Cloning of *Amb a I* (Antigen E), the Major Allergen Family of Short Ragweed Pollen*

(Received for publication, June 22, 1990)

Thorunn Rafnar†, Irwin J. Griffith§, Mei-chang Kuo§, Julian F. Bond§, Bruce L. Rogers§, and David G. Klapper‡

From the †Department of Microbiology and Immunology, University of North Carolina School of Medicine, Chapel Hill, North Carolina 27599 and the §ImmuLogic Pharmaceutical Corporation, Cambridge, Massachusetts 02139

To determine the structure of *Amb a I* (previously called antigen E), the major allergen from short ragweed, cDNA from pollen was cloned into λgt11 and λgt10. One of the three distinct clones isolated from the λgt1 library by screening with anti-denatured *Amb a I* antibodies was used to screen both libraries for other *Amb a I* sequences. Multiple clones were isolated and sequenced and proved to be highly homologous but nonidentical. The clones could be divided into three groups based on sequence similarity, and in accordance with the International Union of Immunological Societies-approved nomenclature (Marsh, D. G., Goodfriend, L., King, T. P., Lowenstein, H., and Platts-Mills, T. A. E. (1986) *Bull. WHO* 64, 767–770) they have been designated *Amb a I.1*, *Amb a I.2*, and *Amb a I.3*. Clones within a group have greater than 99% identity, and similarity among groups is 85–90% at the nucleotide level. The amino acid sequence of four peptides (isolated from antigen E obtained from the Research Resources Branch of the National Institutes of Health) containing 132 amino acids was identical to one of the clones (*Amb a I.1*). The presence of multiple naturally occurring isoelectric forms of *Amb a I* was demonstrated by two-dimensional gel electrophoresis and Western blotting. Southern blot analysis demonstrates the presence of multiple *Amb a I*-related sequences in the ragweed genome. *Amb a I* is therefore not a single molecule but rather a family of closely related proteins.

Of all the seasonal aeroallergens, pollen from short ragweed (*Ambrosia artemisiifolia*) is perhaps the most clinically important. The extremely small particle size allows this pollen to be carried hundreds of miles, and it is the major cause of late summer hay fever in the eastern United States and Canada (2). Of the 52 antigens present in an aqueous extract of pollen, at least 22 are allergens defined by their reactivity with human IgE (3). At least five significant human allergens from short ragweed pollen have been purified to homogeneity and studied with respect to their biochemical and immunological characteristics (2).

Antigen E or *Amb a I* (according to nomenclature in Ref. 1) is considered the most important allergen since 95% of ragweed-sensitive individuals react to it in skin tests and show high IgE antibody titers to it (4, 5). *Amb a I* is highly abundant, comprising about 6% of the total protein in a neutral aqueous extract of pollen (4). It is an acidic, amino-terminal “blocked,” reportedly nonglycosylated single-chain protein of 38 kDa which undergoes proteolysis during chromatographic purification and is cleaved into two chains, α and β, of 26 and 12 kDa, respectively (6). The two-chain form is reported to be allergenically and antigenically indistinguishable from the intact molecule, but modification of the protein, including reduction and alkylation of disulfide bonds, urea denaturation and renaturation, or succinylation of lysine residues, reduces the IgE immunoreactivity of the molecule (2). Recent data with Western blotting demonstrate that *Amb a I* retains the ability to bind antibody from allergic humans and hyperimmunized animals (documented in Fig. 3 and Footnote 1).

It has been reported that immunotherapy utilizing purified *Amb a I* is as effective in alleviating clinical symptoms in allergic patients as is immunotherapy using whole pollen extract (7). Immunizations with modified forms of *Amb a I* have also been tested and shown to be as clinically effective as native *Amb a I* and to cause fewer systemic reactions (8–11). *Amb a I* has been shown to have three nonoverlapping, nonrepeated antigenic sites, as defined by murine monoclonal antibodies, of which at least two represent major human allergenic epitopes (12). In addition, preliminary studies directed at examining T cell epitopes of *Amb a I* suggest that they are linear rather than conformational (13–15).

At present, desensitization immunotherapy for ragweed-allergic individuals relies upon multiple injections of small doses of aqueous pollen extracts. These protocols are not ideal since individuals present with varying sensitivities to each of the multiple components in an extract, and various batches of extract used for diagnostic and therapeutic purposes vary a great deal in their specific allergen content. Furthermore, although immunotherapy offers some improvement to many patients, almost no patients become completely asymptomatic, and a number of patients show no symptomatic improvement at all (16). Identification and characterization of specific epitopes of *Amb a I* might be particularly useful in improving an immunotherapeutic approach to desensitization.

As a first step toward this end, cDNA libraries have been constructed from short ragweed pollen and whole flowers. This report describes the cloning of *Amb a I* and presents the complete nucleotide and deduced amino acid sequences of three clones coding for this 398-amino acid allergen. Comparison of the three cloned sequences shows that they have

*This work was supported by National Institutes of Health Grant AI 14908 (to D. G. K.) and by a grant from the North Carolina Biotechnology Center (to D. G. K.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

† To whom correspondence should be addressed.

‡ Received for publication, June 22, 1990.
several differences, suggesting that Amb a 1 is actually a family of proteins with at least three members. These proteins represent an as yet undescribed family since a search of available data banks shows no significant homology to other proteins from either procaryotes or eucaryotes. Finally, what has been reported in the literature (6), the carboxyl terminus of Amb a 1 appears to be cysteine. During this investigation, a peptide corresponding to the predicted carboxyl terminus of Amb a 1 was isolated from chromatographically purified Amb a 1, confirming the assignment of cysteine as the carboxy-terminal amino acid.

MATERIALS AND METHODS

Antibodies—Mouse monoclonal anti-denatured Amb a 1 antibodies JB4F-3-5, JB3C-9-3, JB1E3-4, 2D6/E6, JB4E3-4, and JB1E2-2 have been described previously (17).

Plant Tissue—Flowers and leaves of short ragweed were picked, frozen immediately in liquid nitrogen, and stored at -70 °C until processing. Defatted short ragweed pollen and meadow rescue pollen were obtained from Greer Laboratories (Lenoir, NC).

Genomic DNA Isolation—Genomic DNA was isolated by a method published previously (18) with the following changes. 50–100 g of frozen ragweed flowers were ground in liquid nitrogen using a mortar and pestle. The dry powder was suspended in homogenization buffer (0.1% SDS, 50 mM KCl, 0.25 M sucrose, 10 mM Tris-HCl, pH 8.5, 0.1 M EDTA) by centrifugation for 5 s in a microcentrifuge. The DNA was then washed twice in CsCl solution and centrifuged for 45 h at 55,000 rpm in an analytical ultracentrifuge. The DNA was then resuspended in 10 mM Tris-HCl, pH 8.5, 0.17 M NaCl, 0.05% Tween 20; Sigma) and stored at -20 °C overnight. The DNA was then isolated using the nonnucleophilic reductant, tributylphosphine, with concomitant alkylation by 4-vinylpyridine in ethylmorpholine buffer (28).

MATERIALS AND METHODS

Antibodies—Mouse monoclonal anti-denatured Amb a 1 antibodies JB4F-3-5, JB3C-9-3, JB1E3-4, 2D6/E6, JB4E3-4, and JB1E2-2 have been described previously (17).

Plant Tissue—Flowers and leaves of short ragweed were picked, frozen immediately in liquid nitrogen, and stored at -70 °C until processing. Defatted short ragweed pollen and meadow rescue pollen were obtained from Greer Laboratories (Lenoir, NC).

Genomic DNA Isolation—Genomic DNA was isolated by a method published previously (18) with the following changes. 50–100 g of frozen ragweed flowers were ground in liquid nitrogen using a mortar and pestle. The dry powder was suspended in homogenization buffer (0.1% SDS, 50 mM KCl, 0.25 M sucrose, 10 mM Tris-HCl, pH 8.5, 0.1 M EDTA) by centrifugation for 5 s in a microcentrifuge. The DNA was then washed twice in CsCl solution and centrifuged for 45 h at 55,000 rpm in an analytical ultracentrifuge. The DNA was then resuspended in 10 mM Tris-HCl, pH 8.5, 0.17 M NaCl, 0.05% Tween 20; Sigma) and stored at -20 °C overnight. The DNA was then isolated using the nonnucleophilic reductant, tributylphosphine, with concomitant alkylation by 4-vinylpyridine in ethylmorpholine buffer (28).

MATERIALS AND METHODS

Antibodies—Mouse monoclonal anti-denatured Amb a 1 antibodies JB4F-3-5, JB3C-9-3, JB1E3-4, 2D6/E6, JB4E3-4, and JB1E2-2 have been described previously (17).

Plant Tissue—Flowers and leaves of short ragweed were picked, frozen immediately in liquid nitrogen, and stored at -70 °C until processing. Defatted short ragweed pollen and meadow rescue pollen were obtained from Greer Laboratories (Lenoir, NC).

Genomic DNA Isolation—Genomic DNA was isolated by a method published previously (18) with the following changes. 50–100 g of frozen ragweed flowers were ground in liquid nitrogen using a mortar and pestle. The dry powder was suspended in homogenization buffer (0.1% SDS, 50 mM KCl, 0.25 M sucrose, 10 mM Tris-HCl, pH 8.5, 0.1 M EDTA) by centrifugation for 5 s in a microcentrifuge. The DNA was then washed twice in CsCl solution and centrifuged for 45 h at 55,000 rpm in an analytical ultracentrifuge. The DNA was then resuspended in 10 mM Tris-HCl, pH 8.5, 0.17 M NaCl, 0.05% Tween 20; Sigma) and stored at -20 °C overnight. The DNA was then isolated using the nonnucleophilic reductant, tributylphosphine, with concomitant alkylation by 4-vinylpyridine in ethylmorpholine buffer (28).

MATERIALS AND METHODS

Antibodies—Mouse monoclonal anti-denatured Amb a 1 antibodies JB4F-3-5, JB3C-9-3, JB1E3-4, 2D6/E6, JB4E3-4, and JB1E2-2 have been described previously (17).

Plant Tissue—Flowers and leaves of short ragweed were picked, frozen immediately in liquid nitrogen, and stored at -70 °C until processing. Defatted short ragweed pollen and meadow rescue pollen were obtained from Greer Laboratories (Lenoir, NC).

Genomic DNA Isolation—Genomic DNA was isolated by a method published previously (18) with the following changes. 50–100 g of frozen ragweed flowers were ground in liquid nitrogen using a mortar and pestle. The dry powder was suspended in homogenization buffer (0.1% SDS, 50 mM KCl, 0.25 M sucrose, 10 mM Tris-HCl, pH 8.5, 0.1 M EDTA) by centrifugation for 5 s in a microcentrifuge. The DNA was then washed twice in CsCl solution and centrifuged for 45 h at 55,000 rpm in an analytical ultracentrifuge. The DNA was then resuspended in 10 mM Tris-HCl, pH 8.5, 0.17 M NaCl, 0.05% Tween 20; Sigma) and stored at -20 °C overnight. The DNA was then isolated using the nonnucleophilic reductant, tributylphosphine, with concomitant alkylation by 4-vinylpyridine in ethylmorpholine buffer (28).

MATERIALS AND METHODS

Antibodies—Mouse monoclonal anti-denatured Amb a 1 antibodies JB4F-3-5, JB3C-9-3, JB1E3-4, 2D6/E6, JB4E3-4, and JB1E2-2 have been described previously (17).

Plant Tissue—Flowers and leaves of short ragweed were picked, frozen immediately in liquid nitrogen, and stored at -70 °C until processing. Defatted short ragweed pollen and meadow rescue pollen were obtained from Greer Laboratories (Lenoir, NC).

Genomic DNA Isolation—Genomic DNA was isolated by a method published previously (18) with the following changes. 50–100 g of frozen ragweed flowers were ground in liquid nitrogen using a mortar and pestle. The dry powder was suspended in homogenization buffer (0.1% SDS, 50 mM KCl, 0.25 M sucrose, 10 mM Tris-HCl, pH 8.5, 0.1 M EDTA) by centrifugation for 5 s in a microcentrifuge. The DNA was then washed twice in CsCl solution and centrifuged for 45 h at 55,000 rpm in an analytical ultracentrifuge. The DNA was then resuspended in 10 mM Tris-HCl, pH 8.5, 0.17 M NaCl, 0.05% Tween 20; Sigma) and stored at -20 °C overnight. The DNA was then isolated using the nonnucleophilic reductant, tributylphosphine, with concomitant alkylation by 4-vinylpyridine in ethylmorpholine buffer (28).
peste in 4 M guanidinium isothiocyanate buffer and total RNA isolated by standard procedures (34).

Priming oligonucleotides (Table I) for DNA amplification (35) were RW38, RW32, RW45, and anchor primer. RW38 corresponded to the amino-terminal coding strand sequence encoding amino acids Leu-Tyr-Phe-Thr-Leu (amino acids 10–14). The RW38 sequence was conserved between Amb a 1.2 and Amb a 1.3. RW32, which corresponded to the noncoding strand sequence between 12 and 39 nucleotides 3′ of the TAA stop codon, was specific to Amb a 1.1. RW45 corresponded to the noncoding strand sequence complementary to amino acids Ile-Lys-Ser-Asn-Asp-Gly (amino acids 181–186) of Amb a 1. Two additional primers, anchor template and anchor linker, were used for linking to cDNA. The anchor primer oligomer sequence was contained within the anchor template sequence. The anchor linker oligomer was phosphorylated (see Table I for specific sequences and restriction sites). These oligonucleotides were purchased from Repligene (Cambridge, MA).

First-strand cDNA was synthesized from 1 μg of total RNA with the cDNA Synthesis System Plus kit (Amersham Corp.) using poly(dT) as a primer. The single-stranded DNA (20 μl) was mixed with 100 pmol of each priming oligo, RW38 and RW32, 10 μl of 10 × reaction buffer (GeneAmp kit, U. S. Biochemicals), and 0.5 μl of Thermus aquaticus polymerase (U. S. Biochemicals). The mixture was brought to 100 μl with distilled water and overlaid with a two-layer oil (Sigma). The sample was amplified with a programmable thermal controller from MJ Research, Inc. (Cambridge, MA). The first five rounds of amplification consisted of denaturation at 94 °C for 1 min, annealing of primers to the template at 45 °C for 1.5 min, and chain elongation at 70 °C for 4 min. The final 20 rounds of amplification consisted of denaturation at 95 °C for 1 min, annealing at 55 °C for 1.5 min, and chain elongation as above.

Amplified DNA was recovered by sequential chloroform, phenol and chloroform extractions followed by overnight precipitation with isopropl alcohol. DNA was digested simultaneously with EcoRI and PstI and electrophoresed on a 1% GTG agarose (FMC, Rockland, ME) preparative gel. The predicted 1.2-kilobase band was isolated and recovered by glass bead adherence (GeneClean kit, BRL, La Jolla, CA). The digested DNA was ligated into EcoRI/PstI-digested M13 for dideoxy sequencing (22).

The very 5′ end sequence of Amb a 1 was determined using a modification of the anchored PCR (36). Double-stranded cDNA was synthesized from 1 μg of RNA with the cDNA Synthesis System Plus kit using poly(dT) as a primer, blunt ended with T4 polynucleotide, and blunt end ligated to self-anealed anchor template and anchor linker primers. Linked cDNA (3 μl) was mixed with 100 pmol of the anchor primer and RW32 primers, 10 μl of 10 × reaction buffer and 0.5 μl of T. aquaticus polymerase. The mixture was brought to 100 μl and amplified as described above. 1% of the volume of the primary PCR was reamplified with anchor primer and RW45 oligomers. RW45 is nested (internal) relative to oligomer RW32 used in the primary PCR. Amplified DNA from the secondary PCR corresponding to the 5′ end RW45 sequence was recovered as above, digested with KpnI, and ligated into KpnI/HindII-digested M13 for dideoxy sequencing (22).

RESULTS

Three clones were isolated from a 1g11 ragweed pollen cDNA library by screening with a pool of seven mouse monoclonal antibodies raised to denatured Amb a 1. The sequences of all three of these clones showed extensive homology to peptides isolated from highly purified Amb a 1. One of these clones (Amb a 1.2) was radiolabeled and used to screen this library and a 1g10 library constructed from cDNA made from whole ragweed flowers, resulting in the isolation of 10 additional clones.

All the clones sequenced could be divided into three groups, Amb a 1.1, Amb a 1.2, and Amb a 1.3, in which clones within a group share greater than 99% identity, and identity among groups ranges from 85 to 90% at the nucleotide level. The DNA sequences of the three largest prototypic clones in each group are shown in Fig. 1. The longest clone, a member of the Amb a 1.3 group, is 1331 bases long, has 13 nucleotides preceding the putative ATG start codon at position 1, an unbroken reading frame of 1190 nucleotides ending with a TAA stop codon, and a 125-nucleotide untranslated AT-rich region before the poly(A) addition site. The other clones start at positions −2 and 44 relative to the Amb a 1.3 prototype start site and have unbroken reading frames extending to a stop site identical to that found in the Amb a 1.3 group. The traditional mammalian consensus sequence for polyadenylation, AAATAA, does not occur in the untranslated regions. The Amb a 1.2 group has a nucleotide triplet at position 118 which is not present in the Amb a 1.3 and Amb a 1.1 groups, and the Amb a 1.1 group lacks an additional triplet at position 101 relative to the other two groups of clones. Since only a single non-full-length clone representing the Amb a 1.1 group was originally isolated from the libraries, the full-length clone was obtained by sequence-specific amplifications of pollen RNA using Taq polymerase and synthetic DNA primers. The amino-terminal primer corresponded to coding sequence nucleotides 25–44 in clone Amb a 1.3 and was conserved between the Amb a 1.2 and Amb a 1.3 groups. The carboxyl-terminal primer was specific to Amb a 1.1 and corresponded to the 3′-noncoding sequence between 12 and 29 nucleotides 3′ of the TAA stop codon in that particular clone. The very 5′ end sequence of Amb a 1.1 was determined using anchored PCR methodology. In this case, the amino-terminal primer corresponded to a synthetic linker sequence. The carboxyl-terminal primer described above was used in a primary amplification whereas a carboxyl-terminal primer corresponding to the noncording strand sequence for amino acids 181–186 was used in a secondary amplification. A secondary amplification using a nested primer was necessary since the primary amplification

<table>
<thead>
<tr>
<th>Table I</th>
<th>Nucleotide sequence of oligomers used in the PCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oligomer</td>
<td>Sequence</td>
</tr>
<tr>
<td>RW32</td>
<td>5′ GGCTTCGATCACCTATTATAAAGCTTTATGG</td>
</tr>
<tr>
<td>RW38</td>
<td>5′ GGAAATTCGTTATTTACCTTACGC</td>
</tr>
<tr>
<td>RW45</td>
<td>5′ ACCATCGTGGACCTAAT</td>
</tr>
<tr>
<td>AL</td>
<td>5′ p-AATTGCGATGCT</td>
</tr>
<tr>
<td>AP</td>
<td>5′ GGCTTCGAACGTATCCGG</td>
</tr>
<tr>
<td>AT</td>
<td>5′ GGCTTCGATTACCGTGCC</td>
</tr>
</tbody>
</table>

The very 5′ end sequence of Amb a 1 was determined using a modification of the anchored PCR (36). Double-stranded cDNA was synthesized from 1 μg of RNA with the cDNA Synthesis System Plus kit using poly(dT) as a primer, blunt ended with T4 polynucleotide, and blunt end ligated to self-anealed anchor template and anchor linker primers. Linked cDNA (3 μl) was mixed with 100 pmol of the anchor primer and RW32 primers, 10 μl of 10 × reaction buffer and 0.5 μl of T. aquaticus polymerase. The mixture was brought to 100 μl and amplified as described above. 1% of the volume of the primary PCR was reamplified with anchor primer and RW45 oligomers. RW45 is nested (internal) relative to oligomer RW32 used in the primary PCR. Amplified DNA from the secondary PCR corresponding to the 5′ end RW45 sequence was recovered as above, digested with KpnI, and ligated into KpnI/HindII-digested M13 for dideoxy sequencing (22).
FIG. 1—continued

Cloning of Amb a 1

FIG. 1. Nucleotide sequences of short ragweed clones Amb a 1.3, Amb a 1.2, and Amb a 1.1. Numbering begins with the adenine of the ATG initiation codon of Amb a 1.3. The start codon and TAA stop codon are underscored. Horizontal bars (——) represent identity to clone Amb a 1.1; a dotted line (-----) indicates no sequence information; asterisks (*) mark positions at which nucleotides are absent when compared with either the other sequences. Lower case letters 3' in the clone Amb a 1.1 indicate the sequence of the 3' oligonucleotide used for its amplification. The poly(A) tails of the clones are not shown. Solid lines (-----) indicate the positions of the oligomers used in the PCR.

did not yield a discrete product band. Sequence identities at the nucleotide level in the coding regions are 91% between clones Amb a 1.3 and Amb a 1.2 and 86% between Amb a 1.2 and Amb a 1.1 and between Amb a 1.3 and Amb a 1.1.

Amino acid sequence data are presented in Fig. 2. Sequences derived from chromatographically purified Amb a 1 accompany translated amino acid sequences of the three prototype clones. Approximately 130 amino acids were identified from four peptides. Two of the peptides were sequenced after a cyanogen bromide cleavage of purified Amb a 1, and it was assumed that each follows a methionine. A third peptide was sequenced after acid cleavage (70% formic acid) of the aspartic acid-proline bond at position 361. For peptides 277-307 and 308-321, the NH2-terminal end of this peptide was obtained from the amino terminal of the peptide 45-91 and the NH2-terminal end obtained from a minor component of the chain band. Thus, 129 amino acids or approximately 45% of the coding region of Amb a 1.1, was sequenced. Assignment of start and stop codons was based on comparison of the nucleotide sequence with the coding region of Amb a 1.1.

The proteins coded for by these sequences differ by as much as 20% from the amino acid sequence data are presented in Fig. 2. Sequences derived from chromatographically purified Amb a 1 accompany translated amino acid sequences of the three prototype clones. Approximately 130 amino acids were identified from four peptides. Two of the peptides were sequenced after a cyanogen bromide cleavage of purified Amb a 1, and it was assumed that each follows a methionine. A third peptide was sequenced after acid cleavage (70% formic acid) of the aspartic acid-proline bond at position 361. For peptides 277-307 and 308-321, the NH2-terminal end of this peptide was obtained from the amino terminal of the peptide 45-91 and the NH2-terminal end obtained from a minor component of the chain band. Thus, 129 amino acids or approximately 45% of the coding region of Amb a 1.1, was sequenced. Assignment of start and stop codons was based on comparison of the nucleotide sequence with the coding region of Amb a 1.1.
Cloning of Amb a I

Fig. 2. Comparison of the deduced amino acid sequences of clones Amb a 1.3, Amb a 1.2, and Amb a 1.1 peptides. Horizontal bars (-----) represent identity to clone Amb a 1.3; a dotted line (-----) indicates no sequence information. Asterisks (*) indicate deletions in the amino acid sequences. Unidentified amino acid residues from the peptide sequencing are represented by u.

performed using programs described previously (37, 38).

The presence of multiple naturally occurring forms of Amb a I in pollen was confirmed by two-dimensional electrophoresis of pollen protein extract (Fig. 3). Pollen proteins were separated first by charge and then by size, blotted onto nitrocellulose, and probed with a goat anti-Amb a I antiserum. Three major and three minor variants of Amb a I at 38 kDa could be detected with isoelectric points of approximately 5.2, 4.9, 4.7, 4.3, 5.9, and 5.4 in order of decreasing intensity. The calculated (from amino acid composition) isoelectric point of 5.2 of the clone representing Amb a I, along with its perfect match to Amb a I peptides, might suggest that this clone represents the major variant of the allergen isolated from pollen collected over a wide geographic area. The DNA and protein sequences were compared with sequences in the EMBL and NBRF computer data bases. No significant homologies were found in either case.

Northern blot analysis of RNA from defatted ragweed and meadow fescue (grass) pollens is depicted in Fig. 4. RNA was probed with a radiolabeled member of the Amb a I group. Strong binding can be seen to a 1.5-kilobase message in the ragweed pollen RNA, and no binding is demonstrable in comparable amounts of RNA isolated from the grass pollen. Similar analyses using RNA derived from other ragweed plant tissues such as leaves and roots also showed no binding to the Amb a I probe (data not shown).

A Southern blot of genomic DNA digested with various restriction enzymes and probed with Amb a I is shown in Fig. 5. The clone binds to multiple bands in all the preparations and, when compared with restriction enzyme maps of the clones, suggests that there are multiple genes encoding Amb a I in an outbred population of ragweed plants. Whether

FIG. 3. Two-dimensional gel electrophoresis and Western blotting of pollen protein. Crude, soluble pollen proteins were subjected to isoelectric focusing (left to right) followed by 10% SDS-PAGE electrophoresis (top to bottom). The proteins were blotted onto nitrocellulose and probed with goat anti-Amb a I antiserum. The 26,000-dalton band presumably represents the α chain of Amb a I.

FIG. 4. RNA blot analysis of total RNA from various tissues. 20 µg of total RNA was electrophoresed, blotted onto nitrocellulose, and hybridized to radiolabeled Amb a I. Lane 1, short ragweed pollen; lane 2, meadow fescue pollen.

FIG. 5. DNA blot analysis of short ragweed genomic DNA. Genomic DNA was digested with BamHI (lane 1), EcoRI (lane 2), HindIII (lane 3), and NcoI (lane 4) and run on a 0.7% agarose gel. The DNA was transferred to nitrocellulose and hybridized to radiolabeled Amb a I.
the observed heterogeneity of Amb a I is due to allelic variation, multiple genes in a single plant, or perhaps both is presently under investigation.

DISCUSSION

The use of molecular biology techniques to study directly the structure-function relationships of allergens of medical importance has only recently begun to be exploited. Thus far three non-plant allergens (white-faced hornet venom antigen 5, bee venom phospholipase A2, and the house dust mite allergen Der p I) have been cloned and used for immunological studies (39-41). In this report, the major allergen of short ragweed pollen has been cloned, and its complete nucleic acid and deduced amino acid sequence are described. This allergen (Amb a I) is shown to be a family of proteins, closely related to each other but unique from other families of proteins.

Proteins in the pollen coat can originate from two different sources. They may be expressed in the pollen cytoplasm itself, or they can be synthesized in the tapetal nursing cells surrounding the pollen and then deposited on the pollen surface (42). cDNA libraries, therefore, were made from poly(A') RNA both from maturing flowers and from commercially available defatted ragweed pollen. The pollen cDNA library was made in λgt11 and was screened with a panel of monoclonal antibodies raised specifically against denatured Amb a I. From the information gained by cloning and sequencing Amb a I from this library, it was possible to probe a flower cDNA library made in λgt11. In addition, once sequence information was derived, it became possible to design oligonucleotide primers to amplify cDNA coding specifically for Amb a I from ragweed flowers by means of the PCR reaction.

Initially, a pool of seven monoclonal antibodies raised against denatured Amb a I (recognizing both α and β chains) was used to screen the pollen cDNA library. Three clones were isolated and fully characterized. One of these, Amb a I.2, contained an almost complete coding sequence for Amb a I, another named Amb a I.3 contained approximately two-thirds of the coding sequence, and the third, Amb a I.1, contained 320 base pairs and coded for only 50 COOH-terminal amino acids of Amb a I. Further screening of the libraries (and PCR amplification of cDNA in the case of Amb a I) has produced full-length sequences of three groups of highly related structures representing Amb a I. A total of 14 clones were completely sequenced. On average, clones within each group differed at less than 1% of the nucleotides coding for the structural protein while the difference between groups was of the order of 10-15% at the nucleotide level. When full-length representative clones from each group are expressed in Escherichia coli and the products blotted onto nitrocellulose membranes for probing with monoclonal antibodies to denatured Amb a I, each gives a unique pattern of reactivity with individual antibodies (data not shown). This suggests that individual antibodies are able to discriminate between clones and demonstrates that all three groups of Amb a I proteins were in the chromatographically prepared Amb a I used to originally immunize the mice. These three groups of Amb a I clones, therefore, are expressed and are not transcripts of nonfunctional genes.

Multiple isoelectric forms and conservative amino acid substitutions are a common feature of pollen allergens that have been well characterized such as those from the grasses and from short ragweed (4, 43). It was shown earlier (4, 6) that Amb a I has four electrophoretic and/or structural forms called A, B, C, and D. These forms are indistinguishable by amino acid compositional analysis and comparison of their antigenic and allergenic properties. In this report, two-dimen-

sional Western analysis of pollen extracts shows three major isoelectric forms of Amb a I and a number of minor forms that differ both in size and charge. This is consistent with the description of three groups of Amb a I clones and the appreciable difference in individual members of a group may differ by as much as 1% from the prototype within a group. Current efforts are directed at assigning specific groups to specific isoelectric forms of Amb a I. It is not clear at this time that the A, B, C, and D forms of Amb a I have any direct relationship to Amb a I.1, Amb a I.2, or Amb a I.3 as described in this report.

Since the pollen used to create the cDNA library was collected from a wide geographic area, it is unlikely that other major Amb a I sequences will be found in short ragweed. Current experiments are under way to determine whether individual short ragweed plants collected from various geographical locations in North America express multiple (or all) members of the Amb a I family or whether individual plants show some restricted variation, multiple structural genes, or both contribute to this phenomenon.

Amino acid sequence analyses of peptides from chromatographically isolated Amb a I have been obtained. It is of great interest that the best match of that amino acid sequence is with the Amb a I.1 group of Amb a I sequences. However, despite exhaustive screening of several independently derived cDNA libraries, only one short (320-base pair) clone belonging to this group was ever isolated, making this clone the most underrepresented member of the Amb a I family. A full-length clone of Amb a I.1 was only obtained by sequence-specific amplification of pollen cDNA using primers specific for the 3' end of the original 320-base pair clone and sequences shared between Amb a I.2 and Amb a I.3 at the 5' end. Three possible explanations for this enigma are currently being considered.

First, if it is the most actively transcribed and/or translated message, it could have a high turnover rate in vivo or be particularly prone to degradation during storage or processing of the pollen. Second, it is possible that its structure could somehow interfere uniquely with cDNA synthesis and cloning efficiency, and its paucity in the libraries merely represents a technical difficulty. Third, the Amb a I.1 protein could in fact be a minor Amb a I variant but be selectively enriched for during the purification of Amb a I from pollen extract.

The amino acid composition of the cloned proteins corresponds well to experimental results reported previously (4). The presence of cysteine as the carboxyl-terminal deduced amino acid in all cDNA clones studied is in contrast to a report in which carboxypeptidase digestion of chromatographically purified Amb a I showed that leucine was the carboxyl-terminal amino acid (6). To settle the question of the carboxyl-terminal residue, the aspartic acid-proline peptide bond (position 360-361, Fig. 2) of chromatographically purified Amb a I was cleaved by 70% formic acid. The peptide sequence starting with proline was obtained in the digestion mixture whereas the nonspecific partially cleaved protein background sequence was suppressed by o-phthalaldehyde treatment. The 38-residue peptide sequence obtained corresponded completely with the deduced sequence of Amb a I.1 up to and including the terminal cysteine.

Since the amino terminus of Amb a I is "blocked" (6) and not amenable to direct Edman degradation, it is important to be able to conclude that the clones reported here contain the full coding sequence for Amb a I. Two lines of evidence support that conclusion. The experimental evidence comes from a
previous study (44) in which primer extension was used to add nucleotides to the 3′ end of an oligonucleotide designed from a peptide sequence using flower mRNA as template. The 5′ end of the oligonucleotide corresponds to nucleotide 249 in clone Amb a 1.3 and produced a cDNA fragment about 250 nucleotides long, suggesting that very little 5′-untranslated sequence is missing from the full-length message. The second piece of evidence that these are full-length clones coding for Amb a 1 comes from a study of the area around the presumed translation initiation codon AUG (45, 46). From a study of 211 eucaryotic messages, a consensus sequence has been identified at the initiation site. The sequence contains a purine (preferably an A) at position −3, a G at position +4, and a predominance of C at positions −1, −2, −4, and −5. Clone Amb a 1.3 has A at −3, G at +4, and C at −2.

There is an obvious hydrophobic stretch of amino acids following the initiating methionine, and this area has been examined to predict the cleavage site between the hydrophobic leader and the secreted protein. The prediction is based upon data collected for 450 secreted eucaryotic proteins (47) with known cleavage sites. Positions −1 and −3 are most critical for signal peptide cleavage, and the alanine at position 26 in Amb a 1 best fulfills the criteria associated with the cleavage site. For instance, if alanine 26 is indeed residue +1, then the −1 and −3 positions are serine and valine, respectively. The frequency of those amino acids in those positions in the data base is two to three the expected values if random amino acids occupied those positions. All amino acids, from −1 through −13, are consistent with the tabulations from the 450 known cleavage sites. Current efforts are under way to address directly the question of how and where the Amb a 1 proteins are blocked.

At the 3′ end, the untranslated region does not have the AATAAA polyadenylation signal observed in most animal sequences located 9–23 bases upstream from the poly(A) tail. However, there is a variant sequence, AATGAA, located 45 bases from the poly(A) addition site, which might serve as the adenylation signal; and another variation of this sequence, AAAAAA, is located 99 bases downstream from the TAA stop codon. A similar sequence, AATAAT is located 101 nucleotides upstream from the poly(A) addition signal. Another pollen-extracted gene, alcohol dehydrogenase, has the AATTAT sequence centered 44 bases upstream from its polyadenylation site.

The data presented in this paper firmly establish that Amb a 1 (antigen E) is a family of closely related proteins. Experiments are currently in progress to address the relative abundance of each family member in pollen, as mentioned above. Just as importantly, it is necessary to determine the immunogenicity and allergenicity of the Amb a 1 family members. It has been possible to express prototypic clones for each family member in proaryctic cells. This material is being used to compare T cell proliferation and IgE binding from individual ragweed allergic patients. This information can potentially be used to devise an effective course of ragweed immunotherapy.

Acknowledgments—We wish to thank Drs. Joseph Olson, James Smith, and Robert Esch for providing reagents; Dr. David Marsh for providing goat anti-Amb a 1 IgG; Tom Wells, Annellise Nault, Robert Schreifels, Andrew Brauer, and Joanne Pollock for technical assistance; and Joyce Bradshaw for preparing the manuscript.

REFERENCES

Cloning of Amb a I

34. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter, W. J. (1979) Biochemistry 18, 5294-5299
42. Knox, R. B. (1979) in Pollen and Allergy, pp. 56-57, University Park Press, Baltimore