Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2049921 C 2004/06/15

(11)(21) 2 049 921

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(86) Date de dépôt PCT/PCT Filing Date: 1990/03/09
(87) Date publication PCT/PCT Publication Date: 1990/10/04
(45) Date de délivrance/Issue Date: 2004/06/15
(85) Entrée phase nationale/National Entry: 1991/09/09
(86) N° demande PCT/PCT Application No.: US 1990/001310
(87) N° publication PCT/PCT Publication No.: 1990/011293
(30) Priorité/Priority: 1989/03/17 (325,365) US

(51) Cl.Int.^{5}/Int.Cl.^{5} C12N 15/29, A61K 39/36, G01N 33/68, C07K 14/415, C07K 7/04, C07K 16/16

(72) Inventeurs/Inventors:
ROGERS, BRUCE, US;
KLAPPER, DAVID G., US;
RAFNAR, THORUNN, US;
KUO, MEI-CHANG, US

(73) Propriétaires/Owners:
UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL, US;
IMMULOGIC PHARMACEUTICAL CORPORATION, US

(74) Agent: SMART & BIGGAR

(54) Titre : PROTEINES ALLERGENES DE L'HERBE A POUX ET LEUR UTILISATION
(54) Title: ALLERGENIC PROTEINS FROM RAGWEED AND USES THEREFOR

(57) Abrégé/Abstract:
Antigen E or Amb a I of ragweed pollen has been shown to be a family or families of proteins. cDNAs encoding Amb a I, the major human allergen of ragweed and Amb a II, peptides derived from Amb a I or Amb a II, antibodies against the peptides; and methods of treating individuals for sensitivity to ragweed are disclosed.

2049921

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 :
C07K 7/06, 7/08, 13/00
A61K 37/02, G01N 33/68

A1

(11) International Publication Number: WO 90/11293

(43) International Publication Date: 4 October 1990 (04.10.90)

(21) International Application Number: PCT/US90/01310

(22) International Filing Date: 9 March 1990 (09.03.90)

(30) Priority data:
325,365 17 March 1989 (17.03.89) US

(71) Applicants: IMMULOGIC PHARMACEUTICAL CORPORATION [US/US]; One Kendall Square, Cambridge, MA 02139 (US). THE UNIVERSITY OF NORTH CAROLINA AT CHAPEL HILL [US/US]; Chapel Hill, NC 27599-4100 (US).

(72) Inventors: ROGERS, Bruce ; 64 Oxford Street, #4, Cambridge, MA 02138 (US). KLAPPER, David, G. ; 501 Sharon Road, Chapel Hill, NC 27514 (US). RAFNAR, Thorunn ; Bolinwood Drive, Chapel Hill, NC 27514 (US). KUO, Mei-Chang ; 5 Cox Road, Winchester, MA 01890 (US).

(74) Agents: GRANAHAN, Patricia et al.; Hamilton, Brook, Smith & Reynolds, Two Militia Drive, Lexington, MA 02173 (US).

(81) Designated States: AT (European patent), BE (European + patent), CA, CH (European patent), DE (European patent), DK (European patent), ES (European patent), FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent).

Published
With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: ALLERGENIC PROTEINS FROM RAGWEED AND USES THEREFOR

(57) Abstract

Antigen E or *Amb a* I of ragweed pollen has been shown to be a family or families of proteins. cDNAs encoding *Amb a* I, the major human allergen of ragweed and *Amb a* II, peptides derived from *Amb a* I or *Amb a* II, antibodies against the peptides; and methods of treating individuals for sensitivity to ragweed are disclosed.

2049921

ALLERGENIC PROTEINS FROM RAGWEED AND USES THEREFOR

Description

Background of the Invention

Genetically predisposed individuals, who make up about 10% of the population, become hypersensitized (allergic) to antigens from a variety of environmental sources to which they are exposed. Those antigens that can induce immediate and/or delayed types of hypersensitivity in people are called allergens. King, T. P., Adv. Immun., 23:77-105 (1976). Anaphylaxis or atopy, which includes the symptoms of hay fever, asthma and hives, is one form of immediate allergy. It can be caused by a variety of atopic allergens, such as products of grasses, trees, weeds, animal dander, insects, and food, drugs and chemicals.

The antibodies involved in atopic allergy belong primarily to the IgE class of immunoglobulins. IgE binds to mast cells and basophils. Upon combination of a specific allergen with IgE bound to mast cells, the IgE is cross-linked on the cell surface, resulting in the physiological effects of IgE-antigen interaction. Degranulation results in release of, among other substances, histamine, heparin, a chemotactic factor for eosinophilic leukocytes and the leukotrienes,

C4, D4 and E4, which cause prolonged constriction of bronchial smooth muscle cells. Hood, L.E. et al., Immunology, (2nd ed.), pp460-462, The Benjamin/Cumming Publishing Co., Inc. (1984). These released substances are the mediators which result in allergic symptoms caused by combination of IgE with a specific allergen. Through them, the effects of an allergen are manifested. Such effects may be systemic or local in nature, depending on the route by which the antigen entered the body and the pattern of deposition of IgE and mast cells. Local manifestations generally occur on epithelial surfaces at the location at which the allergen entered the body. Systemic effects can include anaphylaxis (anaphylactic shock), which is the result of an IgE-basophil response to circulating (intravascular) antigen.

One allergen of particular concern for many people is Antigen E or Amb a I, a poorly-defined constituent (or group of constituents) which is the major allergenic component(s) of short ragweed (Ambrosia artemisiifolia I. or Ambrosia elatior) pollen and the major cause of late summer hayfever in North America and Canada. Smith, J. J., et al., Mol. Immun, 25:355-364 (1988); King, T. P., et al., Biochem., 3:458-468 (1964); King, T.P., Adv. Immun., 23:77-105 (1976). It has been estimated that, on average, as much as 13% of the total serum IgE in ragweed-sensitive individuals is specific for Amb a I. Zeiss, C. R., et al., J. Immun., 110:414-421 (1973).

Amb a I has been claimed to be an acidic, 38,000 molecular weight, non-glycosylated protein which is cleaved during extraction and chromatographic isolation into two non-covalently associated chains: an alpha chain of 26,000 molecular weight and a beta chain of 12,000 molecular weight. Knox, R. B., et al., Nature, 255:1066-1068 (1970); Knox, R. B., and Heslop-Harrison, J., J. Cell Sci., 6:1-27 (1970); King, T. P., Adv. Immun., 23:77-105 (1976); King, T. P., et al., Archs Biochem. Biophys., 212:127-135 (1981). The two-chain and the single chain forms of Amb a I, which are both highly reactive with IgE, are allergenically and antigenically related. King, T. P., et al., Biochemistry, 3:458-468 (1964). It has been shown, however, that several physical and chemical modifications of Amb a I cause a marked loss of antigen and allergenic activity. King, T. P., et al., Archs Biochem. Biophys., 212:127-135 (1981); King, T. P., et al., Immunochemistry, 11: 83-92 (1974).

Because ragweed pollen is the chief causative agent of late-summer hay fever in the eastern United States and Canada, it has been the subject of more studies by different laboratories than any other pollen allergen. King, T. P., Adv. Immun.,. 23:77-105 (1976). Despite extensive study, the immunochemical definition of Amb a I is still far from complete. Smith and co-workers have begun characterization of the epitope structure of Amb a I, using a series of murine monoclonal antibodies raised against purified, native Amb a I. Three non-overlapping, non-repeating antigenic sites

2049921

were defined (sites A, B, and C) and monoclonal antibodies directed to sites A and B together resulted in inhibition of 80% of human IgE binding to Amb a I. The reactivity of the monoclonal antibodies used was greatly diminished when Amb a I was physically or chemically modified. Olsen, Ph.D. thesis, University of North Carolina, Chapel Hill (1986); Olson, J. R., and Klapper, D. G., J. Immun., 136: 2109-2115 (1986). They indicated that the two sites (A and B) are conformationally dependent epitopes. That is, they are either single structures which lose their conformation during modification or composite structures made up of two or more discontinuous peptides which are proximal in the native allergen but separate once the allergen has been modified. Smith, J. J., et al., Mol. Immun., 25:355-365 (1988).

Despite the considerable attention ragweed allergens have received, definition or characterization of the structure(s) or component(s) of the allergen responsible for its adverse effects on people is far from complete and current desentization therapy involves treatment with a complex, ill-defined extract of ragweed pollen.

Summary of the Invention

The present invention relates to allergenic proteins or peptides from ragweed, DNAs encoding all or a portion of such allergenic proteins or peptides; to compositions containing such an allergen(s) or portions of the allergen(s); and to methods of administering the allergen(s) or a

portion thereof or a composition which includes the allergen(s) or portions thereof to reduce or prevent the adverse effects that exposure to the allergen normally has on ragweed-sensitive individuals (i.e., to desensitize individuals to the allergen or block the effects of the allergen). The present invention further relates to methods of diagnosing sensitivity to ragweed pollen.

In particular, the present invention provides an isolated protein or peptide capable of influencing ragweed T-cell responses in ragweed-sensitive individuals, the protein or peptide being selected from:

(a) Amb a IA free of other Amb a I family members and comprising at least a T cell epitope-containing portion of the amino acid sequence shown in Figure 11; or

(b) Amb a IB free of other Amb a I family members and comprising at least a T cell epitope-containing portion of the amino acid sequence shown in Figure 12; or

(c) Amb a IC free of other Amb a I family members and comprising at least a T cell epitope-containing portion of the amino acid sequence shown in Figure 13; or

(d) Amb a ID free of other Amb a I family members and comprising at least a T cell epitope-containing portion of the amino acid sequence shown in Figure 14, for use in therapy, prophylaxis or diagnosis.

It has now been shown that Antigen E or Amb a I is not a single protein but, rather, a family or families of proteins to which ragweed-sensitive individuals react. In particular, the present invention relates to DNA encoding an

C

amino acid sequence or peptide present in allergenic proteins from ragweed pollen. It relates to DNA encoding all or a portion of the ragweed allergen Amb a I or Antigen E preparation which has been isolated. Such ragweed allergen preparations are heterogeneous in nature and may include, in addition to what is currently referred to as Amb a I or Antigen E, other ragweed components which are allergenic (i.e. cause the typical adverse effects observed in a ragweed-sensitive individual upon exposure to ragweed pollen). These may include, for example, what is referred to in the literature as Antigenk and referred to herein as Amb a II. The present invention also relates to DNAs encoding similar amino acid sequences (i.e., DNA encoding amino acid sequences of allergens) in types of ragweed other than short ragweed, such as giant ragweed and western ragweed.

In one aspect of the present invention provides an isolated protein or peptide capable of influencing ragweed T cell responses in ragweed-sensitive individuals, the protein or peptide being selected from: (a) a protein or peptide comprising at least a T cell epitope-containing portion, or a B cell epitope-containing portion, of the amino acid sequence of *Amb a* IA shown in Figure 11; or (b) a protein or peptide comprising at least a T cell epitope-containing portion, of the amino acid sequence of *Amb a* IB shown in Figure 12; or (c) a protein or peptide comprising at least a T cell epitope-containing portion, or a B cell epitope-containing portion, of the amino acid sequence of *Amb a* IC shown in Figure 13; (d) a protein or peptide comprising at least a T cell epitope-containing portion, or a B cell epitope-containing portion, of the amino acid sequence of *Amb a* ID shown in Figure 14; or (e) a protein or peptide

which is a polymorphic variant of any one of (a) to (d); for use in therapy, prophylaxis or diagnosis.

In another aspect the present invention provides a process for producing a ragweed protein/peptide allergen comprising the steps of: (a) identifying a DNA sequence encoding ragweed allergen which hybridizes under conditions of high stringency with the DNA sequence shown in Figure 11, 12, 13 or 14; (b) inserting the DNA identified in step (a) into an expression vector; (c) inserting the expression vector of step (b) into a host cell; (d) culturing the host cell of step (c) under conditions appropriate for expression of the DNA inserted in step (b); and (e) recovering the expressed product.

In another aspect the present invention provides an isolated protein or peptide or a polymorphic variant thereof, capable of influencing ragweed T cell response in ragweed-sensitive individuals, the protein or peptide comprising at least a T cell epitope-containing portion, or a B cell epitope-containing portion, of the amino acid sequence of *Amb a* II shown in Figure 15.

In another aspect the present invention provides a process for producing a ragweed protein/peptide allergen comprising the steps of: (a) identifying a DNA sequence encoding ragweed allergen which hybridizes under conditions of high stringency with the DNA sequence shown in Figure 15; (b) inserting the DNA identified in step (a) into an expression vector; (c) inserting the expression vector of step (b) into a host cell; (d) culturing the host cell of step (c) under conditions appropriate for expression of the DNA inserted in step (b); and (e) recovering the expressed product.

According to another aspect of the present invention, there is provided an isolated protein or peptide capable of influencing ragweed T cell responses in ragweed-sensitive individuals, the protein or peptide being selected from: (a) a protein or peptide comprising at least a T cell epitope-containing portion, or a B cell epitope-containing portion, of the amino acid sequence of *Amb a* IA shown in Figure 11; or (b) a protein or peptide comprising at least a T cell epitope-containing portion, or a B cell epitope-containing portion, of the amino acid sequence of *Amb a* IB shown in Figure 12; or (c) a protein or peptide comprising at least a T cell epitope-containing portion, or a B cell epitope-containing portion, of the amino acid sequence of *Amb a* IC shown in Figure 13; (d) a protein or peptide comprising at least a T cell epitope-containing portion, or a B cell epitope-containing portion, of the amino acid sequence of *Amb a* ID shown in Figure 14; or (e) a protein or peptide which is a polymorphic variant of any one of (a) to (d).

According to still another aspect of the present invention, there is provided a protein or peptide capable of influencing ragweed T cell responses in ragweed-sensitive individuals, wherein the protein or peptide comprises an amino acid sequence which is a modification of an amino acid sequence selected from: (a)EFPILGGITEVKDNDNSVDFC; (b) YTVTSDKDDDVANC; (c) GKADWAENRC; and (d) GFFQVVNNNYDRWGTYA; wherein the modification is an amino acid substitution, addition or deletion to alter immunogenicity or reduce allergenicity.

2049921

Brief Description of the Drawings

Figure 1 is a schematic representation of several routes by which assessment of proteins or peptides present in Amb a I and DNA encoding such proteins or peptides have been identified isolated and characterized.

Figure 1A is a schematic representation of screening of an Amb a I or Antigen E preparation, using monoclonal antibodies and oligoprobes.

Figure 1B is a schematic representation of screening of a ragweed flowerhead λgt10 library. It also illustrates the use of cross-hybridization and polymerase chain reaction (PCR) methods to obtain full-length cDNA clones encoding Amb a I and Amb a II.

Figure 2 is the nucleotide sequence of the DNA insert of UNC Clone 1 (referred to as Amb a IA), which was isolated from a λgt11 library by screening with monoclonal antibodies specific for components of an Amb a I preparation.

Figure 3 is the nucleotide sequence of the DNA insert of UNC Clone 6 (referred to as Amb a IB), which was isolated from a λgt11 library by screening with monoclonal antibodies specific for components of an Amb a I preparation.

Figure 4 is the nucleotide sequence of the DNA insert of UNC Clone 15 (referred to as Amb a IC), which was isolated from a λgt11 library by screening with monoclonal antibodies specific for components of an Amb a I preparation.

Figure 5 is the nucleotide sequence of the cDNA insert of IPC Clone 1, which was isolated from a λgt10 cDNA library using an oligonucleotide probe

whose sequence was deduced from an amino acid sequence known to be present in the ragweed allergen preparation Amb a I. The location of the sequence from which the sequence of the oligonucleotide probe was deduced is underlined.

Figure 6 is the nucleotide sequence of the cDNA insert of IPC Clone 5, which was isolated from a λgt10 cDNA library using an oligonucleotide probe whose sequence was deduced from an amino acid sequence known to be present in the ragweed allergen preparation Amb a I. The location of the sequence from which the sequence of the oligonucleotide probe was deduced is underlined.

Figure 7 is the nucleotide sequence of the cDNA insert of IPC Clone 6, which was isolated from a λgt10 cDNA library using an oligonucleotide probe whose sequence was deduced from an amino acid sequence known to be present in the ragweed allergen preparation Amb a I. The location of the sequence from which the sequence of the oligonucleotide probe was deduced is underlined.

Figure 8 is a schematic representation of open reading frame analysis of the DNA insert of IPC Clone 1.

Figure 9 is a schematic representation of open reading frame analysis of the DNA insert of IPC Clone 5.

Figure 10 is a schematic representation of open reading frame analysis of the DNA insert of IPC Clone 6.

Figure 11 is the nucleotide sequence and deduced amino acid sequence of a full length Amb a IA clone (related to UNC clone 1).

Figure 12 is the nucleotide sequence and deduced amino acid sequence of a full length Amb a IB clone (related to UNC clone 6).

Figure 13 is the nucleotide sequence and deduced amino acid sequence of a full length Amb a IC clone (related to UNC clone 15).

Figure 14 is the nucleotide sequence and deduced amino acid sequence of a full length Amb a ID clone.

Figure 15 is the nucleotide sequence and deduced amino acid sequence of a full length Amb a II clone.

Figure 16 is the composite amino acid sequences of the Amb a I and Amb a II multigene family showing regions of similarity as well as regions of disagreements.

Figure 17 is a photograph of a Western blot of affinity purified Amb a I treated with rabbit anti-Amb a I polyclonal antibody, JB1E3-4 anti-Amb aI monoclonal antibody or ragweed allergic patient sera.

Figure 18 is a photograph of a two dimensional gel of an aqueous extract of short ragweed pollen, separated on the basis of size and charge and stained with T.P. King's antibody, which recognizes Amb a I (goat polyclonal anti-Amb a I).

Figure 19 is a photograph of a Western blot of several E. coli-expressed recombinant Amb a I cDNAs treated with goat anti-Amb a I antibody.

Figure 20 is a photograph of a Western blot of several E. coli expressed recombinant Amb a I cDNAs treated with human allergic sera strained with anti-human IgE.

Figure 21 is a graphic representation of T cell proliferation responses of ragweed allergic patient PBMC toward an aqueous extract of short ragweed pollen, affinity purified Amb a I (B7) chromatographically purified Amb a I and E. coli lysate containing expressed recombinant Amb a I proteins.

Detailed Description of the Invention

The present invention is based on an investigation of ragweed pollen allergens, particularly the preparation known as Amb a I (or Antigen E) from short ragweed, using several inter-related approaches, each described below. The terms Amb a I and Antigen E are used interchangeably. Such a preparation, obtained from ragweed pollen, is likely to contain other ragweed allergens, such as Antigen K or Amb a II. The possibility that such a preparation does contain other such allergens has been assessed and results demonstrate that this is the case.

Results of work described herein show that Amb a I is not a single protein or peptide but is, in fact, heterogeneous in nature. That is, what is presently referred to as Antigen E (or Amb a I) appears to be a family or families of proteins or to be polymorphic in nature. The work described herein has resulted in identification and isolation of DNAs

encoding peptides or amino acid sequences present in a ragweed allergen. As described, full-length cDNAs encoding Amb a IA, Amb a IB, Amb a IC and Amb a ID and Amb a II have been isolated and sequenced. It has also resulted in isolation and purification from an Amb a I preparation of a protein shown to bind human ragweed IgE and to bind rabbit Amb a I antisera produced using a purified Amb a I preparation. Interrelationships among DNAs and proteins or peptides identified and isolated using the approaches described in the following section have been demonstrated. For ease of presentation, the several approaches used are represented schematically in Figure 1A and 1B, to which reference is made in the following discussion.

As a result of the work described herein, DNAs encoding proteins or peptides present in ragweed allergens have been identified and isolated and the amino acid sequence of the encoded product has been deduced. In addition, through the use of monoclonal antibodies specific for Amb a I or Antigen E, a protein has been obtained from an Amb a I preparation. This protein, referred to as affinity purified Amb a I, has been shown to have biological activity (human IgE binding ability and ability to bind rabbit Amb a I antisera) and, thus, is highly likely to be an allergen. It has also been shown to be encoded by a region of the nucleotide sequences present in two of the isolated DNAs.

The following is a description of several approaches which have been used to identify and isolate DNAs encoding proteins or peptides from

Amb a I or Antigen E preparations, as well as to isolate from an Amb a I preparation a protein shown to have Amb a I activity. As represented in Figure 1A, an Amb a I or Antigen E preparation, which was prepared from pollen extract by a method based on the method of T.P. King and co-workers, was produced. King, T.P. et al, Arch. Bioch. and Biophys., 212:127-135 (1981). A panel of monoclonal antibodies produced by Klapper and co-workers was used to identify proteins in the preparation. Smith, J.J. et al., Mol. Immun., 25:355-365 (1988). Sequences of several peptides from an Antigen E preparation were determined by conventional techniques.

The following sections describe: 1) use of a pool of these monoclonal antibodies (i.e., a pool of monoclonal antibodies reactive with Amb a I) to identify clones containing DNA inserts encoding the reactive product and 2) use of an oligonucleotide probe, constructed from an amino acid sequence present in the Amb a I preparation to identify clones containing DNA inserts encoding the amino acid sequence. Each approach resulted in identification of three clones containing DNA encoding an amino acid sequence present in the Amb a I or Antigen E preparation. The two sets of clones isolated as described below have been shown to be different from each other.

Use of Monoclonal Antibodies to Identify Clones Containing DNA Inserts Encoding Ragweed Protein

2049921

A pool of seven monoclonal antibodies specifically reactive with components of the Amb a I preparation was used to screen a ragweed pollen λgt11 library, using a known method. Young, R.A. and R.W. Davis, Proceedings of the National Academy of Sciences, USA, 80: 1194-1198 (1983). This resulted in identification of three clones, initially designated UNC Clones 1, 6 and 15 and referred to herein as Amb a IA, IB and IC, respectively, which expressed a product recognized by at least one of the monoclonal antibodies in the panel. The nomenclature of cDNAs encoding the allergens Amb a I and Amb a II have been named according to the recommendations of the International Union of Immunological Societies Sub-Committee for Allergen Nomenclature (Marsh et al., Annals of Allergy, 60:499-504 (1988)).

DNA isolated from the three reactive clones was sequenced, using the method of Sanger, F. et al. Sanger, F. et al., Proc. Natl. Acad. Sci., USA, 74:5463 (1977). The nucleotide sequences of the three clones are presented in Figures 2-4.

Using the partial cDNA sequences presented in Figures 2-4, cross-hybridization (as described in Example 2) and PCR methods (as described in Example 3) were used to isolate full-length cDNAs encoding Amb a IA (Figure 11), Amb a IB (Figure 12), Amb a IC (Figure 13) and Amb a ID (Figure 14).

In the course of DNA sequencing of cross-hybridizing cDNAs from a separately constructed λgt10 ragweed flowerhead library, a new cDNA was derived that shared sequence with Amb a II peptide

sequence (Figure 15 and Figure 16). Construction of this library and isolation of the new cDNA are described in Example 2. The composite amino acid sequences of the Amb a I and Amb a II multigene family are shown in Figure 16, with the regions of similarity and of disagreement represented. In Figure 16, the sequence of Amb a I is given in standard one-letter code. Sequences for the other Amb a I family members are given relative to that of Amb a I, with only differences being shown. A dash indicates identity between the two sequences. An asterisk indicates a break in the sequence introduced to maintain maximal alignment. Amino acid numbering is based on the Amb a IB sequence. 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. Polymorphisms in a given family member occur as independent events, except for amino acids 183-189 of Amb a ID, in which the polymorphism occurs as a block.

Use of an Oligonucleotide Probe to Identify Clones Containing DNA Inserts Encoding Ragweed Protein

As also represented in Figure 1A, an amino acid sequence (WENFK) in the Amb a I preparation, which was identified and sequenced by conventional techniques, was used to deduce the sequence of an oligonucleotide probe (oligoprobe) encoding the amino acid sequence. The amino acid sequence used to deduce the oligonucleotide sequence was VWVKPWENFK. A portion of that amino acid sequence

(WENFK) was used to deduce the sequence of the oligoprobe, designated AGE#1. AGE#1 was used, as described in Example 1, to screen a cDNA library constructed in λgt10 using polyA$^+$ enriched RNA from pooled short ragweed flower heads. Screening with this oligoprobe resulted in identification of ten duplicated signals. These duplicated signals (clones) were subjected to a secondary screening with the same AGE #1 oligonucleotide probe. Three of the positives (referred to as secondary positives) were clearly detected in duplicate. The clones (designated IPC Clone 1, IPC Clone 5 and IPC Clone 6) identified in this manner were grown under appropriate conditions and verified as positive, by Southern blot analysis.

The cDNA insert from each of the three clones was isolated and cloned into M13mp18 and sequenced (Figures 5-7). The amino acid sequence was also deduced (Figures 8-10). Open reading frames in the sequenced cDNAs were examined (Figures 8-10) and the sequence (from which the sequence of the oligonucleotide probe had been deduced) was identified. That the cDNA inserts encode a portion of translated protein was supported by the fact that the surrounding amino acid sequence deduced from the DNA sequence (VWVKP) agreed with the amino acid sequence initially used to deduce the sequence of the oligoprobe (Figure 8-10). T cells from allergic patients could be stimulated by a synthetic peptide RAE4 (Table 5). The RAE4 sequence was deduced from IPC Clone 5 (Figure 8).

As is evident from a comparison of the two "sets" of nucleotide sequences (i.e., set 1, which are the DNAs isolated through use of monoclonal antibodies, and set 2, which are the DNAs isolated through use of the oligoprobe), there is homology among sequences within a set (i.e., within Figures 2-4 and within Figures 5-7) but little similarity in sequences between sets.

Thus, it is apparent that the Amb a I or Antigen E preparation is heterogenous in nature and represents a family (or families) of proteins or that there is considerable polymorphism in Amb a I-encoding DNA. This is in contrast to present literature descriptions of Amb a I or Antigen E, which refer to Antigen E as a protein, rather than as a group or groups of allergenic proteins, present in ragweed pollen, to which ragweed-sensitive individuals respond.

Additional Demonstration of Isolation of Antigenic Peptides and DNAs of Amb a I

Additional results further demonstrate that antigenic peptides of Amb a I and DNAs encoding them have been identified and isolated. As represented in Figure 1A, a selected monoclonal antibody (designated 4B5/B7) which recognizes an Amb a I preparation unsubjected to denaturing conditions was used to affinity purify from pollen extract a single protein, which is referred to as affinity purified Amb a I. This was carried out, using known techniques, by producing the desired monoclonal antibody, isolating it in large quantities from

ascites and immobilizing it on Sepharose* (Pharmacia). Aqueous pollen extract was passed over the monoclonal antibody-containing column and a protein species was eluted. Antigen E isolated in this manner was shown, using both Western blot (Figure 17) and ELISA techniques, to bind human IgE, thus demonstrating biological activity expected of an Amb a I protein or peptide.

Peptide sequence analysis was carried out as follows: Two peptides were isolated from partial tryptic digestion or cyanogen bromide (CNBr) cleavage of affinity purified Amb a I, respectively, and then subjected to peptide sequencing. Because the N-terminal of Amb a I is blocked, no amino acid sequence can be obtained from direct N-terminal protein sequence analysis. The result of the sequence analysis of the tryptic peptide demonstrated that the major portion of its amino acid sequence agreed with peptide sequence 45 to 77 encoded by the Amb a IA cDNA (Table 1). Table 1 is a comparison of the amino acid sequences of Amb a I protein, determined by protein sequence analysis, with the amino acid sequence deduced from Amb a I cDNA. The CNBr cleavage peptide sequencing demonstrated that the CNBr cleavage peptide was similar to the peptide sequence 171 to 184 encoded by the Amb a IA cDNA (Table 1).

Further peptide sequence analysis was performed from the protein cleavage mixture without isolating individual peptides. The techniques employed involved specific hydrolysis (with 70% formic acid or CNBr) of the putative Asp-Pro and Met-Pro bonds deduced from the cDNA

*Trade-mark

sequences of Amb a I. Any primary amino groups were then blocked by reaction with o-phthalaldehyde prior to conventional sequencing from any available N-terminal proline residue.

TABLE 1

Amb a I PROTEIN SEQUENCES[a] COMPARED TO PROTEIN SEQUENCE DEDUCED FROM Amb a IA cDNA SEQUENCES

PARTIAL TRYPTIC DIGEST[b]

		45 50 55 60 65 70 75 80 85 90
Amb a IA	cDNA	T S G A Y N 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
Amb a I[c]	MAJOR[d]	(T)S(G)A Y N 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)
	MINOR[e]	(D) (S R)
AgE[f]		(T)S G A Y N 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)

CNBr CLEAVAGE

		175 180 185
Amb a IA	cDNA	H D V K V N P G G L I K S N D G
Amb a I[g,h]	MAJOR	F D L K V N I G Q L I K(S)N
	MINOR	(A)F N Y(I)P L (N)
AgE[i]		(H D V K V)P G G L I K( )N( )G

		280 285 290 295 300 305 310 315 320
Amb a IA	cDNA	P R C R H G F F Q V V N N N Y D R W G S Y A I G G S A S P T I L S Q G N R F C A P D E R S
Amb a I[i]	MAJOR	P R C R H G F F Q V V N N N Y D R W G(S)Y A I G G S(A)P T I L S Q G M( )F(C)A P(D G Y)
	MINOR#1[j]	F I F D(H) (N) V
	MINOR#2[k]	P V L(T)P E(Q)S A(G M)
	MINOR#3[l]	T S G A Y N I I D G C W R G(K)A(D W)A
AgE	MAJOR	P R( )R H G F F Q V V N N N Y D(E W)G S Y A I G G S A S P T I
	MINOR#1[m]	A(W)N(W)R(T E K)D L
	MINOR#2[n]	V(I)N L(D Q)E I(F V)

TABLE 1 con't.

70% FORMIC ACID HYDROLYSIS OF ASP-PRO PEPTIDE BOND[1]

		365 370 375 380 385 390 395
Amb a IA	cDNA	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
Amb a I	MAJOR[o]	P V L(N P)E( )N A G M I Q A E(P G)E A
	MINOR	I
AgE		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(C)Q P(G A)P
35kD[p,q]		P V L T P V Q S A G M I P A E P G E A A I(K)L T S S

a. the amino acids are presented in single letter code; uncertain residues are included in paranthesis
b. the peptides were separated by SDS-PAGE then Western blotted on PVDF membrane for sequence analysis
c. IPC's affinity purified Amb a I preparation
d. major sequence determined in protein sequence analysis
e. minor sequence determined in protein sequence analysis
f. T. P. King's Amb a I preparation
g. the cleavage mixture was separated by SDS-PAGE then Western blotted on PVDF membrane
h. the sequence is most similar to a IA cDNA sequence among all the cloned cDNA sequence
i. the primary amine of the cleavage mixture was blocked by o-phthalaldehyde on the 7th step of sequence analysis
j. similar to the a IIA cDNA sequence 277-315
k. similar to the a IA cDNA sequence 361-371
l. similar to the a IA cDNA sequence 45-63
m. similar to the a IC cDNA sequence 338-347
n. similar to the a IC cDNA sequence 126-135
o. matches to a IC cDNA sequence 363
p. IPC's Amb a I preparation with molecular weight of 35,000 dalton
q. matches to a IC cDNA sequence 361-366

2049921

PCT/US90/01310

Results of these assessment (shown in Table 1) demonstrated that two peptide sequences determined from the affinity purified Amb a I preparation agreed with that encoded by two portions of Amb a IA DNA sequence (277-321 or 361-397). The minor sequences detected in the peptide sequence analysis also corresponded to a portion of peptide sequence encoded by cDNA's of Amb a I or Amb a II. The above peptide sequence analyses provided strong support that Amb a I or Antigen E-encoding DNA had been isolated.

An Antigen E preparation obtained from Dr. T.P. King was also subjected to peptide sequencing. The same peptide sequencing techniques were employed. Four peptides sequences were identified which agreed with the same four segments of peptide sequence encoded by Amb a IA DNA (45-92, 171-186, 277-321 and 361-397 in Table 1). This provided additional proof that Amb a I or Antigen E-encoding DNA had been isolated.

The same techniques were used with purified Antigen K (Amb a II) from Dr. T.P. King. Results demonstrated that two peptide sequences agreed with two portions of peptide sequence encoded by DNA of Amb a II (Table 2, see Example 2; Figure 15). Table 2 is a comparison of the amino acid sequences of Amb a II protein, determined by protein sequence analysis, with the amino acid sequence deduced from Amb a II cDNA. This finding provided support that ragweed pollen allergen encoding DNA had been isolated.

TABLE 2

Amb a II PROTEIN SEQUENCES[a] COMPARED TO PROTEIN SEQUENCE DEDUCED FROM Amb a IIA c DNA SEQUENCE

CNBr CLEAVAGE[b]

		280 285 290 295 300 305 310 315 320
Amb a IIA cDNA		P R C R F G F F Q I V N N F Y D R W D K Y A I G G S S N P T I L S Q G N K F V A P D F I Y
AgK[c]	MAJOR[d]	P R( )R F G F F Q I V N N F Y D R W D(H)Y A I G G S S N P T I L S Q G N(R)F V A P(D)I(Y)
	MINOR[e,f]	P V L T P E Q N A G M
Amb a II[g]		P(R R)F G F F Q I V N N F Y D

705 FORMIC ACID HYDROLYSIS OF ASP-PRO PEPTIDE BOND[b]

		365 370 375 380 385 390 395
Amb a IIA cDNA		P V L T A E Q N A G M M Q A E P G D M V P Q L T M N A G V L T C S P G A P
AgK		P V L T A E Q N A G M M Q A E P G D M V P Q L T M N A(G)V(L S)P G A P
Amb a II	MAJOR[h]	P V L T A E Q N A G M M Q A E P G D M V P Q L T M N A G V L T( )S P G A P
	MINOR	P S I P E S A L S S (S)

a. the amino acids are presented in single letter code; uncertain residues are included in paranthesis
b. o-phthalaldehyde is reacted with peptide mixture prior to conventional peptide sequence analysis
c. T.P. King's Amb a II preparation
d. major sequence determined in protein sequence analysis
e. minor sequence determined in protein sequence analysis
f. matches the a IIA cDNA sequence 361-371
g. matches the a IIA cDNA sequence 361-371
h. matches to a IA cDNA sequence 361-397

It has been previously reported that Amb a I and Amb a II share some antigenic determinants using rabbit and human antisera (King, T.P., Adv. Immun., 23:77-105 (1976)). However, the exact relationship between the two antigens, until the present invention, has remained unclear. King and colleagues have also reported that different isoforms of antigen E and K (Amb a I and Amb a II) can be isolated by ion-exchange chromatography (King, T.P. et al., Ach. Biochem. Biophys., 212:127-135 (1981)). The different isoforms described, designated A, B, C and D, were interpreted to be produced by limited proteolysis of the intact Amb a I and Amb a II species. It should be noted that these isoforms, designated A, B, C, etc., have no direct relationship with the nomenclature outlined in this invention (i.e., Amb a IA, Amb a IB, etc.).

A 35,000 dalton species coprecipitates from ragweed pollen extract with Amb a II in 45% saturation of ammonium sulfate. Most of these proteins are shown to be agregated by gel filtration chromatography. Some monomeric forms of these proteins were separated from Amb a II by ion exchange chromatography. The sequencing technique, which involved 70% formic acid hydrolysis of putative Asp-Pro bound and o-phthalaldehyde blocking of primary amino groups, demonstrated that the predominant protein corresponds to that encoded by the DNA sequence of Amb a IC. This peptide sequence is referred to as 35kD in Table 1. This result provided additional support that Amb a I proteins

are heterogeneous in nature and are encoded by closely related DNA's.

As is also represented in Figure 1A, rabbit polyclonal antibodies were produced using the King Antigen E preparation (King, T.P. et al. (1981) Arch. Biochem. Biophys. 212:127-135). These antibodies were shown to identify a 38kd protein species on a Western blot of pollen extracts (Figure 17). A two-dimensional gel of ragweed pollen extract, electrophoresed in one dimension on the basis of charge and in the other dimension on the basis of size and treated with goat anti-Amb antibodies is shown in Figure 18. Results demonstrate binding to several proteins present in ragweed pollen extract with a relative molecular weight of 38kd, corresponding to differently charged forms of what was formerly referred to as Amb a I protein. These antibodies were also shown, using a similar technique, to bind to the affinity purified Amb a I described previously (Figure 17).

It is clear from the antibody reactivity that the 4B5/B7 affinity purified Amb a has a recognition pattern similar to that of the Amb a I of pollen and skin test reagent with both rabbit polyclonal anti-Amb a I and JB1E3-4 anti-Amb a I monoclonal antibody (Figure 17). It also has readily detectable IgE reactivity on a Western blot (Figure 17; patient number 155). It is also clear that chromatographically purified Amb a II (Antigen K) has cross-reactive B-cell epitopes with the affinity purified Amb a I (Figure 17); anti-Amb a I polyclonal).

As a result of the work described herein, cDNAs encoding allergenic peptides of proteins from a preparation of Amb a I, the major human allergen of ragweed and a preparation of Amb a II, have been cloned, isolated and sequenced; the encoded amino acid sequences (of the allergen(s)) have been deduced and peptides derived from Amb a I and Amb a II have been identified and isolated.

Furthermore, full-length and truncated cDNAs encoding several members of the Amb a I multigene family, as well as Amb a II, were cloned in-frame into the expression vector pTrc99 (Amann et al. Gene, 69: 301-315, (1988)) and transformed into the JM109 host. Expression of recombinant Amb a I and Amb a II protein was induced by 1mM isoprophyl-β-D-thiogalactopyranoside, cells were harvested, lysozyme treated, sonicated and insoluble inclusion bodies recovered by a low speed centrifugation. Recombinant Amb a I and Amb a II protein present in the recovered pellet was solubilized in buffer containing 8M urea, 50mM Tris HCl pH8.0, 50mM NaCl, 1mM EDTA, 1mM dithiothreitol, 1mM phenylmethylsulfonyl fluoride. After solubilization, the crude urea lysate was dialyzed at 4°C against PBS. The expressed recombinant Amb a I and Amb a II proteins were Western blotted and results are shown in Figures 19 and 20. Results demonstrate (Figure 19) that goat anti-Amb a I antibody binds specifically to several forms of Amb a I (A, B and C), as well as to Amb a II (antigen K). This antigenic cross-reactivity is consistent with the observed sequence homology of the cDNAs (see Figure 16). They further

demonstrate (Figure 20) that allergic human IgE binds specifically to some members of the Amb a I multigene family. In the case of patient #295, Amb a IA (full-length) and Amb a IC are bound specifically by IgE to a far greater extent than Amb a IB or Amb a II. A high level of variability in the patterns of IgE binding is seen (Table 3 and data not shown), suggesting that different patients respond to the different Amb a I proteins to different extents.

TABLE 3

SUMMARIZED WESTERN BLOT DATA*

Patient	Antigen					
	Pollen	IA(t)	IA	IB	IC	IIA
151	+	-	+	-	+	+
222	+-	+-	+-	+-	+-	+
291	+++	+	+++	-	+++	+-
295	+++	+	+++	+	+++	-
296	++		++		++	-

\- no signal over background
+- barely discernable over background
\+ clearly positive
++ strongly positive
+++ highly positive

* selected from the total of ten patients screened to date.

An analysis of SDS-PAGE Western blot of IgE binding to several recombinant forms of Amb a I and Amb a II has demonstrated that there is considerable

variation in the pattern observed with different patients. Of the ten ragweed allergic patients examined, all possess serum IgE that binds to at least one recombinant Amb a I or Amb a II, with some patient's IgE binding several different recombinant species (summarized in Table 3). Comparison of human IgE binding to recombinant Amb a I and Amb a II proteins with anti-peptide and monoclonal anti-Amb a I antibodies have provided data consistent with the conclusion that the N-terminal portion (historically referred to as the β-region) of Amb a IA includes the major IgE epitope(s). This data (Table 3) is based on the observation that Amb a IA(t) (truncated Amb a IA; amino acid 70-398) binds ragweed allergic patient IgE less well than the full-length Amb a IA (amino acid 10-398). It is expected that the other Amb a I and Amb a II forms possess the same IgE binding properties (see Figure 20, for example).

T cells from patients allergic to ragweed, previously stimulated with a mixed ragweed pollen extract, can recognize and proliferate in response to pollen extract, ragweed skin test reagent (RWST), affinity purified Amb a I protein and crude bacterial lysates containing recombinant Amb a I gene products IA, IB and IC (Table 4). T cells from these patients do not proliferate in the presence of an equivalent amount of control bacterial lysate, JM109. These results demonstrate that each gene product can stimulate some T cell reactivity. The use of crude bacterial lysates as antigens precludes a firm conclusion from the negative responses, since

the relative levels of recombinant proteins in lysate have not been determined.

Table 4 STIMULATORY RESPONSE[a] OF THE HUMAN T CELL TO RECOMBINANT RAGWEED PROTEINS

PATIENT #		POLLEN	RWST[b]	*Amb a* I[c]	*Amb a* IB LYSATE	*Amb a* IC LYSATE	*Amb a* IA(t) LYSATE	*Amb a* IA LYSATE	JM109 LYSATE
151	2°	+++	+	(+)	+	+	+	(+)	-
222	2°	+++	+++	+++	-	++	++	++	-
274	2°	++	++	++	+	++			-
295	2°	+++			+	+	+		-
296	2°	+++	+++	+++	+++	+++	+++		-
314	2°	+++	+++	+++	+++	+++	+++		-
316	2°	+++	+++	+++	++	+++	+++		-
319	2°	+++	+++	++	(+)	++	-		-
320	2°	++	++	+++	+	-	++		-
321	2°	+++	+++	+	++	+	++		-

[a] proliferation responses as compared to medium control: (+) 2 fold
+ 2-4 fold
++ 4-10 fold
+++ >10 fold

[b] ragweed skin test reagent from Hollister-Stier

[c] affinity purified *Amb a* I

2049921

PCT/US90/01310

Uses of the Subject Allergenic Proteins/Peptides and DNA Encoding Same

The materials resulting from the work described herein, as well as compositions containing these materials, can be used in methods of diagnosing, treating and preventing ragweed allergy. In addition, the cDNA (or the mRNA from which it was transcribed) can be used to identify similar sequences in any variety or type of ragweed and, thus, to identify or "pull out" sequences which have sufficient homology to hybridize to, for example, DNA from short ragweed pollen. This can be carried out, for example, under conditions of low stringency; those sequences which have sufficient homology (generally greater than 40%) can be selected for further assessment using the method described herein. Alternatively, high stringency conditions can be used.

In this manner, DNA of the present invention can be used to identify, in other types of ragweed (such as giant ragweed or Western ragweed) sequences encoding peptides having amino acid sequences similar to that of Amb a I and, thus, to identify allergens in such other types of ragweed. Thus, the present invention includes not only Amb a I and other ragweed allergens (e.g., Amb a II or Antigen K) encoded by the present DNA sequences, but also other ragweed allergens encoded by DNA which hybridizes to DNA of the present invention.

Proteins or peptides encoded by the cDNA of the present invention can be used, for example, as "purified" allergens. Such purified allergens are

useful in the standardization of allergen extracts which are key reagents for the diagnosis and treatment of ragweed allergy. Furthermore, by using peptides based on the sequences listed in Figures 2 through 16, anti-peptide antisera or monoclonal antibodies can be made using standard methods. Such reagents can be specifically directed against individual isoforms of Amb a I or Amb a II (i.e., directed against divergent regions/epitopes of the molecule) or can be specific for all forms of Amb a I or Amb a II (i.e., directed against common sequences/epitopes). These sera or monoclonal antibodies, directed against Amb a I or Amb a II, can be used to standardize allergen extracts. One such monospecific anti-peptide antisera has already been successfully produced. This rabbit antisera, directed against an Amb a II sequence (amino acid 326-338; designated RAE 50.K with the sequence: CLRTGAQEPEWMT) binds specifically on Western blots to recombinant Amb a II but not Amb a IA, B or C (data not shown).

Through use of the peptides of the present invention, allergen preparations of consistent, well-defined composition and biological activity can be made and administered for therapeutic purposes (e.g., to modify the allergic response of a ragweed-sensitive individual to a ragweed pollen). Such peptides or proteins (or modified versions thereof, such as are described below) may, for example, modify B-cell response to a ragweed allergen, T-cell response to a ragweed allergen or both responses. Purified allergens can also be used to study the

mechanism of immunotherapy of ragweed allergy and to design modified derivatives or analogues which are more useful in immunotherapy than are the unmodified ("naturally-occurring") peptides.

Work by others has shown that high doses of allergens generally produce the best results (i.e., best symptom relief). However, many people are unable to tolerate large doses of allergens because of allergic reactions to the allergens. Modification of naturally-occurring allergens can be designed in such a manner that modified peptides or modified allergens which have the same or enhanced therapeutic properties as the corresponding naturally-occurring allergen but have reduced side effects (especially anaphylactic reactions) can be produced. These can be, for example, a peptide of the present invention (e.g., one having all or a portion of the amino acid sequence of a peptide derived from the DNA insert of Clone Amb a IA, Clone Amb a IB, Clone Amb a IC, Amb a II, IPC Clone 1, IPC Clone 5 or IPC Clone 6, or their full-length cDNAs) or a modified peptide or peptide analogue (e.g., a peptide in which the amino acid sequence has been altered to modify immunogenicity and/or reduce allergenicity or to which a component has been added for the same purpose). For example, Amb a I peptides can be modified using the polyethylene glycol method of A. Sehon and co-workers (Wie et al. (1981) International Archives of Allergy and Applied Immunology 64:84-99).

Administration of the peptides of the present invention to an individual to be desensitized can be carried out using known techniques. A peptide or combination of different peptides can be

2049921

administered to an individual in a composition which includes, for example, an appropriate buffer, a carrier and/or an adjuvant. Such compositions will generally be administered by injection, oral administration, inhalation, transdermal application or rectal administration. Using the structural information now available, it is possible to design a ragweed pollen peptide which, when administered to a ragweed-sensitive individual in sufficient quantities, will modify the individual's allergic response to a ragweed allergen. This can be done, for example, by examining the structures of the ragweed proteins, producing peptides to be examined for their ability to influence B-cell and/or T-cell responses in ragweed-sensitive individuals and selecting appropriate epitopes recognized by the cells. Synthetic amino acid sequences which mimic those of the epitopes and which are capable of down regulating allergic response to ragweed allergen can also be used. Proteins, peptides or antibodies of the present invention can also be used for detecting and diagnosing ragweed allergy. For example, by combining blood or blood products obtained from an individual to be assessed for sensitivity to ragweed allergen with an isolated allergenic peptide of ragweed pollen, under conditions appropriate for binding of components (e.g., antibodies, T cells, B cells) in the blood with the peptide and determining the extent to which such binding occurs.

It is now also possible to design an agent or a drug capable of blocking or inhibiting the ability of ragweed allergens to induce an allergic reaction

in ragweed-sensitive individuals. Such agents could be designed, for example, in such a manner that they would bind to relevant anti-ragweed IgEs, thus preventing IgE-allergen binding and subsequent mast cell degranulation. Alternatively, such agents could bind to cellular components of the immune system, resulting in suppression or desensitization of the allergic response to ragweed allergens. A non-restrictive example of this is the use of appropriate B- and T-cell epitope peptides, or modifications thereof, based on the cDNA/protein structures of the present invention to suppress the allergic response to ragweed allergens. This can be carried out by defining the structures of B- and T-cell epitope peptides which affect B- and T-cell function in in vitro studies with blood cells from ragweed-sensitive individuals.

The cDNA encoding an allergenic protein or peptide from ragweed can be used to produce additional peptides, using known techniques such as gene cloning. A method of producing a protein or a peptide of the present invention can include, for example, culturing a host cell containing an expression vector which, in turn, contains DNA encoding all or a portion of a selected allergenic protein or peptide (e.g., Amb a I protein or peptide). Cells are cultured under conditions appropriate for expression of the DNA insert (production of the encoded protein or peptide). The expressed product is then recovered, using known techniques. Alternatively, the Amb a I allergen or portion thereof can be synthesized using known

2049921

mechanical or chemical techniques. As used herein, the term protein or peptide refers to proteins or peptides made by any of these techniques. The resulting peptide can, in turn, be used as described previously.

DNA to be used in any embodiment of this invention can be cDNA obtained as described herein or, alternatively, can be any oligodeoxynucleotide sequence having all or a portion of a sequence represented herein (See Figures 2-16), or their functional equivalents. Such oligodeoxynucleotide sequences can be produced chemically or mechanically, using known techniques. A functional equivalent of an oligonucleotide sequence is one which is capable of hybridizing to a complementary oligonucleotide sequence to which the sequence (or corresponding sequence portions) of Figures 2-16 hybridizes and/or which encodes a product (e.g., a polypeptide or peptide) having the same functional characteristics of the product encoded by the sequence (or corresponding sequence portion) of Figures 2-16. Whether a functional equivalent must meet one or both criteria will depend on its use (e.g., if it is to be used only as an oligoprobe, it need meet only the first criterion and if it is to be used to produce an Amb a I allergen, it need only meet the second criterion).

Antibodies against Amb a I peptides can be used to isolate additional components of ragweed allergens which can be used for further definition of the characteristics of the Amb a I family. Furthermore, anti-peptide sera or monoclonal

antibodies directed against Amb a I and/or Amb a II can be used to standardize and define the content of ragweed skin test reagents (RWST). This use would include RWST other than those derived from Ambrosia artemisiifolia I. (e.g., Western, Desert, Giant ragweeds, etc.).

The structural information now available (e.g., DNA, protein/peptide sequences) can also be used to identify or define T cell epitope peptides and/or B cell epitope peptides which are of importance in ragweed allergic reactions and to elucidate the mediators or mechanisms (e.g., interleukin-2, interleukin-4, gamma interferon) by which these reactions occur. This knowledge should make it possible to design peptide-based ragweed therapeutic agents or drugs which can be used to modulate these responses.

The present invention will now be further illustrated by the following Examples, which are not intended to be limiting in any way.

EXAMPLE 1 Screening of a λgt10 cDNA Library Using an Oligonucleotide Probe

PolyA$^+$ enriched RNA extracted from pooled short ragweed flower heads was used to construct a cDNA library in the vector λgt10. The cDNA library was constructed using the Gubler and Hoffman method and the kit supplied by Amersham. Gubler, U. and B.J. Hoffman, Gene, 25:263 (1983). A library of 1.4 x 10^5 plaques constituting approximately 7 x 10^4 recombinants was constructed. This library was plated out and screened according to the method of

Benton and Davis. Benton, W.D. and R.W. Davis, Science, 196:180 (1977).

Screening of the cDNA Library:

An amino acid sequence (VWVKPWENFKK), thought to be derived from antigen E was used to deduce an oligonucleotide probe with which to screen the cDNA library.

Amino Acid Sequence	W	E	N	F	K	K
Deduced Nucleotide Sequence	TGG	GAA	AAT	TTC	AAA	AAA
		G	C	T	G	G

AGE #1 OLIGOPROBE

The AGE #1 oligoprobe was end-labeled with ^{32}P and used to probe 70,000 recombinants using the hybridization conditions listed below:

6 x SSC	
1 x Denhardt's	at 30 degrees C for
50 μg/ml E. coli tRNA	22 hours

Ten duplicated signals were detected and these clones were subjected to a secondary screening with the same AGE #1 oligoprobe. It was subsequently discovered that the correct amino acid sequence is WENFKE.

A summary of the cloning procedure is listed below:

Primary Screen: 70,000 plaques
^{32}P - AGE #1 oligoprobe

Numerous spots with 10 signals were clearly seen on duplicated filters.

Secondary Screen:

Plaques from the 10 duplicate signals were picked and plated out at low density and rescreened using methods outlined in Clontech's catalog.

Tertiary Screening:

Three secondary positives numbered #1, #5, and #6 were clearly detected in duplicate. Each clone was grown up and verified as positive by Southern Blot analysis.

Sequencing of Positive Clones:

cDNA inserts from each of the clones were isolated then cloned into M13mp18. Each clone was sequenced using the Sanger dideoxy method and the deduced amino acid sequence was determined. Sanger, F. et al., Proc. Natl. Acad. Sci., USA, 74:5463 (1977).

Identification of WENFKK and Surrounding Sequence:

The DNA sequences of the cDNA clones are presented in Figures 5, 6 and 7. The cDNA clones are not full-length and are less than 500 nucleotides in length. The AGE#1 oligoprobe nucleotide sequence is underlined in Figures 5, 6 and 7. Open reading frames in the sequenced cDNAs

were examined and are presented in Figures 8, 9 and 10. The translated amino acid sequence (WENFK) used to deduce AGE#1 oligoprobe sequence is underlined as well as the N-terminal surrounding sequence (VWVKPWENFK; see Figures 8, 9 and 10). IPC clones 1 and 5 disagree with the amino acid sequence at only one out of ten residues (i.e., L instead of P). The presence of the correct surrounding sequence (VWVKP) verifies that the cDNAs encode protein in pollen. Furthermore, a synthetic peptide based on the cDNA sequence designated RAE 4, which has the sequence EFPILGGITEVKDNDNSVDFC, stimulates ragweed allergic patient T cells, in in vitro proliferation assays (see Table 5 and sequences in Figures 8, 9 and 10).

EXAMPLE 2 Cross-hybridization Methods Used to Obtain Full-Length cDNAs

Antigen E is reported to be a protein of approximately 38,000 molecular weight and consequently a full-length cDNA encoding this protein must be at least 1.1 Kb in length (King, T.P et al., Arch. Biochem. Biophys., 212:127 (1981)). Consequently, IPC clones 1, 5 and 6 as well as UNC clones 1, 6 and 15 (designated Amb a IA, IB and IC, respectively) are not full-length.

In order to isolate full-length clones, nick-translated ^{32}p-labelled Amb a I cDNA probes were used to screen the ragweed flowerhead λgt10 (see Example 1) and the ragweed pollen λgt11 library using standard methods (Maniatis et al., Molecular Cloning, Cold Spring Harbor Laboratory, (1982)). Full-length or near full-length cDNAs encoding Amb a

IB (Figures 12 and 16) and Amb a IC (Figures 13 and 16) were isolated using this method (Figure 1B). One cross-hybridizing cDNA clone (called K6-5), which has an open reading frame of approximately 145 amino acids (amino acids 253-398; Figure 15), was found to be significantly divergent from the previously isolated Amb a IA, Amb a IB, Amb a IC and Amb a ID and showed complete agreement (Table 2) with a peptide sequence derived from conventionally purified antigen K (a gift from T.P. King, New York). Consequently, this partial cDNA was designated as Amb a II (see Figure 15 and below).

EXAMPLE 3 Polymerase Chain Reaction (PCR) Methods Used to Obtain Full-length cDNAs

PCR methods can be successfully used to isolate both rare message cDNA as well as genomic clones of known sequence (Mullis et al., Cold Spring Harbor Symposium Quant. Biol., 51:263-273 (1986)). 5' and 3' oligonucleotide primers were synthesized and used in a PCR experiment with ragweed pollen cDNA serving as template. The 5' primers were deduced from N-terminal conserved regions of Amb a IB (Figure 12) and Amb a IC (Figure 13). The 3' primers were deduced from Amb a IA specific (UNC clone 1, designated Amb a IA, Figure 2) and Amb a II specific (clone K6-5, partial 3' sequence of Figure 15) non-coding strand sequences at the 3' end of the cDNA. A third 3' primer used to PCR clone Amb a ID was derived from a conserved region of the C-terminal end of Amb a IA, B and C (amino-acids 395-398 corresponding to GAPC.stop). The

2049921

oligonucleotide primers used to amplify and clone Amb a IA, Amb a ID and Amb a II cDNAs are listed below:

N-terminal primers used to produce full-length Amb a IA and Amb a II (amino acids 10-15)

	ECORI	L	Y	F	T	L	A
IG38							
	GGGAATTC	TTG	TAT	TTT	ACC	TTA	GC
5'							3'

N-terminal primer used to produce truncated Amb a IA and Amb a II (amino acids 70-75)

	ECORI	D	C	A	Q	G	F
IG33							
	GGGAATTC	GAC	TGT	GCC	CAA	GGT	TTT	G

C-terminal primer used to produce full-length and truncated Amb_a IA (12-29 nucleotides of the noncoding strand 3' of the TAA stop codon; see Figure 2).

Pst I
IG32
GGGCTGCAG TCATTATAAGTGCTTAGT
5' 3'

C-terminal primer used to produce full-length Amb_a ID (corresponding to the C-terminal conserved GAPC encoding region). The primer is of the non-coding strand and includes the stop codon and an artifically introduced Pst I cloning site (see Figure 15).

Pst I
IG49
GGGCTGCAG TGC TTA GCA AGG TGC TCC
5' 3'

C-terminal primer used to produce full-length and truncated Amb_a II (44-76 nucleotides of the noncoding strand 3' of the TAA stop codon; see Figure 15).

Pst I
AgK2
GGGCTGCAG CGT GTC CAA ATC TAA TCA AAT GAA CAC TTA TGG
5' 3'

First strand cDNA was synthesized from 1 μg RNA with the cDNA synthesis system plus kit*(Amersham) using poly dT as primer. This single stranded cDNA was amplified using sets of primers (IG38 plus IG32;

*Trade-mark

IG33 plus IG32; IG38 plus IG49; IG38 plus AgK2; IG33 plus AgK2) according to methods recommended in the GeneAmp kit* (US Biochemicals, Cleveland, OH). The samples were amplified with a programmable thermal controller; the first five rounds of amplification consisted of denaturation at 94° for 30 sec., annealing of primers to the template at 45° for 1 min. 30 sec., and chain elongation at 70° for 4 min. The final 20 rounds of amplification consisted of denaturation as above, annealing at 55° for 1 min. 30 sec. and elongation as above. The PCR generated bands corresponding to the predicted size on an analytical gel and DNA sequencing confirmed that the cDNAs corresponded to full-length and truncated Amb a IA and Amb a II (Figures 11 and 15, respectively) and full-length Amb a ID (Figure 14).

The near full-length cDNAs presented in Figures 11 through 15, have their nucleotide sequences numbered such that the first nucleotide is designated number 1. Although some cDNAs start at what is probably the N-terminal methionine (Amb a IB, Figure 12; Amb a IC, Figure 13), some do not (Amb a IA, Figure 11; Amb a ID, Figure 14; Amb a II, Figure 15). Consequently, since the cDNAs are of different lengths, their nucleotide numbers do not necessarily correspond from one sequence to another. The universal genetic code is used to deduce the amino acid sequences from the cDNA sequences and the complete amino acid sequence comparisons between the clones are presented in Figure 16. In Figure 16, the amino acids are numbered sequentially from the

*Trade-mark

probably N-terminal methionine (designated number 1) of the Amb a IB sequence.

EXAMPLE 4 T Cell Responses to Ragweed Proteins and Peptides

Peripheral blood mononuclear cells (PBMC) were purified from 60 ml of heparinized blood from ragweed-allergic patients. PBMC were subsequently treated as described below, although in individual cases, the length of time of cultivation with IL-2 and IL-4 and the specific ragweed proteins and peptides used for stimulation varied. As an example, ten ml of patient 222 PBMC at 10^6/ml were cultured at 37°C for 7 days in the presence of 20 micrograms aqueous ragweed pollen extract/ml RPMI-1640 supplemented with 5% pooled human AB serum. Viable cells were purified by Ficoll-Hypaque centrifugation and cultured for three weeks at 5 units recombinant human IL-2/ml and 5 units recombinant human IL-4/ml. The resting T cells were then restimulated (secondary) with 20 micrograms aqueous ragweed pollen extract/ml at a density of 2 x 10^5 T cells/ml in the presence of X-irradiated (3500 RADS) autologous PBMC (5 x 10^5/ml) for three days, purified by Ficoll-Hypaque centrifugation and grown in 5 units IL-2/ml and 5 units IL-4/ml for two weeks. For assay, 2 x 10^4 resting secondary T cells were restimulated (tertiary) in the presence of 5 x 10^4 X-irradiated (3500 RADS) autologous PBMC or 2 x 10^4 autologous Epstein-Barr virus-transformed B cells (20,000 RADS) with various concentrations of allergen or their fragments in a volume of 200

microliters in 96-well round bottom assay plates for 3 days. Each well then received 1 microCurie tritiated (methyl) thymidine for 16 hours. The counts incorporated were collected onto glass fiber filters and processed for liquid scintillation counting. Figure 21 shows the results of a representative assay, demonstrating the reactivity and specificity of the T cell culture to ragweed pollen proteins. Antigens used: IPC aqueous pollen extract (pollen), Hollister-Stier ragweed skin test extract (RWST), ALK cat epithelium skin test extract (CST), affinity 4B5/B7 antibody purified (dialyzed) Amb a I (B7), and chromatographically purified Amb a I (Amb a I). Medium only control is shown as a line with no symbol. Alternatively, PBMC were sometimes carried only into a secondary assay (as outlined above for a tertiary assay) with 20 micrograms aqueous pollen extract for 7 days, followed by culture in 5 units IL-2/ml and 5 units IL-4/ml for 2-3 weeks. One ragweed allergic patient's T cells in secondary assay responded to pollen extract, RWST, B7 or Amb a I, but did not respond to CST or medium only (Figure 21). Secondary and tertiary assays of a panel of ragweed allergic patients were performed using synthetic peptides derived from the sequences of various ragweed pollen proteins. The results of several experiments are shown in Table 5. Three peptides (RAE16.6, RAE45.15, RAE24.E) which are derived from the sequence of three different Amb a I cDNA's could not stimulate any of the patients' T cells. Another four peptides (RAE15.6, RAE3.D, RAE28.1, RAE26.15) which are also dervied from the

sequence of the same three cDNA's could stimulate 35 to 58% of the patients' T cells. One peptide (RAE4) which is derived from the cDNA of IPC Clone 5 could also stimulate 25% of the patients' T cells. These results are consistent with the above cDNA's encoding ragweed pollen proteins. They further demonstrate the opportunity offered by knowledge of the protein structures of the Amb I/II family/ies to identify peptidic fragments which stimulate a response in T cells from ragweed allergic patients and others which do not. By this method it is possible to identify novel therapeutic and diagnostics entities for use in the treatment and the diagnosis of ragweed allergy.

2049921

TABLE 5

Human Ragweed-Allergic T Cell Responses to Ragweed Peptides

PEPTIDE[b] NAME	SEQUENCE BASED ON	NO. PATIENTS TESTED	NUMBER POSITIVE	POSITIVE %
05 RAE 16.6	Amb a IB	7	0	0
RAE 45.15	Amb a IC	2	0	0
RAE 24.E	Amb a IA	9	0	0
RAE 4	Clone #5	28	7	25
RAE 15.6	Amb a IB	20	7	35
10 RAE 3.D	Amb a IA	35	13	37
RAE 28.1	Amb a IA	33	17	52
RAE 26.15	Amb a IC	24	14	58

[a] Responses were scored as positive when the T cell proliferative response of ragweed pollen-specific T cells was greater than 2-fold above the culture medium control.

[b] Sequence of named peptide is as follows:

RAE 16.6	RTDKDLLENGAIC
RAE 45.15	LNQELVVNSDKTIDGRGVK
RAE 24.E	ETRRSLKTSGAYNIIDGCWRGKAD
20 RAE 4	EFPILGGITEVKDNDNSVDFC
RAE 15.6	YTVTSDKDDDVANC
RAE 3.D	GKADWAENRC
RAE 28.1	LENGAIFVASGVDPVLTPEQ
RAE 26.15	GFFQVVNNNYDRWGTYA

25 EXAMPLE 5 Antibody Binding to Recombinant Affinity Purified Amb a I, and Pollen Extract Derived Amb a I and Amb a II

Affinity purified Amb a I was electrophoresed, Western transfered (Towbin et al., Proc. Natl. Acad. Sci. USA, 76:4350 (1979)) and probed with a variety

of antibodies, including IgE from an allergic patient (Figure 17). In pollen extract Amb a I is not only present as an intact 38-KD species, but also characterized by its component 26-KD alpha chains and 12-KD beta chains which are formed by enzymatic cleavage. The intact 38-KD species and the alpha chain are clearly detected using rabbit anti-Amb a I, polyclonal affinity purified anti-RAE 16 and monoclonal anti-Amb a I JBIE3-4 (Figure 17; RAE 16 peptide has the sequence RTDKDLLENGAIC derived from amino-acids 342-353 of Amb a IB, Figure 16). Affinity purified Amb a I (partial sequence presented in Table 1) as well as chromatographically purified Amb a II (partial sequence presented in Table 2) are bound on Western blots by these antibodies as well as by patient IgE (Figure 17). The goat anti-Amb a I polyclonal antibody also binds multiple Amb a I and Amb a II species on a two dimensional Western blot of pollen extract (Figure 18). The Western blot was performed as outlined below.

Isoelectric focusing was done on a Hoeffer gel apparatus with 15μg of crude soluble pollen protein. The gel consisted of 7.5% acrylamide with 3.5% Pharmalytes* pH 4.5-5.3 (Pharmacia) and 3.5% Ampholines* pH 3.5-10 (LKB), run at 13W for 3.5 hours until a constant voltage was reached. The gel section was placed on a slab of 10% acrylamide SDS-PAGE and electrophoresed for 3.5 hours at 40 mA according to the protocol cited. The proteins were transferred overnight in phosphate buffer to 0.1

*Trade-mark

B

micron nitrocellulose (Schleicher and Schuell) at 0.2A. The blot was rinsed in blot solution (25mM Tris-HCl pH 7.5, 0.171 M NaCl, 0.05% Tween-20*; Sigma). The first antibody incubation was overnight at room temperature with a 1:000 dilution of goat anti-Amb a I IgG (obtained from Dr. David Marsh) in blot solution. The excess first antibody was removed with three 15 minute rinses with blot solution. The second antibody was a 1:2,500 dilution of biotinylated swine anti-goat IgG (Boehringer-Manneheim) in blot solution for two hours. The blot was then rinsed with blot solution three times for 15 minutes and incubated for 1 hr in blot solution with 2μCi I^{125} streptavidin (Amersham). The blots were rinsed with blot solution until the waste wash was down to background. The blot was then exposed to film at -80°C overnight. In the case of one-dimensional SDS-PAGE Western blots (Figures 17, 19 and 20) the isoelectric focusing step was omitted. When human sera was used to probe the Western blots (Figures 17 and 20), 10% human plasma in 1% milk in blot solution was incubated overnight with the blot prior to using as second antibody biotinylated goat anti-human IgE.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described specifically

*Trade-mark

B

herein. Such equivalents are intended to be encompassed in the scope of the following claims.

CLAIMS:

1. An isolated protein or peptide capable of influencing ragweed T cell responses in ragweed-sensitive individuals, the protein or peptide being selected from:

(a) a protein or peptide comprising at least a T cell epitope-containing portion, or a B cell epitope-containing portion, of the amino acid sequence of *Amb a* IA shown in Figure 11; or

(b) a protein or peptide comprising at least a T cell epitope-containing portion, or a B cell epitope-containing portion, of the amino acid sequence of *Amb a* IB shown in Figure 12; or

(c) a protein or peptide comprising at least a T cell epitope-containing portion, or a B cell epitope-containing portion, of the amino acid sequence of *Amb a* IC shown in Figure 13;

(d) a protein or peptide comprising at least a T cell epitope-containing portion, or a B cell epitope-containing portion, of the amino acid sequence of *Amb a* ID shown in Figure 14; or

(e) a protein or peptide which is a polymorphic variant of any one of (a) to (d).

2. The isolated protein or peptide according to claim 1, produced by chemical synthesis.

3. The protein or peptide according to claim 1 or 2 wherein:

(a) the protein or peptide is *Amb a* IA as shown in Figure 11; or

(b) the protein or peptide is *Amb a* IB as shown in Figure 12; or

(c) the protein or peptide is *Amb a* IC as shown in Figure 13; or

(d) the protein or peptide is *Amb a* ID as shown in Figure 14; or

(e) the protein or peptide is a polymorphic variant of any one of (a)-(d).

4. The protein or peptide according to claim 1 or 2 wherein the protein or peptide comprises an amino acid sequence selected from:

(a) EFPILGGITEVKDNDNSVDFC;

(b) YTVTSDKDDDVANC;

(c) GKADWAENRC; and

(d) GFFQVVNNNYDRWGTYA.

5. The protein or peptide according to claim 1 or 2 wherein the *Amb a* I protein or peptide is a giant, desert or western ragweed *Amb a* I variant.

6. The protein or peptide according to any one of claims 1 to 3 and 5 wherein the protein or peptide has been modified by amino acid substitution, addition or deletion to alter immunogenicity or reduce allergenicity.

7. A protein or peptide capable of influencing ragweed T cell responses in ragweed-sensitive individuals, wherein the protein or peptide comprises an amino acid

sequence which is a modification of an amino acid sequence selected from:

(a) EFPILGGITEVKDNDNSVDFC;

(b) YTVTSDKDDDVANC;

(c) GKADWAENRC; and

(d) GFFQVVNNNYDRWGTYA;

wherein the modification is an amino acid substitution, addition or deletion to alter immunogenicity or reduce allergenicity.

8. A composition capable of influencing T cell responses in ragweed-sensitive individuals comprising one or more isolated protein or peptide according to any one of claims 1-7 and a buffer, carrier or adjuvant.

9. The composition according to claim 8 which comprises a combination of different *Amb a* I proteins or peptides.

10. A process for producing a ragweed protein/peptide allergen comprising the steps of:

(a) identifying a DNA sequence encoding ragweed allergen which hybridizes under conditions of high stringency with the DNA sequence shown in Figure 11, 12, 13 or 14;

(b) inserting the DNA identified in step (a) into an expression vector;

(c) inserting the expression vector of step (b) into a host cell;

(d) culturing the host cell of step (c) under conditions appropriate for expression of the DNA inserted in step (b); and

(e) recovering the expressed product.

11. The process according to claim 10 wherein the DNA sequence identified in step (a) has at least 40% homology with the DNA sequences shown in Figure 11, 12, 13 or 14.

12. A ragweed protein/peptide allergen produced by the process according to claim 10 or 11.

13. A process for preparing a modified ragweed protein, peptide, or allergen comprising the steps of:

(a) providing the protein or peptide according to any one of claims 1 to 7 or the allergen according to claim 12;

(b) modifying the protein, peptide or allergen of step (a) by amino acid substitution, addition or deletion to produce a modified protein, peptide, or allergen.

14. Modified ragweed protein, peptide or allergen produced by the process according to claim 13.

15. An isolated nucleic acid comprising a nucleotide sequence which encodes the protein or peptide according to any one of claims 1 to 7 or the allergen according to claim 12 or 14.

16. The nucleic acid according to claim 15 comprising the coding region shown in Figure 11, 12, 13 or 14.

17. The nucleic acid according to claim 15 comprising the nucleotide sequence shown in Figure 11, 12, 13 or 14.

18. A composition comprising the allergen according to claim 12 or 14 and a buffer, carrier or adjuvant.

19. Use of the protein or peptide according to any one of claims 1 to 7, the allergen according to claim 12 or 14, or the composition according to any one of claims 8, 9 and 18, for ragweed allergy therapy, or for the treatment of ragweed pollen sensitivity in an individual.

20. Use of the protein or peptide according to any one of claims 1 to 7, the allergen according to claim 12 or 14 or the composition according to any one of claims 8, 9 and 18 for *in vivo* diagnosis of ragweed allergy, or for detecting ragweed pollen sensitivity in an individual.

21. A method of detecting sensitivity in an individual to a ragweed pollen protein/peptide, comprising combining a blood sample of the individual with the protein or peptide according to any one of claims 1 to 7, the allergen according to claim 12 or 14, or the composition according to any one of claims 8, 9 and 18, under conditions appropriate for binding of blood components with the protein, the peptide, or the allergen, and determining the extent to which binding occurs.

22. The method according to claim 21 wherein the extent to which binding occurs is determined by assessing T cell function, T cell proliferation, B cell function, binding of the protein or the peptide to antibodies present in the blood, or a combination thereof.

23. An isolated protein or peptide or a polymorphic variant thereof, capable of influencing ragweed T cell response in ragweed-sensitive individuals, the protein or peptide comprising at least a T cell epitope-containing

portion, or a B cell epitope-containing portion, of the amino acid sequence of *Amb a* II shown in Figure 15.

24. The isolated protein or peptide or polymorphic variant according to claim 23 for use in therapy, prophylaxis or diagnosis.

25. The protein or peptide according to claim 23 or 24 wherein the protein or peptide is *Amb a* II as shown in Figure 15 or a polymorphic variant thereof.

26. The protein or peptide according to any one of claims 23 to 25 wherein the protein or peptide has been modified by amino acid substitution, addition or deletion to alter immunogenicity or reduce allergenicity.

27. A process for producing a ragweed protein/peptide allergen comprising the steps of:

(a) identifying a DNA sequence encoding ragweed allergen which hybridizes under conditions of high stringency with the DNA sequence shown in Figure 15;

(b) inserting the DNA identified in step (a) into an expression vector;

(c) inserting the expression vector of step (b) into a host cell;

(d)culturing the host cell of step (c) under conditions appropriate for expression of the DNA inserted in step (b); and

(e) recovering the expressed product.

28. The process according to claim 27 wherein the DNA sequence identified in step (a) has at least 40% homology with the DNA sequence shown in Figure 15.

29. A process for preparing a modified ragweed protein or peptide comprising the steps of:

(a) providing the protein or peptide according to any one of claims 23 to 25;

(b) modifying the protein or peptide of step (a) by amino acid substitution, addition or deletion to produce a modified protein or peptide.

30. Modified ragweed protein or peptide produced by the process according to claim 29.

31. An isolated nucleic acid comprising a nucleotide sequence which encodes the protein or peptide according to any one of claims 24 to 26.

32. The nucleic acid according to claim 31 comprising the coding region shown in Figure 15.

33. The nucleic acid according to claim 31 comprising the nucleotide sequence shown in Figure 15.

34. A composition comprising the protein or peptide according to any one of claims 23 to 26 and a buffer, carrier or adjuvant.

35. The protein or peptide according to any one of claims 23 to 26 for use in therapy, or for use in the treatment of ragweed pollen sensitivity in an individual.

36. The protein or peptide according to any one of claims 23 to 26 for use in *in vivo* diagnosis, or for use in detecting ragweed pollen sensitivity in an individual.

37. A method of detecting sensitivity in an individual to a ragweed pollen protein/peptide, comprising combining a blood sample of the individual with the protein or peptide

according to any one of claims 23 to 26 under conditions appropriate for binding of blood components with the protein or the peptide, and determining the extent to which binding occurs.

38. The method according to claim 37 wherein the extent to which binding occurs is determined by assessing T cell function, T cell proliferation, B cell function, binding of the protein or the peptide to antibodies present in the blood, or a combination thereof.

SMART & BIGGAR

OTTAWA, CANADA

PATENT AGENTS

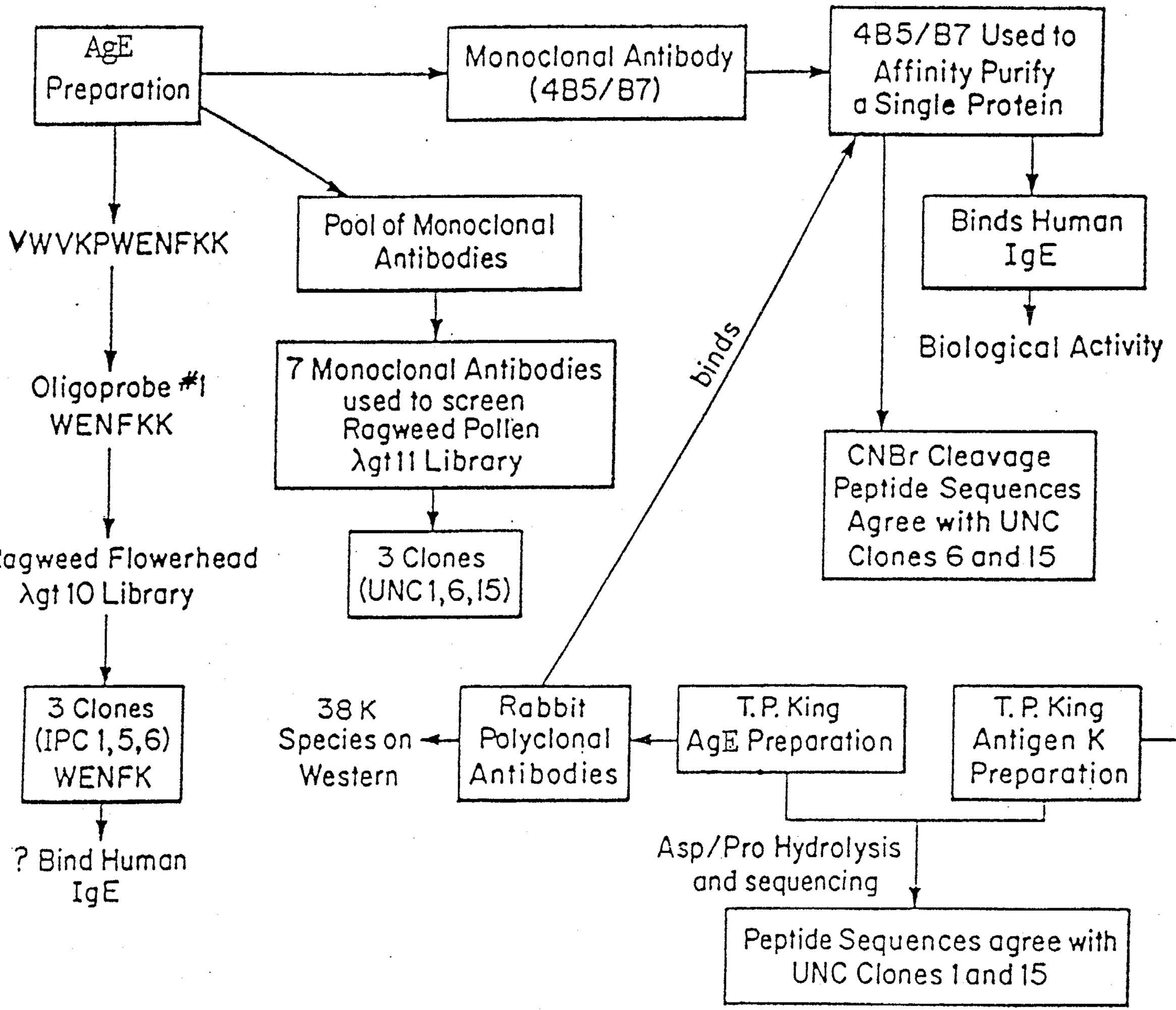


FIG. 1A

2049921

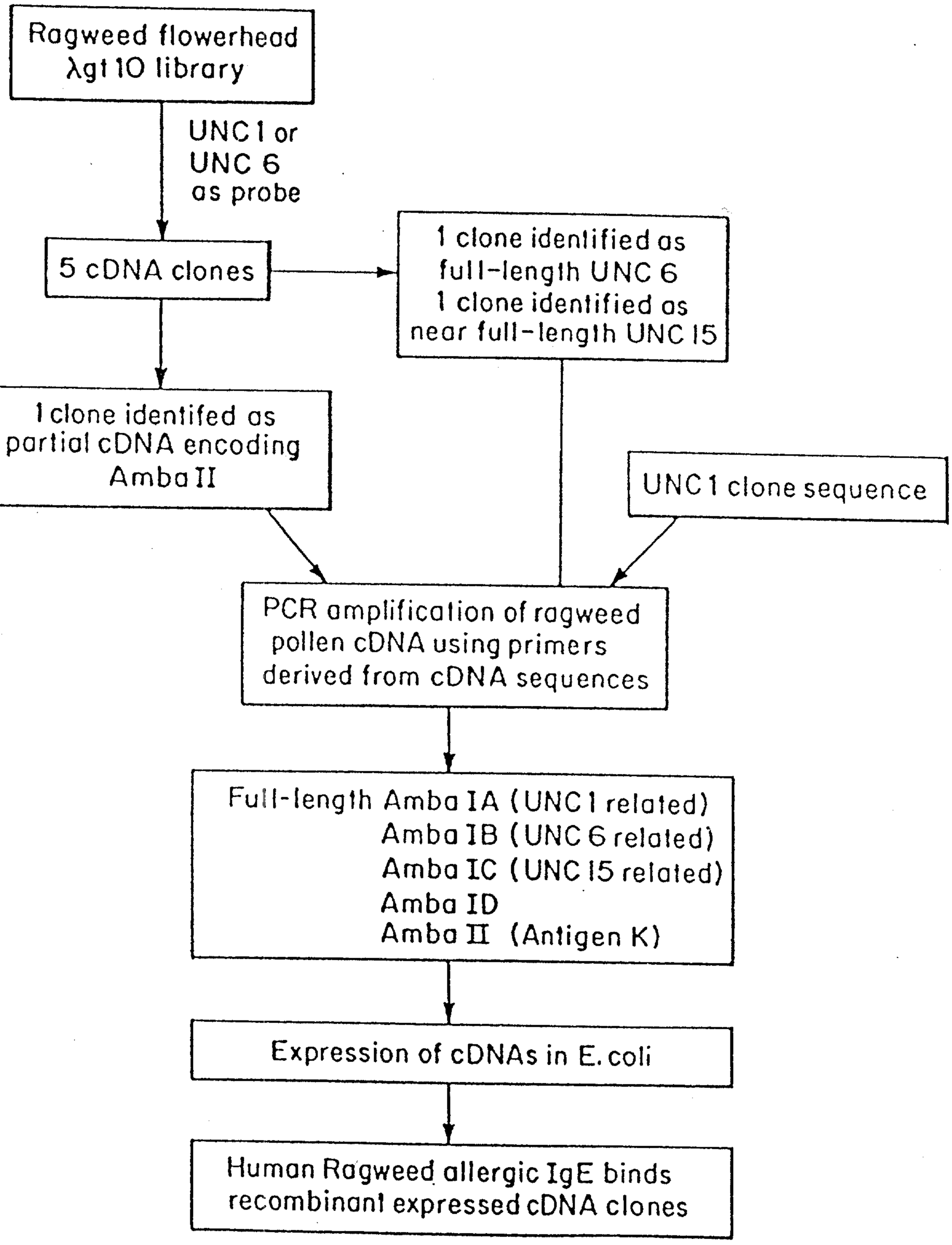

Ragweed flowerhead
λgt 10 library
UNC 1 or
UNC 6
as probe
5 cDNA clones
1 clone identified as
full-length UNC 6
1 clone identified as
near full-length UNC 15
1 clone identifed as
partial cDNA encoding
Amba II
UNC 1 clone sequence
PCR amplification of ragweed
pollen cDNA using primers
derived from cDNA sequences
Full-length Amba IA (UNC 1 related)
Amba IB (UNC 6 related)
Amba IC (UNC 15 related)
Amba ID
Amba II (Antigen K)
Expression of cDNAs in E. coli
Human Ragweed allergic IgE binds
recombinant expressed cDNA clones

FIG. 1B

DE UNC CLONE 1

FIG.2

```
          10         20         30         40         50         60
          |          |          |          |          |          |
GAATTCGGCTGGAGAACGAATAAAGACGTGCTTGAAAATGGTGCTATTTTTGTTGCATCC

 E  F  G  W  R  T  N  K  D  V  L  E  N  G  A  I  F  V  A  S

          70         80         90        100        110        120
          |          |          |          |          |          |
GGGGTCGATCCAGTGCTAACCCCTGAGCAAAGCGCAGGGATGATTCCAGCCGAACCAGGA

 G  V  D  P  V  L  T  P  E  Q  S  A  G  M  I  P  A  E  P  G

          130        140        150        160        170        180
          |          |          |          |          |          |
GAGTCCGCTCTAAGCCTCACTAGTAGTGCTGGTGTACTCTCATGCCAACCCGGAGCACCT

 E  S  A  L  S  L  T  S  S  A  G  V  L  S  C  Q  P  G  A  P

          190        200        210        220        230        240
          |          |          |          |          |          |
TGCTAAGCACCCGACCAATTACTAAGCACTTATAATGATCATTAATACTTTTTTTTATTT

 C  -  A  P  D  Q  L  L  S  T  Y  N  D  H  -  Y  F  F  L  F

          250        260        270        280        290        300
          |          |          |          |          |          |
TATTTTTGATATTTTATATGTACTAAGGTAATGGAAATGAACCTTTACCTTCTAGTACTC

 Y  F  -  Y  F  I  C  T  K  V  M  E  M  N  L  Y  L  L  V  L

          310        320
          |          |
TAAAAAAAAAAAAAACCGAATTC   (SEQ. ID NO:56)

 -  K  K  K  K  P  N      (SEQ. ID NO:57)
```

C

FIG.3A

```
**************************************************
* TRANSLATION OF A NUCLEIC ACID SEQUENCE *
**************************************************
Done on DNA sequence AMB_A_IB_K.
DE   SEQUENCE OF AMB A IB CLONE.

Total number of bases is: 1328.
Analysis done on the complete sequence.
Done on (absolute) phase(s): 1.
Using the Universal genetic code.

                                        10                  20                  30                  40                  50                  60
                                        |                   |                   |                   |                   |                   |
TACATCTTGTATTTTACCTTAGCCCTTGTCACTTTGCTGCAACCTGTTCGTTCTGCAGAA
 Y   I   L   Y   F   T   L   A   L   V   T   L   L   Q   P   V   R   S   A   E

                                        70                  80                  90                  100                 110                 120
                                        |                   |                   |                   |                   |                   |
GATGTTGAAGAATTCTTACCTTCAGCTAACGAAACAAGGAGGAGCCTGAAAGCATGTGAA
 D   V   E   E   F   L   P   S   A   N   E   T   R   R   S   L   K   A   C   E

                                        130                 140                 150                 160                 170                 180
                                        |                   |                   |                   |                   |                   |
GCACACAACATTATAGACAAGTGCTGGAGGTGCAAAGCCGATTGGGCGAATAACCGACAA
 A   H   N   I   I   D   K   C   W   R   C   K   A   D   W   A   N   N   R   Q

                                        190                 200                 210                 220                 230                 240
                                        |                   |                   |                   |                   |                   |
GCGTTAGCCGATTGTGCCCAAGGTTTTGCAAAGGGAACCTACGGTGGAAAACATGGTGAT
 A   L   A   D   C   A   Q   G   F   A   K   G   T   Y   G   G   K   H   G   D

                                        250                 260                 270                 280                 290                 300
                                        |                   |                   |                   |                   |                   |
GTCTACACGGTCACCAGTGATAAAGATGATGATGTTGCAAATCCAAAAGAAGGCACACTC
 V   Y   T   V   T   S   D   K   D   D   D   V   A   N   P   K   E   G   T   L

                                        310                 320                 330                 340                 350                 360
                                        |                   |                   |                   |                   |                   |
CGGTTTGCTGCTGCCCAAAACAGGCCCTTGTGGATCATTTTTAAAAGAAATATGGTGATT
 R   F   A   A   A   Q   N   R   P   L   W   I   I   F   K   R   N   M   V   I

                                        370                 380                 390                 400                 410                 420
                                        |                   |                   |                   |                   |                   |
CATTTGAATCAAGAGCTTGTCGTAAACAGCGACAAGACCATCGATGGCCGAGGGGTGAAA
 H   L   N   Q   E   L   V   V   N   S   D   K   T   I   D   G   R   G   V   K
```

C

FIG.3B

```
          430        440        450        460        470        480
GTTAACATCGTTAACGCCGGTCTCACCCTCATGAATGTCAAGAATATAATCATTCATAAC
 V  N  I  V  N  A  G  L  T  L  M  N  V  K  N  I  I  I  H  N

          490        500        510        520        530        540
ATAAATATCCATGATATTAAAGTTTGTCCAGGAGGCATGATTAAGTCCAACGATGGTCCA
 I  N  I  H  D  I  K  V  C  P  G  G  M  I  K  S  N  D  G  P

          550        560        570        580        590        600
CCAATTTTAAGACAACAAAGTGATGGTGATGCTATAAATGTTGCTGGTAGTTCACAAATA
 P  I  L  R  Q  Q  S  D  G  D  A  I  N  V  A  G  S  S  Q  I

          610        620        630        640        650        660
TGGATCGACCATTGCTCGCTCAGTAAGGCTTCCGATGGGCTGCTCGATATCACCCTCGGC
 W  I  D  H  C  S  L  S  K  A  S  D  G  L  L  D  I  T  L  G

          670        680        690        700        710        720
AGCTCACACGTGACCGTTTCCAACTGCAAATTCACCCAACACCAATTTGTATTATTGCTC
 S  S  H  V  T  V  S  N  C  K  F  T  Q  H  Q  F  V  L  L  L

          730        740        750        760        770        780
GGGGCTGATGACACCCATTATCAAGATAAAGGCATGCTAGCAACGGTAGCATTCAACATG
 G  A  D  D  T  H  Y  Q  D  K  G  M  L  A  T  V  A  F  N  M

          790        800        810        820        830        840
TTCACCGATCACGTTGACCAAAGAATGCCTAGATGTAGATTTGGGTTTTTCCAAGTCGTT
 F  T  D  H  V  D  Q  R  M  P  R  C  R  F  G  F  F  Q  V  V

          850        860        870        880        890        900
AACAACAACTACGACAGATGGGGAACGTACGCCATCGGTGGTAGCTCGGCCCCAACTATA
 N  N  N  Y  D  R  W  G  T  Y  A  I  G  G  S  S  A  P  T  I

          910        920        930        940        950        960
CTCAGCCAAGGGAACAGATTCTTCGCCCCCGATGATATCATCAAGGAAAATGTCTTAGCG
 L  S  Q  G  N  R  F  F  A  P  D  D  I  I  K  E  N  V  L  A
```

C

FIG.3C

970 980 990 1000 1010 1020
AGGACTGGTACTGGCAACGCAGAGTCGATGTCGTGGAACTGGAGAACAGATAAAGACTTG
R T G T G N A E S M S W N W R T D K D L

1030 1040 1050 1060 1070 1080
CTTGAAAATGGTGCTATTTTTCTCCCATCCGGGTCTGATCCAGTGCTAACCCCTGAGCAA
L E N G A I F L P S G S D P V L T P E Q

1090 1100 1110 1120 1130 1140
AAAGCAGGGATGATTCCAGCTGAACCAGGAGAAGCCGTTCTAAGACTCACTAGTAGTGCT
K A G M I P A E P G E A V L R L T S S A

1150 1160 1170 1180 1190 1200
GGTGTACTCTCATGCCATCAAGGAGCACCTTGCTAAGCACCTGGCCAATTCCTAAGCTTT
G V L S C H Q G A P C - A P G Q F L S F

1210 1220 1230 1240 1250 1260
TATAATAATCATAAATACTTATTTTATTTTATTTTTGATATTTTATATGAACCATTACGT
Y N N H K Y L F Y F I F D I L Y E P L R

1270 1280 1290 1300 1310 1320
TCAAGTACTCTATTAACATGTTTTAAATTCATAAGAGTTTATTGATAAAAAAAAAAAAAA
S S T L L T C F K F I R V Y - - K K K K

CCGAATTC (SEQ. ID NO:58)
P N (SEQ. ID NO:59)

C

FIG.4A

```
**************************************************
* TRANSLATION OF A NUCLEIC ACID SEQUENCE *
**************************************************

Done on DNA sequence KKLAPPER1.

DE   UNC CLONE 1

Total number of bases is: 323.
Analysis done on the complete sequence.
Done on (absolute) phase(s): 1.
Using the Universal genetic code.

                                        ----------

          10        20        30        40        50        60
          |         |         |         |         |         |
GAATTCGGCTGGAGAACGAATAAAGACGTGCTTGAAAATGGTGCTATTTTTGTTGCATCC

  E   F   G   W   R   T   N   K   D   V   L   E   N   G   A   I   F   V   A   S

          70        80        90       100       110       120
          |         |         |         |         |         |
GGGGTCGATCCAGTGCTAACCCCTGAGCAAAGCGCAGGGATGATTCCAGCCGAACCAGGA

  G   V   D   P   V   L   T   P   E   Q   S   A   G   M   I   P   A   E   P   G

          130       140       150       160       170       180
          |         |         |         |         |         |
GAGTCCGCTCTAAGCCTCACTAGTAGTGCTGGTGTACTCTCATGCCAACCCGGAGCACCT

  E   S   A   L   S   L   T   S   S   A   G   V   L   S   C   Q   P   G   A   P

          190       200       210       220       230       240
          |         |         |         |         |         |
TGCTAAGCACCCGACCAATTACTAAGCACTTATAATGATCATTAATACTTTTTTTTTATTT

  C   -   A   P   D   Q   L   L   S   T   Y   N   D   H   -   Y   F   F   L   F

          250       260       270       280       290       300
```

FIG.4B

```
                |         |         |         |         |         |
TATTTTTGATATTTTATATGTACTAAGGTAATGGAAATGAACCTTTACCTTCTAGTACTC
 Y  F  -  Y  F  I  C  T  K  V  M  E  M  N  L  Y  L  L  V  L   (SEQ. ID NO:61)

          310       320
           |         |
TAAAAAAAAAAAAAACCGAATTC   (SEQ. ID NO:60)
```

2049921

FIG.5

IPC CLONE #1

From:

```
                10        20        30        40        50        60
                |         |         |         |         |         |
  1 GAATTCCGAT TCTTGGAGGA ATTACCGAAG TTAAAGACAA TGATAACAGC GTCGATTTCG
    CTTAAGGCTA AGAACCTCCT TAATGGCTTC AATTTCTGTT ACTATTGTCG CAGCTAAAGC

 61 ACGAGCTTGC TAAATTCGCC ATCGCTGAAC ACAACAAGAA GGAGAATGCT GCTCTGGAGT
    TGCTCGAACG ATTTAAGCGG TAGCGACTTG TGTTGTTCTT CCTCTTACGA CGAGACCTCA

121 TTGGAAAAGT AATAGAAAAA AAGCAGCAGG CGGTACAGGG CACCATGTAT TATATAAAAG
    AACCTTTTCA TTATCTTTTT TTCGTCGTCC GCCATGTCCC GTGGTACATA ATATATTTTC

181 TGGAAGCAAA TGATGGTGGT GAGAAGAAAA CTTATGAAGC CAAGGTGTGG GTTAAGCTAT
    ACCTTCGTTT ACTACCACCA CTCTTCTTTT GAATACTTCG GTTCCACACC CAATTCGATA

241 GGGAAAATTT CAAGGAATTG CAGGAACTCA AACTTGTTTG ATGGACGGGT GTGTGCTATG
    CCCTTTTAAA GTTCCTTAAC GTCCTTGAGT TTGAACAAAC TACCTGCCCA CACACGATAC

301 ACAAAATAGC TCGAGCAGGT GAAGCATGAA TGTATAAATA TTCTTTTTAA GTTTAATAAT
    TGTTTTATCG AGCTCGTCCA CTTCGTACTT ACATATTTAT AAGAAAAATT CAAATTATTA

361 AAACATTTCT TGTAATATGG TACAGGTTTA TGTACTTTGG TATGTATAAC AGAAAACATA
    TTTGTAAAGA ACATTATACC ATGTCCAAAT ACATGAAACC ATACATATTG TCTTTTGTAT

421 TCATAAATTC AAACTTAGAA TTTTGGGAAT TC   (SEQ. ID NO:62)
    AGTATTTAAG TTTGAATCTT AAAACCCTTA AG   (SEQ. ID NO:63)
```

Total number of bases is: 452.
DNA sequence composition: 162 A; 59 C; 107 G; 124 T;

Sequence name: NIPC_CLONE1.

FIG.6

IPC CLONE #5

From:

```
                10         20         30         40         50         60
                |          |          |          |          |          |
  1 GAATTCCCGA TTCTTGGAGG AATTACCGAA GTTAAAGACA ATGATAACAG CGTCGATTTC
    CTTAAGGGCT AAGAACCTCC TTAATGGCTT CAATTTCTGT TACTATTGTC GCAGCTAAAG

 61 GACGAGCTTG CTAAATTCGC CATCACTGAA CACAACAAGA AGGAGAATGC TGCTCTGGAG
    CTGCTCGAAC GATTTAAGCG GTAGTGACTT GTGTTGTTCT TCCTCTTACG ACGAGACCTC

121 TTTGGAAAAG TAATAGAAAA AAAGCAGCAG GCGGTACAGG GCACCATGTA TTATATAAAA
    AAACCTTTTC ATTATCTTTT TTTCGTCGTC CGCCATGTCC CGTGGTACAT AATATATTTT

181 GCGGAAGCAA ATGATGGTGG TGAGAAGAAA ACTTATGAAG CCAAGGTGTG GGTTAAGCTA
    CGCCTTCGTT TACTACCACC ACTCTTCTTT TGAATACTTC GGTTCCACAC CCAATTCGAT

241 TGGGAAAATT TCAAGGAATT TGCAAGGAAC TCAAACCTTG TTTGATGATG CCACCTCACC
    ACCCTTTTAA AGTTCCTTAA ACGTTCCTTG AGTTTGGAAC AAACTACTAC GGTGGAGTGG

301 TTAACTCCAT ATGGACGGTG TGCTATGACA AAATAGCTCA AGGAGGTGAA GCATAAATGT
    AATTGAGGTA TACCTGCCAC ACGATACTGT TTTATCGAGT TCCTCCACTT CGTATTTACA

361 ATAAATATTC TTTTTAAGTT TAATAATAAA CATTTCTTGT AATATAGTAC AAGTTTATGT
    TATTTATAAG AAAAATTCAA ATTATTATTT GTAAAGAACA TTATATCATG TTCAAATACA

421 ACTTTGGTAT GTATAACAGA AAACATATCA TAAATTCAAA CTTAATGTTT TTTTTTCTCG
    TGAAACCATA CATATTGTCT TTTGTATAGT ATTTAAGTTT GAATTACAAA AAAAAAGAGC

481 CGGAATTC   (SEQ. ID NO:64)
    GCCTTAAG   (SEQ. ID NO:65)
```

Total number of bases is: 488.
DNA sequence composition: 174 A; 74 C; 103 G; 137 T;

Sequence name: NIPC_CLONE5.

C

IPC CLONE #6

FIG.7

From:

```
                10         20         30         40         50         60
                |          |          |          |          |          |
     1 TCGATTCGCT GTCGATGAAC ACAACAAGAA GCAGAATACC CTGCTGGAAT TTAAGAAGGT
       AGCTAAGCGA CAGCTACTTG TGTTGTTCTT CGTCTTATGG GACGACCTTA AATTCTTCCA

    61 ACTGAATACA AAGGAGCAGG TAGTAGCTGG TATAATGTAT TATATCACAC TTGAAGCAAC
       TGACTTATGT TTCCTCGTCC ATCATCGACC ATATTACATA ATATAGTGTG AACTTCGTTG

   121 TGATGGTGGT GAGAAAAAGA CTTATGAAGC CAAGGTTTGG GTTAAGCCAT GGGAAAACTT
       ACTACCACCA CTCTTTTTCT GAATACTTCG GTTCCAAACC CAATTCGGTA CCCTTTTGAA

   181 CAAAGAATTC  (SEQ. ID NO:66)
       GTTTCTTAAG  (SEQ. ID NO:67)
```

Total number of bases is: 190.
DNA sequence composition: 69 A; 29 C; 47 G; 45 T;

Sequence name: NIPC_CLONE6.

20429921

FIG.8

IPC CLONE 1

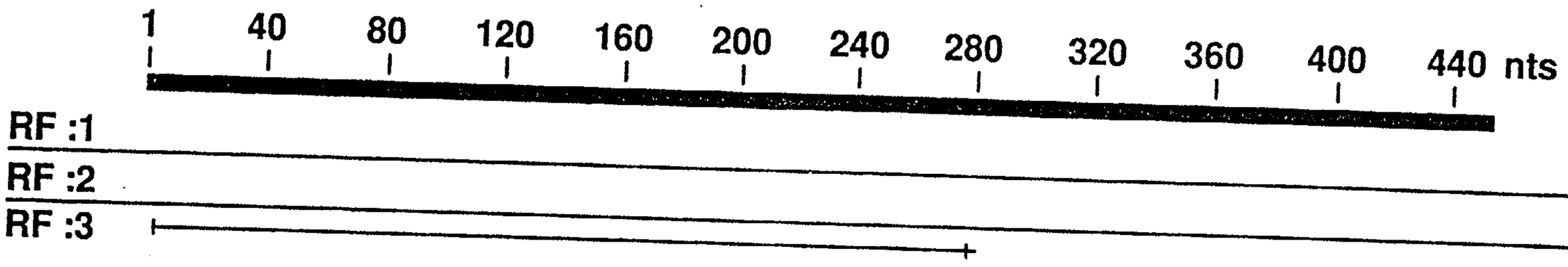
The DNA inspector Ile

Open Reading Frame Analysis

DNA > CLONE #1, FINAL
Minimum analysis length: 80 amino acids

DNA Length: 452 nts
not starting with ATG

1 open reading frame found.



The DNA Inspector Ile

Open Reading Frame Analysis

Analysis of peptide # 1 Reading frame: 3
starts at nt #: +3
number of amino acids: 92

One letter representation :

1 IPILGGITEV KDNDNSVDFD ELAKFAIAEH NKKENAALEF GKVIEKKQQA VQGTMYYIKV EANDGGEKKT
71 YEAKVWVKLW ENFKELQELK LV* (SEQ ID NO:68)

20499921

FIG.9
IPC Clone 5

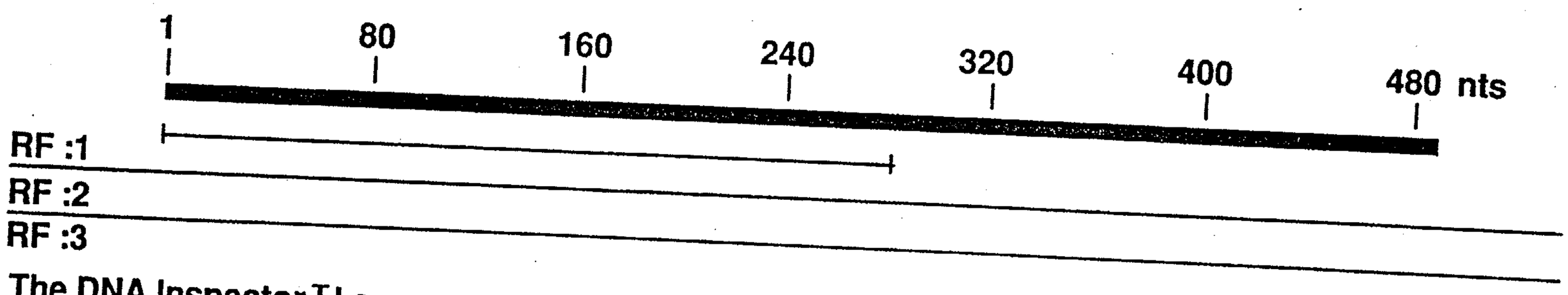
The DNA Inspector IIe

Open Reading Frame Analysis

DNA > CLONE #5, FINAL
Minimum analysis length: 80 amino acids

DNA Length: 488 nts
not starting with ATG

1 open reading frame found.



The DNA Inspector IIe

Open Reading Frame Analysis

Analysis of peptide # 1 Reading frame: 1
starts at nt #1: +1
number of amino acids: 94

One letter representation :

1 EFPILGGITE VKDNDNSVDF DELAKFAITE HNKKENAALE FGKVIEKKQQ AVQGTMYYIK AEANDGGEKK
71 TYEAKVWVKL WENFKEFARN SNLV* (SEQ ID NO:69)

2049921

FIG.10

IPC Clone 6

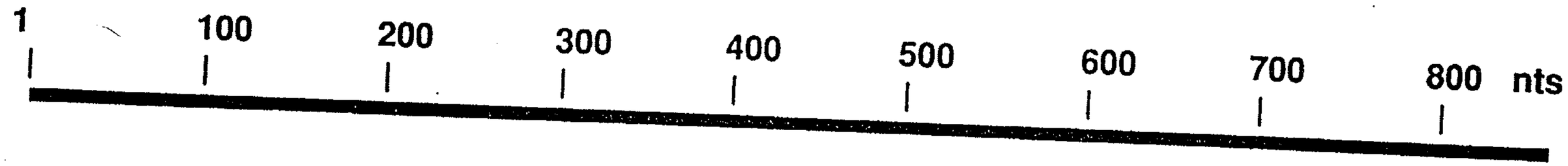
The DNA Inspector IIe

Open Reading Frame Analysis

DNA > #6, FINAL
Minimum analysis length: 80 amino acids

DNA Length: 855 nts
not starting with ATG

1 open reading frame found.



The DNA Inspector IIe

Open Reading Frame Analysis

Analysis of peptide # 1 Reading frame: 2
starts at nt #: +311
number of amino acids: 99

One letter representation :

1
71 EKKTYEAKVW VKPWENFKEF VDEHNKKQNT LLEFKKVLNT KEQVVAGIMY YITLEATDGG
(SEQ ID NO:70)

2049921

```
**************************************************
* TRANSLATION OF A NUCLEIC ACID SEQUENCE *
**************************************************
Done on DNA sequence AMB_A_IA.

DE   SEQUENCE OF AMB A IA CLONE.

Total number of bases is: 1196.
Analysis done on the complete sequence.
Done on (absolute) phase(s): 1.
Using the Universal genetic code.
```

FIG.11A

```
                                                            10                  20                  30                  40                  50                  60
TTGTATTTTACCTTAGCCCTTGTCACTTTGCTGCAACCTGTTCGTTCTGCCGAAGATCTC
 L   Y   F   T   L   A   L   V   T   L   L   Q   P   V   R   S   A   E   D   L

                                                            70                  80                  90                  100                 110                 120
CAGGAAATCTTACCAGTTAACGAAACAAGGAGGCTGACAACAAGTGGAGCATACAACATT
 Q   E   I   L   P   V   N   E   T   R   R   L   T   T   S   G   A   Y   N   I

                                                            130                 140                 150                 160                 170                 180
ATAGACGGGTGCTGGAGGGGCAAAGCCGATTGGGCGGAAAACCGAAAAGCGTTAGCCGAT
 I   D   G   C   W   R   G   K   A   D   W   A   E   N   R   K   A   L   A   D

                                                            190                 200                 210                 220                 230                 240
TGTGCCCAAGGTTTTGGGAAGGGAACAGTGGGCGGAAAAGATGGTGATATATACACGGTC
 C   A   Q   G   F   G   K   G   T   V   G   G   K   D   G   D   I   Y   T   V

                                                            250                 260                 270                 280                 290                 300
ACCAGTGAGCTAGATGATGATGTTGCAAATCCAAAAGAAGGCACACTCCGGTTTGGTGCC
 T   S   E   L   D   D   D   V   A   N   P   K   E   G   T   L   R   F   G   A

                                                            310                 320                 330                 340                 350                 360
GCCCAAAACAGGCCCTTGTGGATCATTTTTGAAAGAGATATGGTGATTCGTTTGGATAAA
 A   Q   N   R   P   L   W   I   I   F   E   R   D   M   V   I   R   L   D   K

                                                            370                 380                 390                 400                 410                 420
GAGATGGTGGTAAACAGTGACAAGACCATCGATGGCCGAGGGGCGAAAGTTGAAATCATT
 E   M   V   V   N   S   D   K   T   I   D   G   R   G   A   K   V   E   I   I
```

C

FIG.11B

```
                                                                    430
                                                                     |
AACGCTGGTTTCACCCTTAATGGTGTCAAGAATGTAATCATTCATAACATAAATATGCAT
 N  A  G  F  T  L  N  G  V  K  N  V  I  I  H  N  I  N  M  H

                                                                    490
                                                                     |
GATGTTAAAGTGAATCCAGGAGGCCTGATTAAGTCCAACGATGGTCCAGCAGCTCCAAGA
 D  V  K  V  N  P  G  G  L  I  K  S  N  D  G  P  A  A  P  R

                                                                    550
                                                                     |
GCTGGTAGTGATGGTGATGCTATAAGTATTTCTGGTAGTTCACAAATATGGATCGACCAT
 A  G  S  D  G  D  A  I  S  I  S  G  S  S  Q  I  W  I  D  H

                                                                    610
                                                                     |
TGTTCGCTCAGTAAGTCTGTTGATGGGCTGGTAGATGCCAAGCTCGGCACCACACGCTTA
 C  S  L  S  K  S  V  D  G  L  V  D  A  K  L  G  T  T  R  L

                                                                    670
                                                                     |
ACCGTTTCCAACAGCTTATTCACCCAACACCAGTTTGTACTATTATTCGGGGCTGGTGAC
 T  V  S  N  S  L  F  T  Q  H  Q  F  V  L  L  F  G  A  G  D

                                                                    730
                                                                     |
GAAAATATTGAAGATAGAGGCATGCTAGCAACGGTCGCTTTCAACACGTTCACTGATAAC
 E  N  I  E  D  R  G  M  L  A  T  V  A  F  N  T  F  T  D  N

                                                                    790
                                                                     |
GTTGACCAAAGAATGCCTAGATGTCGACATGGGTTTTTCCAAGTCGTTAACAACAACTAT
 V  D  Q  R  M  P  R  C  R  H  G  F  F  Q  V  V  N  N  N  Y

                                                                    850
                                                                     |
GATAAATGGGGATCGTATGCCATCGGTGGTAGCGCGTCCCCAACCATACTCAGCCAAGGG
 D  K  W  G  S  Y  A  I  G  G  S  A  S  P  T  I  L  S  Q  G

                                                                    910
                                                                     |
AACAGATTCTGCGCCCCCCGATGAACGCAGCAAGAAAAATGTCCTAGGAAGGCATGGTGAA
 N  R  F  C  A  P  D  E  R  S  K  K  N  V  L  G  R  H  G  E
```

C

FIG.11C

```
          970        980        990       1000       1010       1020
GCCGCCGCAGAGTCGATGAAGTGGAACTGGAGAACGAATAAAGACGTGCTTGAAAATGGT
 A  A  A  E  S  M  K  W  N  W  R  T  N  K  D  V  L  E  N  G

          1030       1040       1050       1060       1070       1080
GCTATTTTTGTTGCATCCGGGGTCGATCCAGTGCTAACCCCTGAGCAAAGCGCAGGGATG
 A  I  F  V  A  S  G  V  D  P  V  L  T  P  E  Q  S  A  G  M

          1090       1100       1110       1120       1130       1140
ATTCCAGCCGAACCAGGAGAGTCCGCTCTAAGCCTCACTAGTAGTGCTGGTGTACTCTCA
 I  P  A  E  P  G  E  S  A  L  S  L  T  S  S  A  G  V  L  S

          1150       1160       1170       1180       1190
TGCCAACCCGGAGCACCTTGCTAAGCACCCGACCAATTACTAAGCACTTATAATGA (SEQ. ID NO:71)
 C  Q  P  G  A  P  C  -  A  P  D  Q  L  L  S  T  Y  N    (SEQ. ID NO:72)
```

2049921

FIG.12A

```
**********************************************
* TRANSLATION OF A NUCLEIC ACID SEQUENCE *
**********************************************
Done on DNA sequence AMB_A_IB.

DE    SEQUENCE OF AMB A IB CLONE.

Total number of bases is: 1349.
Analysis done on the complete sequence.
Done on (absolute) phase(s): 1.
Using the Universal genetic code.

                                        10                  20                  30                  40                  50                  60
ATGGGGATCAAACACTGTTGTTACATCTTGTATTTTACCTTAGCCCTTGTCACTTTGCTG
 M   G   I   K   H   C   C   Y   I   L   Y   F   T   L   A   L   V   T   L   L

                                        70                  80                  90                  100                 110                 120
CAACCTGTTCGTTCTGCAGAAGATGTTGAAGAATTCTTACCTTCAGCTAACGAAACAAGG
 Q   P   V   R   S   A   E   D   V   E   E   F   L   P   S   A   N   E   T   R

                                        130                 140                 150                 160                 170                 180
AGGAGCCTGAAAGCATGTGAAGCACACAACATTATAGACAAGTGCTGGAGGTGCAAAGCC
 R   S   L   K   A   C   E   A   H   N   I   I   D   K   C   W   R   C   K   A

                                        190                 200                 210                 220                 230                 240
GATTGGGCGAATAACCGACAAGCGTTAGCCGATTGTGCCCAAGGTTTTGCAAAGGGAACC
 D   W   A   N   N   R   Q   A   L   A   D   C   A   Q   G   F   A   K   G   T

                                        250                 260                 270                 280                 290                 300
TACGGTGGAAAACATGGTGATGTCTACACGGTCACCAGTGATAAAGATGATGATGTTGCA
 Y   G   G   K   H   G   D   V   Y   T   V   T   S   D   K   D   D   D   V   A

                                        310                 320                 330                 340                 350                 360
AATCCAAAAGAAGGCACACTCCGGTTTGCTGCTGCCCAAAACAGGCCCTTGTGGATCATT
 N   P   K   E   G   T   L   R   F   A   A   A   Q   N   R   P   L   W   I   I

                                        370                 380                 390                 400                 410                 420
TTTAAAAGAAATATGGTGATTCATTTGAATCAAGAGCTTGTCGTAAACAGCGACAAGACC
 F   K   R   N   M   V   I   H   L   N   Q   E   L   V   V   N   S   D   K   T
```

FIG.12B

```
                430                 440                 450                 460                 470                 480
ATCGATGGCCGAGGGGTGAAAGTTAACATCGTTAACGCCGGTCTCACCCTCATGAATGTC
 I  D  G  R  G  V  K  V  N  I  V  N  A  G  L  T  L  M  N  V

                490                 500                 510                 520                 530                 540
AAGAATATAATCATTCATAACATAAATATCCATGATATTAAAGTTTGTCCAGGAGGCATG
 K  N  I  I  I  H  N  I  N  I  H  D  I  K  V  C  P  G  G  M

                550                 560                 570                 580                 590                 600
ATTAAGTCCAACGATGGTCCACCAATTTTAAGACAACAAAGTGATGGTGATGCTATAAAT
 I  K  S  N  D  G  P  P  I  L  R  Q  Q  S  D  G  D  A  I  N

                610                 620                 630                 640                 650                 660
GTTGCTGGTAGTTCACAAATATGGATCGACCATTGCTCGCTCAGTAAGGCTTCCGATGGG
 V  A  G  S  S  Q  I  W  I  D  H  C  S  L  S  K  A  S  D  G

                670                 680                 690                 700                 710                 720
CTGCTCGATATCACCCTCGGCAGCTCACACGTGACCGTTTCCAACTGCAAATTCACCCAA
 L  L  D  I  T  L  G  S  S  H  V  T  V  S  N  C  K  F  T  Q

                730                 740                 750                 760                 770                 780
CACCAATTTGTATTATTGCTCGGGGCTGATGACACCCATTATCAAGATAAAGGCATGCTA
 H  Q  F  V  L  L  L  G  A  D  D  T  H  Y  Q  D  K  G  M  L

                790                 800                 810                 820                 830                 840
GCAACGGTAGCATTCAACATGTTCACCGATCACGTTGACCAAAGAATGCCTAGATGTAGA
 A  T  V  A  F  N  M  F  T  D  H  V  D  Q  R  M  P  R  C  R

                850                 860                 870                 880                 890                 900
TTTGGGTTTTTCCAAGTCGTTAACAACAACTACGACAGATGGGGAACGTACGCCATCGGT
 F  G  F  F  Q  V  V  N  N  N  Y  D  R  W  G  T  Y  A  I  G

                910                 920                 930                 940                 950                 960
GGTAGCTCGGCCCCAACTATACTCAGCCAAGGGAACAGATTCTTCGCCCCCGATGATATC
 G  S  S  A  P  T  I  L  S  Q  G  N  R  F  F  A  P  D  D  I
```

FIG.12C

```
          970        980        990       1000       1010       1020
ATCAAGAAAAATGTCTTAGCGAGGACTGGTACTGGCAACGCAGAGTCGATGTCGTGGAAC
 I  K  K  N  V  L  A  R  T  G  T  G  N  A  E  S  M  S  W  N

         1030       1040       1050       1060       1070       1080
TGGAGAACAGATAGAGACTTGCTTGAAAATGGTGCTATTTTTCTCCCATCCGGGTCTGAT
 W  R  T  D  R  D  L  L  E  N  G  A  I  F  L  P  S  G  S  D

         1090       1100       1110       1120       1130       1140
CCAGTGCTAACCCCCTGAGCAAAAAGCAGGGATGATTCCAGCTGAACCAGGAGAAGCCGTT
 P  V  L  T  P  E  Q  K  A  G  M  I  P  A  E  P  G  E  A  V

         1150       1160       1170       1180       1190       1200
CTAAGACTCACTAGTAGTGCTGGTGTACTCTCATGCCATCAAGGAGCACCTTGCTAAGCA
 L  R  L  T  S  S  A  G  V  L  S  C  H  Q  G  A  P  C  -  A

         1210       1220       1230       1240       1250       1260
CCTGGCCAATTCCTAAGCTTTTATAATAATCATAAATACTTATTTTATTTTATTTTTGAT
 P  G  Q  F  L  S  F  Y  N  N  H  K  Y  L  F  Y  F  I  F  D

         1270       1280       1290       1300       1310       1320
ATTTTATATGAACCATTACGTTCAAGTACTCTATTAACATGTTTTAAATTCATAAGAGTT
 I  L  Y  E  P  L  R  S  S  T  L  L  T  C  F  K  F  I  R  V

         1330       1340
TATTGATAAAAAAAAAAAAAAACCGAATTC    (SEQ. ID NO:73)
 Y  -  -  K  K  K  K  P  N        (SEQ. ID NO:74)
```

```
**************************************************
* TRANSLATION OF A NUCLEIC ACID SEQUENCE *
**************************************************
Done on DNA sequence AMB_A_IC.
DE   SEQUENCE OF AMB A IC CLONE.

Total number of bases is: 1320.
Analysis done on the complete sequence.
Done on (absolute) phase(s): 1.
Using the Universal genetic code.
```

FIG.13A

```
                                                            10                  20                  30                  40                  50                  60
ATGGGGATCAAACAATGTTGTTACATCTTGTATTTTACCTTAGCACTTGTCGCTTTGCTG
 M   G   I   K   Q   C   C   Y   I   L   Y   F   T   L   A   L   V   A   L   L

                                                            70                  80                  90                 100                 110                 120
CAACCTGTTCGTTCTGCCGAAGGTGTCGGGGAAATCTTACCTTCAGTTAACGAAACGAGG
 Q   P   V   R   S   A   E   G   V   G   E   I   L   P   S   V   N   E   T   R

                                                           130                 140                 150                 160                 170                 180
AGCCTGCAAGCATGTGAAGCACTCAACATTATAGACAAGTGCTGGAGGGGCAAAGCCGAT
 S   L   Q   A   C   E   A   L   N   I   I   D   K   C   W   R   G   K   A   D

                                                           190                 200                 210                 220                 230                 240
TGGGAGAACAACCGACAAGCGTTAGCCGACTGTGCCCAAGGTTTTGCAAAGGGAACCTAC
 W   E   N   N   R   Q   A   L   A   D   C   A   Q   G   F   A   K   G   T   Y

                                                           250                 260                 270                 280                 290                 300
GGCGGAAAATGGGGTGATGTCTACACGGTCACCAGCAATCTAGATGATGATGTTGCAAAT
 G   G   K   W   G   D   V   Y   T   V   T   S   N   L   D   D   D   V   A   N

                                                           310                 320                 330                 340                 350                 360
CCAAAAGAAGGCACACTCCGGTTTGCTGCCGCCCAAAACAGGCCCTTGTGGATCATTTTT
 P   K   E   G   T   L   R   F   A   A   A   Q   N   R   P   L   W   I   I   F

                                                           370                 380                 390                 400                 410                 420
AAAAATGATATGGTGATTAATTTGAATCAAGAGCTTGTCGTAAACAGCGACAAGACCATC
 K   N   D   M   V   I   N   L   N   Q   E   L   V   V   N   S   D   K   T   I
```

C

FIG.13B

```
                430                 440                 450                 460                 470                 480
GATGGCCGAGGGGTGAAAGTTGAAATCATTAACGGAGGTCTCACCCTCATGAATGTCAAG
 D   G   R   G   V   K   V   E   I   I   N   G   G   L   T   L   M   N   V   K

                490                 500                 510                 520                 530                 540
AATATAATCATTCATAACATAAATATCCATGATGTTAAAGTGCTTCCAGGAGGCATGATT
 N   I   I   I   H   N   I   N   I   H   D   V   K   V   L   P   G   G   M   I

                550                 560                 570                 580                 590                 600
AAGTCCAACGATGGTCCACCAATTTTAAGACAAGCAAGTGATGGGGATACTATAAATGTT
 K   S   N   D   G   P   P   I   L   R   Q   A   S   D   G   D   T   I   N   V

                610                 620                 630                 640                 650                 660
GCTGGTAGTTCCCAAATATGGATAGACCATTGCTCACTCAGCAAGTCTTTCGATGGGCTG
 A   G   S   S   Q   I   W   I   D   H   C   S   L   S   K   S   F   D   G   L

                670                 680                 690                 700                 710                 720
GTCGATGTCACCCTCGGTAGCACACACGTGACCATTTCCAACTGCAAATTCACCCAACAG
 V   D   V   T   L   G   S   T   H   V   T   I   S   N   C   K   F   T   Q   Q

                730                 740                 750                 760                 770                 780
TCAAAAGCAATATTGTTGGGAGCAGATGACACCCATGTTCAAGATAAAGGAATGCTAGCA
 S   K   A   I   L   L   G   A   D   D   T   H   V   Q   D   K   G   M   L   A

                790                 800                 810                 820                 830                 840
ACGGTCGCTTTCAACATGTTCACCGATAACGTTGACCAAAGAATGCCTAGATGTCGATTT
 T   V   A   F   N   M   F   T   D   N   V   D   Q   R   M   P   R   C   R   F

                850                 860                 870                 880                 890                 900
GGGTTTTTCCAAGTTGTTAACAACAACTACGACAGATGGGGAACGTACGCCATAGGTGGT
 G   F   F   Q   V   V   N   N   N   Y   D   R   W   G   T   Y   A   I   G   G

                910                 920                 930                 940                 950                 960
AGCTCGGCCCCAACTATACTCTGCCAAGGGAACAGATTCTTGGCCCCTGATGATCAGATC
 S   S   A   P   T   I   L   C   Q   G   N   R   F   L   A   P   D   D   Q   I
```

C

FIG.13C

```
                970                980                990               1000               1010               1020
AAGAAAAATGTCCTAGCGAGGACTGGTACAGGCGCTGCTGAGTCGATGGCGTGGAACTGG
 K  K  N  V  L  A  R  T  G  T  G  A  A  E  S  M  A  W  N  W

               1030               1040               1050               1060               1070               1080
AGATCTGATAAAGACTTGCTTGAAAATGGTGCTATTTTTGTTACATCTGGGTCTGATCCA
 R  S  D  K  D  L  L  E  N  G  A  I  F  V  T  S  G  S  D  P

               1090               1100               1110               1120               1130               1140
GTGCTAACCCCTGTTCAAAGCGCAGGGATGATTCCAGCTGAACCAGGAGAAGCCGCTATA
 V  L  T  P  V  Q  S  A  G  M  I  P  A  E  P  G  E  A  A  I

               1150               1160               1170               1180               1190               1200
AAACTCACTAGTAGTGCTGGTGTATTCTCATGCCGTCCTGGAGCACCTTGCTAAGCACCC
 K  L  T  S  S  A  G  V  F  S  C  R  P  G  A  P  C  -  A  P

               1210               1220               1230               1240               1250               1260
TGCCAATTCTCCTAAGCTTTTGCAATGATCAAAAATACTTTTTTATTTTATTTTTAATAT
 C  Q  F  S  -  A  F  A  M  I  K  N  T  F  L  F  Y  F  -  Y

               1270               1280               1290               1300               1310               1320
TTTATATGTACTGGAAATGAACCATTACCTTCTAGTACTCTATAACATGTTTTGCATTTA
 F  I  C  T  G  N  E  P  L  P  S  S  T  L  -  H  V  L  H  L
```

(SEQ. ID NO:75)
(SEQ. ID NO:76)

2049921

```
**********************************************
* TRANSLATION OF A NUCLEIC ACID SEQUENCE *
**********************************************
Done on DNA sequence AMB_A_ID.
DE   SEQUENCE OF AMB A ID CLONE.

Total number of bases is: 1160.
Analysis done on the complete sequence.
Done on (absolute) phase(s): 1.
Using the Universal genetic code.
```

FIG.14A

```
                                                                  10         20         30         40         50         60
                                                                  |          |          |          |          |          |
TTGTATTTTACCTTAGCCCTTGTCACTTTGCTGCAACCTGTTCGTTCTGCCGAAGATCTC
 L  Y  F  T  L  A  L  V  T  L  L  Q  P  V  R  S  A  E  D  L

          70         80         90         100        110        120
          |          |          |          |          |          |
CAGGAAATCTTACCTTCAGCTAACGAAACAAGGAGCCTGACAACATGTGGAACATACAAC
 Q  E  I  L  P  S  A  N  E  T  R  S  L  T  T  C  G  T  Y  N

          130        140        150        160        170        180
          |          |          |          |          |          |
ATTATAGACGGGTGCTGGAGGGGCAAAGCCGATTGGGCGGAAAACCGAAAAGCGTTAGCC
 I  I  D  G  C  W  R  G  K  A  D  W  A  E  N  R  K  A  L  A

          190        200        210        220        230        240
          |          |          |          |          |          |
GATTGTGCCCAAGGTTTTGCAAAGGGAACAATCGGCGGAAAAGATGGTGATATATACACG
 D  C  A  Q  G  F  A  K  G  T  I  G  G  K  D  G  D  I  Y  T

          250        260        270        280        290        300
          |          |          |          |          |          |
GTCACCAGTGAGCTAGATGATGATGTTGCAAATCCAAAAGAAGGCACACTCCGGTTTGGT
 V  T  S  E  L  D  D  D  V  A  N  P  K  E  G  T  L  R  F  G

          310        320        330        340        350        360
          |          |          |          |          |          |
GCCGCCCAAAACAGGCCCTTGTGGATTATTTTTGAAAGAGATATGGTGATTCGTTTGGAT
 A  A  Q  N  R  P  L  W  I  I  F  E  R  D  M  V  I  R  L  D

          370        380        390        400        410        420
          |          |          |          |          |          |
AGAGAGTTGGCTATAAACAACGACAAGACCATCGATGGCCGAGGGGCGAAAGTTGAAATC
 R  E  L  A  I  N  N  D  K  T  I  D  G  R  G  A  K  V  E  I
```

FIG.14B

```
          430         440         450         460         470         480
           |           |           |           |           |           |
ATTAACGCTGGTTTCGCCATCTATAATGTCAAGAATATAATCATTCATAACATAATTATG
 I   N   A   G   F   A   I   Y   N   V   K   N   I   I   I   H   N   I   I   M

          490         500         510         520         530         540
           |           |           |           |           |           |
CATGATATTGTAGTGAATCCAGGAGGCCTGATTAAGTCCCACGATGGTCCACCAGTTCCA
 H   D   I   V   V   N   P   G   G   L   I   K   S   H   D   G   P   P   V   P

          550         560         570         580         590         600
           |           |           |           |           |           |
AGAAAGGGTAGTGATGGTGATGCTATAGGTATTTCTGGTGGTTCACAAATATGGATCGAC
 R   K   G   S   D   G   D   A   I   G   I   S   G   G   S   Q   I   W   I   D

          610         620         630         640         650         660
           |           |           |           |           |           |
CATTGCTCCCTCAGTAAGGCTGTTGATGGGCTAATCGATGCTAAACACGGCAGCACACAC
 H   C   S   L   S   K   A   V   D   G   L   I   D   A   K   H   G   S   T   H

          670         680         690         700         710         720
           |           |           |           |           |           |
TTCACCGTTTCTAACTGCTTATTCACCCAACACCAATATTTATTATTGTTCTGGGATTTT
 F   T   V   S   N   C   L   F   T   Q   H   Q   Y   L   L   L   F   W   D   F

          730         740         750         760         770         780
           |           |           |           |           |           |
GACGAGCGAGGCATGCTATGTACGGTCGCATTCAACAAGTTCACTGATAACGTTGACCAA
 D   E   R   G   M   L   C   T   V   A   F   N   K   F   T   D   N   V   D   Q

          790         800         810         820         830         840
           |           |           |           |           |           |
AGAATGCCTAACTTACGACATGGGTTTGTCCAAGTCGTTAACAACAACTACGAAAGATGG
 R   M   P   N   L   R   H   G   F   V   Q   V   V   N   N   N   Y   E   R   W

          850         860         870         880         890         900
           |           |           |           |           |           |
GGATCGTACGCCCTCGGTGGTAGCGCAGGCCCAACCATACTTAGCCAAGGGAACAGATTC
 G   S   Y   A   L   G   G   S   A   G   P   T   I   L   S   Q   G   N   R   F

          910         920         930         940         950         960
           |           |           |           |           |           |
TTAGCCTCCGATATCAAGAAAGAGGTCGTAGGGAGGTATGGTGAATCCGCCATGTCAGAG
 L   A   S   D   I   K   K   E   V   V   G   R   Y   G   E   S   A   M   S   E
```

C

FIG.14C

```
          970        980        990       1000       1010       1020
TCGATTAATTGGAACTGGAGATCGTATATGGACGTATTTGAAAATGGTGCTATTTTTGTT
 S  I  N  W  N  W  R  S  Y  M  D  V  F  E  N  G  A  I  F  V

         1030       1040       1050       1060       1070       1080
CCATCCGGGGTTGATCCAGTGCTAACCCCTGAGCAAAACGCAGGGATGATTCCAGCCGAA
 P  S  G  V  D  P  V  L  T  P  E  Q  N  A  G  M  I  P  A  E

         1090       1100       1110       1120       1130       1140
CCAGGAGAAGCCGTTCTAAGACTCACTAGTAGTGCTGGTGTCCTCTCATGCCAACCTGGA
 P  G  E  A  V  L  R  L  T  S  S  A  G  V  L  S  C  Q  P  G

         1150       1160
GCACCTTGCTAAGCACTGCA
 A  P  C  -  A  L
```

(SEQ. ID NO:77)
(SEQ. ID NO:78)

2049921

```
**************************************************
* TRANSLATION OF A NUCLEIC ACID SEQUENCE *
**************************************************
Done on DNA sequence AMB_A_II.

DE   SEQUENCE OF AMB A II CLONE.

Total number of bases is: 1368.
Analysis done on the complete sequence.
Done on (absolute) phase(s): 1.
Using the Universal genetic code.
```

FIG.15A

```
                                                            10                  20                  30                  40                  50                  60
TTGTATTTTACCTTAGCACTTGTCACTTTGGTGCAAGCTGGACGTCTTGGCGAAGAGGTC
 L   Y   F   T   L   A   L   V   T   L   V   Q   A   G   R   L   G   E   E   V

                                                            70                  80                  90                 100                 110                 120
GACATCTTACCTTCACCTAACGATACAAGGAGGAGCCTGCAAGGATGTGAAGCACACAAC
 D   I   L   P   S   P   N   D   T   R   R   S   L   Q   G   C   E   A   H   N

                                                           130                 140                 150                 160                 170                 180
ATTATAGACAAGTGTTGGAGGTGCAAACCCGATTGGGCGGAGAACCGACAAGCGTTAGGC
 I   I   D   K   C   W   R   C   K   P   D   W   A   E   N   R   Q   A   L   G

                                                           190                 200                 210                 220                 230                 240
GATTGTGCGCAAGGTTTTGGAAAGGCAACTCACGGCGGAAAATGGGGTGATATCTACATG
 D   C   A   Q   G   F   G   K   A   T   H   G   G   K   W   G   D   I   Y   M

                                                           250                 260                 270                 280                 290                 300
GTCACAAGTGATCAGGATGATGATGTTGTAAATCCAAAAGAAGGCACACTCCGGTTCGGT
 V   T   S   D   Q   D   D   D   V   V   N   P   K   E   G   T   L   R   F   G

                                                           310                 320                 330                 340                 350                 360
GCTACCCAGGACAGGCCCTTGTGGATCATTTTTCAAAGAGATATGATTATTTATTTGCAA
 A   T   Q   D   R   P   L   W   I   I   F   Q   R   D   M   I   I   Y   L   Q

                                                           370                 380                 390                 400                 410                 420
CAAGAGATGGTCGTAACCAGCGACACGACCATTGATGGTCGAGGGGCGAAAGTTGAGCTC
 Q   E   M   V   V   T   S   D   T   T   I   D   G   R   G   A   K   V   E   L
```

C

FIG.15B

```
                                                                    430
                                                                     |
GTTTATGGAGGTATCACCCTCATGAATGTCAAGAATGTAATCATTCACAACATAGATATC
 V  Y  G  G  I  T  L  M  N  V  K  N  V  I  I  H  N  I  D  I

                                                                    490
                                                                     |
CATGATGTTAGAGTGCTTCCAGGAGGTAGGATTAAGTCCAATGGTGGTCCAGCCATACCA
 H  D  V  R  V  L  P  G  G  R  I  K  S  N  G  G  P  A  I  P

                                                                    550
                                                                     |
AGACATCAGAGTGATGGTGATGCTATCCATGTTACGGGTAGTTCAGACATATGGATCGAC
 R  H  Q  S  D  G  D  A  I  H  V  T  G  S  S  D  I  W  I  D

                                                                    610
                                                                     |
CATTGCACGCTCAGTAAGTCATTTGATGGGCTCGTCGATGTCAACTGGGGCAGCACAGGA
 H  C  T  L  S  K  S  F  D  G  L  V  D  V  N  W  G  S  T  G

                                                                    670
                                                                     |
GTAACCATTTCCAACTGCAAATTCACCCACCACGAAAAAGCTGTTTTGCTCGGGGCTAGT
 V  T  I  S  N  C  K  F  T  H  H  E  K  A  V  L  L  G  A  S

                                                                    730
                                                                     |
GACACGCATTTTCAAGATCTGAAAATGCATGTAACGCTTGCATACAACATCTTCACCAAT
 D  T  H  F  Q  D  L  K  M  H  V  T  L  A  Y  N  I  F  T  N

                                                                    790
                                                                     |
ACCGTTCACGAAAGAATGCCCAGATGCCGATTTGGGTTTTTCCAAATCGTTAACAACTTC
 T  V  H  E  R  M  P  R  C  R  F  G  F  F  Q  I  V  N  N  F

                                                                    850
                                                                     |
TACGACAGATGGGATAAGTACGCCATCGGTGGTAGCTCGAACCCTACTATTCTCAGCCAA
 Y  D  R  W  D  K  Y  A  I  G  G  S  S  N  P  T  I  L  S  Q

                                                                    910
                                                                     |
GGGAACAAATTCGTGGCCCCCGATTTCATTTACAAGAAAAACGTCTGTCTAAGGACTGGT
 G  N  K  F  V  A  P  D  F  I  Y  K  K  N  V  C  L  R  T  G
```

C

FIG.15C

```
      970        980        990       1000       1010       1020
       |          |          |          |          |          |
GCACAGGAGCCAGAATGGATGACTTGGAACTGGAGAACACAAAACGACGTGCTTGAAAAT
 A  Q  E  P  E  W  M  T  W  N  W  R  T  Q  N  D  V  L  E  N

     1030       1040       1050       1060       1070       1080
       |          |          |          |          |          |
GGTGCTATCTTTGTGGCATCTGGGTCTGATCCAGTGCTAACCGCTGAACAAAATGCAGGC
 G  A  I  F  V  A  S  G  S  D  P  V  L  T  A  E  Q  N  A  G

     1090       1100       1110       1120       1130       1140
       |          |          |          |          |          |
ATGATGCAAGCTGAACCGGGAGATATGGTTCCACAACTCACCATGAATGCAGGTGTACTC
 M  M  Q  A  E  P  G  D  M  V  P  Q  L  T  M  N  A  G  V  L

     1150       1160       1170       1180       1190       1200
       |          |          |          |          |          |
ACATGCTCGCCTGGAGCACCTTGCTAAGCACCTGGCCAATTCCTATGCAACGATCATAAA
 T  C  S  P  G  A  P  C  -  A  P  G  Q  F  L  C  N  D  H  K

     1210       1220       1230       1240       1250       1260
       |          |          |          |          |          |
TACTTGCTCACCATAAGTGTTCATTTGATTAGATTTGGACACGAATGATGTAACCGATTC
 Y  L  L  T  I  S  V  H  L  I  R  F  G  H  E  -  C  N  R  F

     1270       1280       1290       1300       1310       1320
       |          |          |          |          |          |
GTCTGAATTATGATTTGTTTTGATTCTCAGTTTCATAATATGGCTTCTTGAGAGCAAAAT
 V  -  I  M  I  C  F  D  S  Q  F  H  N  M  A  S  -  E  Q  N

     1330       1340       1350       1360
       |          |          |          |
TAGAGAAGAGTGTCTTTGATCAACTACATTTTATGGTTTTTATATTAA  (SEQ. ID NO:79)
 -  R  R  V  S  L  I  N  Y  I  L  W  F  L  Y  -  (SEQ. ID NO:80)
```

C

FIG.16A

COMPOSITE *Amb a* I and *a* II SEQUENCES RELATIVE TO *Amb a* IA

```
                  10                  20                  30                  40                  50
                  [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   IA
- - - - Q - - - - - - - - - - - - - - - - - - - - - - - - - - V E - F - - S A - - - - - - S - K A C E - H -   IB
- - - - Q - - - - - - - - - - - - - A - - - - - - - - - - G V G - - - - S - - - - - - * S - Q A C E - -L -   IC
                  [- - - - -]- - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - - -   ID
                  [- - - - -]- - - - - V - A G - L G - E V * D - - - S P - D - - - S - Q G C E - H -   IIA

                  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   IA
- - - K - - - C - - - - - - N - - Q - - - - - - - - - - - A - - - Y - - - H - - V - - - - - D K - - - - -   IB
- - - K - - - - - - - - - E N - - Q - - - - - - - - - - - A - - - Y - - - H - - V - - - - - N - - - - - -   IC
- - - - - - - - - - - - - - - - - - - - - - - - - - - - - A - - - I - - - - - - - - - - - - - - - - - - -   ID
- - - K - - - C - P - - - - - - - Q - - G - - - - - - - - - - A - H - - - W - - - - M - - - D Q - - - - V   IIA
                                    N

                  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   IA
- - - - - - - - - - A - - - - - - - - - - - - - K - N - - - H - N Q - L - - - - - - - - - - - - V - - N -   IB
- - - - - - - - - - A - - - - - - - - - - - - - K N - - - - N - N Q - L - - - - - - - - - - - - V - - - -   IC
- - - - - - - - - - - - - - - - - - - - - - - - A - - - - - - - - - - R - L A I - N - - - - - - - - - - - -   ID
                                                -
- - - - - - - - - - - T - D - - - - - - - Q - - - I - Y - Q Q - - - - T - - - - - - - - - - - - - - - - L   IIA
                                                T

                  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   IA
- - - - L - - M N - - - I - - - - - - I - - I - - C - - - M - - - - - - - - P I L - Q Q - - - - - - - N   IB
- - G - L - - M N - - - I - - - - - - I - - - - - L - - - M - - - - - - - - P I L - Q A - - - - T - N   IC
- - - - - A I Y N - - - I - - - - - I - - - I V - - - - - - - - - C - - - - P - - - - - - - - - - - - -   ID
V Y G - I - - M N - - - - - - - - - - D I - - - R - L - - - R - - - - G - - - I - - H Q - - - - - - - H   IIA
                                                - H - - - P V - - K - - - - - - - G
```

20499921

FIG.16B

```
                    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   IA
V A - - - - - - - - - - - - - - - - A S - - - L - I T - - S S H V - - - - C K - - - - - - - - - - L - - D   IB
V A - - - - - - - - - - - - - - - - F - - - - - V T - - S - H V - I - - C K - - - Q S K A I - L -R- - D   IC
- - - G - - - - - - - - - - - - - - A - - - - I - - - H - S - H F T - - - C - - - - - - - Y L - - - W D F   ID
V T - - - D - - - - - - - T - - - - F - - - - - V N W - S - G V - I - - C K - - H - E K A V - L - - S   IIA

                    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   IA
- T H Y Q - K - - - - - - - - - - M - - - H - - - - - - - - - - F - - - - - - - - - - - - R - - T - - - -   IB
- T H V Q - K - - - - - - - - - - M - - - - - - - - - - - - - -Q- F - - - - - - - - - - - - R - - T - - - -   IC
- - * * * * - - - - C - - - - - - K - - - - - - - - - - - N L - - - - V - - - - - - - - E R - - - - - - L -   ID
- T H F Q - L K - H V - L - Y - I - - N T - H E - - - - - - F - - - - I - - - F - - R - D K - - - -   IIA

                    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   IA
- - S A - - - - - - - - - - - - F - - - D I I -G- - - - A - T - T * G N - - - - S - - - - - - D -R- - L - - -   IB
- - S A - - - - C - - - - - - L - - - D Q I - - - - - - A - T - T * G - - - - - A - - - - - S D - - L - - -   IC
- - - G - - - - - - - - - - - - L - S - * * I - - E - V - - Y - - S - M S - - I N - - - - - S Y M - - F - -   ID
- - S N - - - - - - - - - K - V - - - F I Y -R- - - C L - T - A * Q E P - W - T - - - - - - Q N - - - - - -   IIA

                    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       IA
- - - - - L P - - S - - - - - - - - - K - - - - - - - - - - - A V -I- R - - - - - - - - - - - H Q - - - -       IB
- - - - - - T - - S - - - - - - - V - - - - - - - - - - - - - - A - I K - - - - - - - -F- - - R - - - - -       IC
- - - - - - P - - - - - - - - - - - N - - - - - - - - - - - A V - R - - - - - - - - - - - - - - - - - -       ID
- - - - - - - - - S - - - - - - A - - N - - - M Q - - - - D M V P Q - - M N - - - - - T - S - - - - - -       IIA
```

20499921

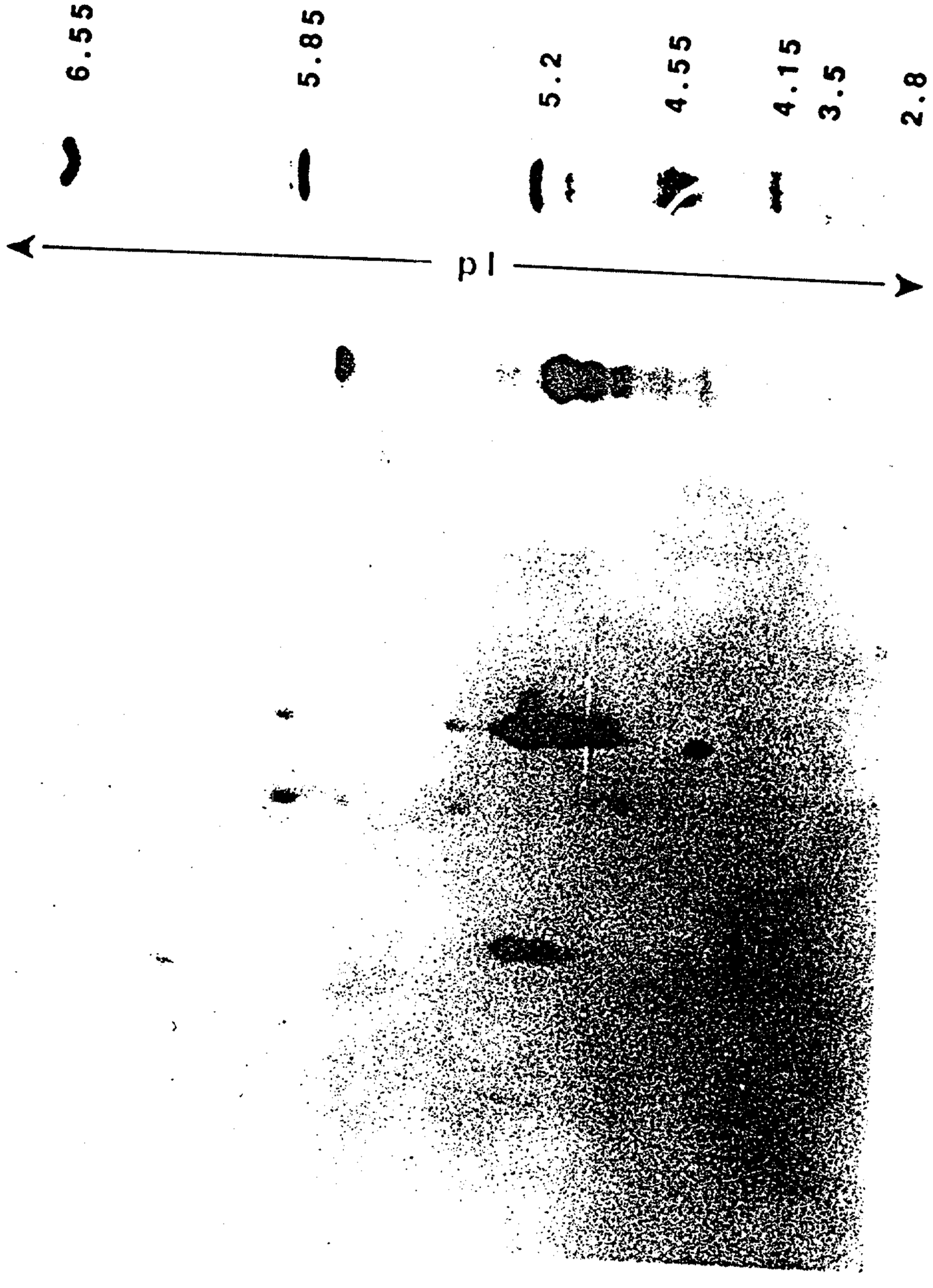
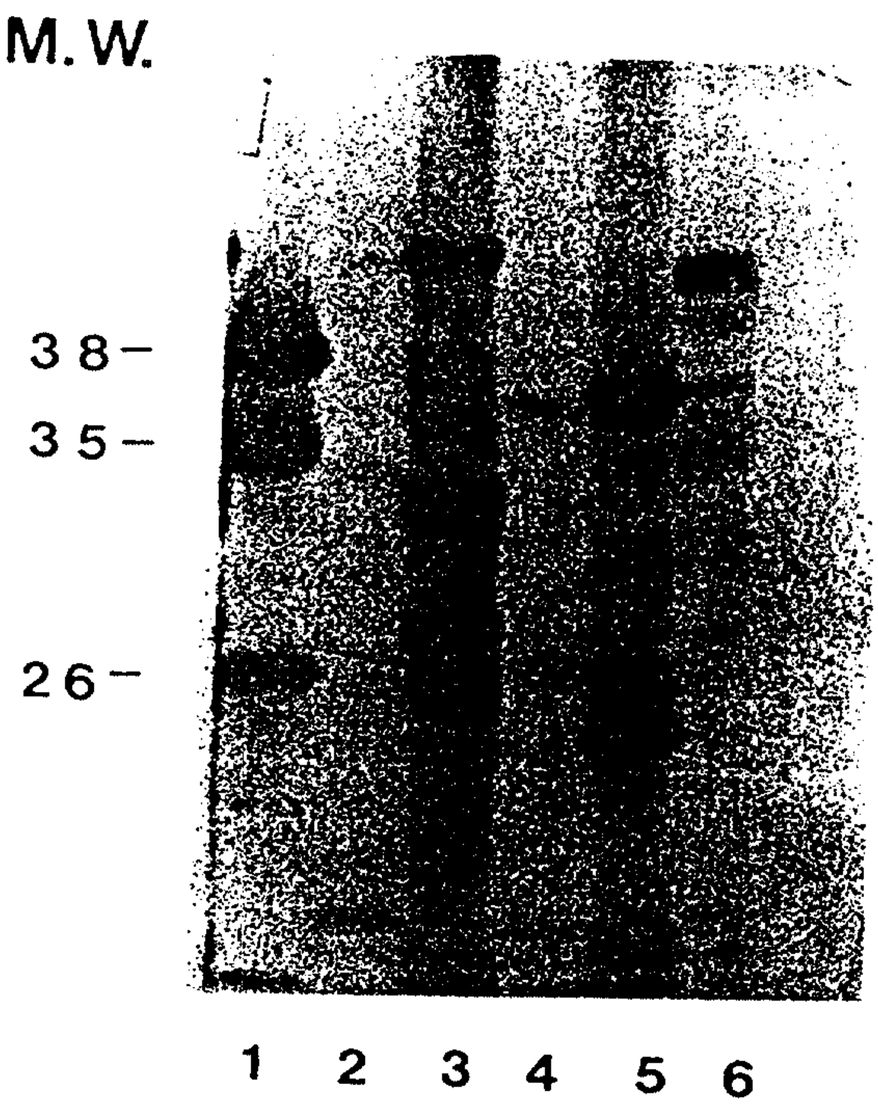


FIG.18

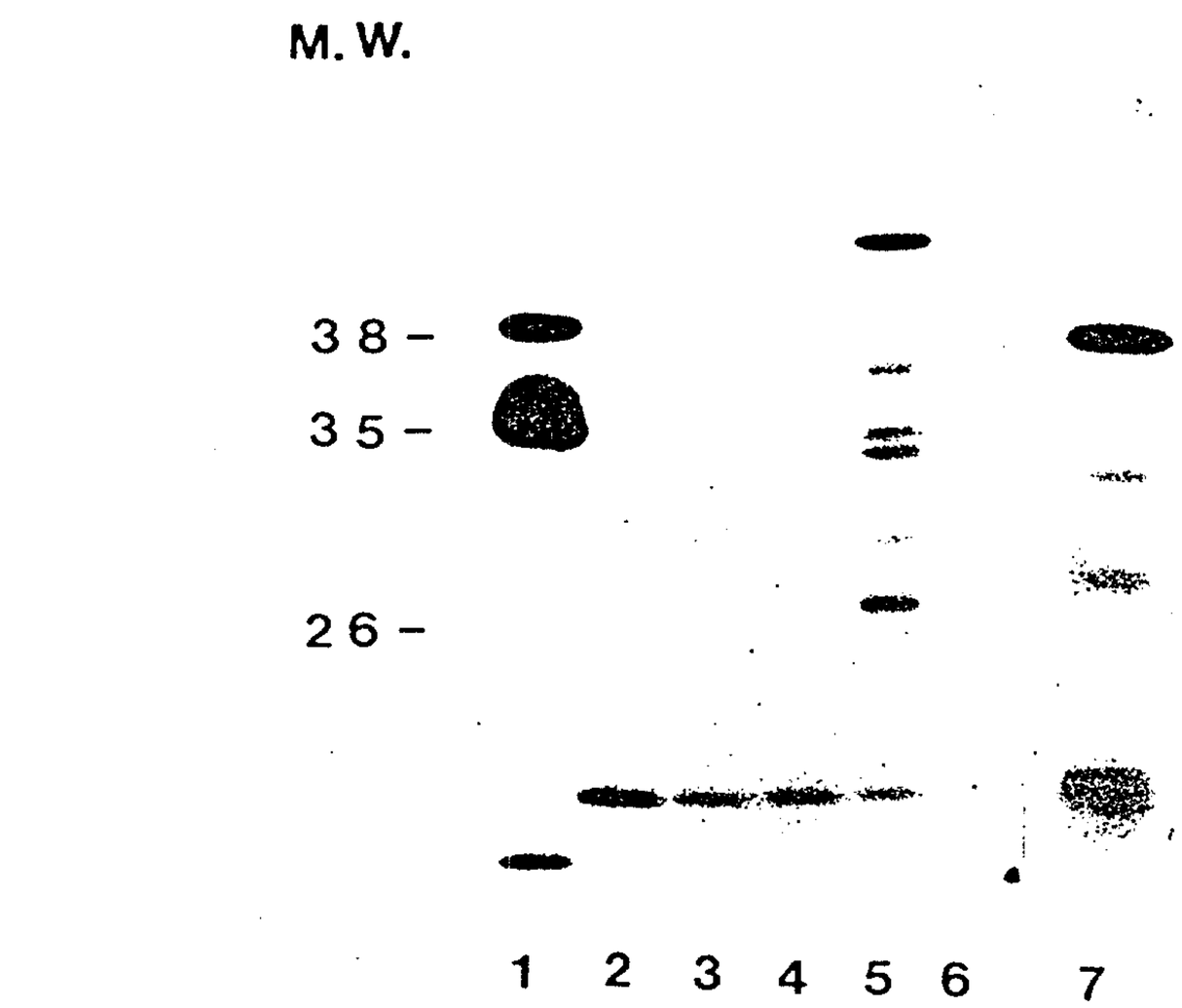
FIG.19



Sample Loaded

1- Pollen Extract
2- JM 109
3- Amb a I A
4- Amb a I B
5- Amb a I C
6- Amb a II A

FIG.20



ALLERGIC PATIENT No. 295

1- POLLEN EXTRACT
2- JM 109
3- Amb a I B
4- Amb a I A(t)
5- Amb a I A
6- Amb a II
7- Amb a I C

C

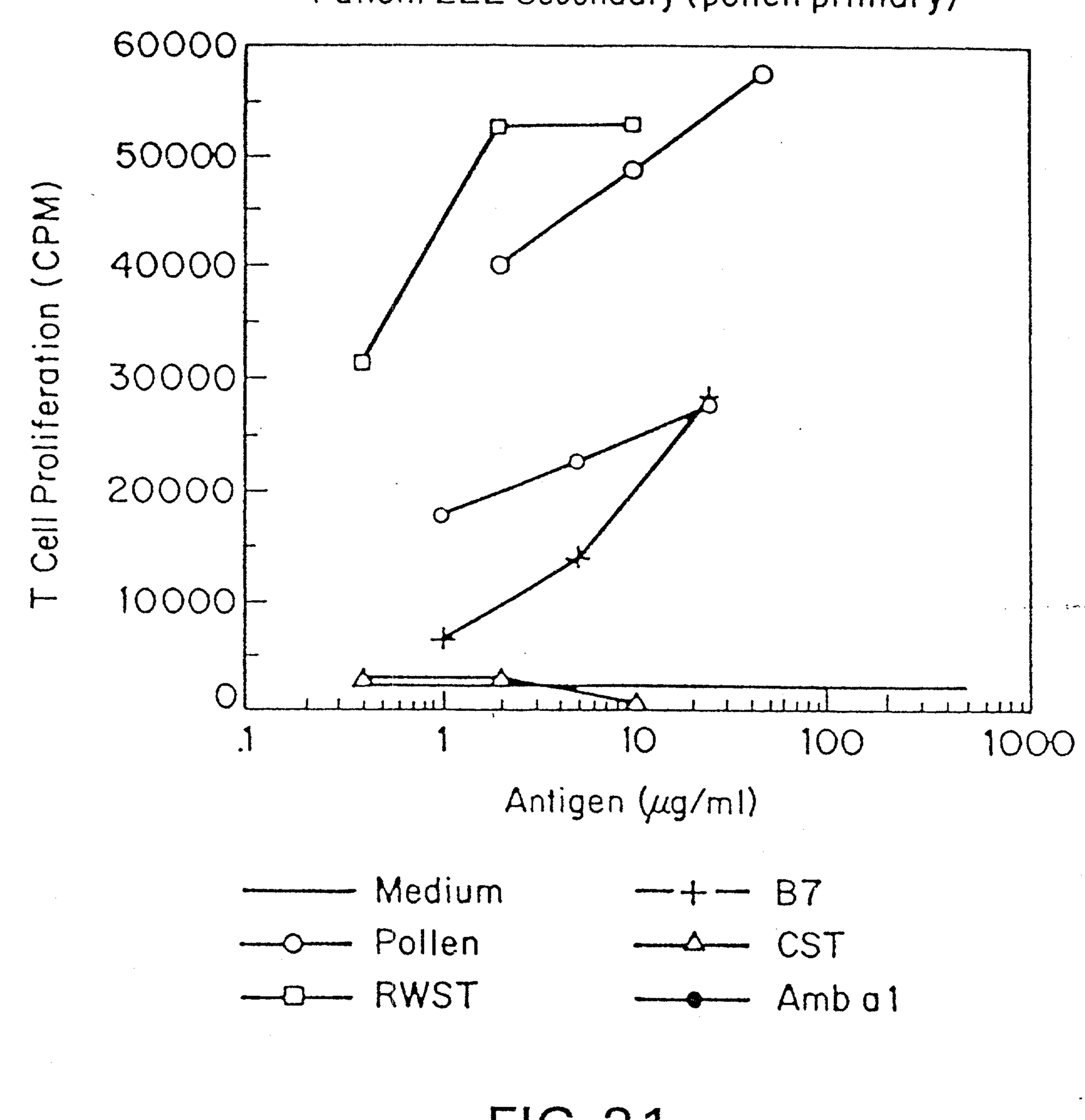

Patient 222 Secondary (pollen primary)
T Cell Proliferation (CPM)
60000
50000
40000
30000
20000
10000
0
.1
1
10
100
1000
Antigen (μg/ml)
Medium
Pollen
RWST
B7
CST
Amb a1

FIG.21