussions on the analysis of the 3-dimensional structure. We also thank Sabi Airouche, who contributed to the basophil activation tests.

Clinical implications: On the basis of IgE reactivity documented in a majority of American and European patients and given its high homology with known allergenic cysteine proteases, Amb a 11 should be considered an essential component for the diagnosis of and specific immunotherapy for common ragweed allergy.

REFERENCES

- Arbes SJ Jr, Gergen PJ, Elliott L, Zeldin DC. Prevalences of positive skin test responses to 10 common allergens in the US population: results from the third National Health and Nutrition Examination Survey. *J Allergy Clin Immunol* 2005;116:377-83.
- D'Amato G, Cecchi L, Bonini S, Nunes C, Annesi-Maesano I, Behrendt H, et al. Allergic pollen and pollen allergy in Europe. *Allergy* 2007;62:976-90.
- Ziska L, Knowlton K, Rogers C, Dalan D, Tierney N, Elder MA, et al. Recent warming by latitude associated with increased length of ragweed pollen season in central North America. *Proc Natl Acad Sci U S A* 2011;108:4248-51.
- Creticos PS, Schroeder JT, Hamilton RG, Balcer-Whaley SL, Khattignavong AP, Lindblad R, et al. Immunotherapy with a ragweed-toll-like receptor 9 agonist vaccine for allergic rhinitis. *N Engl J Med* 2006;355:1445-55.
- Skoner D, Gentile D, Bush R, Fasano MB, McLaughlin A, Esch RE. Sublingual immunotherapy in patients with allergic rhinoconjunctivitis caused by ragweed pollen. *J Allergy Clin Immunol* 2010;125:660-6.
- Moingeon P. Update on immune mechanisms associated with sublingual immunotherapy: practical implications for the clinician. *J Allergy Clin Immunol Pract* 2013;1:228-41.
- Creticos PS, Maloney J, Bernstein DI, Casale T, Kaur A, Fisher R, et al. Randomized controlled trial of a ragweed allergy immunotherapy tablet in North American and European adults. *J Allergy Clin Immunol* 2013;131:1342-9.e6.
- Creticos PS, Esch RE, Couroux P, Gentile D, D'Angelo P, Whitlow B, et al. Randomized, double-blind, placebo-controlled trial of standardized ragweed sublingual-liquid immunotherapy for allergic rhinoconjunctivitis. *J Allergy Clin Immunol* 2014;133:751-8.
- Akdis M, Akdis CA. Mechanisms of allergen-specific immunotherapy: multiple suppressor factors at work in immune tolerance to allergens. *J Allergy Clin Immunol* 2014;133:621-31.
- Gadermaier G, Hauser M, Ferreira F. Allergens of weed pollen: an overview on recombinant and natural molecules. *Methods* 2014;66:55-66.
- Gadermaier G, Wopfner N, Wallner M, Egger M, Didierlaurent A, Regl G, et al. Array-based profiling of ragweed and mugwort pollen allergens. *Allergy* 2008;63:1543-9.
- Asero R, Wopfner N, Gruber P, Gadermaier G, Ferreira F. *Artemisia* and *Ambrosia* hypersensitivity: co-sensitization or co-recognition? *Clin Exp Allergy* 2006;36:658-65.
- Jahn-Schmid B, Hauser M, Wopfner N, Briza P, Berger UE, Asero R, et al. Humoral and cellular cross-reactivity between Amb a 1, the major ragweed pollen allergen, and its mugwort homolog Art v 6. *J Immunol* 2012;188:1559-67.
- Hunt LT, George DG, Yeh LS. Ragweed allergen Ra3: relationship to some type 1 copper-binding proteins. *J Mol Evol* 1984;21:126-32.
- Leonard R, Wopfner N, Pabst M, Stadlmann J, Petersen BO, Duss JO, et al. A new allergen from ragweed (*Ambrosia artemisiifolia*) with homology to art v 1 from mugwort. *J Biol Chem* 2010;285:27192-200.
- Ghosh B, Rafnar T, Perry MP, Bassolino-Klimas D, Metzler WJ, Klapper DG, et al. Immunologic and molecular characterization of Amb p V allergens from *Ambrosia psilostachya* (western ragweed) pollen. *J Immunol* 1994;152:2882-9.
- Wopfner N, Gruber P, Wallner M, Briza P, Ebner C, Mari A, et al. Molecular and immunological characterization of novel weed pollen pan-allergens. *Allergy* 2008;63:872-81.
- Matsudaira P. Sequence from picomole quantities of proteins electroblotted onto polyvinylidene difluoride membranes. *J Biol Chem* 1987;262:10035-8.
- Gouet P, Courcelle E, Stuart DI, Metoz F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* 1999;15:305-8.
- Combet C, Jambon M, Deleage G, Geourjon C. Geno3D: automatic comparative molecular modelling of protein. *Bioinformatics* 2002;18:213-4.
- Than ME, Helm M, Simpson DJ, Lottspeich F, Huber R, Gietl C. The 2.0 Å crystal structure and substrate specificity of the KDEL-tailed cysteine endopeptidase functioning in programmed cell death of *Ricinus communis* endosperm. *J Mol Biol* 2004;336:1103-16.
- Klapper DG, Goodfriend L, Capra JD. Amino acid sequence of ragweed allergen Ra3. *Biochemistry* 1980;19:5729-34.
- Grudkowska M, Zagdanska B. Multifunctional role of plant cysteine proteinases. *Acta Biochim Pol* 2004;51:609-24.

24. Wilson IB, Altmann F. Structural analysis of N-glycans from allergenic grass, ragweed and tree pollens: core a1,3-linked fucose and xylose present in all pollens examined. *Glycoconj J* 1998;15:1055-70.
25. Chruszcz M, Chapman MD, Vailes LD, Stura EA, Saint-Remy JM, Minor W, et al. Crystal structures of mite allergens Der f 1 and Der p 1 reveal differences in surface-exposed residues that may influence antibody binding. *J Mol Biol* 2009;386:520-30.
26. Takai T, Kato T, Yasueda H, Okumura K, Ogawa H. Analysis of the structure and allergenicity of recombinant pro- and mature Der p 1 and Der f 1: major conformational IgE epitopes blocked by prodomains. *J Allergy Clin Immunol* 2005;115:555-63.
27. Chruszcz M, Pomes A, Glesner J, Vailes LD, Osinski T, Porebski PJ, et al. Molecular determinants for antibody binding on group 1 house dust mite allergens. *J Biol Chem* 2012;287:7388-98.
28. Nandy A AS, Wald M, Pump L, Hermann A, Treder S, Hagen S, et al. Recombinant major ragweed allergen Amb a 1: physicochemical characterization and immunologic comparison of five recombinant ragweed isoallergens Amb a 1.01 to Amb a 1.05 [abstract]. *J Allergy Clin Immunol* 2013;131:AB16.
29. Wopfner N, Gadermaier G, Egger M, Asero R, Ebner C, Jahn-Schmid B, et al. The spectrum of allergens in ragweed and mugwort pollen. *Int Arch Allergy Immunol* 2005;138:337-46.
30. Malandain H, Giroux F, Cano Y. The influence of carbohydrate structures present in common allergen sources on specific IgE results. *Eur Ann Allergy Clin Immunol* 2007;39:216-20.
31. Le TM, Bublin M, Breiteneder H, Fernandez-Rivas M, Asero R, Ballmer-Weber B, et al. Kiwifruit allergy across Europe: clinical manifestation and IgE recognition patterns to kiwifruit allergens. *J Allergy Clin Immunol* 2013;131:164-71.
32. Kamijo S, Takeda H, Tokura T, Suzuki M, Inui K, Hara M, et al. IL-33-mediated innate response and adaptive immune cells contribute to maximum responses of protease allergen-induced allergic airway inflammation. *J Immunol* 2013;190:4489-99.
33. Takai T, Ikeda S. Barrier dysfunction caused by environmental proteases in the pathogenesis of allergic diseases. *Allergol Int* 2011;60:25-35.
34. Kikuchi Y, Takai T, Kuhara T, Ota M, Kato T, Hatanaka H, et al. Crucial commitment of proteolytic activity of a purified recombinant major house dust mite allergen Der p1 to sensitization toward IgE and IgG responses. *J Immunol* 2006;177:1609-17.
35. Wan H, Winton HL, Soeller C, Tovey ER, Gruenert DC, Thompson PJ, et al. Der p 1 facilitates transepithelial allergen delivery by disruption of tight junctions. *J Clin Invest* 1999;104:123-33.
36. Chapman MD, Wunschmann S, Pomes A. Proteases as Th2 adjuvants. *Curr Allergy Asthma Rep* 2007;7:363-7.
37. Gunawan H, Takai T, Kamijo S, Wang XL, Ikeda S, Okumura K, et al. Characterization of proteases, proteins, and eicosanoid-like substances in soluble extracts from allergenic pollen grains. *Int Arch Allergy Immunol* 2008;147:276-88.

Online Repository**Identification of the cysteine protease Amb a 11 as a novel major allergen from short ragweed**

Julien Bouley*, PhD, Rachel Groeme*, MSc, Maxime Le Mignon, PhD, Karine Jain, MSc, Henri Chabre PhD, Véronique Bordas-Le Floch, PhD, Marie-Noëlle Couret, PhD, Laetitia Bussieres MSc, Aurélie Lautrette, PhD, Marie Naveau, PhD, Véronique Baron-Bodo, PhD, Vincent Lombardi, PhD, Laurent Mascarell, PhD, Thierry Batard PhD, Emmanuel Nony, PhD, Philippe Moingeon, PhD.

From Stallergenes, 6 rue Alexis de Tocqueville, 92183 Antony cedex, France

* These authors contributed equally to this work

Correspondence:

Dr. Moingeon Philippe, Senior VP, Research and Pharmaceutical Development, Stallergenes, 6 rue Alexis de Tocqueville, 92183 Antony cedex, France. Tel: 0033 1 55 59 25 35, fax: 0033 1 55 59 21 02, e-mail: pmoingeon@stallergenes.com

Disclosure of potential conflict of interest:

All authors are employed by Stallergenes.

Online Repository Methods

Protein extraction, 1D-, 2D-PAGE, western-blot and dot-blot analysis

Total proteins were extracted from ragweed pollen (GREER Laboratories, Lenoir, USA). Briefly, ragweed pollen grains were homogenized in 50 mM NH_4HCO_3 followed by centrifugation at 15000g for 30 min at 4°C. For 1D-PAGE (n=23 donors), proteins were analyzed on NuPAGE 4-12% Midi gels (Life Technologies, Saint-Aubin, France) according to the manufacturer's instructions. For 2D-PAGE (n=69 donors), proteins were precipitated with PerfectFocus Kit from G-Biosciences (AGRO-BIO, La-Ferté-Saint-Aubin, France) and re-suspended in a buffer containing 7 M urea, 2 M thiourea, 4% CHAPS and 30 mM Tris pH 8.8. Proteins were quantified using a Bradford assay (Bio-Rad, Marnes-la-Coquette, France), labelled with a CyDye™ (GE Healthcare, Velizy-Villacoublay, France) and then fractionated by 2D-PAGE (3–10 NL pH range, 12.5 % DALT gel, GE Healthcare) according to the manufacturer's instructions.

Proteins were blotted onto a nitrocellulose membrane, the latter was then incubated with 3% BSA in Tris-buffered saline (TBS)/0.1% Tween 20 to block nonspecific binding and subsequently probed with individual sera from ragweed-allergic patients (n=92) at a dilution of 1:10 in 0.2% BSA in TBS/0.1% Tween 20. IgE-binding was detected by using rabbit anti-human IgE (Dako, Les-ULis, France), then anti-rabbit IgG-HRP (Merck Millipore, Molsheim, France) and visualized by chemiluminescence (Advansta WesternBright Quantum kit, Diagomics, Blagnac, France). Alternatively, proteins were probed with anti-Amb a 11 rabbit pAb (directed against the Amb a 11 N-terminus) at a dilution of 1:10000 in 0.2% BSA in TBS/0.1% Tween 20, detected by using anti-rabbit IgG-HRP and visualized by chemiluminescence.

For dot-blot analysis, allergens (including purified natural or recombinant Amb a 11, as well as purified natural Der p 1 included as a negative control) were spotted on a 0.45 µm nitrocellulose membrane, using 100 ng of proteins in 1 µL. The membrane was blocked in TBS, 0.1% Tween-20, 3% BSA. Sera from ragweed pollen allergic patients (n=67) were diluted 1/10 in TBS, 0.1% Tween-20, 0.2% BSA. Rabbit anti-human IgE followed by HRP-conjugated goat anti-rabbit antibodies diluted at 1/1000 and 1/10000, respectively, in TBS, 0.1% Tween-20, 0.2% BSA, were used for revealing.

Identification of proteins reactive with human IgE

Gel plugs (ie, 2D PAGE spots) were digested with trypsin and peptides were separated by reversed-phase chromatography using an Ultimate 3000 RS nano LC system (Thermo scientific, Villebon-sur-Yvette Courtaboeuf, France). The nanoLC system was coupled to a Maxis ESI-Qq-TOF mass spectrometer (Bruker Daltonics, Wissembourg, France). NanoLC-MS/MS data were analyzed using an in-house Mascot server (Matrix Science Ltd, London, UK) or PEAKS software (Bioinformatics Solutions Inc., Waterloo, Canada) to search a custom database containing translated nucleotide sequences from ragweed pollen transcriptome (ie, RNA-Seq) or public databases, such as the Uniprot/Swiss-Prot database and the nonredundant National Center for Biotechnology Information (nrNCBI) database, assuming tryptic digestion. Precursor mass and fragment mass were searched with initial mass tolerance of 8 ppm and 0.05 Da, respectively. The search included fixed modification of carbamidomethyl cysteine. Minimal peptide length was set to 6 amino acids and a maximum of one miscleavage was allowed. Peptide identifications were accepted if established at a $\geq 95\%$ probability, as specified by Mascot or PEAKS software.

Ragweed pollen RNA-Seq and Amb a 11 cDNA cloning

Briefly, total RNA was extracted from defatted ragweed pollen (GREER Laboratories). Pollen grains were first re-suspended in water and ground in liquid nitrogen using a small mortar and pestle before RNA extraction using the RNEasy kit (Qiagen, Courtaboeuf, France) according to the manufacturer's instructions. Total RNA was then used as a template for 5' and 3' RACE-PCR experiments (Clontech, Saint-Germain-en-Laye, France). Reverse transcription and subsequent PCR amplifications were performed starting from 1 μ g total RNA. The obtained cDNAs were used as template for a first round of PCR amplification using universal primers and primers derived from Amb a 11 amino acid sequence information. Specifically, peptide sequences obtained by N-terminal sequencing and 2D-PAGE-nanoLC-MS/MS were used to design a set of non-degenerated primers. Codons were chosen upon frequency tables. Amplification cycles were performed with 50-65°C hybridization steps. First round PCR products were then diluted 1/10 and used in a second PCR amplification with two additional primers designed from Amb a 11 amino acid information. Amplified DNA fragments were subsequently loaded on a 1%

agarose gel stained with ethidium bromide and separated by electrophoresis. Bands of interest were excised and purified DNA fragments were then cloned in pCR4-TOPO vector (Life Technologies) using the T/A cloning for sequencing kit and TOP10 competent cells (Life Technologies). In order to identify missing nucleotide sequences at the 5' and 3' ends (ie, translated and untranslated regions), a set of overlapping primers was subsequently designed based on sequencing data. These specific primers were used in combination with a UPM primer (Universal Primer Mix, Life Technologies) and ragweed pollen cDNAs to perform PCR cycles with an annealing temperature between 50-65°C. Amplified DNA fragments were purified and cloned as described above. 5', internal and 3' sequences were then assembled to obtain a consensus Amb a 11 pre-pro-sequence covering putative amino acid substitutions ([Fig E1](#)).

Purification and MS analysis of natural Amb a 11

Ragweed pollen proteins were precipitated with PerfectFocus kit from G-Biosciences (AGRO-BIO), re-suspended in OFFGEL buffer containing pH 4-7 ampholytes and focused using the high-resolution kit and the Agilent 3100 OFFGEL Fractionator as described in the user manual (Off Gel Electrophoresis, OGE, Agilent Technologies, Les-Ulis, France). Runs were stopped after achieving 74 kVh and the Amb a 11-containing fraction recovered in the liquid phase was acidified with 1% acetic acid and subsequently fractionated using a high-recovery macroporous reversed-phase column (mRPC₁₈, Agilent Technologies).

To determine the intact mass of Amb a 11, the protein was analyzed by nanoLC-MS in positive polarity mode using an ESI-Qq-TOF (Maxis 4G, Bruker Daltonics) coupled to an Ultimate 3000 RSLC (Thermo Scientific). The protein was separated using a C₄ column (300 µm ID, 15 cm, 5 µm particles, and pore size of 300 Å, Thermo Scientific Dionex) equipped with a µ-precolumn C₄ PepMap300 (Thermo Scientific). Solvent A (aqueous 0.15 % (v/v) FA) was used for equilibration and a gradient containing up to 60 % solvent B (0.15 % FA in 80 % ACN) was applied. Data were analyzed using Data Analysis software (Bruker Daltonics).

To confirm the protein sequence of Amb a 11 by MS/MS, purified Amb a 11 was solubilized in a buffer containing 8 M urea, 75 mM Tris pH 8.5 and 5 mM TCEP. Proteins were then alkylated with iodoacetamide (10 mM) for 20 min and submitted to a LysC digestion *in-solution* (with a protein/protease mass ratio 50/1) for 15 h at

37°C. 3 % FA was added to the mixture to quench enzymatic activity and peptides were stored at -80°C until analysis. Peptide samples were then spun at 18000 g and the peptide mixture was analyzed by MALDI-MS and nanoLC-MS/MS in a positive polarity mode using an Autoflex Speed MALDI-TOF and a Maxis ESI-Qq-TOF (Bruker Daltonics) coupled to an Ultimate 3000 RSLC (Thermo Scientific), respectively. MALDI-TOF mass spectra of peptides were obtained following desalting using a C₁₈ ZipTip (Millipore, Molsheim, France) according to the manufacturer's instructions. Data were analyzed using Flex Analysis (Bruker Daltonics) and an in-house Mascot server. Precursor mass and fragment mass were searched with initial mass tolerance of 40 ppm and 0.25 Da, respectively. The search included fixed modification of carbamidomethyl cysteine. Peptide identifications were accepted if established at a $\geq 95\%$ probability, as specified by Mascot software. Peptide mixtures were separated by nanoLC-MS/MS using a C₁₈ column (Acclaim® PepMap RSLC 75 μ m ID, 25 cm, 2 μ m particles, and pore size at 100 Å, Thermo Scientific). Data were analyzed using an in-house Mascot server and PEAKS software as above.

Recombinant protein expression and purification

The sequence coding pro-Amb a 11 was cloned in a pET15b vector (Novagen, Merck Millipore, Molsheim, France). The vector adds 21 residues comprising a hexahistidine tag. The resulting plasmid was used to transform BL21 (DE3) *E. coli* cells (Agilent Technologies). Expression and inclusion bodies preparation were done as previously described.^{E1} Inclusion bodies were washed 3 times with lysis buffer (50 mM Tris pH 8, 50 mM NaCl and 1 mM EDTA), lysis buffer containing 200 mM urea and lastly lysis buffer containing 1 M urea. Inclusion bodies were then solubilized overnight on a rotator at 4°C in 50 mM Tris pH 8, 8 M Urea buffer and centrifuged for 30 min at 15000g. The soluble fraction was diluted in buffer containing 50 mM Tris pH 8, 1M NaCl, 8 M Urea and 40 mM imidazole (1/2, v/v). The refolding protocol was adapted from Choudhury *et al.*^{E2} Recombinant pro-Amb a 11 was purified by nickel affinity chromatography on HisTrap FF column (GE Healthcare) in denaturing conditions (50 mM Tris pH 8, 0.5 M NaCl, 20 mM imidazole and 7 M Urea) and eluted with a imidazole gradient. Elution fractions containing pro-Amb a 11 were pooled and diluted at 1/100 in refolding buffer (50 mM Tris pH 8, 0.5 M NaCl, 5 mM EDTA, 1 mM GSH, 0.1 mM GSSG, 0.5 M arginine and 10 % glycerol) and incubated overnight on

a rotator at 4°C. Samples were then centrifuged for 30 min at 10000g, concentrated on Amicon ultra-15 centrifugal filter unit with 10 kDa MWCO (Millipore) and dialyzed against PBS.

To determine the intact mass of rAmb a 11, the protein was analyzed by MALDI-MS in positive ion and linear mode using an Autoflex Speed MALDI-TOF (Bruker Daltonics). Briefly, MALDI-TOF mass spectra of rAmb a 11 (with or without alkylation with 55 mM iodoacetamide) were obtained following desalting using a C₄ ZipTip (Millipore) according to the manufacturer's instructions and spotting onto an AnchorChipTM target with 2,5-Dihydroxybenzoic acid (SDHB, Bruker Daltonics).

Basophil activation test

Allergen-induced basophil activation was monitored using the Allergenicity Kit (Beckman Coulter, Villepinte, France), according to the manufacturer's recommendations. All assays were performed on blood samples from 5 ragweed-sensitized patients, obtained following informed consent. Briefly, whole blood was incubated for 15 minutes at 37°C with phycoerythrin-cyanine 7 (PE-C7)-labeled anti-CD3, fluorescein isothiocyanate (FITC)-labeled anti-CRTH2 and phycoerythrin (PE)-labeled anti-CD203c monoclonal antibodies, in the presence of serial concentrations (ie 10 µg/mL, 1 µg/mL and 0.1 µg/mL) of either an aqueous ragweed pollen extract, purified nAmb a 1 or purified nAmb a 11. PBS and an anti-IgE antibody were used as negative and positive controls, respectively. Basophil activation was analyzed by flow cytometry (FACSVerse, BD Biosciences, Rungis, France) and basophils were recognized as cells with a granulocyte morphology based on FSC/SSC parameters. Activated basophils were identified as CD3⁻CRTH2⁺ CD203c^{bright} cells. Results were expressed for each condition as percentages of CD203c^{bright} basophils.

***In vitro* activation of recombinant pro-Amb a 11**

The pro-rAmb a 11 zymogen was dialyzed overnight against 20 mM Na acetate buffer (pH 5.0), incubated 2 h at 40°C and further analyzed by SDS-PAGE. Amb a 11 protease activity was measured with the Enzcheck protease assay kit (Life Technologies) using green-fluorescent casein-Bodipy FL substrate, according to the

supplier's instructions. Activity assays were performed on zymogen and mature forms with or without 100 μ M E-64 cysteine protease inhibitor (Sigma, Saint-Quentin Fallavier, France). The fluorescence was monitored using a FluoroMax-4 (Horiba, Les Ulis, France) with excitation and emission wavelengths at 485 and 530 nm, respectively.

Circular dichroism analysis of natural and recombinant Amb a 11

Circular dichroism (CD) spectra of both rAmb a 11 produced in *E. coli* and natural purified molecule (see Fig. E5) were acquired. Spectra were obtained using a Jasco J-815 CD spectrophotometer with a 1 mm cell and the following parameters: band width 2.00 nm, response 1 sec, data pitch 1 nm, scanning speed 100 nm/min, accumulation 8 and temperature 20°C. Protein concentration was approx. 100 μ g/mL.

Online Repository Figure Legends

FIG E1. Predicted amino acid sequences of pollen allergen Amb a 11.

A, Deduced amino acid sequence obtained by 5' and 3' RACE-PCRs, including putative amino acid substitutions. EDMAN sequencing indicated that the N-terminal amino acid residue is G₁₀₉ (highlighted in yellow, [table E1](#)). **B**, Main deduced-mature protein sequences from cDNA clones. The residue numbering refers to the zymogen form of Amb a 11. **C**, Sequence alignment of the residues of Amb a 11 predicted amino acid sequence obtained by 5' and 3' RACE-PCRs and high-throughput sequencing (HTS). Annotations used were as follows: identical residues (*) and missing residues (-).

FIG E2. Sequence coverage of Amb a 11 in mass spectrometry.

Covered regions are highlighted in gray. The residue numbering corresponds to the mature form. Sequence coverages were obtained after *in-solution* digestion with LysC endoproteinase and MALDI-MS (77 %, **A**) or nanoLC-MS/MS (91%, **B**) and after *in-gel* digestion with trypsin endoproteinase and nanoLC-MS/MS (52%, **C**). For complete details, see [tables E2-E4](#).

FIG E3. Glycosylation of Amb a 11 is not essential for IgE-binding.

A, Six donors (070, 005, 041, 065, 93 and 110) with IgE-reactivity to nAmb a 11. Proteins from the ragweed pollen extract (**37 kDa-mature nAmb a 11, nAmb a 1 (arrow), lane 1) and non-glycosylated rAmb a 11 expressed in *E. coli* (°45 kDa-proform, lane 2) were resolved on a 1D-gel. M, molecular weight markers. **B**, Mass determination of the intact rAmb a 11 by MALDI-TOF MS. The protein was then alkylated. After this treatment, all free cysteine residues should be found alkylated (+ 57 Da/free cysteine). **C**, rAmb a 11 was cut out from Coomassie-stained gel ([Fig 6, A](#)), trypsin digested and analyzed by MS/MS. Fragmentation of peptide G₁SAPGSIDTDPNPKDFIYANVT_K₂₂ showing the absence of N-glycosylation at position N₁₉. The residue numbering corresponds to the mature protein.

FIG E4. Secondary structure estimation of natural and recombinant Amb a 11.

Overlay of CD spectra from natural (green spectrum) and recombinant (blue spectrum) Amb a 11. The overlay of the CD spectra of these 2 molecules shows curves that are close to each other in the 210-260 nm range. Some differences are observed in the far UV CD (195-205 nm), possibly due to the fact that natural Amb a 11 encompasses multiple proteoforms and glycoforms, whereas rAmb a 11 represents a single non glycosylated proteoform.

FIG E5. 3D structure model of cysteine protease Amb a 11.

Structure-based sequence alignment between Amb a 11 from *Ambrosia artemisiifolia* and the Vignain cysteine endoproteinase from *Ricinus communis* (*ie*, template structure). The structure of Vignain is experimentally determined (PDB-ID 1S4V) using X-ray diffraction. The 3D model was built by using Geno 3D ([Fig 8](#)). The residues forming the active site, the important residues within conserved peptidase C 1 domain and the cysteine residues involved in disulfide bonds formation are highlighted. Residues forming the active site (C₄₇, H₁₈₁, N₂₀₂) and within conserved Peptidase C1 domain are colored in green and orange, respectively. Cysteine residues forming the disulfide bridges C₄₄₋₈₅, C₇₈₋₁₁₈ and C₁₇₅₋₂₂₆ (by homology with

250 Vignain) are highlighted in red, blue and magenta, respectively. The residue
251 numbering corresponds to the mature protein.
252

Online Repository Table Legends

Table E1. List of IgE-reactive proteins identified from short ragweed pollen.

Excised protein spots were analyzed by nanoLC-MS/MS and proteins were identified by searching Uniprot/Swiss-Prot, NCBI nr databases and translated nucleotide sequences from ragweed pollen RNA-Seq. Homologous sequences were then obtained by running BLAST against the nrNCBI database. The GSAPGSIDTDPNKDF amino acid sequence (Spot # 14) was obtained by Edman degradation. Numbers refer to [Fig 1, A](#). Accession number (Acc.No.).

Table E2. Summary of the main Amb a 11 LysC-peptides (*in-solution* digestion) identified by MALDI-MS.

Modified amino acid residues are shown in bold type. The monoisotopic mass of each N-glycan is indicated. ^aThe zymogen consists of 386 amino acids, including 108 additional N-terminal residues. Miscleavage (MisC.). Measured (Meas.). Calculated (Calc.). Carbamidomethyl Cys (CAM).

Table E3. Summary of the main Amb a 11 LysC-peptides (*in-solution*) identified by nanoLC-MS/MS.

Modified amino acid residues are shown in bold type. The monoisotopic mass of each N-glycan is indicated. ^aThe zymogen consists of 386 amino acids, including 108 additional N-terminal residues. Miscleavage (MisC.). Measured (Meas.). Calculated (Calc.). Carbamidomethyl Cys (CAM).

Table E4. Summary of the main Amb a 11 tryptic peptides (*in-gel* digestion) identified by nanoLC-MS/MS.

Modified amino acid residues are shown in bold type. ^aThe zymogen consists of 386 amino acids, including 108 additional N-terminal residues. Miscleavage (MisC.). Measured (Meas.). Calculated (Calc.). Carbamidomethyl Cys (CAM).

Online Repository References

- E1. Bussi res L, Bordas-Le Floch V, Bulder I, Chabre H, Nony E, Lautrette A, et al. Recombinant Fusion Proteins Assembling Der p 1 and Der p 2 Allergens from *Dermatophagoides pteronyssinus*. *Int Arch Allergy Immunol* 2010; 153:141-51.
- E2. Choudhury D, Roy S, Chakrabarti C, Biswas S, Dattagupta JK. Production and recovery of recombinant propain with high yield. *Phytochemistry* 2009; 70:465-72.

Online Repository Tables

Table E1. List of IgE-reactive proteins identified from short ragweed pollen.

Excised protein spots were analyzed by nanoESI-Qq-TOF MS/MS and proteins were identified by searching Uniprot/Swiss-Prot, NCBI nr databases and translated nucleotide sequences from ragweed pollen (homologous sequences were then obtained by running BLAST against the nrNCBI database). The GSAPGSIDTDPNKDF amino acid sequence (Spot # 14) was obtained by Edman degradation. Numbers refer to Fig 1, A. Accession number (Acc.No.).

Identification of the cysteine protease Amb a 11 as a novel major allergen from short ragweed

Protein spot #	Measured m/z	z	Peptide Sequence	Protein name / Organism	Acc. No.	ORF sequence	BLAST result	IUIS allergen nomenclature
1	540.7711	2	K.SDGDATYIAR.A	unknown	unknown	MARRSITKNNNNHMTGQKHCCYLYFTLALVLLGPVRSADLHEILPSNGTRKGLQACEANNIVDQWRCKADWAEINRQALADCAPGFAKGTFFGKLGDIYTVTNEDDDVANPEGSLRFGAALERPWIIFERDMVHLNDELINTDITDGRGANVEINAGLHNNVNNVHNNHNDKPKNEGKVKYNDGPARKRPKSDGDATYIARASQVWIDHCSLMTADELQVKGSTMVTSNCKFSQSQFVLLLGANDEQVEDKAMM/VNAFTNFDQVQGRMPRCRHGFQVNNKYDRWGTYALSGSAAPTILSGNRFFAGDEPSKSVVMRSQPGTSGMSWNNWRSRDRLLENGATFEPGSDPVLTPQNGAGMPAEPGEAALTFTAGVLSCGPQVPC	Pectate lyase Amb a 1.01-like	
	606.6742	3	R.FGAALERPWIIFER.D	unknown	unknown			
	624.9071	2	R.QALADCAQGFAG.K	unknown	unknown			
	776.3360	2	R.SGPGTEGSMWNNWR.S	unknown	unknown			
	779.7745	3	K.NVIHNNIHDKPNEGKVK.Y	unknown	unknown			
	805.3804	2	R.HGFFQVNNVNYDR.W	unknown	unknown			
	888.4934	2	R.GANVEINAGLHNNVKN.N	unknown	unknown			
	1025.5218	2	R.WGTYALSGSAAPTILSGNR.F	unknown	unknown			
	1127.5481	2	K.FSOSQFVLLGANDEQVEDK.A	unknown	unknown			
	1303.6166	2	K.LGDVYTVTNEDDDVANPEGSLR.F	unknown	unknown			
2, 3	516.2769	3	K.NVIHNNIMHDVK.V	Pollen allergen Amb a 1.1 / Ambrosia artemisiifolia	MPA11_AMBAR			Amb a 1.01
	569.2731	2	K.ALADCAQGFAG.K	Pollen allergen Amb a 1.1 / Ambrosia artemisiifolia	MPA11_AMBAR			
	737.9139	2	K.VEINAGFTLNGVK.N	Pollen allergen Amb a 1.1 / Ambrosia artemisiifolia	MPA11_AMBAR			
	791.3774	2	R.HGFFQVNNVNYDR.W	Pollen allergen Amb a 1.1 / Ambrosia artemisiifolia	MPA11_AMBAR			
	856.0760	3	K.GTVGGKDGDIYTVTSELDDVANPK.E	Pollen allergen Amb a 1.1 / Ambrosia artemisiifolia	MPA11_AMBAR			
	864.4047	3	R.AGSDGDAISIGSSQIWDHCSLSK.S	Pollen allergen Amb a 1.1 / Ambrosia artemisiifolia	MPA11_AMBAR			
	909.4952	2	R.FGAAGNRRLWIIFER.D	Pollen allergen Amb a 1.1 / Ambrosia artemisiifolia	MPA11_AMBAR			
	1000.4809	2	R.GMLATVAFNTFTDNVDQR.M	Pollen allergen Amb a 1.1 / Ambrosia artemisiifolia	MPA11_AMBAR			
	1011.5064	2	K.WGSAIGGSSAPTILSGNR.F	Pollen allergen Amb a 1.1 / Ambrosia artemisiifolia	MPA11_AMBAR			
	1017.5125	3	R.LTVNSLSLTQHQPVLFGAGDENIEDR.G	Pollen allergen Amb a 1.1 / Ambrosia artemisiifolia	MPA11_AMBAR			
	1033.9729	2	K.DGDIYTVTSELDDVANPK.E	Pollen allergen Amb a 1.1 / Ambrosia artemisiifolia	MPA11_AMBAR			
4, 5	409.8984	3	K.FVAPDFIYKK.N	Pollen allergen Amb a 2 / Ambrosia artemisiifolia	MPAA2_AMBAR			Amb a 1.05
	550.2966	2	K.FVAPDFIYK.K	Pollen allergen Amb a 2 / Ambrosia artemisiifolia	MPAA2_AMBAR			
	556.2718	3	R.FGFFQVNNFYDR.W	Pollen allergen Amb a 2 / Ambrosia artemisiifolia	MPAA2_AMBAR			
	556.6608	3	R.HGSDGDAIHVTSIGSSQIWDHCTLSK.S	Pollen allergen Amb a 2 / Ambrosia artemisiifolia	MPAA2_AMBAR			
	616.6691	3	R.FGATQDRPLWIIFOR.D	Pollen allergen Amb a 2 / Ambrosia artemisiifolia	MPAA2_AMBAR			
	712.6954	3	R.WDKYAGGSSNPILSQGNK.F	Pollen allergen Amb a 2 / Ambrosia artemisiifolia	MPAA2_AMBAR			
	732.9968	3	K.WGDIYMYTSDDDVDVNPKE	Pollen allergen Amb a 2 / Ambrosia artemisiifolia	MPAA2_AMBAR			
	768.4283	2	K.VELVYGGITLMNVK.N	Pollen allergen Amb a 2 / Ambrosia artemisiifolia	MPAA2_AMBAR			
	779.4297	2	K.NVIHNDHVDVR.V	Pollen allergen Amb a 2 / Ambrosia artemisiifolia	MPAA2_AMBAR			
	807.9309	2	K.AVILGASDTHFDQK.M	Pollen allergen Amb a 2 / Ambrosia artemisiifolia	MPAA2_AMBAR			
	853.9398	2	K.YAIGGSSNPILSQGNK.F	Pollen allergen Amb a 2 / Ambrosia artemisiifolia	MPAA2_AMBAR			
	1128.5333	2	K.SFDGLVDVNWGSTGYTISNCK.F	Pollen allergen Amb a 2 / Ambrosia artemisiifolia	MPAA2_AMBAR			
6, 7	576.2797	2	K.ALADCAQGFAG.K	Pollen allergen Amb a 1.4 / Ambrosia artemisiifolia	MPA14_AMBAR			Amb a 1.04
	603.0079	3	R.GAKVEINAGFAIYNNK.N	Pollen allergen Amb a 1.4 / Ambrosia artemisiifolia	MPA14_AMBAR			
	664.6699	3	R.WGSAIGGSSAPTILSQGNR.F	Pollen allergen Amb a 1.4 / Ambrosia artemisiifolia	MPA14_AMBAR			
	729.8823	2	R.ELANNKTIQGR.D	Pollen allergen Amb a 1.4 / Ambrosia artemisiifolia	MPA14_AMBAR			
	775.9326	2	K.VEINAGFAIYNNK.N	Pollen allergen Amb a 1.4 / Ambrosia artemisiifolia	MPA14_AMBAR			
	788.3879	2	R.HGFFQVNNVNYDR.W	Pollen allergen Amb a 1.4 / Ambrosia artemisiifolia	MPA14_AMBAR			
	820.7194	3	K.QSDGDAIGSSQIWDHCSLSK.S	Pollen allergen Amb a 1.4 / Ambrosia artemisiifolia	MPA14_AMBAR			
	860.7491	3	K.GTGGKDGDIYTVTSELDDVANPK.E	Pollen allergen Amb a 1.4 / Ambrosia artemisiifolia	MPA14_AMBAR			
	863.4187	3	R.KGSDGDAIGSSQIWDHCSLSK.S	Pollen allergen Amb a 1.4 / Ambrosia artemisiifolia	MPA14_AMBAR			
	1033.9746	2	K.DGDIYTVTSELDDVANPK.E	Pollen allergen Amb a 1.4 / Ambrosia artemisiifolia	MPA14_AMBAR			
	499.2634	3	K.AILLGADTHVQDK.G	Pollen allergen Amb a 1.3 / Ambrosia artemisiifolia	MPA13_AMBAR			Amb a 1.03
	514.9634	3	K.NIHHNNIHQDK.V	Pollen allergen Amb a 1.3 / Ambrosia artemisiifolia	MPA13_AMBAR			
	587.8279	2	R.FLAPDDQIKK.N	Pollen allergen Amb a 1.3 / Ambrosia artemisiifolia	MPA13_AMBAR			
	640.3084	2	R.QALADCAQGFAG.K	Pollen allergen Amb a 1.3 / Ambrosia artemisiifolia	MPA13_AMBAR			
	703.6771	3	R.WGTYAIGGSSAPTILCQGNR.F	Pollen allergen Amb a 1.3 / Ambrosia artemisiifolia	MPA13_AMBAR			
	810.3836	2	R.FGFFQVNNVNYDR.W	Pollen allergen Amb a 1.3 / Ambrosia artemisiifolia	MPA13_AMBAR			
	897.0882	3	R.QASDGDITINAGSSQIWDHCSLSK.S	Pollen allergen Amb a 1.3 / Ambrosia artemisiifolia	MPA13_AMBAR			
	1054.9921	2	K.WGSDVYTVTSELDDVANPK.E	Pollen allergen Amb a 1.3 / Ambrosia artemisiifolia	MPA13_AMBAR			
	392.2220	3	R.FLAPDDQIKK.N	Pollen allergen Amb a 1.3 / Ambrosia artemisiifolia	MPA13_AMBAR			
	523.7804	2	R.FLAPDDQIKK.N	Pollen allergen Amb a 1.3 / Ambrosia artemisiifolia	MPA13_AMBAR			
8, 9	640.3101	2	R.QALADCAQGFAG.K	Pollen allergen Amb a 1.3 / Ambrosia artemisiifolia	MPA13_AMBAR			Amb a 1.03
	703.6656	3	K.WGSDVYTVTSELDDVANPK.E	Pollen allergen Amb a 1.3 / Ambrosia artemisiifolia	MPA13_AMBAR			
	748.3928	2	K.AILLGADTHVQDK.G	Pollen allergen Amb a 1.3 / Ambrosia artemisiifolia	MPA13_AMBAR			
	771.9424	2	K.NIHHNNIHQDK.V	Pollen allergen Amb a 1.3 / Ambrosia artemisiifolia	MPA13_AMBAR			
	810.3842	2	R.HGFFQVNNVNYDR.W	Pollen allergen Amb a 1.3 / Ambrosia artemisiifolia	MPA13_AMBAR			
	897.0882	3	R.QASDGDITINAGSSQIWDHCSLSK.S	Pollen allergen Amb a 1.3 / Ambrosia artemisiifolia	MPA13_AMBAR			
	1055.0136	2	R.WGTYAIGGSSAPTILCQGNR.F	Pollen allergen Amb a 1.3 / Ambrosia artemisiifolia	MPA13_AMBAR			
	1125.5578	2	K.SFDGLVDVTLGSHVTISNCK.F	Pollen allergen Amb a 1.3 / Ambrosia artemisiifolia	MPA13_AMBAR			
10, 12, 13	533.2870	2	R.FFAPDDIIK.K	Pollen allergen Amb a 1.2 / Ambrosia artemisiifolia	MPA12_AMBAR			Amb a 1.02
	594.7962	4	K.FTQHGFVLLLGADTHYQDK.G	Pollen allergen Amb a 1.2 / Ambrosia artemisiifolia	MPA12_AMBAR			
	597.3351	2	R.FFAPDDIIK.N	Pollen allergen Amb a 1.2 / Ambrosia artemisiifolia	MPA12_AMBAR			
	640.3105	2	R.QALADCAQGFAG.K	Pollen allergen Amb a 1.2 / Ambrosia artemisiifolia	MPA12_AMBAR			
	679.3459	3	R.WGTYAIGGSSAPTILSQGNR.F	Pollen allergen Amb a 1.2 / Ambrosia artemisiifolia	MPA12_AMBAR			
	703.8642	4	R.TDRDLLENGAIFLPSGSDPVLTPPEQK.A	Pollen allergen Amb a 1.2 / Ambrosia artemisiifolia	MPA12_AMBAR			
	705.8744	2	K.AGMPAEPGEAVLR.L	Pollen allergen Amb a 1.2 / Ambrosia artemisiifolia	MPA12_AMBAR			
	778.9501	2	K.NIHHNNIHQDK.V	Pollen allergen Amb a 1.2 / Ambrosia artemisiifolia	MPA12_AMBAR			
	798.8466	2	R.TGTGNAESMSWNNWR.T	Pollen allergen Amb a 1.2 / Ambrosia artemisiifolia	MPA12_AMBAR			
	814.0914	3	R.DLENGAIFLPSGSDPVLTPPEQK.A	Pollen allergen Amb a 1.2 / Ambrosia artemisiifolia	MPA12_AMBAR			
	822.8775	2	R.LTSSAGVLSCHQGAP.-	Pollen allergen Amb a 1.2 / Ambrosia artemisiifolia	MPA12_AMBAR			
	906.0934	3	R.QQSDGDAINAGSSQIWDHCSLSK.A	Pollen allergen Amb a 1.2 / Ambrosia artemisiifolia	MPA12_AMBAR			
	1038.4827	2	K.HGSDVYTVTSDKDDVANPK.E	Pollen allergen Amb a 1.2 / Ambrosia artemisiifolia	MPA12_AMBAR			

Identification of the cysteine protease Amb a 11 as a novel major allergen from short ragweed

Protein spot #	Measured m/z	z	Peptide Sequence	Protein name / Organism	Acc. No.	ORF sequence	BLAST result	IUIS allergen nomenclature
11	416.9047	3	K.SKPLWITFEK.D	unknown	unknown	MEEKLYLVFFTFSLVSVQSVRNAASVDDEQLAFSGGNSRRGLTECVASNIDKCVWRCKPDWDQDRQALAKCAAGFATGTTGGSGGEFYVVTNPSEDA ANPAPGSLRAGVYKSKPLWITFEKDMVKLTQELVINSOKTIDGRGAKEITGGLTLYNVKNVHGLNHDVKAAPGGMIKSSDGPPLRGKVEGGDGVCTGSS KIWDHCTFSKGLDGLIDVNFPS.TDVTISNCKFTNHOKAMLLGANDAHSDKDMRVTLAFNWF.SKCDQRLPRCLGFFQVNNDEYEGWGT.YAIGGSANPTLS HGNRF.SAPFDPKKKNVCMRVDAPEAEWKKWNRSGDVLNGAIFYETGVDPSLTFEGKGLIAEFGERVPKLTSCAGVLSVCPGEPC	Pectate lyase Amb a 1.05-like	
	436.2139	3	K.IWDHCTFSK.G	unknown	unknown			
	550.3123	2	R.FSAPPDPKK.N	unknown	unknown			
	595.8256	2	K.GLDGLIDVNFK.S	unknown	unknown			
	601.3273	3	K.LTQELVINSOKTIDGR.G	unknown	unknown			
	606.8241	2	R.VTLAFNWSK.C	unknown	unknown			
	630.3463	2	K.LTQELVINSOK.T	unknown	unknown			
	697.1188	4	R.SKQDVLNGAIFVETGVDPDLTPEQK.G	unknown	unknown			
	697.3844	2	K.GGLIAEFGERVPK.L	unknown	unknown			
	736.4240	2	K.NVIHGLNIHDVK.A	unknown	unknown			
	752.4260	2	K.VEITGPGTLTYNVK.N	unknown	unknown			
	853.8801	2	K.LTSCAGVLSVCPGEPC.-	unknown	unknown			
	892.6630	4	K.CAAGFATGTTGGSGGEIYVVTNPSEDAANPAPGSLR.A	unknown	unknown			
	892.9373	4	R.LGFFQIVNNDEYEGWGT.YAIGGSANPTLSHGNR.F	unknown	unknown			
14, 15	326.8637	3	K.IDPKVDWR.E	unknown	unknown	MEINKLYCFBSLVLLGLVESFHYHERELEESEEGFMGMYDRWREQHNIEMRSPERFNVFKYNVRIHESNMOKPYKLVNEFADMTNLEFVNTYANSKISH FQALRG.SAPGSIDTDPNKDFITYANVTKPDKVDVREKNNAVTDVKGGGGCGSCWAFAAVALEGINAIR.TGKLVKFSQQLVDCDMTNAAGCDGGLMEPAFTYI KHGGIAPEASYPYVQKRETCDAKNDVLDKDRGRNVPLDEALRAAHGPNATGICLSGHGLQFYSEG.YTTGDCGTEPHNGVGVYGENEKGFVTV KNSWGPTWGEKGYHLQRGARKGLCQVAMHSSFPIMNDPNPKDQPNVPKDQDPAKDPKFTQRLQGRIT	Cysteine protease Act d 1-like (Kiwi) Ana c 2-like (Pineapple) Ale o 1-like (Mite) Car p 1-like (Papaya) Der fp 1-like (Mite)	Amb a 11
	862.9330	4	R.EKNVTDVKGGGGCGSCWAFAAVALEGINAIR.T	unknown	unknown			
	1064.5294	3	K.NAVTDVKGGGGCGSCWAFAAVALEGINAIR.T	unknown	unknown			
	822.0667	3	K.GGGGCGSCWAFAAVALEGINAIR.T	unknown	unknown			
	773.3887	2	K.HGGIAPEASYPYVQK.R	unknown	unknown			
	851.4404	2	K.HGGIAPEASYPYVQKR.E	unknown	unknown			
	637.3419	3	K.IDGRONVPLDEEALRK.A	unknown	unknown			
	670.8440	2	R.QNVPLDEEALR.K	unknown	unknown			
	734.8918	2	R.QNVPLDEEALR.K	unknown	unknown			
	1000.8772	5	K.AVAHPVATGIQLSGHGLQFYSEG	unknown	unknown			
	326.8637	3	K.GIKFWTVK.N	unknown	unknown			
	581.2698	2	K.NSWGPTWGEK.G	unknown	unknown			
	508.0064	4	K.NSWGPTWGEKGYIHLQR.G	unknown	unknown			
	EDMAN (Spot # 17)		GSAPGSIDTDPNKDF	unknown	unknown			
16	791.3765	2	R.HGFFQVNNNYDK.W	Pollen allergen Amb a 1.1 / Ambrosia artemisiifolia	MPA11_AMBAR			Amb a 1.01
	864.4058	3	R.AGSDGDAISISQSIQIWDHCSLSK.S	Pollen allergen Amb a 1.1 / Ambrosia artemisiifolia	MPA11_AMBAR			
17	1011.5059	2	K.WGSYAIGGSAPTLISQGNRF	Pollen allergen Amb a 1.1 / Ambrosia artemisiifolia	MPA11_AMBAR	gi 62249512 gi 62249512 gi 62249512 gi 62249512		Amb a 8 Isoform 01.02
	641.3359	3	K.SSSFPEFKPDEINAIK.E	Profilin isoallergen 2 / Ambrosia artemisiifolia				
	716.8842	2	K.YMVIQEGPGAIR.G	Profilin isoallergen 2 / Ambrosia artemisiifolia				
	938.4974	2	K.EFSEPGALAPTGLFLAGAK.Y	Profilin isoallergen 2 / Ambrosia artemisiifolia				
	945.4956	4	K.SSSFPEFKPDEINAIKEFSEPGALAPTGLFLAGAK.Y	Profilin isoallergen 2 / Ambrosia artemisiifolia				
	716.8842	2	K.YMVIQEGPGAIR.G	Profilin isoallergen 1 / Ambrosia artemisiifolia	gi 62249502			
18	925.4886	2	K.EFDAAGTLAPTGLFLAGAK.Y	Profilin isoallergen 1 / Ambrosia artemisiifolia	gi 62249502	gi 62249502 gi 62249502 gi 62249502		Amb a 8 Isoform 01.01
	946.4944	2	K.SGSFPEFKPDEINAIK.E	Profilin isoallergen 1 / Ambrosia artemisiifolia				
	417.7220	2	K.FVANVTPS.-	unknown	unknown			
	458.9198	3	K.VYLVGGPELGGWK.L	unknown	unknown			
	469.2492	2	R.AYALWSAR.Q	unknown	unknown			
	572.9394	3	K.EEAYHACDIKPIRL.F	unknown	unknown			
19	590.3083	3	R.FTLTTPGSHFICTK.F	unknown	unknown	MYQHKQGVKRSMAFYKLNILVVMVVMQFHYTVAGKYLVGGPELGGWKLOSDPRAYALWSARQOFKTTDLVWFNFTTGEDSVAEVEKEAYHACDIKPI RLPEGGPDRFTLLTPGSHFICTKFDHCQRGQKQFVANVTPS	Amb a 3-like	
	732.3128	3	R.YCPWOVVVCYESEICS.K	Pollen allergen Amb a 5 / Ambrosia artemisiifolia	MPAA5_AMBEL			
	780.8680	2	-LPCAWAGNVCGEK.R	Pollen allergen Amb a 5 / Ambrosia artemisiifolia	MPAA5_AMBEL			

Table E2. Summary of the main Amb a 11 LysC-peptides (*in-solution* digestion) identified by MALDI-MS.

Modified amino acid residues are shown in bold type. The monoisotopic mass of each N-glycan is indicated. ^aThe zymogen consists of 386 amino acids, including 108 additional N-terminal residues. Miscleavage (MisC.). Measured (Meas.). Calculated (Calc.). Carbamidomethyl Cys (CAM).

Identification of the cysteine protease Amb a 11 as a novel major allergen from short ragweed

Meas. m/z	Calc. M+H ⁺	Δ m/z (ppm)	MisC.	Sequence	Modifications	^a Range	Range	^a Protein	Protein
3348.600	3348.500	30	1	-GSAPGSIDTDPNKFDFIYANVT.K	Glycosylation: 19 (1038.375 Da, MMF ³)	109 - 130	01 - 22		
3345.612	3345.588	7	2	K.DFIYANVTKIPDKVDWREK.N	Glycosylation: 6 (1008.364 Da, M0XF ³)	122 - 140	14 - 32		
3507.690	3507.641	14	2	K.DFIYANVTKIPDKVDWREK.N	Glycosylation: 6 (1170.417 Da, MMXF ³)	122 - 140	14 - 32		
1285.732	1285.690	32	1	K.IPDKVDWREK.N		131 - 140	23 - 32		
3817.997	3817.984	3	2	K.NAVTDVKGQGGCGSCWAFAAVVA LEGINAIRTGKLV.K.F	CAM: 12, 15	141 - 174	33 - 69		
3477.778	3477.737	12	1	K.NAVTDVKGQGGCGSCWAFAAVVA LEGINAIRTGK.L	CAM: 12, 15	141 - 171	33 - 66		
2750.387	2750.350	13	0	K.GQGGCGSCWAFAAVVA LEGINAIRTGK.L	CAM: 12, 15	148 - 171	40 - 66		
3090.516	3090.5978	26	1	K.GQGGCGSCWAFAAVVA LEGINAIRTGKLV.K.F	CAM: 12, 15	148 - 174	40 - 69		
3410.559	3410.5018	17	0	K.FSEQQLVDCDMTNAGCDGGLMEP AFTYIIK.H	CAM: 9, 16 ; Ox	175 - 207	70 - 99	Amb a 11 I ₂₀₅	Amb a 11 I ₉₇
3396.562	3396.489	21	0	K.FSEQQLVDCDMTNAGCDGGLMEP AFTYVIK.H	CAM: 9, 16	175 - 207	70 - 99	Amb a 11 V ₂₀₅	Amb a 11 V ₉₇
2335.144	2335.114	13	1	K.HGGIAPEASYPYVGKRETC.DK.A	CAM: 19	208 - 228	100 - 120		
2805.621	2805.594	10	3	K.AKIKDVLKIDGRQNVPLDDEEALRK.A		229 - 253	121 - 145		
2606.499	2606.462	14	2	K.IKDVLKIDGRQNVPLDDEEALRK.A		231 - 253	123 - 145		
2365.315	2365.283	14	1	K.DVLKIDGRQNVPLDDEEALRK.A		233 - 253	125 - 145		
1910.033	1910.009	13	0	K.IDGRQNVPLDDEEALRK.A		237 - 253	129 - 145		
5000.385	5000.354	6	0	K.AVAHQPVATGIQLSGHGLQFYSEG VYTGDCGTEPNHGVGIVGYGENEK.G	CAM: 30	254 - 301	146 - 193		
1161.556	1161.532	21	0	K.NSWGPTWGEK.G		310 - 319	202 - 211		
1298.775	1298.744	24	0	K.GYIHLQRGARK.E		320 - 330	212 - 222		

Table E3. Summary of the main Amb a 11 LysC-peptides (*in-solution*) identified by nanoESI-Qq-TOF MS/MS.

Modified amino acid residues are shown in bold type. The monoisotopic mass of each N-glycan is indicated. ^aThe zymogen consists of 386 amino acids, including 108 additional N-terminal residues. Miscleavage (MisC.). Measured (Meas.). Calculated (Calc.). Carbamidomethyl Cys (CAM).

Identification of the cysteine protease Amb a 11 as a novel major allergen from short ragweed

Meas. m/z	Calc. M+H ⁺	z	Δ m/z (ppm)	MisC.	Sequence	Modifications	^a Range	Range	^a Protein	Protein
1062.8215	3186.4489	3	-1.0	1	-GSAPGSIDTPNKFDFIYANVT.K.I	Glycosylation: 19 (876.322 Da. M0F ³)	109 - 130	01 - 22		
1106.8359	3318.4888	3	-1.0	1	-GSAPGSIDTPNKFDFIYANVT.K.I	Glycosylation: 19 (1008.364 Da. M0XF ³)	109 - 130	01 - 22		
1116.8437	3348.5155	3	5.0	1	-GSAPGSIDTPNKFDFIYANVT.K.I	Glycosylation: 19 (1038.375 Da. MMF ³)	109 - 130	01 - 22		
1160.8539	3480.5418	3	1.2	1	-GSAPGSIDTPNKFDFIYANVT.K.I	Glycosylation: 19 (1170.417 Da. MMXF ³)	109 - 130	01 - 22		
1179.8572	3537.5588	3	-0.8	1	-GSAPGSIDTPNKFDFIYANVT.K.I	Glycosylation: 19 (1227.434 Da. MGnX)	109 - 130	01 - 22		
1228.5462	3683.623	3	0.0	1	-GSAPGSIDTPNKFDFIYANVT.K.I	Glycosylation: 19 (1373.496 Da. MGnXF ³)	109 - 130	01 - 22		
950.2476	4747.2139	5	-1.5	2	-GSAPGSIDTPNK K.DFIYANVT.KIPDKVDWREK.N	Glycosylation: 6 (1170.417 Da. MMXF ³)	109 - 140	01 - 32		
837.1528	3345.5878	4	0.0	1	K.DFIYANVT.KIPDKVDWREK.N	Glycosylation: 6 (1008.364 Da. M0XF ³)	122 - 140	14 - 32		
877.6660	3507.6408	4	-0.1	1	K.DFIYANVT.KIPDKVDWREK.N	Glycosylation: 6 (1170.417 Da. MMXF ³)	122 - 140	14 - 32		
973.9399	1946.8738	2	-0.9	0	K.DFIYANVT.K.I	Glycosylation: 6 (876.322 Da. M0F ³)	122 - 130	14 - 22		
1039.9614	2078.9161	2	-0.5	0	K.DFIYANVT.K.I	Glycosylation: 6 (1008.364 Da. M0XF ³)	122 - 130	14 - 22		
1054.9656	2108.9267	2	-0.5	0	K.DFIYANVT.K.I	Glycosylation: 6 (1038.375 Da. MMF ³)	122 - 130	14 - 22		
1120.9880	2240.9690	2	-0.3	0	K.DFIYANVT.K.I	Glycosylation: 6 (1170.417 Da. MMXF ³)	122 - 130	14 - 22		
955.2513	3817.984	4	-0.1	2	K.NAVTDVKGQGGCGSCWAFAAVVA LEGNAIRTGKLVK.F	CAM: 12, 15	141 - 174	33 - 69		
870.1893	3477.737	4	-0.3	1	K.NAVTDVKGQGGCGSCWAFAAVVA LEGNAIRTGK.L	CAM: 12, 15	141 - 171	33 - 66		
917.4547	2750.350	3	-0.6	0	K.GQGGCGSCWAFAAVVA LEGNAIRTGK.L	CAM: 12, 15	148 - 171	40 - 66		
773.4047	3090.5978	4	-0.7	1	K.GQGGCGSCWAFAAVVA LEGNAIRTGKLVK.F	CAM: 12, 15	148 - 174	40 - 69		
1137.5057	3410.5053	3	-0.8	0	K.FSEQLVDCDMTNAGCDGGLMEP AFTYIIK.H	CAM: 9, 16	175 - 207	70 - 99	Amb a 11 I ₂₀₅	Amb a 11 I ₉₇
1142.8376	3426.5002	3	-0.6	0	K.FSEQLVDCDMTNAGCDGGLMEP AFTYIIK.H	CAM: 9, 16; Ox : 11	175 - 207	70 - 99	Amb a 11 I ₂₀₅	Amb a 11 I ₉₇
1142.8370	3426.5002	3	-0.7	0	K.FSEQLVDCDMTNAGCDGGLMEP AFTYIIK.H	CAM: 9, 16; Ox : 21	175 - 207	70 - 99	Amb a 11 I ₂₀₅	Amb a 11 I ₉₇
1132.8348	3396.489	3	0	0	K.FSEQLVDCDMTNAGCDGGLMEP AFTYVIK.H	CAM: 9, 16	175 - 207	70 - 99	Amb a 11 V ₂₀₅	Amb a 11 V ₉₇
1138.1655	3412.4839	3	-1.0	0	K.FSEQLVDCDMTNAGCDGGLMEP AFTYVIK.H	CAM: 9, 16; Oxidation (Ox) : 11	175 - 207	70 - 99	Amb a 11 V ₂₀₅	Amb a 11 V ₉₇
1138.1652	3412.4839	3	-1.0	0	K.FSEQLVDCDMTNAGCDGGLMEP AFTYVIK.H	CAM: 9, 16; Ox : 21	175 - 207	70 - 99	Amb a 11 V ₂₀₅	Amb a 11 V ₉₇
773.3889	1545.7696	2	0.6	0	K.HGGIAPEASYPYVGK.R		208 - 222	100 - 114		
779.0438	2335.114	2	1.4	1	K.HGGIAPEASYPYVGKRETC.DK.A	CAM: 19	208 - 228	100 - 120		
561.9244	2805.594	5	-1.0	3	K.AKIDVLKIDGRQNVPLDEEALRK.A		229 - 253	121 - 145		
652.3717	2606.462	4	0.9	2	K.IKIDVLKIDGRQNVPLDEEALRK.A		231 - 253	123 - 145		
478.2582	1910.009	4	1.0	0	K.IDGRQNVPLDEEALRK.A		237 - 253	129 - 145		
834.2332	5000.354	6	1.7	0	K.AVAHQPVATGIQLSGHGLQFYSEG VYTGDCGTEPNHGVGIVGYGENEK.G	CAM: 30	254 - 301	146 - 193		
581.2709	1161.532	2	1.8	0	K.NSWGPTWGEK.G		310 - 319	202 - 211		
609.7803	1218.5544	2	-0.3	1	K.NSWGPTWGEK.G.Y		310 - 320	202 - 212		
932.9034	1864.8026	2	-1.7	0	K.EGLCGVAMHSSFPIMND.P		331 - 347	223 - 239		
870.2034	4346.9791	5	1.5	4	K.EGLCGVAMHSSFPIMNDPNPPKDDPNPKDDPDAPKDPKF.-	CAM: 4	331 - 370	223 - 262	Amb a 11 G ₃₅₇ D ₃₆₀	Amb a 11 G ₂₄₉ D ₂₅₂
873.0042	4360.9947	5	-1.4	4	K.EGLCGVAMHSSFPIMNDPNPPKDDPNAPKDDPDAPKDPKF.-	CAM: 4	331 - 370	223 - 262	Amb a 11 A ₃₅₇ D ₃₆₀	Amb a 11 A ₂₄₉ D ₂₅₂
881.4113	4403.0417	5	-3.3	4	K.EGLCGVAMHSSFPIMNDPNPPKDDPNVPKEDPDAPKDPKF.-	CAM: 4	331 - 370	223 - 262	Amb a 11 V ₃₅₇ E ₃₆₀	Amb a 11 V ₂₄₉ E ₂₅₂

Table E4. Summary of the main Amb a 11 tryptic peptides (*in-gel* digestion) identified by nanoESI-Qq-TOF MS/MS.

Modified amino acid residues are shown in bold type. ^aThe zymogen consists of 386 amino acids, including 108 additional N-terminal residues. Miscleavage (MisC.). Measured (Meas.). Calculated (Calc.). Carbamidomethyl Cys (CAM).

Identification of the cysteine protease Amb a 11 as a novel major allergen from short ragweed

Meas. m/z	Calc. M+H ⁺	z	Δ m/z (ppm)	MisC.	Sequence	Modifications	*Range	Range	*Protein	Protein
326.8637	978.5771	3	-0.58	0	K.IPDKVDWR.E		131 - 138	23 - 30		
862.9330	3448.7100	4	0.04	2	R.EKNAVTDVKGGGCGSCWAFAAVVALEGINAIR.T		132 - 171	23 - 63		
1064.5294	3191.5725	3	0.37	1	K.NAVTDVKGGGCGSCWAFAAVVALEGINAIR.T		141 - 171	33 - 63		
822.0667	2464.1860	3	-0.19	0	K.GQGGCGSCWAFAAVVALEGINAIR.T		148 - 171	40 - 63		
773.3887	1545.7696	2	0.33	0	K.HGGIAPEASYPYVGK.R		208 - 222	100 - 114		
851.4404	1701.8707	2	1.64	1	K.HGGIAPEASYPYVGK.R		208 - 223	100 - 115		
637.3419	1910.0090	3	1.12	1	K.IDGRQNVPLDDEALRK.A		237 - 253	129 - 145		
670.8440	1340.6805	2	0.19	0	R.QNVPLDDEALR.K		241 - 252	133 - 144		
734.8918	1468.7754	2	0.61	1	R.QNVPLDDEALR.K		241 - 253	133 - 145		
1000.8772	5000.3544	5	0.47	0	K.AVAHQPVATGIQLSGHGLQFYSEG VYTGDGCTEPNHGVGIVGYGENEK.G	CAM: 30	254 - 301	146 - 193		
326.8637	978.5771	3	-0.58	1	K.GIKFWTVK.N		302 - 309	194 - 201		
581.2698	1161.5323	2	-0.02	0	K.NSWGPTWGEK.G		310 - 319	202 - 211		
508.0064	2029.0039	4	-0.05	1	K.NSWGPTWGEKGYIHLQR.G		310 - 326	202 - 218		

Meas. m/z	Calc. M+H ⁺	z	Δ m/z (ppm)	MisC.	Sequence	Modifications	*Range	Range
326.8637	978.5771	3	-0.58	0	K.IPDKVDWR.E		131 - 138	23 - 30
862.9330	3448.7100	4	0.04	2	R.EKNAVTDVKGGGCGSCWAFAAVVALEGINAIR.T		132 - 171	23 - 63
1064.5294	3191.5725	3	0.37	1	K.NAVTDVKGGGCGSCWAFAAVVALEGINAIR.T		141 - 171	33 - 63
822.0667	2464.1860	3	-0.19	0	K.GQGGCGSCWAFAAVVALEGINAIR.T		148 - 171	40 - 63
773.3887	1545.7696	2	0.33	0	K.HGGIAPEASYPYVGK.R		208 - 222	100 - 114
851.4404	1701.8707	2	1.64	1	K.HGGIAPEASYPYVGK.R		208 - 223	100 - 115
637.3419	1910.0090	3	1.12	1	K.IDGRQNVPLDDEALRK.A		237 - 253	129 - 145
670.8440	1340.6805	2	0.19	0	R.QNVPLDDEALR.K		241 - 252	133 - 144
734.8918	1468.7754	2	0.61	1	R.QNVPLDDEALR.K		241 - 253	133 - 145
1000.8772	5000.3544	5	0.47	0	K.AVAHQPVATGIQLSGHGLQFYSEG VYTGDGCTEPNHGVGIVGYGENEK.G	CAM: 30	254 - 301	146 - 193
326.8637	978.5771	3	-0.58	1	K.GIKFWTVK.N		302 - 309	194 - 201
581.2698	1161.5323	2	-0.02	0	K.NSWGPTWGEK.G		310 - 319	202 - 211
508.0064	2029.0039	4	-0.05	1	K.NSWGPTWGEKGYIHLQR.G		310 - 326	202 - 218

A

10	20	30	40	50	60
MEINKLVCFS	FSLVLILGLG	ESFHYHEREL	ESEEGFMGM	YDRWREQHNIE	MRSPEFNFVF
F S	A/V				L→F ₆ , C→S ₈ , G→A/V ₂₀
70	80	90	100	110	120
KYNVRRIHES	NKMDKPYKLK	VNEFADMTNL	EFVNTYANSK	ISHFQALRGS	APGSIDTDPN
130	140	150	160	170	180
KDFIYANVT	KIPDKVDWREK	NAVTDVKGQG	GCGSCWAFAA	VVALEGINAI	RTGKLVKFSE
190	200	210	220	230	240
QQLVDCDMTN	AGCDGGLMEP	AFTYVIKHGG	IAPEASYPYV	GKRETCDAK	IKDVLKIDGR
	I				V→I ₂₀₅
250	260	270	280	290	300
QNVPLDDEEA	LRKAVAHQPV	ATGIQLSGHG	LQFYSEGVYT	GDCGTEPNHG	VGIVGYGENE
310	320	330	340	350	360
KGIKFWTVKN	SWGPTWGEKG	YIHLQRGARK	EGLCGVAMHS	SFPIMNDPNP	PKDDPN GPKD
			A/V	G→A/V ₃₅₇	
			E	D→E ₃₆₀	
370	380	386			
DPDAPKDPKF	KTTQRLQGIR	TKLLEL			

B

V ₂₀₅ G ₃₅₇	I ₂₀₅ G ₃₅₇	V ₂₀₅ A ₃₅₇	I ₂₀₅ A ₃₅₇	V ₂₀₅ V ₃₅₇	I ₂₀₅ V ₃₅₇
D ₃₆₀	D ₃₆₀	D ₃₆₀	D ₃₆₀	E ₃₆₀	E ₃₆₀

C

HTS	MEINKLVCFSFSLVLILGLVESFHYHERELESEEGFMGM	YDRWREQHNIE	MRSPEFNFVF	60
RACE	MEINKLVCFSFSLVLILGLGESFHYHERELESEEGFMGM	YDRWREQHNIE	MRSPEFNFVF	60

HTS	KYNVRRIHESNKMDKPYKLKVNEFADMTNLEFVNTYANSK	ISHFQALRGS	APGSIDTDPN	120
RACE	KYNVRRIHESNKMDKPYKLKVNEFADMTNLEFVNTYANSK	ISHFQALRGS	APGSIDTDPN	120

HTS	KDFIYANVTKIPDKVDWREKNAVTDVKGQGGCGSCWAF	AAVVALEGINAI	RTGKLVKFSE	180
RACE	KDFIYANVTKIPDKVDWREKNAVTDVKGQGGCGSCWAF	AAVVALEGINAI	RTGKLVKFSE	180

HTS	QQLVDCDMTNAGCDGGLMEPAFTYVIKHGGIAPEASYP	YVGKRETCDAK	IKDVLKIDGR	240
RACE	QQLVDCDMTNAGCDGGLMEPAFTYVIKHGGIAPEASYP	YVGKRETCDAK	IKDVLKIDGR	240

HTS	QNVPLDDEEA	LRKAVAHQPV	ATGIQLSGHGLQFYSEGVYT	GDCGTEPNHG
RACE	QNVPLDDEEA	LRKAVAHQPV	ATGIQLSGHGLQFYSEGVYT	GDCGTEPNHG

HTS	KGIKFWTVKN	SWGPTWGEKGY	IHLQRGARKEGLCGVAMHSS	SFPIMNDPNPPKDDPNVPKD
RACE	KGIKFWTVKN	SWGPTWGEKGY	IHLQRGARKEGLCGVAMHSS	SFPIMNDPNPPKDDPNVPKD

HTS	DPDAPKDPKF	KTTQRLQGIR	TKLLEL	381
RACE	DPDAPKDPKF	KTTQRLQGIR	TKLLEL	386

Fig E1



Fig E2

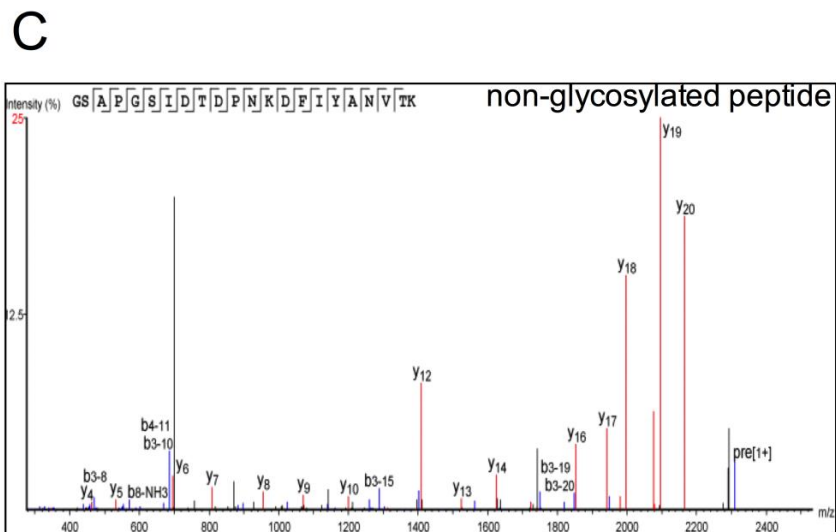
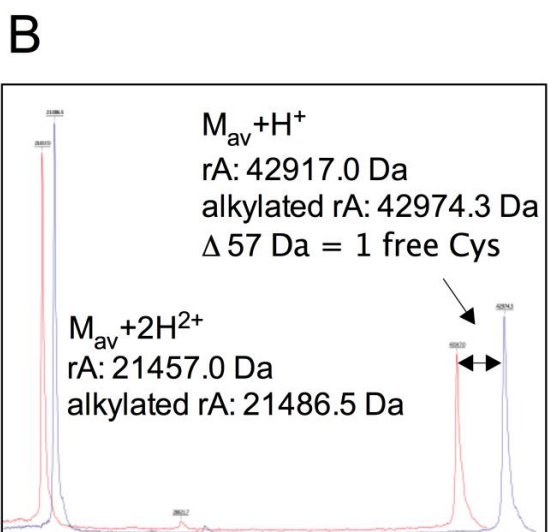
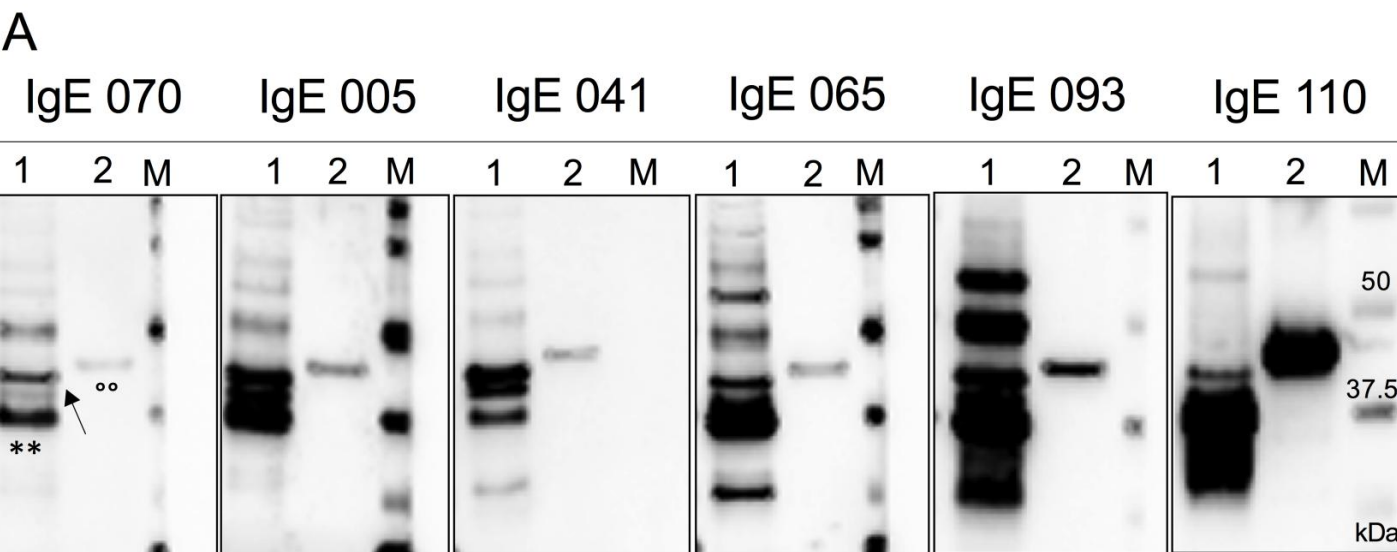


Fig E3

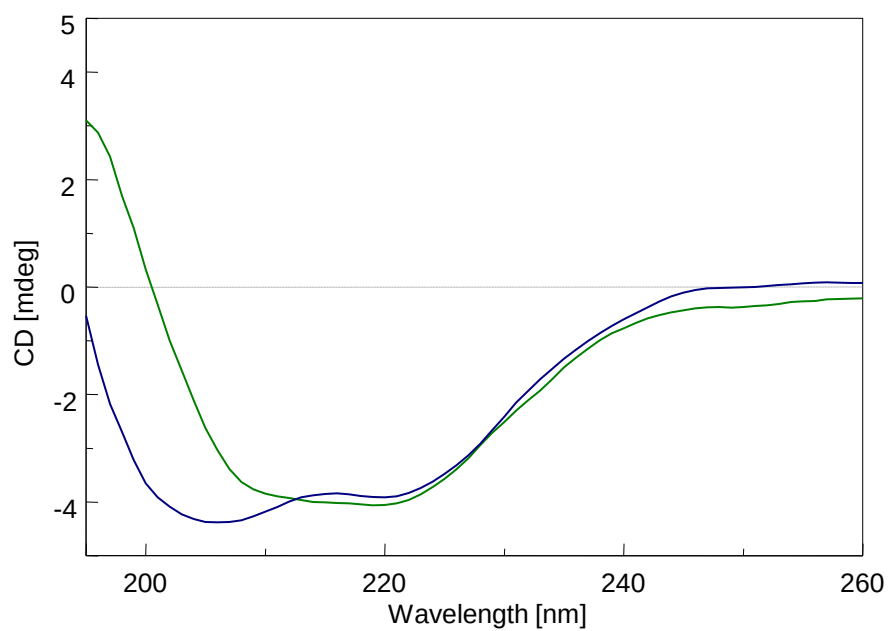


Fig E4

```
Amb_a_11_V97G249D252_mature      GSAPGSIDTDPNKFYIANVTKIPDKVDWREKNAVTDVKGGGGSWAF 50
1S4V_B|PDBID|CHAIN|SEQUENCE      -----TVPASVDWRKKGAVTSVKDQGQCGSWAF 29
                                   .:* .***:*.***.*.* *****

Amb_a_11_V97G249D252_mature      AAVVALEGINAIRTGKLVKFSEQQLVDCDMT-NAGCDGGLMEPAFTYVIK 99
1S4V_B|PDBID|CHAIN|SEQUENCE      STIVAVEGINQIKTNKLVSLSQELVDCDTDQNGGGLMDYAFEFIKQ 79
                                   ::*:***.*.*.***:***:***** * **:***: * * ::

Amb_a_11_V97G249D252_mature      HGGIAPEASYPYVGKRECDKAKIKD-VLKIDGRQNPVGLDEEALRKAVA 148
1S4V_B|PDBID|CHAIN|SEQUENCE      RGGITTEANYPYEAYDGTCDVSKENAPAVSIDGHENVPENDENALLKAVA 129
                                   :***:*.***.*.* . *** : * : .:***:*** **:** ****

Amb_a_11_V97G249D252_mature      HQPVATGIQLSGHGLQFYSEGVTGDCGTEPNHGVGIVGYGENEGIKIKFW 198
1S4V_B|PDBID|CHAIN|SEQUENCE      NQPVSVDAIDAGGSDFQFYSEGVTGSCGTELDHGVAIVGYTTIDGTKYW 179
                                   :***:..*: . * :*****:*.*** :***.***** . * *:

Amb_a_11_V97G249D252_mature      TVKNSWGPTWGEKGYIHLQRGAR-KEGLCGVAMHSSFPIMNDPNPPKDDP 247
1S4V_B|PDBID|CHAIN|SEQUENCE      TVKNSWGPEWGEKGYIRMERGISDKEGLCGIAMEASYPIKKSSN----NP 225
                                   ***** *****::** *****:***:*** :..* :*

Amb_a_11_V97G249D252_mature      NGPKDDPDAPKDPKF 262
1S4V_B|PDBID|CHAIN|SEQUENCE      SGIK----- 229
                                   .* *
```

Template Sequence (pdb ID 1S4V)

```
>1S4V:A|PDBID|CHAIN|SEQUENCE (Vignain, Ricinus communis)
TVPASVDWRKKGAVTSVKDQGQCGSCWAFSTIVAVEGINQIKTNKLVSLSQELVDCDTDQNGGGLMDYAFEFIKQR
GGITTEANYPYEAYDGTCDVSKENAPAVSIDGHENVPENDENALLKAVANQPVSVDAIDAGGSDFQFYSEGVTGSCGTEL
DHGVAIVGYGTTIDGTKYWTVKNSWGPEWGEKGYIRMERGISDKEGLCGIAMEASYPIKKSSNNPSGIK
>1S4V:B|PDBID|CHAIN|SEQUENCE (Vignain, Ricinus communis)
TVPASVDWRKKGAVTSVKDQGQCGSCWAFSTIVAVEGINQIKTNKLVSLSQELVDCDTDQNGGGLMDYAFEFIKQR
GGITTEANYPYEAYDGTCDVSKENAPAVSIDGHENVPENDENALLKAVANQPVSVDAIDAGGSDFQFYSEGVTGSCGTEL
DHGVAIVGYGTTIDGTKYWTVKNSWGPEWGEKGYIRMERGISDKEGLCGIAMEASYPIKKSSNNPSGIK
>1S4V:C|PDBID|CHAIN|SEQUENCE
VLKX
>1S4V:D|PDBID|CHAIN|SEQUENCE
VLKX
```

Model energy (kcal/mol) :

Stereochemical quality of models with PROCHECK (Roman A. Laskowski) :

	core	allowed	generously	disallowed
Model 3	74.0%	24.9%	0.6%	0.6%

Number of intrachain restraints deduced from templates :

- . 7257 on MChain atoms (CA,C,N)
- . 6432 on other atoms (CB,CE,OG,...)

Number of violation of these intrachain restraints :

	MChain	Other
Model 3	1 (0.00%)	0 (0.00%)