

31374

Sequence Polymorphism of *Amb a I* and *Amb a II*, the Major Allergens in *Ambrosia artemisiifolia* (Short Ragweed)

I.J. Griffith^a, J. Pollock^a, D.G. Klapper^b, B.L. Rogers^a, A.K. Nault^a

^a ImmuLogic Pharmaceutical Corporation, Cambridge, Mass.; ^b Department of Microbiology and Immunology, University of North Carolina School of Medicine, Chapel Hill, N.C., USA

Abstract. Two of the major allergens in the pollen of short ragweed are *Amb a I* and *Amb a II* (formerly antigen E and antigen K, respectively). The genes for *Amb a I* and *Amb a II* have recently been cloned, and it was shown that *Amb a I* is a family of proteins with at least three distinct polymorphic family members. This study addresses the number of individual *Amb a I* and *Amb a II* family members, the polymorphism in each family member and the expression of these genes in ragweed plants from different geographical locations. This work led to the cloning and characterization of a fourth *Amb a I* family member, designated *Amb a I.4*.

Introduction

Short ragweed (*Ambrosia artemisiifolia*) is perhaps the major clinical cause of seasonal pollen allergy in North America. The two major allergens in the pollen of short ragweed are *Amb a I* (formerly antigen E) and *Amb a II* (formerly antigen K). It has been shown that at least 95% of ragweed-allergic patients have IgE that binds to *Amb a I* [1, 2], making *Amb a I* the major allergen in short ragweed pollen. Similarly, 50–90% of ragweed-allergic patients have IgE that binds to *Amb a II* [1, 2].

These allergens were first separated using biochemical purification methods [3–5] and shown to be nonglycosylated acidic proteins with molecular weights of approximately 38,000 [2]. A partial sequence of *Amb a I* was obtained by protein sequence analysis and used, in conjunction with antibodies raised against purified and denatured *Amb a I*, to identify a series of cDNA clones encoding a family of related proteins [6]. The three identified *Amb a I* family members, designated *Amb a I.1*, *Amb a I.2* and *Amb a I.3* in accordance with the nomenclature recommended by the International Union of Immunological Societies [7], have more than 80% amino acid homology with each other [6]. Recombinant proteins

expressed from each of these cloned cDNAs can bind IgE from pooled sera of allergic patients, indicating that all *Amb a I* family members are capable of eliciting disease symptoms in ragweed-allergic individuals [8]. The cDNA encoding *Amb a II* has also been recently cloned [9]. *Amb a II* has between 70 and 80% homology with the individual *Amb a I* family members.

Several pollen proteins, including *Amb a V* and *Amb a VI* from ragweed [10, 11], *Poa p I* [12] and *Poa p IX* from Kentucky bluegrass [13] and *Lol p I* from rye grass [14–16], have been shown to be polymorphic by direct protein sequence analysis [10–12, 16] or comparison of predicted protein sequence of isolated cDNA clones [13–15]. Comparison of several independent *Amb a I.3* clones showed that this *Amb a I* family member is also polymorphic [B.L.R. and D.G.K., unpubl. data]. This is consistent with the observation that multiple *Amb a I* and *Amb a II* protein spots can be detected on two-dimension Western blots using *Amb-a-I*- or *Amb-a-II*-specific reagents [8, 9]. The number of *Amb a I* and *Amb a II* genes and the extent of the polymorphism for each gene have not been defined.

This study was undertaken to determine whether there are additional *Amb a I* or *Amb a II* family mem-

Table 1. Oligonucleotides used in PCR and sequencing

Oligomer	Nucleotide sequence	Strand	Amino acids
RW32	5'- <u>GGGCTGCAGT</u> CATTATAAGTGCTTAGT	noncoding	3'-UT
RW33	5'- <u>GGGAATTC</u> GACTGTGCCCAAGGTTTGG	coding	DCAQGF 71→76
RW34	5'- <u>AGATCTGCAGT</u> ACCACCGATGGCGTAC	noncoding	YAIGGS 297←302
RW38	5'- <u>GGGAATTC</u> TGTATTTTACCTTAGC	coding	LYFTL 10→14
RW40	5'- <u>GGGCTGCAGT</u> CATAGCAAAAGCTTAGG	noncoding	3'-UT
RW41	5'-GATGATGATGTTGCAAAT	coding	DDDVAN 96→101
RW42	5'-ATTAAGTCCAACGATGGT	coding	IKSNDG 181→186
RW43	5'-CAAGTCGTTAACAACAAC	coding	QVVNNN 285→290
RW44	5'-ATTTGCAACATCATCATC	noncoding	DDDVAN 96←101
RW45	5'-ACCATCGTTGGACTTAAT	noncoding	IKSNDG 181←186
RW46	5'-GTTGTTGTTAACGACTTG	noncoding	QIVNNN 285←290
RW48	5'- <u>GGGAATTC</u> CAAAAGAAGGCACACT	coding	PKEGTL 102→107
RW49	5'- <u>GGGCTGCAGT</u> GCTTAGCAAGGTGCTCC	noncoding	GAPC 394←3'-UT
RW50	5'- <u>GGGAATTC</u> GACAGAGAGTTGGCTA	coding	DRELA 130→134
RW51	5'-CAAATCGTTAACAACCTTC	coding	QIVNNF 285→290
AGK2	5'- <u>GGGCTGCAGC</u> GTGTCCAAATCTAATCAAATGAACACTTATGG	noncoding	3'-UT <i>Amb a</i> II
AL	5'-p-AATGATCGATGCT	-	-
AP	5'-GGGTCTAGAGGTACCGTCCG	-	-
AT	5'-GGGTCTAGAGGTACCGTCCGATCGATCATT	-	-

The oligomers match coding or noncoding strand sequences as noted. RW32 is *Amb a* I.1 specific, RW40 is *Amb a* I.3 specific, RW50 is *Amb a* I.4 specific and RW51 and AGK2 are *Amb a* II specific. AP, AL and AT do not correspond to *Amb a* I or II sequences. Amino acid numbering corresponds to that used in figure 2. Nucleotide sequences added for cloning purposes are underlined.

bers and to define the extent of polymorphism for each *Amb a* I and *Amb a* II gene. Questions concerning the distribution of gene expression in individual plants and in different geographic locations were also addressed. RNA samples prepared from the flowers of individual plants, from flowers of small pools of plants and from pooled pollen were examined for the presence of *Amb a* I and *Amb a* II mRNA using the polymerase chain reaction (PCR). The presence of mRNA for a given *Amb a* I or *Amb a* II gene, and the polymorphism of that gene (and hence protein polymorphism), was determined by the direct sequence analysis of PCR-generated cDNA clones.

Materials and Methods

Plant Tissue

Flowers of short ragweed were picked, frozen immediately in liquid nitrogen, and stored at -70 °C until processed. Flowers from individual plants were either stored separately or mixed into pools of 3–5 plants each. Defatted short ragweed pollen from *A. artemisiifolia* was obtained from Greer Laboratories (Lenoir, N.C., USA).

Preparation of Ragweed RNA from Pollen and Whole Flowers

RNA was isolated from defatted ragweed pollen or whole flowers as previously described [6]. Poly A⁺ RNA was isolated after binding to oligo-thymidylic acid (oligo dT)-cellulose (Collaborative Research, Inc.) [17].

Amplification of cDNA by the PCR

Priming oligonucleotides for DNA amplification by PCR [18] were RW32, RW33, RW34, RW38, RW40, RW45, RW48, RW49, RW50, AGK2 and anchor primer (AP). These oligonucleotides are described in table 1. RW32 is *Amb a* I.1 specific, RW40 is *Amb a* I.3 specific, RW50 is *Amb a* I.4 specific and AGK2 is *Amb a* II specific. All other RW primers are shared between *Amb a* II and all *Amb a* I family members. RW49 corresponds to the noncoding strand sequence complementary to amino acids 394 through the TAA stop codon and the first 4 nucleotides of the 3' untranslated region of *Amb a* I. All RW oligomers corresponding to the coding strand sequence contained an *Eco* RI restriction site, while all oligomers corresponding to the noncoding strand sequence, except for RW45, contained a *Pst* I restriction site. Oligomers RW32, RW38, RW45, anchor linker (AL), anchor template (AT) and AP have been previously described [6]. AGK2 has also been described [9]. All oligomers were purchased from Research Genetics (Huntsville, Ala., USA).

First-strand cDNA was synthesized from 1 µg of total RNA with the cDNA Synthesis System Plus Kit (Amersham) using oligo dT as a primer [19]. The single-strand DNA (20 µl) was mixed with 100 pmol of each priming oligomer and amplified using the Gene

Table 2. Comparison of identity and between different *Amb a I* and *Amb a II* members: homology matrix

	I.1	I.2	I.3	I.4	II
I.1	—	83.4	83.8	88.5	73.3
I.2	76.8	—	90.7	82.7	78.6
I.3	77.0	86.9	—	81.2	78.6
I.4	82.4	74.5	73.0	—	72.0
II	65.7	69.5	70.2	61.4	—

Comparisons were made between the amino acid sequences reported for *Amb a I.1*, *I.2* and *I.3* reported in Rafnar et al. [6], *Amb a I.4*, described here, and *Amb a II* reported in Rogers et al. [9]. The lower left half of the table gives percent identity, and the upper right half gives percent homology for each pair of sequences. For example, the sequence of *Amb a I.4* has 82.4% identity and 88.5% homology with the sequence of *Amb a I.1*. Comparison of a sequence with itself would give 100% identity. Similar amino acids used in the calculation of homology; A, S, T; D, E; N, Q; R, K; I, L, M, V; and F, Y, W.

Amp kit (US Biochemicals) as previously described [6]. The first 5 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 1–4 min. The final 20 rounds of amplification consisted of denaturation as above, annealing at 55°C for 1.5 min and elongation as above. Amplified DNA was recovered by sequential chloroform, phenol and chloroform extractions followed by overnight precipitation at 4°C with 0.5 vol of 7.5 M ammonium acetate and 1.5 vol of isopropanol. DNA was simultaneously digested with *Eco* RI and *Pst* I and electrophoresed on a 1% Low Melt agarose (FMC, Rockland, Me., USA) preparative gel. The predicted size DNA band was isolated and ligated into M13 for sequencing [20].

The amino-terminal sequences of *Amb a I* family members were determined using a modification of the anchored PCR [21, 22] as previously described [6]. Briefly, double-stranded cDNA was synthesized from 1 µg RNA with the cDNA System Plus Kit using oligo dT as a primer [19], blunt ended with T4 polymerase and blunt end ligated to self-annealed AT and AL oligonucleotides using standard techniques [23]. The cDNA (3 µl from a 20-µl reaction) was mixed with 100 pmol each of AP and RW49 primers and amplified as described. 0.01 vol of the primary PCR was then reamplified with oligomers AP and RW45. RW45 is nested (internal) relative to oligomer RW49 used in the primary PCR. DNA from the secondary PCR was recovered as above, digested with *Kpn* I and ligated into *Kpn*-I/*Hinc*-II-digested M13 for sequencing.

Sequence Analysis of *Amb a I* and *Amb a II*

M13 clones were sequenced (Sequenase kit, US Biochemicals) using the dideoxy chain termination method of Sanger et al. [20]. Clones were sequenced with standard forward and reverse primers. RW41–46 and RW51, purchased from Research Genetics for

internal sequencing, are described in table 1. RW41, RW42 and RW44–46 primed *Amb a I* and *Amb a II* clones equally well. RW43 primed all *Amb a I* clones, while the corresponding primer, RW51, was specific for *Amb a II*. Computer programs contained in PCGENE (IntelliGenetics, Inc., Mountain View, Calif., USA) were used for both protein and nucleotide sequence analysis.

Results

Identification of a Novel *Amb a I* Family Member, *Amb a I.4*

Analysis of cDNA clones generated by PCR amplification from ragweed RNA revealed a novel *Amb a I* sequence, designated *Amb a I.4*, that was more closely related to *Amb a I.1* than to any other family member (table 2). This fourth *Amb a I* family member was almost as closely related to *Amb a I* (82.4% identity and 88.5% homology) as *Amb a I.2* was to *Amb a I.3* (86.9% identity and 90.7% homology). The complete nucleotide and predicted amino acid sequence of *Amb a I.4* are shown in figure 1. Originally detected in RNA from the flowers of a single ragweed plant collected in North Carolina, *Amb a I.4* mRNA was detected in the flowers of plants from Massachusetts and in pollen pooled from plants over a wide geographic area (table 3).

Amb a I.4 has not been detected by amino acid sequence analysis of purified *Amb a I* or *Amb a II* preparations. Furthermore, the *Amb a I.4* message appears to be only a minor component (estimated to be 3% based on the number of *Amb a I.4* clones isolated) of the total *Amb a I* family in pollen and pools of flowers. The inability to detect *Amb a I.4* by protein sequencing would be consistent with its apparent amount of mRNA.

IgE from the sera of ragweed-allergic patients binds on a Western blot to recombinant *Amb a I.4* expressed in *Escherichia coli* [Bond et al., unpubl. data]. In conjunction with results obtained by Bond et al. [8] and Rogers et al. [unpubl. data], the binding of IgE to recombinant *Amb a I.4* suggests that all *Amb a I* and *Amb a II* family members are capable of eliciting disease symptoms in ragweed-allergic individuals.

Polymorphism of *Amb a I* and *Amb a II*

The complete or partial nucleotide sequence was determined for clones generated in PCR amplifications with primers AP/RW45, RW33/49, RW33/32, RW33/40 and RW50/34. The amino acid polymorphisms predicted from this sequence analysis of 66

ATGGGGATCAAACTGTTGTTACATCTTGTATTTTACCTTAGCCCTTGTCACCTTGCTGCAACCTGTTTCGTTCT	75
M G I K H C C Y I L Y F T L A L V T L L Q P V R S	25
GCCGAAGATCTCCAGCAAATCTTACCTTCAGCTAACGAAACAAGGAGCCTGACAACATGTGGAACATACAACATT	150
TA E D L Q Q I L P S A <u>N E T</u> R S L T T C G T Y N I	50
ATAGACGGGTGCTGGAGGGGCAAAGCCGATTGGGCGGAAAACCGAAAAGCGTTAGCCGATTGTGCCCAAGGTTTT	225
I D G C W R G K A D W A E N R K A L A D C A Q G F	75
GCAAAGGGAACAATCGGCGGAAAAGATGGTGATATATACACGGTCACCACTGAGCTAGATGATGATGTTGCAAAT	300
A K G T I G G K D G D I Y T V T S E L D D D V A N	100
CCAAAAGAAGGCACACTCCGGTTTGGTGCCGCCAAAACAGGCCCTTGTGGATTATTTTGAAGAGATATGGTG	375
P K E G T L R F G A A Q N R P L W I I F A R D M V	125
ATTCGTTTGGATAGAGAGTTGGCTATAAACAACGACAAGACCATCGATGGCCGAGGGGCGAAAAGTTGAAATCATT	450
I R L D R E L A I N N D K T I D G R G A K V E I I	150
AACGCTGGTTTCGCCATCTATAATGTCAAGATATAATCATTACATAAATTATGCATGATATTGTAGTGAAT	525
N A G F A I Y N V K N I I I H N I I M H D I V V N	175
CCAGGAGGCTGATTAAAGTCCCACGATGGTCCACCACTTCCAAGAAAGGGTAGTGATGGTGATGCTATAGGTATT	600
P G G L I K S H D G P P V P R K G S D G D A I G I	200
TCTGGTGGTTACAAATATGGATCGACCATTCCTCCCTCAGTAAGGCTGTTGATGGGCTAATCGATGCTAAACAC	675
S G G S Q I W I D H C S L S K A V D G L I D A K H	225
GGCAGCACACACTTCACCGTTTCTAACTGCTTATTCACCCAACCAATATTTATTATTGTTCTGGGATTTTGAC	750
G S T H F T V S N C L F T Q H Q Y L L L F W D F D	250
GAGCGAGGCATGCTATGTACGGTCGCATTCAACAAGTTCAGTGATAACGTTGACCAAAGATGCCTAACTTACGA	825
E R G M L C T V A F N K F T D N V D Q R M P N L R	275
CATGGGTTTGTCCAAGTCGTTAACAACAACACTACGAAAGATGGGGATCGTACGCCCTCGGTGGTAGCGAGGCCCA	900
H G F V Q V V N N N Y E R W G S Y A L G G S A G P	300
ACCATACTTAGCCAAGGGAACAGATTCTTAGCCTCCGATATCAAGAAAGAGGTCGTAGGGAGGTATGGTGAATCC	975
T I L S Q G N R F L A S D I K K E V V G R Y G E S	325
GCCATGTGAGAGTCGATTAATTGGAAGTGGAGATCGTATATGGACGTATTTGAAAATGGTGCTATTTTGTTC	1050
A M S E S I N W N W R S Y M D V F E N G A I F V P	350
TCCGGGGTTGATCCAGTGCTAACCCCTGAGCAAAACGCGGGATGATTCCAGCCGAACCAGGAGAAGCCGTTCTA	1125
S G V D P V L T P E Q N A G M I P A E P G E A V L	375
AGACTCACTAGTAGTGCTGGTGTCTCTCATGCCAACCTGGAGCACCTTGCTAAGCA	1183
R L T S S A G V L S C Q P G A P C -	392

Fig. 1. Nucleotide and predicted amino acid sequences for *Amb a* I.4. A potential cleavage site between the hydrophobic leader and mature protein is indicated by a vertical arrow. Underlined amino acids correspond to the potential N-linked glycosylation site. Underlined nucleotides are complementary to the oligomer used in cloning the carboxy terminus of *Amb a* I.4

Amb a I.1, 28 *Amb a* I.2, 22 *Amb a* I.3, 28 *Amb a* I.4 and 10 *Amb a* II clones are shown in figure 2. Due to the inherent error rate of Taq polymerase [18], only those polymorphisms detected in at least two independent amplification reactions are shown. The most prevalent amino acid at a given polymorphic site is depicted in superscript, while the minor amino acid is depicted in subscript.

There were 8 polymorphic sites detected in *Amb a* I and 3 polymorphic sites detected in *Amb a* II by this PCR-dependent analysis. All polymorphisms appeared to occur independently of each other, except for amino acids 183–189 of *Amb a* I.4 where 3 polymorphisms are inherited as a block. This indicates that there are at least 2 isoforms of *Amb a* I.1 (a single polymorphic site), 4 isoforms of *Amb a* I.2 (2 poly-

Table 3. Presence of mRNA for *Amb a I* and *Amb a II*

RNA source	Specific <i>Amb a I</i> and <i>Amb a II</i> member				
	I.1	I.2	I.3	I.4	II
Pollen (pool)	+	+	+	[+]	-
Flower (pool-NC)	+	+	+	[+]	+
Flower (pool-MA)	+	+	+	+	+
Flower (NC)	+	+	-	+	+
Flower (MA)	+	+	+	?	?

RNA was prepared from pooled pollen (Greer Laboratories), pools of flowers collected from a small number of plants in North Carolina (pool-NC) or Massachusetts (pool-MA) and the flowers of single plants collected in North Carolina (NC) or Massachusetts (MA).

+ = The presence of message was detected by amplification with conserved primers; [+] = the presence of message was detected only after member-specific amplification; - = the absence of message was confirmed in multiple member-specific amplifications on different first-strand syntheses; ? = messages for *Amb a I.4* and *Amb a II* were not detected in the small number of clones examined; member-specific amplifications were not done with this RNA preparation, and the presence of *Amb a I.4* and *Amb a II* messages in the flowers from the Massachusetts plant remains unknown.

morphic sites), 2 isoforms of *Amb a I.3* (a single polymorphic site), 4 isoforms of *Amb a I.4* (2 independent polymorphic sites) and 8 isoforms of *Amb a II* (3 polymorphic sites). All *Amb a I* isoforms were detected in both pollen and flower RNA (data not shown). As indicated in table 3, *Amb a II* isoforms were only detected in flower RNA. It should also be noted that all isoforms of a given *Amb a I* or *Amb a II* family member had greater than 99% identity with each other, while different family members shared 70–90% homology (table 2).

Expression of mRNA for Individual Amb a I and Amb a II Family Members in Pools of Pollen and Flowers

Polymorphism and presence of each individual family member was determined in RNA prepared from the flowers of small pools of plants and from pollen pooled from a large number of plants over a wide geographic area. Although initial primers were chosen because of sequence conservation between *Amb a I.1*, *Amb a I.2* and *Amb a I.3*, cDNAs of *Amb a I.4* and *Amb a II* were also amplified. The results, shown in table 3, indicate that only mRNAs for *Amb a*

I.1, *Amb a I.2* and *Amb a I.3* were detected in pollen RNA using conserved primers. *Amb a I.4* and *Amb a II* cDNAs were not observed among the pollen-derived clones examined. In contrast, mRNA for *Amb a I.1*, *Amb a I.2*, *Amb a I.3*, *Amb a I.4* and *Amb a II* were detected in flower RNA. It was subsequently shown in PCR reactions employing family-member-specific primers (RW50/34 and RW33/AGK2) that *Amb a I.4* mRNA could be detected in pollen but that *Amb a II* mRNA could not. This indicates that the *Amb a I.4* mRNA is only a minor component of pollen RNA, while *Amb a II* mRNA is either totally absent from pollen or present at such low levels that it is not detectable by PCR (table 3).

Expression of mRNA for Individual Amb a I and Amb a II Family Members in the Flowers of Single Plants

RNA from the flowers of individual ragweed plants collected in North Carolina and Massachusetts were also examined for expression of *Amb a I* and *Amb a II* family member mRNAs (table 3). Messages for *Amb a I.1*, *Amb a I.2*, *Amb a I.4* and *Amb a II* were detected in 1 North Carolina plant examined. However, it was not possible to detect mRNA for *Amb a I.3* in this plant, even when using primers RW33/40 in a family-member-specific amplification (table 3). Since mRNA for *Amb a I.3* was detected in flower RNA from pools of 3–5 plants collected either in North Carolina or in Massachusetts (table 3), this suggests that *Amb a I.3* was not expressed in the individual North Carolina plant examined. Significantly, 7 different mRNA transcripts were present in flowers of the individual North Carolina plant. These transcripts encoded 1 isoform

Fig. 2. Composite *Amb a I* and *Amb a II* sequences relative to *Amb a I.1*. The sequence of *Amb a I.1* is given in the standard one-letter code. Sequences for the other *Amb a I* family members are given relative to *Amb a I.1* with only differences being shown. A dash indicates identity, and an asterisk indicates a break introduced to maintain maximal alignment. Amino acid numbering is based on the *Amb a I.2* sequence, the longest *Amb a I* family member. Wherever sequence polymorphism has been observed in a given family member, the dominant sequence is given in superscript, and the minor sequence is given in subscript at the site of polymorphism. Polymorphisms in a given family member occur as independent events, except for amino acids 183–189 of *Amb a I.4* where 3 amino acid differences occur as a block. The bracketed region in *Amb a I.4* indicates the oligomer sequence used in cloning that has not been independently verified.

10										20										30										40										50												
M	G	I	K	H	C	C	Y	I	L	Y	F	T	L	A	L	V	T	L	L	Q	P	V	R	S	A	E	D	L	Q	E	I	L	P	*	V	N	E	T	R	R	*	L	T	T	S	G	A	Y	N	I.1		
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	I.2		
-	-	-	-	Q	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	G	V	G	-	-	-	-	S	-	-	-	-	*	S	-	Q	A	C	E	-	L	-	I.3	
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Q	-	-	-	S	A	-	-	-	*	S	-	-	-	C	-	T	-	-	I.4			
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V	-	A	G	-	L	G	-	-	-	-	E	V	*	D	-	-	-	S	P	-	D	-	-	-	S	-	Q	G	C	E	-	H	-	II
60										70										80										90										100												
I	I	D	G	C	W	R	G	K	A	D	W	A	E	N	R	K	A	L	A	D	C	A	Q	G	F	G	K	G	T	V	G	G	K	D	G	D	I	Y	T	V	T	S	E	D	L	D	D	D	V	A	I.1	
-	-	-	K	-	-	-	C	-	-	-	-	-	-	N	-	-	Q	-	-	-	-	-	-	-	-	-	A	-	-	Y	-	-	-	H	-	-	V	-	-	-	-	-	D	K	-	-	-	-	-	I.2		
-	-	-	K	-	-	-	-	-	-	-	-	-	-	E	N	-	Q	-	-	-	-	-	-	-	-	-	A	-	-	Y	-	-	-	W	-	-	V	-	-	-	-	-	-	N	-	-	-	-	-	I.3		
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	I.4			
-	-	-	K	-	-	-	C	-	P	-	-	-	-	-	-	Q	-	G	N	-	-	-	-	-	-	-	-	A	-	H	-	-	-	W	-	-	-	M	-	-	-	D	Q	-	-	-	-	V	-	II		
110										120										130										140										150												
N	P	K	E	G	T	L	R	F	G	A	A	Q	N	R	P	L	W	I	I	F	E	R	D	M	V	I	R	L	D	K	E	M	V	V	N	S	D	K	T	I	D	G	R	G	A	K	V	E	I	I.1		
-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	K	-	N	-	-	-	-	H	-	N	Q	-	L	-	-	-	-	-	-	-	-	-	-	-	V	-	-	N	-	-	I.2		
-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	K	N	-	-	-	-	N	-	N	Q	-	L	-	-	-	-	-	-	-	-	-	-	-	-	V	-	-	-	-	I.3		
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	I.4		
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	R	-	L	A	I	-	N	-	-	-	-	-	-	-	-	-	-	-	-	-	II		
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	Q	-	-	-	-	-	I	-	Y	-	Q	Q	-	-	-	-	T	-	-	T	-	-	-	-	-	-	-	-	-	L	-	II
160										170										180										190										200												
I	N	A	G	F	T	L	N	G	V	K	N	V	I	I	H	N	I	N	M	H	D	V	K	V	N	P	G	G	L	I	K	S	N	D	G	P	A	A	P	R	A	G	S	D	G	D	A	I	S	I.1		
-	-	-	-	L	-	-	M	N	-	-	-	I	-	-	-	-	-	I	-	-	I	-	-	-	-	-	-	-	M	-	-	-	-	-	-	-	P	I	L	-	Q	Q	-	-	-	-	-	N	-	-	I.2	
-	-	-	G	-	L	-	-	M	N	-	-	-	I	-	-	-	-	-	I	-	-	-	-	-	-	-	-	-	-	M	-	-	-	-	-	-	P	I	L	-	Q	A	-	-	-	-	T	-	N	-	I.3	
-	-	-	-	-	A	I	Y	N	-	-	-	-	I	-	-	-	-	I	-	-	-	I	V	-	-	-	-	-	-	-	-	-	-	-	-	C	-	-	-	P	-	-	-	-	-	-	-	G	-	I.4		
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	H	-	-	P	V	-	-	K	-	-	-	-	-	-	II		
V	Y	G	-	I	-	-	M	N	-	-	-	-	-	-	-	-	-	D	I	-	-	-	R	-	L	-	-	R	-	-	-	-	-	-	G	-	-	-	I	-	-	H	Q	-	-	-	-	-	H	-	II	
210										220										230										240										250												
I	S	G	S	S	Q	I	W	I	D	H	C	S	L	S	K	S	V	D	G	L	V	D	A	K	L	G	T	T	R	L	T	V	S	N	S	L	F	T	Q	H	Q	F	V	L	L	F	G	A	G	I.1		
V	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	A	S	-	-	-	L	-	I	T	-	-	S	S	H	V	-	-	-	-	-	C	K	-	-	-	-	-	-	-	-	L	-	-	D	I.2	
V	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	F	-	-	-	-	V	T	-	-	S	-	H	V	-	I	-	-	-	-	C	K	-	-	-	Q	S	K	A	I	-	L	-	-	D	I.3
-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	A	-	-	-	-	I	-	-	-	H	-	S	-	H	F	T	-	-	-	C	-	-	-	-	-	Y	L	-	-	-	W	D	F	-	I.4	
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	V	N	W	-	S	-	G	V	-	I	-	-	-	C	K	-	-	H	-	E	K	A	V	-	L	-	-	S	II	
260										270										280										290										300												
D	E	N	I	E	D	R	G	M	L	A	T	V	A	F	N	T	F	T	D	N	V	D	Q	R	M	P	R	C	R	H	G	F	F	Q	V	V	N	N	N	Y	D	K	W	G	S	Y	A	I	G	I.1		
-	T	H	Y	Q	-	K	-	-	-	-	-	-	-	-	-	-	M	-	-	H	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	I.2		
-	T	H	V	Q	-	K	-	-	-	-	-	-	-	-	-	-	M	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	I.3		
-	-	*	*	*	*	-	-	-	-	C	-	-	-	-	-	-	K	-	-	-	-	-	-	-	-	-	-	N	L	-	-	-	-	V	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	I.4	
-	T	H	F	Q	-	L	K	-	H	V	-	L	-	Y	-	I	-	-	N	T	-	H	E	-	-	-	-	-	-	-	-	-	-	F	-	-	-	I	-	-	-	F	-	-	-	-	-	-	-	-	II	
310										320										330										340										350												
G	S	A	S	P	T	I	L	S	Q	G	N	R	F	C	A	P	D	E	R	S	K	K	N	V	L	G	R	H	G	E	*	A	A	A	E	S	M	K	W	N	W	R	T	N	K	D	V	L	E	N	I.1	
-	-	S	A	-	-	-	-	-	-	-	-	-	-	-	-	-	F	-	-	-	D	I	I	-	-	-	-	A	-	T	-	T	*	G	N	-	-	-	S	-	-	-	-	-	-	-	-	-	-	-	-	I.2
-	-	S	A	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	I.3	
-	-	-	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	I.4	
-	-	S	N	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	II	
360										370										380										390																						
G	A	I	F	V	A	S	G	V	D	P	V	L	T	P	E	Q	S	A	G	M	I	P	A	E	P	G	E	S	A	L	S	L	T	S	S	A	G	V	L	S	C	Q	P	G	A	P	C	I.1				
-	-	-	-	-	L	P	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	I.2	
-	-	-	-	-	T	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	I.3	
-	-	-	-	-	P	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	I.4	
-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	II	

of *Amb a* I.1, 2 isoforms of *Amb a* I.2, 3 isoforms of *Amb a* I.4 and 1 isoform of *Amb a* II.

Analysis of RNA obtained from the flowers of an individual Massachusetts plant showed the presence of mRNA for *Amb a* I.1, *Amb a* I.2 and *Amb a* I.3 (table 3). The inability to detect *Amb a* I.4 and *Amb a* II in this plant may reflect the relatively small number of clones examined rather than a lack of expression. Family-member-specific amplifications of *Amb a* I.4 or *Amb a* II were not attempted from this plant. Nevertheless, *Amb a* I.4 and *Amb a* II mRNAs were detected in the RNA of flowers pooled from 3–5 plants collected nearby, showing that *Amb a* I.4 and *Amb a* II are expressed in Massachusetts.

Discussion

PCR analysis of flower mRNA led to the identification of a new member of the *Amb a* I family, *Amb a* I.4 (fig. 1), which is highly homologous to *Amb a* I.1 (81% identity, 88% homology, table 2). The N-terminus for mature *Amb a* I is blocked [6] and has not been experimentally determined. A predicted cleavage site, VRS↓A at amino acid 26, was identified in *Amb a* I [6] based upon empirical observation of a large number of defined cleavage sites [24]. This predicted cleavage site in *Amb a* I.1, I.2 and I.3 is conserved in *Amb a* I.4 (fig. 1, 2). The single potential N-linked glycosylation site N-X-T present in *Amb a* I and *Amb a* II at amino acids 37–39 has also been conserved. However, this potential glycosylation site may not be utilized, since *Amb a* I has been reported to be nonglycosylated [2].

Amb a I.4 clones represented approximately 3% of the total *Amb a* I cDNA clones amplified from pools of flower or pollen RNA. Assuming that reverse transcription and PCR amplification do not distort the relative abundance of RNAs, this infers a low mRNA level for *Amb a* I.4 and suggests that its protein level would also be less than those of the other *Amb a* I family members. This is consistent with *Amb a* I.4 not being identified in purified pollen protein by sequence analysis. However, given the homology of *Amb a* I.4 with other *Amb a* I family members, it is not surprising that *Amb a* I.4 is an allergen and can bind IgE from the sera of allergic patients [Bond et al., unpubl. data].

Comparison of a large number of independently isolated clones derived from multiple RNA sources

identified 8 polymorphic sites in *Amb a* I (1 in *Amb a* I.1, 2 in *Amb a* I.2, 1 in *Amb a* I.3 and 4 in *Amb a* I.4) and 3 polymorphic sites in *Amb a* II. It should be stressed that a polymorphism had to be observed in at least two independent amplification reactions before it was considered authentic. This is due to the inherently high error rate (approx. 2.0×10^{-4}) [18] in Taq-polymerase-mediated DNA synthesis. Although nucleotide changes only observed in single amplifications may represent rare polymorphisms, one cannot exclude the possibility that they represent incorporation errors by the Taq polymerase. Only 11 nucleotide changes leading to productive amino acid changes were observed in single amplification reactions in the 61 kb of the sequence analyzed. If these 11 changes are indeed in vitro artifacts, this would give a Taq polymerase error rate of approximately 1.8×10^{-4} in our experiments, which is in close agreement with the reported error rate [18].

It should be noted that the polymorphisms given in figure 2 are not all inclusive and additional, less frequent amino acid alterations are expected to be found. For example, 3 polymorphisms were detected in independent λ gt10- and λ gt11-cloned sequences of *Amb a* I.3 that have not yet been observed in the PCR-generated sequences [Rafnar et al., unpubl. data]. Further, other *Amb a* I or *Amb a* II family members may remain undetected. However, these yet-to-be-discovered family members would be expressed at very low levels (<3% of the clones amplified from mRNA, the level at which *Amb a* I.4 can be detected). This suggests that the major *Amb a* I and *Amb a* II members have been identified.

Analysis of mRNA from the flowers of single plants demonstrated that each individual plant can express multiple forms of *Amb a* I and *Amb a* II (table 3). Indeed, 7 unique mRNA transcripts for *Amb a* I.1 (1 form), *Amb a* I.2 (2 forms), *Amb a* I.4 (3 forms) and *Amb a* II (1 form) were expressed in a North Carolina plant. This plant did not appear to have mRNA for *Amb a* I.3. These results clearly show that a single plant can express multiple *Amb a* I and *Amb a* II proteins. Further, deficiency in expression of an *Amb a* I or *Amb a* II gene seems confined to individual plants rather than to geographic regions (compare gene expression in single and small pools of plants from Massachusetts and North Carolina, table 3).

The number of different mRNAs that have been detected in single ragweed plants indicates that every plant has at least 1 gene for each *Amb a* I and *Amb a*

II family member that it expresses. Different isoforms of each family member most likely reflect alleles of that gene. The 3 isoforms of *Amb a I.4* expressed in the individual North Carolina plant would therefore imply that this plant has a second copy of the *Amb a I.4* gene. This would allow the expression of 4 alleles (3 detected) of *Amb a I.4*.

Message encoding *Amb a I* and *Amb a II* can be detected in flower RNA (table 3). This flower mRNA encodes the *Amb a I* and *Amb a II* proteins contained in pollen. What role *Amb a I* and *Amb a II* play in ragweed fertilization is not currently known. Both *Amb a I* and *Amb a II* are ~40% homologous to cDNAs Lat56/Lat59 and 9612 found in the male (anther) [25] and female (pistil) [26] sex organs of *Lycopersicon esculentum* (tomato plant), respectively. It has been speculated that these tomato proteins may serve as self-compatibility recognition molecules during pollination [25, 26]. Both the *Lycopersicon* and *Ambrosia* proteins are also homologous to pectate lyases [25], which are pectin-degrading enzymes of *Aspergillus* and the plant pathogen *Erwinia*. This would suggest that *Amb a I* and *Amb a II* might have enzymatic properties necessary for digestion of pectin in the plant wall during formation of the pollen tube in pollination. It is also interesting to speculate on the significance of the differential expression of *Amb a I* (present) and *Amb a II* (absent) mRNAs in pollen. This difference may suggest that *Amb a I* and *Amb a II* proteins play different roles in flower pollination, even though they are approximately 75% homologous.

This study has led to the identification of a fourth member of the *Amb a I* family as well as to the identification of a limited number of polymorphic residues in each *Amb a I* and *Amb a II* family member. Other polymorphisms or family members may exist, although at very low levels. Further, all *Amb a I* and *Amb a II* family members are expressed by plants in widely separated geographic areas. It may now be possible to use this information to develop an improved immunotherapy for ragweed allergy based upon known protein sequences.

Acknowledgments

The authors would like to thank Dr. T. Briner and Dr. J. Bond for the kind gift of RNA, and A. Brauer, Dr. T. Briner, Dr. J. Bond, Dr. M. Geftter, Dr. J. Greenstein and Dr. J. Morgenstern for their critical evaluation of this manuscript.

Note added in proof

Amb a I and *Amb a II* family members have been assigned GenBank accession numbers M80558 (*Amb a I.1*), M80559 (*Amb a I.2*), M80560 (*Amb a I.3*), M80561 (*Amb a II*) and M80562 (*Amb a I.4*).

References

- 1 Adolphson C, Goodfriend L, Gleich GJ: Reactivity of ragweed allergens with IgE antibodies. Analyses by leukocyte histamine release and the radioallergosorbent test and determination of cross-reactivity. *J Allergy Clin Immunol* 1978;62:197-210.
- 2 King TP: Chemical and biological properties of some atopic allergens. *Adv Immunol* 1976;23:77-105.
- 3 King TP, Norman PS, Connell JJ: Isolation and characterization of allergens from ragweed pollen. Part IV. *Biochemistry* 1964;3:458-468.
- 4 King TP, Norman PS, Lichtenstein LM: Isolation and characterization of allergens from ragweed pollen. *Biochemistry* 1967;6:1992-2000.
- 5 King TP: Separation of proteins by ammonium sulfate gradient solubilization. *Biochemistry* 1972;11:367-371.
- 6 Rafnar T, Griffith IJ, Kuo MC, Bond JF, Rogers BL, Klapper DG: Cloning of *Amb a I* (antigen E), the major allergen family of short ragweed pollen. *J Biol Chem* 1991;266:1229-1236.
- 7 Marsh DG, Goodfriend L, King TP, Lowenstein H, Platts-Mills TAE: Allergen nomenclature. *Int Arch Allergy Appl Immunol* 1988;85:194-200.
- 8 Bond JF, Garman RG, Keating KM, Briner TJ, Rafnar T, Klapper DG, Rogers BL: Multiple *Amb a I* allergens demonstrate specific reactivity with IgE and T cells from ragweed allergic patients. *J Immunol* 1991;146:3380-3385.
- 9 Rogers BL, Counsell C, King TP, Morgenstern JM, Griffith IG, Kuo MC: Complete sequence of the allergen *Amb a II*: Recombinant expression and reactivity with T cells from ragweed allergic patients. *J Immunol* 1991;147:2547-2552.
- 10 Roebber M, Hussain R, Klapper D, Marsh D: Isolation and properties of a short ragweed pollen allergen, Ra6. *J Immunol* 1983;131:706-711.
- 11 Mole LE, Goodfriend L, Lapkoff CB, Kehoe JM, Capra JD: The amino acid sequence of ragweed pollen allergen Ra5. *Biochemistry* 1975;14:1216-1220.
- 12 Nayak BN, Kisil FT: NH₂-terminal amino acid sequences of *Poa p Ia* and *Poa p Ib* allergens of Kentucky bluegrass pollen (abstract). *J Allergy Clin Immunol* 1991;87:325.
- 13 Silvanovich A, Astwood J, Zhang L, Olsen E, Kisil F, Sehon A, Mohapatra S, Hill R: Nucleotide sequence analysis of three cDNAs coding for *Poa p IX* isoallergens of Kentucky bluegrass pollen. *J Biol Chem* 1991;266:1204-1210.
- 14 Griffith IJ, Smith PA, Pollock J, Theerakulpisut P, Avjioglu A, Davies S, Hough T, Singh MB, Simpson RJ, Ward LD, Knox RB: Cloning and sequencing of *Lol p I*, the major allergenic protein of rye-grass pollen. *FEBS Lett* 1991;279:210-215.
- 15 Perez M, Ishioka GY, Walker LE, Chestnut RW: cDNA cloning and immunological characterization of the rye grass allergen *Lol p I*. *J Biol Chem* 1990;265:16210-16215.
- 16 Johnson P, Marsh DG: Allergens from common rye grass pol-

- len (*Lolium perenne*). II. The allergenic determinants and carbohydrate moiety. *Immunochemistry* 1966;3:101.
- 17 Aviv H, Leder P: Purification of biologically active globin messenger RNA by chromatography on oligo-thymidylic acid-cellulose. *Proc Natl Acad Sci USA* 1972;69:1408-1412.
- 18 Saiki RK, Gelfand DH, Stoffen S, Scharf SH, Higuchi R, Horn GT, Mullis KB, Ehrlich HA: Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 1988;239:487-491.
- 19 Gubler U, Hoffman BJ: A simple and very efficient method for generating cDNA libraries. *Gene* 1983;25:263-269.
- 20 Sanger F, Nicklen S, Coulson AR: DNA sequencing with chain-terminating inhibitors. *Proc Natl Acad Sci USA* 1977;74:5463.
- 21 Roux KH, Dhanarajan P: A strategy for single site PCR amplification of dsDNA: Priming digested cloned or genomic DNA from an anchor-modified restriction site and a short internal sequence. *BioTech* 1990;8:48-57.
- 22 Frohman MA, Dush MK, Martin GR: Rapid production of full-length cDNAs from rare transcripts: Amplification using a single gene-specific oligonucleotide primer. *Proc Natl Acad Sci USA* 1990;85:8998-9002.
- 23 Sambrook J, Fritsch EF, Maniatis T: *Molecular Cloning: A Laboratory Manual*, ed 2. Cold Spring Harbor, Cold Spring Harbor Press, 1989.
- 24 von Heijne G: How signal sequences maintain cleavage specificity. *J Mol Biol* 1984;173:243-251.
- 25 McCormack S: Sex and sneezing - Are they related? *Trends Genet*, in press.
- 26 Budelier KA, Smith AG, Gasser CS: Regulation of a stylar transmitting tissue-specific gene in wild-type and transgenic tomato and tobacco. *Mol Gen Genet* 1990;224:183-192.

Received: May 31, 1991

Accepted: July 31, 1991

Correspondence to: Dr. Irwin Griffith
ImmuLogic Pharmaceutical Corporation
One Kendall Square
Cambridge, MA 02139 (USA)