

GENE 04057

Cloning the cDNA encoding the AmbtV allergen from giant ragweed (*Ambrosia trifida*) pollen

(Recombinant DNA; polymerase chain reaction; degenerate oligodeoxyribonucleotide; M13 vector; cDNA sequence; immune recognition; allergic response)

Balaram Ghosh, Michael P. Perry and David G. Marsh

Division of Clinical Immunology, Johns Hopkins Asthma and Allergy Center, Johns Hopkins University, School of Medicine, Baltimore, MD 21224 (U.S.A.)

Received by G.N. Godson: 26 September 1990
Revised: 19 November 1990
Accepted: 20 November 1990

SUMMARY

Ragweed (*Ambrosia*) pollens contain a number of proteins that cause allergic disease in ragweed-sensitive people. The cloning of the *AmbtV* cDNA is important, since the 4.4-kDa AmbtV, one of the allergens in giant ragweed (*Ambrosia trifida*) pollen, serves as a simple model system to study the basic structural requirements for immune recognition of foreign protein allergens. We report the cloning of the *AmbtV* cDNA by means of the polymerase chain reaction (PCR) using degenerate primers. We generated three sets of overlapping cDNA clones by a combination of PCR and anchored-PCR, and determined the complete nucleotide (nt) sequence. From the nt sequence, the amino acid (aa) sequence of the protein was confirmed and the leader sequence was deduced. This general approach can be used to clone allergen and other cDNAs from complex biological sources provided partial aa sequence information is available. It may be the best available approach in cases where the isolation of clones from a cDNA library is difficult, which proved to be the case for *AmbtV*.

INTRODUCTION

Substantial proportions (15–20%) of the populations in the developed countries suffer from IgE-mediated atopic

diseases such as allergic rhinitis and bronchial asthma (Young, 1980; Smith, 1988). Ragweed pollens represent major sources of allergenic proteins in the United States (Lewis et al., 1983). The isolation and characterization of several protein allergens from ragweed pollens have been reported in previous studies (Goodfriend et al., 1985; Roebber et al., 1983; 1985; King et al., 1964; 1967; Underdown and Goodfriend, 1969). AmbtV from giant ragweed (*A. trifida*) pollen has been purified and its aa sequence has been determined (Goodfriend et al., 1985; Roebber et al., 1985). AmbtV (4.4 kDa; 40 aa) and its homologues from other ragweed species have served as important models in studies of the mechanisms of human immune response. In Caucasoid populations, striking associations between HLA-DR2 and Dw2 and immune responsiveness toward AmbtV and its homologues from other *Ambrosia* species have been established (Roebber et al., 1985; Marsh et al., 1982a,b). Recently, by DNA sequencing and DNA typing of *HLA-D*-region genes, and by isolating AmbaV-specific T cells from allergic individuals, it has been shown that responsiveness to the AmbV homologues is primarily

Correspondence to: Dr. B. Ghosh, Johns Hopkins Asthma and Allergy Center, Johns Hopkins University School of Medicine, 301 Bayview Blvd., Unit Office 1, Baltimore, MD 21224 (U.S.A.) Tel. (301) 550-2022; Fax (301) 550-2090.

Abbreviations: *A.*, *Ambrosia*; aa, amino acid(s); AmbaV, allergen from *A. artemisiifolia*; AmbtV, allergen from *A. trifida* pollen; *AmbtV*, gene (cDNA) encoding AmbtV (provisional nomenclature); AMV, avian myeloblastosis virus; bp, base pair(s); BSA, bovine serum albumin; cDNA, DNA complementary to RNA; cpm, counts/min; DIF, differentiation-inducing factor; dNTP, deoxyribonucleoside triphosphate; EtBr, ethidium bromide; HLA, human leukocyte antigen; IgE, immunoglobulin E; IPTG, isopropyl- β -D-thiogalactopyranoside; kb, kilobase(s) or 1000 bp; MHC, major histocompatibility complex; nt, nucleotide(s); oligo, oligodeoxyribonucleotide; PCR, polymerase chain reaction; RF, replicative form; RT, reverse transcriptase; SDS, sodium dodecyl sulfate; ss, single strand(ed); SSC, 0.15 M NaCl/0.015 M Na₂·citrate pH 7.6; TE, 10 mM Tris·HCl pH 7/1 mM EDTA; XGal, 5-bromo-4-chloro-3-indolyl- β -D-galactopyranoside.

restricted to a particular HLA-DR2-associated MHC class-II molecule, HLA-DR α 2.2 (Zwollo et al., 1991; Marsh et al., 1990). These results provide evidence for a causal relationship between the presence of a particular HLA class-II molecule and the expression of a specific immune response, and demonstrate the influence of specific HLA-linked *Ir* genes in human responsiveness toward inhaled allergens.

The molecular cloning of *AmbtV* (and its homologues) is, therefore, relevant in exploring this model system further; specifically, this will be important for (a) the physicochemical characterization of the allergen, (b) the production of large quantities of allergen for basic and clinical research, and (c) the mapping of epitopes which interact with antibodies, T-cell receptors and class-II molecules. Although a number of allergen cDNAs from other species have been cloned and sequenced (Fang et al., 1988; Thomas et al., 1988; Chua et al., 1988; Breiteneder et al., 1989; Perez et al., 1990; Rafnar et al., 1990), the cloning of an allergen cDNA from giant ragweed is reported here for the first time.

RESULTS AND DISCUSSION

(a) Northern-blot analysis

To check whether the *AmbtV*-specific RNA was present in the total RNA preparation isolated from giant ragweed pollen, we performed a Northern-blot analysis (Fig. 1). An oligo probe, BG4 (Table I), was synthesized based on a portion of the internal aa sequence of the AmbtV protein (aa 21–30). The ragweed RNA was able to hybridize with this probe (lane 1), whereas control RNA similarly prepared from rye-grass pollen failed to hybridize (lane 2), indicating that the *AmbtV*-specific mRNA was present in the ragweed RNA preparation.

(b) Amplification of a partial coding segment of *AmbtV*

To obtain the full-length cDNA sequence of *AmbtV*, we followed the overall strategy summarized in Fig. 2 and detailed below. The ss cDNA was synthesized from giant ragweed total RNA using the AMV RT and an oligo(dT) primer (Fig. 2A). The cDNA was then amplified by the PCR utilizing the BG1 and BG2 degenerate oligo primers. As shown in Fig. 3A, a large number of amplified DNA bands was observed in the agarose gel (lane 2), whereas control amplification without DNA showed no bands (lane 1). A band of 119 bp was amplified by the PCR (arrow in Fig. 3A, lane 2), corresponding possibly to *AmbtV* cDNA (40 codons less one base, Table I). To confirm this, the internal oligo probe, BG4, was used for Southern hybridization. Fig. 3B shows that this probe specifically hybridizes with the 119-bp band. The DNA was isolated from the agarose gel by electroelution and subcloned into



Fig. 1. Northern-blot analysis. Lanes 1 and 2 represent RNA from giant ragweed pollen and from rye-grass pollen (control), respectively. 28S and 18S RNAs are shown as markers. Total RNA was isolated from giant ragweed pollen (Greer Labs., North Carolina and kept dry at -70°C until used) using a modified procedure of plant RNA isolation (B.G. and D.G.M., in preparation). 15 μg of total RNA was separated on a 2.0% agarose (GTG, NuSieve, Rockland, ME)-formaldehyde (2.2 M) gel and transferred to a nylon membrane using the standard protocol (Maniatis et al., 1982). The blots were prehybridized in a buffer containing $10\times$ Denhardt's solution ($50\times$ Denhardt's: 1% Ficoll, 1% PVP and 1% BSA)/ $5\times$ SSC/20 mM Na $\cdot$ phosphate buffer pH 6.5/50 μg per ml denatured salmon-sperm DNA/0.5% SDS for 4 h at 42°C with constant gentle shaking. The blots were allowed to hybridize at 42°C for 12–18 h in a solution containing 20% formamide/ $5\times$ SSC/20 mM Na $\cdot$ phosphate buffer pH 6.5/2.5 $\times$ Denhardt's solution/0.5 $\times$ SDS/50 μg per ml salmon sperm DNA/1 mM EDTA/ ^{32}P -labeled degenerate oligo probe BG4 (Table I, specific activity of 7.6×10^7 cpm/ μg) with constant shaking. The oligo was end-labeled by [γ - ^{32}P]ATP and T4-poly-nucleotide kinase as described (Maniatis et al., 1982). The blot was washed with $2\times$ SSC/0.1% SDS at room temperature for 2×15 min and then with $1\times$ SSC/0.1% SDS at 37°C for 2×15 min. The membrane was then exposed to x-ray film (Kodak XAR 5) for autoradiography.

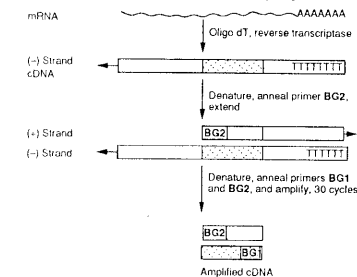
the M13 vector for sequencing (legend to Fig. 5). Four of the sequences corresponded to the known AmbtV aa sequence. Incorporation of the 'neutral' nt, inosine, at positions 9 and 12 in the primer, BG2 (Table I), where any nt could be present reduces the degeneracy from 512- to 32-fold and may account for the success of this procedure. These cDNA sequence data (except the positions corresponding to the two inosine residues in primer BG2) were subsequently used for the synthesis of *AmbtV*-specific oligos (i.e., BG19, BG25, BG26 and BG27), which were used for obtaining the full-length cDNA sequence (see sections c and d).

(c) 3'-End amplification of *AmbtV*

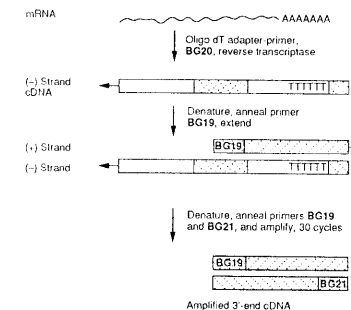
To obtain the 3'-end of *AmbtV* cDNA, total pollen RNA was reverse-transcribed with the BG20 primer, a 35-mer oligo with 17 dT residues and an adapter sequence con-

taining three endonuclease-recognition sites (*Eco*RI, *Sal*I and *Cla*I; Table I). The resultant cDNA was then used for PCR using BG21 (adapter-primer) and BG19 (*AmbtV* gene-specific primer). No DNA bands were observed when cDNA was omitted from the PCR (data not shown), indicating that the amplification was not contaminated with exogenous DNA. When primer BG21 was used alone, some nonspecific amplification was observed (Fig. 4A, lane 1). Certain secondary structures of ss cDNAs and/or some nonspecific priming may account for this result. However, when both the primers were used, a quite different pattern of DNA bands was observed (Fig. 4A, lane 2). To obtain more specific amplification (lanes 3–5), we changed

A: PCR amplification of partial coding segment



B: 3'-end amplification by anchored-PCR



C: 5'-end amplification by anchored-PCR

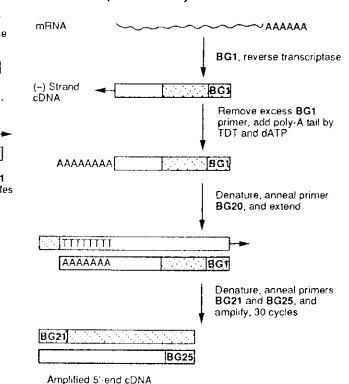


Fig. 2. Schematic representation of the PCR protocols. The stippled areas represent the portion of *AmbtV* where the aa sequence is known. (Part A) Amplification of the partial coding segment. Based on the known N- and C-terminal sequences of AmbtV, two degenerate primers BG2 and BG1, respectively, were synthesized (Table I). To reduce the degeneracy, 'inosine' was introduced at positions 9 and 12 of BG2 primer where any base could be present; also, the third nt of the C-terminal Thr was excluded in the BG1 primer. Total RNA was reverse-transcribed with oligo(dT) primer to create a (-) strand of cDNA and the BG1 and BG2 primers were used to amplify the partial coding segment. From the nt sequence obtained, a number of gene-specific primers were synthesized and used for 3'- and 5'-end amplifications. (Part B) Amplification of the 3'-end by anchored-PCR. Total RNA was reverse-transcribed with a (dT) $_{17}$ adapter-primer (BG20, Table I) to create a (-) strand of cDNA. Then a gene-specific primer, BG19 was annealed with the (-) strand cDNA and extended to generate the (+) strand. The resulting cDNA was amplified with BG19 (gene-specific) and BG21 (adapter) primers. The gene-specific primer and an adapter-primer were used for amplification as suggested previously (Frohman et al., 1988). The gene-specific amplified band was analyzed by Southern blot and by employing a second PCR with an internal oligo. (Part C) Amplification of the 5'-end by anchored-PCR. Total RNA was reverse-transcribed with a gene-specific primer, BG1, to create a (-) strand of cDNA. A poly(A) tail was added, annealed with the (dT) $_{17}$ adapter primer (BG20) and extended to generate the (+) strand. Then, a gene-specific internal oligo primer (BG25) and the adapter-primer (BG21) were used for amplification. The 5'-end specific amplified band was analyzed by Southern blot and by employing a second PCR with another internal oligo. See sections (b), (c), and (d) for further details.

TABLE I

Oligos used for reverse transcription and/or PCR amplification

Name ^a	Sequence ^b	Derived from aa positions ^c
BG1	5'-GTRCARTTYTTRTRCA	40-35
BG2	5'-GAYGAYGGICTITGYTAYGARGG	1- 8
BG4	5'-TCRTARCAIACRCARTAYTTICCDATIGG	30-21
BG19	5'-AGgaattcGTGGTAAAGTGGGCAAATAC	11-17
BG20	5'-GAgaaattcGTCGACATCGA(T) ₁₇	—
BG21	5'-GAgaaattcGTCGACATCGAT	—
BG25	5'-AGgaattcTGCCTTGGGAATCATAACAGAC	33-27
BG26	5'-TGTAGCCCCATTGGGAAGTACTGT	19-26
BG27	5'-ACAGTACTTCCCAATGGGGCTACA	26-19

^a The oligos are designated by the initials of the designer (B. Ghosh) according to our laboratory coding system. BG1, BG2 and BG4 are degenerate primers, where R = A or G; Y = C or T; D = G, A or T; I = inosine. Oligos were synthesized and purified by HPLC at the Johns Hopkins School of Hygiene and Public Health.

^b Sequences of oligos used in this study. *Eco*RI restriction sites are shown in lower-case letters.

^c The respective aa positions of AmbtV from where the oligos are synthesized. For the oligos BG1, BG2, BG4 and BG19 part of the codon of the aa at positions 40, 8, 30 and 11 are used during synthesis (see Fig. 5).

template for a second PCR using primers BG26 and BG21 (Table I). Since the BG26 primer is 23 nt internal to BG19 (used in the first PCR), DNA band 23 nt shorter than the template was seen by agarose-gel electrophoresis (data not shown).

The electroeluted DNA segments (approx. 190 pp), obtained from the PCR utilizing primers BG19 and BG21, were subcloned into the M13 vector and ten recombinant clones were sequenced. None of these clones were found to correspond to the *AmbtV* coding sequence, presumably due to significant heterogeneity of the electroeluted DNA fragments. Therefore, to increase our chance of obtaining *AmbtV*-specific clones, we screened about 2000 recombinant M13 clones with BG26 and BG27 probes (BG27 is complementary to BG26, Table I) separately and 19 positive clones were obtained (Fig. 5 legend). Both BG26 and BG27 probes, which are complementary to each other were used to obtain M13 recombinants containing inserts in both orientations. The 3'-end cDNA sequence obtained from these clones was found to correspond to part of the known *AmbtV* aa sequence (22 aa). Moreover, it was also found to overlap with the cDNA sequence (65 nt) obtained from the initial PCR using degenerate primers BG1 and BG2.

(d) 5'-End amplification of *AmbtV*

Total RNA was reverse-transcribed with a gene-specific degenerate primer (BG1) and excess primer was removed using a Centricon 30 spin-filter. The cDNA was 'tailed' with poly(A) using terminal deoxynucleotidyl transferase in the presence of dATP, and was used as a template for PCR amplification with BG20, BG21 and BG25 primers (Table I). The PCR products were analyzed by agarose gel

electrophoresis (Fig. 4A, lanes 6 and 7) which showed that a DNA band of about 281 bp had been amplified. The amplification was more prominent at a lower annealing temperature (55°C, lane 7) than at a higher temperature (60°C, lane 6). To confirm that the amplified band was *AmbtV*-specific, a Southern-blot analysis was performed using BG26 as a probe (Fig. 4B, lanes 6 and 7). In addition, the first PCR product was used as a template for a second PCR using oligos BG27 and BG21 as primers (Table I). Since the BG27 primer is 21 nt internal to the BG25 primer (used in the first PCR), a DNA band 21 nt shorter than the template DNA was seen in an agarose gel (data not shown).

The 281-bp DNA segment was then digested with *Eco*RI and subcloned into the M13 vector. Of 30 recombinant M13 clones sequenced, ten sequences (both orientations) were found to correspond to the known *AmbtV* N-terminal aa sequence (27 aa) plus additional nt corresponding to the leader and 5'-untranslated region. Moreover, the 5'-end cDNA sequence was also found to overlap with the cDNA sequences obtained by the PCRs of 3'-end (47 bp) and degenerate primers (78 bp).

(e) Assembly of the overlapping sequences

Having three sets of overlapping sequences, we compiled the full-length *AmbtV* cDNA sequence (Fig. 5), which is compatible with the known aa sequence of *AmbtV* (Goodfriend et al., 1985). From the aa sequence deduced from the cDNA sequence, it is apparent that the *AmbtV* protein has a 28- or 33-aa signal sequence which is comprised mostly of hydrophobic aa. Of the two possible translation-initiation codons, it is not known which is actually used in *AmbtV* protein synthesis. Also, there are two

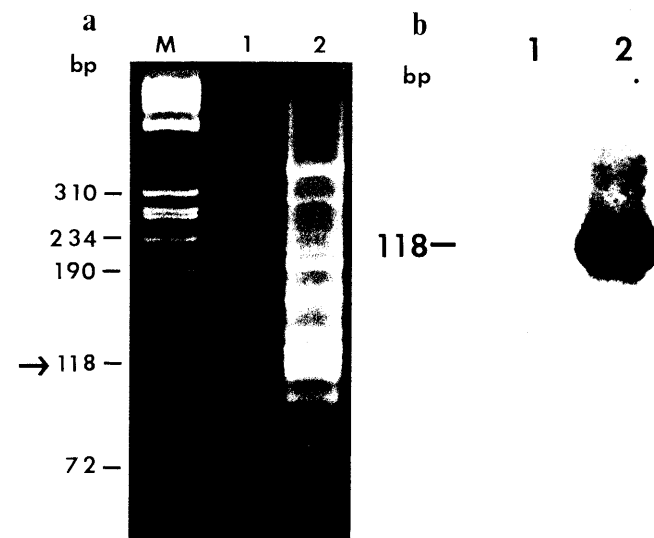


Fig. 3. Amplification of *AmbtV* cDNA using degenerate primers BG1 and BG2 (Table I). (Panel a) Agarose gel (4.5%) electrophoresis of PCR products derived from a total RNA preparation from *A. trifida* pollen. **Reverse transcription:** 12 µg of total RNA in 40 µl of water was mixed with 1 µl (0.5 µg/µl) of oligo(dT) primer, heated at 65°C for 2 min and quenched on ice. 10 × RT buffer (100 mM Tris · HCl pH 8.2/60 mM MgCl₂/400 mM KCl/10 mM dithiothreitol)/1 µl (40 units) of RNasin/6 µl of 10 × dNTPs (5 mM of each dNTP)/1.5 µl (36 units) of AMV RT and 4.5 µl of water were added to this tube. The mixture was then incubated at 42°C for 2 h. The reaction mixture was phenol/chloroform-extracted, ethanol-precipitated and resuspended in 50 µl of TE. **PCR:** 1 µl of cDNA was amplified through 30 PCR cycles in the presence of 75 pmol each of BG1 and BG2 primers/10 µl of 10 × PCR buffer (500 mM KCl/100 mM Tris · HCl pH 8.4/40 mM MgCl₂/0.1% gelatin)/15 µl of dNTPs (5 mM of each dNTP)/Taq polymerase as described (Saiki et al., 1988). The cDNA was initially denatured for 5 min at 95°C and cooled to 72°C. Then 2.5 units of Taq polymerase (Perkin-Elmer/Cetus) were added and the mixture was overlaid with 40 µl of mineral oil. The conditions for PCR amplification were as follows: denaturation at 92°C for 1.5 min, annealing at 47°C for 2 min, and extension at 72°C for 1.5 min. At the end of the last cycle, the extension was carried out for an additional 10 min at 72°C, the DNA was chloroform-extracted and the aqueous phase transferred to another tube. Of the sample 5 µl was taken for electrophoresis on a 4.5% agarose gel (GTG agarose, NuSieve) and stained with EtBr. Lane 'M' represents *Hae*III-digested φX174 RF DNA marker, lane 1 is a control (no DNA), lane 2 represents the PCR products obtained from 1 µl of cDNA. The putative band of *AmbtV* is indicated by an arrow. (Panel b) Southern-blot analysis of the DNAs shown in panel a. The DNAs were denatured and transferred to a nylon membrane (GeneScreen Plus) as described (Maniatis et al., 1982). The blot was dried, baked for 2 h at 80°C, prehybridized and hybridized with the BG4 oligo probe (specific activity, 7.6 × 10⁷ cpm/µg) under conditions described for the Northern-blot analysis. The filter was washed with 1 × SSC/0.1% SDS at 37°C for 2 × 30 min and then with 0.5 × SSC/0.1% SDS at 54°C for 2 × 30 min. The membrane was exposed to x-ray film. Lanes 1 and 2 represent PCR products as described in panel a. The positions of the DNA markers are shown on the left.

in-frame stop codons (UAA and UGA) immediately after an ACU codon, which encodes the published C-terminal aa, Thr (Goodfriend et al., 1985).

(f) Comparison of the sequence with the protein database

Comparing the *AmbtV* aa (including leader) sequence with the EMBL-protein sequence database revealed the expected homologies with AmbaV allergen (49% identity) from *A. artemisiifolia* pollen (Mole et al., 1975). It shows weak but significant homology with a protozoan, *Paramecium primaurelia* G surface protein (47.8% identity in a 23-aa overlap; Prat, 1990), a DIF-induced prestalk pDd63 precursor protein of the slug, *Dictyostelium discoideum*

(27.8% identity in a 79-aa overlap; McRobbie et al., 1988), a von Willebrand factor precursor (63.6% identity in an 11-aa overlap; Sporn et al., 1986), a herpesvirus probable membrane antigen p140 (25% identity in a 56-aa overlap; Kikuchi et al., 1990) and a maize α-amylase/trypsin inhibitor (29.6% identity in a 27-aa overlap), also known to be an allergen (Barbar et al., 1989). Other than the known homology with short ragweed AmbaV, it is premature to speculate whether any of these homologies are of biological significance.

(g) Conclusions

(1) Using a combination of PCR and anchored-PCR

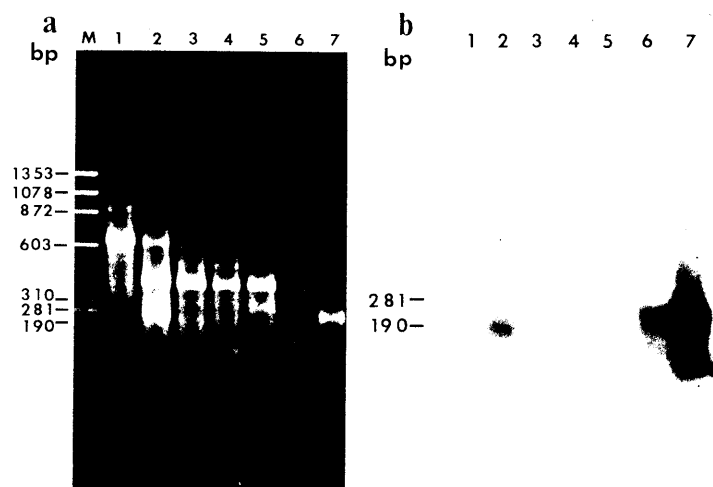


Fig. 4. Amplification of 3' and 5' ends of *AmbtV* cDNAs. (Panel a) Analysis of 3' and 5' ends of *AmbtV* cDNAs in an agarose gel. **3'-End amplification by anchored-PCR.** Total RNA (12 µg in 40 µl of water) was mixed with 1 µl (0.5 µg/µl) (dT)₁₇-adapter-primer (BG20), heated at 65°C for 2 min and quenched in ice. The reverse transcription was performed as described in Fig. 3. The reaction mixture was diluted to 500 µl with TE and stored at 4°C. The cDNA pool (2 µl)/50 pmol each of BG21 (adapter) and BG19 (*AmbtV*-specific) primers/10 µl of 10 × PCR buffer/15 µl of dNTPs (5 mM of each dNTP) in a total volume of 100 µl were denatured at 95°C for 5 min and cooled at 72°C. 2.5 units of *Taq* polymerase were added and the cDNA was amplified in a DNA Thermal Cycler as described below. **5'-End amplification by anchored-PCR.** Total RNA (12 µg) was reverse-transcribed using the *AmbtV* gene-specific degenerate primer BG1 as described in panel a, and the resultant cDNA was diluted to 2 ml with 0.1 × TE. The excess primer was removed using a Centricon 30 spin-filter (Amicon Corp.) for 20 min at 1000 × g following the manufacturer's protocol. The residual material was collected by centrifugation for 2 min at 1000 × g, resuspended in 2 ml deionized water, concentrated using a Centricon 30 unit and collected in a final volume of 200 µl. Of the cDNA 45 µl was mixed with 3 µl of 6 mM dATP, 12 µl of 5 × 'tailing' buffer (BRL) and 30 units of terminal deoxynucleotidyl-transferase (BRL). The mixture was incubated for 30 min at 37°C, heated at 65°C for 5 min, diluted to 300 µl with 0.1 × TE and stored at -20°C. The 5' cDNA pool (5 µl)/5 pmol of BG20 [(dT)₁₇ adapter; Table I]/30 pmol of BG21 (adapter-primer)/30 pmol of BG25 (*AmbtV* specific primer)/10 µl of PCR buffer (500 mM KCl/100 mM Tris · HCl pH 8.4/15 mM MgCl₂/0.1% gelatin) and 10 µl of 10 × dNTPs (2 µM each) were mixed in a total volume of 100 µl and denatured for 10 min at 95°C. *Taq* polymerase (2.5 units) was added and the cDNA amplified as described below. Amplified DNA (5 µl) was run on a 1.5% agarose gel and stained with EtBr. Lane 'M' is as in Fig. 3. Lanes 1-5 represent 3'-end amplification. Lanes 1, amplified products with primer BG21 alone, extended initially at 72°C for 30 min, followed by 30 cycles of 93°C, 1 min; 54°C, 1 min, 72°C, 1.5 min and a final extension of 10 min at 72°C; 2, products amplified with BG19 and BG21 using same conditions as lane 1; 3, as lane 2 except that the initial 30-min extension was omitted and the annealing time was reduced to 45 s; 4, as lane 3 except 10% (v/v) dimethylsulfoxide was included in the PCR buffer; 5, as lane 3 except the annealing temperature was increased to 60°C. Lanes 6 and 7 represent the 5'-end amplification. Lanes 6 amplified products using primers BG20, BG21 and BG25 using a cycling conditions of 94°C, 2 min; 60°C, 45 s, 72°C, 45 s for 30 cycles, followed by a final extension of 10 min at 72°C; 7, same as lane 6 except the annealing temperature was decreased to 55°C. (Panel b) Southern-blot analysis of gels shown in panel a. The blot was used for hybridization with the oligo probe, BG26 (Table I, specific activity of 2.0 × 10⁸ cpm/µg). After overnight hybridization at 60°C, the filter was washed with 2 × SSC/0.1% SDS at 23°C for 2 × 15 min and then with 1 × SSC/0.1% SDS at 45°C for 2 × 15 min. The blot was then exposed to x-ray film for autoradiography. Lanes 1 through 7 are as described in panel a.

approach, we have cloned and sequenced the full length *AmbtV* cDNA. The approach described above can be used as a general method to clone other allergen cDNAs provided some aa sequence information is available. We originally attempted to clone *AmbtV* cDNA by constructing a cDNA library. The library was initially screened extensively by using human sera from *AmbtV*-allergic patients and also by appropriate (degenerate) oligo probes derived from the *AmbtV* aa sequence. We screened more than 2 × 10⁶ recombinant clones without success. A number of reasons could account for these results: (a) the represen-

tation of the *AmbtV* clone in the cDNA library may be very low so that one would have to screen several million clones to get a single positive clone; (b) the *AmbtV* protein may be toxic to the cell and as a result clones can not be maintained and propagated in *Escherichia coli* (this protein contains eight half-Cys residues out of 40 total aa, which may render it toxic); (c) the antibody we were using may not be able to detect the fusion protein even if it were expressed in the clone; (d) the codon usage of this plant species is not known and as a result the effective concentration of the degenerate oligo probe used for screening may have been

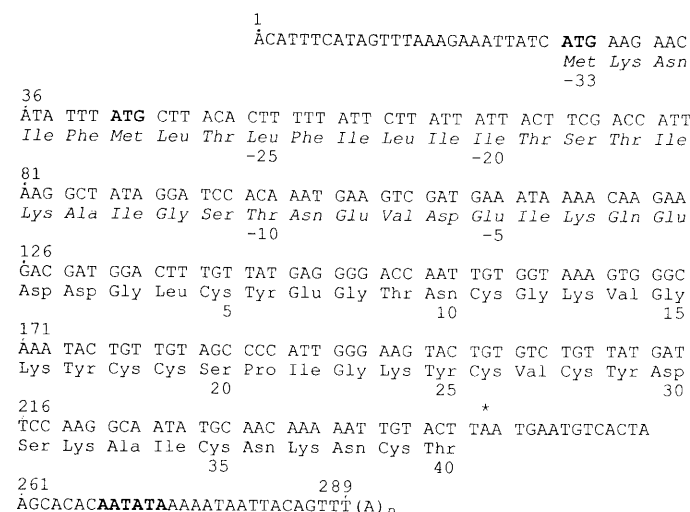


Fig. 5. Nucleotide sequence of *AmbtV* cDNA. Numbers above the sequences indicate nt and those below the aa sequences. The aa sequence deduced only from the nt sequence is shown in italics. The stop codon is marked by an asterisk. The two possible start codons and the polyadenylation signal are shown in bold type. In some 3'-end clones, additional sequences, 5'-CAAAATGCTT (25%) or 5'-AAATGCT (25%) are present immediately before the poly(A) sequence. The GenBank accession number is M38782. **Sequencing of PCR products.** The PCR products were digested with *EcoRI* and separated by preparative agarose gel electrophoresis. Regions of the gel containing specific bands were isolated by electroelution (IBI, electroeluter apparatus) and precipitated by adding two volumes of ethanol. The DNA which was amplified with the degenerate BG1 and BG2 primers (Fig. 2A) had no restriction sites at the ends. For subcloning into M13 vectors, the DNA was phosphorylated with T4-polynucleotide kinase (New England Biolabs), treated with *E. coli* DNA polymerase and ligated with an *EcoRI/NotI* linker-adaptor (5'-AATTCGGCGCCG) as described (Maniatis et al., 1982). The ligated DNA was phosphorylated with T4-polynucleotide kinase and the excess linker-adaptor was removed from the reaction mixture by Sephadex G-150 column chromatography as described (Maniatis et al., 1982). The peak fraction from the column was pooled, ethanol-precipitated, resuspended in sterile water and digested with *EcoRI*. The PCR DNA fragment was ligated into an *EcoRI*-digested dephosphorylated M13mp10 vector. CaCl₂-treated competent *E. coli*, strain JM103, was transfected with the ligated DNA and plated appropriately to obtain well-separated plaques. The white plaques were selected on XGal/ITP plates as described (Maniatis et al., 1982). For the isolation of 3'-end specific clones, the M13 plaques were screened as described below. Biotrans nylon membranes (ICN) were placed on the plates containing overnight grown M13 plaques. After 30 s of incubation, the filters were removed, dried and baked for 2 h at 80°C. The filters were washed in 5 × SSPE, prehybridized for 3-4 h at 60°C (for prehybridization and hybridization solutions, see Northern blot) and hybridized overnight at 60°C (without formamide) using probes BG26 (specific activity 2.0 × 10⁸ cpm/µg) and BG27 (specific activity, 7.2 × 10⁷ cpm/µg) separately. The filters were washed with an excess of 2 × SSC/0.1% SDS at room temperature for 2 × 15 min and then with 1 × SSC/0.1% SDS at 40°C for 2 × 15 min, and exposed to x-ray film. The positive plaques were identified by aligning the autoradiograms with the plates. The ss DNA was isolated from M13 clones and sequenced using Sequenase (U.S. Biochemicals, Cleveland, OH), according to the supplier's protocol.

too low to detect any positive clones. To improve our chances of finding a positive clone, we used the PCR-amplified cDNA as well as oligos representing codons 1-31 and 20-31 to screen the cDNA library (5 × 10⁵ recombinant clones), but we were still unable to pick up a single positive clone. In consideration of all these problems, our PCR approach is straightforward and much less time-consuming than conventional cDNA cloning. The overall procedure should also be generally useful. We think this is perhaps the best, and possibly the only, approach to clone cDNAs expressed at very low levels in a cDNA library. The expression of *AmbtV* cDNA will be tested in a variety of host cells. After successful expression of the cDNA, site-

directed mutants will be generated to map the epitopes which interact with class-II MHC molecules and T-cell receptors. Shorter fragments generated from this DNA can be similarly expressed for mapping the epitopes recognized by human T and B lymphocytes (cf., Mehra et al., 1986; Lamb et al., 1987).

(2) Finally, cloning of allergen-encoding cDNAs and determination of their sequences and the encoded aa sequences is a prerequisite to understanding the fundamental mechanisms of allergic disorders in humans. The sequence information may also enable us to map the epitopes involved in the interactions with human antibodies, class-II MHC molecules and T-cell receptors.

ACKNOWLEDGEMENTS

We thank Dr. Kunal Sen for helpful discussions. This work was supported by NIH grant No. AI-19727, and is publication No. 018 from the Johns Hopkins Asthma and Allergy Center.

REFERENCES

- Barbar, D., Sanchez-Monge, R., Gomez, L., Carpizo, J., Armentia, A., Lopez-Otin, C., Juan, F. and Salcedo, G.: A barley flour inhibitor of insect α -amylase is a major allergen associated with baker's asthma disease. *FEBS Lett.* 248 (1989) 119–122.
- Breiteneder, H., Pettenburger, K., Bito, A., Valenta, R., Kraft, D., Rumpold, H., Scheiner, O. and Breitenbach, M.: The gene coding for the major birch pollen allergen *Ber1*, is highly homologous to a pea disease resistance response gene. *EMBO J.* 8 (1989) 1935–1938.
- Chua, K.Y., Stewart, G.A., Thomas, W.R., Simpson, R.J., Dilworth, R.J., Plozza, T.M. and Turner, K.J.: Sequence analysis of cDNA coding for a major house dust mite allergen, *Der p 1*. *J. Exp. Med.* 167 (1988) 175–182.
- Fang, K.S.F., Vitale, M., Fehlner, P. and King, T.P.: cDNA cloning and primary structure of a white-face hornet venom allergen, antigen 5. *Proc. Natl. Acad. Sci. USA* 85 (1988) 895–899.
- Frohman, M.A., Dush, M.K. and Martin, G.R.: Rapid production of full-length cDNAs from rare transcripts: amplification using a single gene-specific oligonucleotide primer. *Proc. Natl. Acad. Sci. USA* 85 (1988) 8998–9002.
- Goodfriend, L., Choudhury, A.M., Klapper, D.G., Coulter, K.M., Dorval, G., DelCarpio, J. and Osterland, C.K.: Ra5G, a homologue of Ra5 in giant ragweed pollen: isolation, HLA-DR-associated activity and amino acid sequence. *Mol. Immunol.* 22 (1985) 899–906.
- Kikuchi, G.E., Glorioso, J.C. and Nairn, R.: Cross-linking studies show that herpes simplex virus type 1 glycoprotein C molecules are clustered in the membrane of infected cells. *J. Gen. Virol.* 71 (1990) 455–458.
- King, T.P., Norman, P.S. and Connell, J.J.: Isolation and characterization of allergens from ragweed pollen, II. *Biochemistry* 3 (1964) 458–468.
- King, T.P., Norman, P.S. and Lichtenstein, L.M.: Isolation and characterization of allergens from ragweed pollen, IV. *Biochemistry* 6 (1967) 1992–2000.
- Lamb, J.R., Ivanyi, J., Rees, A.D.M., Rothbard, J.B., Howland, K., Young, R.A. and Young, D.B.: Mapping of T cell epitopes using recombinant antigens and synthetic peptides. *EMBO J.* 6 (1987) 1245–1249.
- Lewis, W.H., Vinay, P. and Zenger, V.E.: Airborne and Allergenic Pollen of North America. Johns Hopkins Univ. Press, Baltimore, 1983.
- Maniatis, T., Fritsch, E.F. and Sambrook, J.: Molecular Cloning. A Laboratory Manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY, 1982.
- Marsh, D.G., Hsu, S.H., Roebber, M., Kautzky, E.E., Freidhoff, L.R., Meyers, D.A., Pollard, M.K. and Bias, W.B.: HLA-Dw2: a genetic marker for human immune response to short ragweed allergen Ra5, I. Response resulting primarily from natural antigenic exposure. *J. Exp. Med.* 155 (1982a) 1439–1451.
- Marsh, D.G., Meyers, D.A., Freidhoff, L.R., Kautzky, E.E., Roebber, M., Norman, P.S., Hsu, S.H. and Bias, W.B.: HLA-Dw2: a genetic marker for human immune response to short ragweed allergen Ra5, II. Response after ragweed immunotherapy. *J. Exp. Med.* 155 (1982b) 1452–1463.
- Marsh, D.G., Zwoilo, P., Huang, S.K., Ghosh, B. and Ansari, A.A.: Molecular studies of human response to allergens. Cold Spring Harbor Symp. Quant. Biol. 54 (1990) 459–470.
- McRobbie, S.J., Tilly, R., Blight, K., Ceccarelli, A. and Williams, J.G.: Identification and localization of proteins encoded by two DIF-inducible genes of *Dictyostelium*. *Develop. Biol.* 125 (1988) 59–63.
- Mehra, V., Sweetser, D. and Young, R.A.: Efficient mapping of protein antigenic determinants. *Proc. Natl. Acad. Sci. USA* 83 (1986) 7013–7017.
- Mole, L.E., Goodfriend, L., Lapkoff, C.B., Kehoe, J.M. and Capra, J.D.: The amino acid sequence of allergen Ra5. *Biochemistry* 14 (1975) 1216–1220.
- Perez, M., Ishioka, G.Y., Walker, L.E. and Chesnut, R.W.: cDNA cloning and immunological characterization of the rye grass allergen, *Lol p 1*. *J. Biol. Chem.* 265 (1990) 16210–16215.
- Prat, A.: Conserved sequences flank variable tandem repeats in two alleles of the G surface protein of *Paramecium primaurelis*. *J. Mol. Biol.* 211 (1990) 521–535.
- Rafnar, T., Rogers, B.L., Bond, J.E., Kuo, M. and Klapper, D.G.: Isolation of cDNA clones coding for the major short ragweed allergen *Amb a 1* (antigen E). *J. Biol. Chem.* 266 (1991) 1229–1236.
- Roebber, M., Hussain, R., Klapper, D.G. and Marsh, D.G.: Isolation and properties of a new short ragweed pollen allergen, Ra6. *J. Immunol.* 131 (1983) 706–711.
- Roebber, M., Klapper, D.G., Goodfriend, L., Bias, W.B., Hsu, S.H. and Marsh, D.G.: Immunochemical and genetic studies of *Amb1V* (Ra5G), an Ra5 homologue from giant ragweed pollen. *J. Immunol.* 134 (1985) 3062–3069.
- Saiki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Erlich, H.A.: Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science* 239 (1988) 487–491.
- Smith, J.M.: Epidemiology and natural history of asthma, allergic rhinitis, and atopic dermatitis (eczema). In: Middleton, E. (Ed.), *Allergy: Principles and Practice* (3rd ed.) Mosby, St. Louis, MO, 1988, pp. 891–929.
- Sporn, L.A., Marder, V.J. and Wagner, D.D.: Inducible secretion of large, biologically potent von Willebrand factor multimers. *Cell* 46 (1986) 185–190.
- Thomas, W.R., Stewart, G.A., Simpson, R.J., Chua, K.Y., Plozza, T.M., Dilworth, R.J., Nisbet, A. and Turner, K.J.: Cloning and expression of DNA coding for the major house dust mite allergen *Der p 1* in *E. coli*. *Int. Arch. Allergy Appl. Immunol.* 85 (1988) 127–129.
- Underdown, B.J. and Goodfriend, L.: Isolation and characterization of an allergen from short ragweed pollen. *Biochemistry* 8 (1969) 980–989.
- Young, P.: Asthma and Allergies, an Optimistic Future. NIH Publ. No. 80-388, Washington, DC, 1980, pp. 14–26.
- Zwoilo, P., Ehrlich-Kautzky, E., Ansari, A.A., Scharf, S.J., Erlich, H.A. and Marsh, D.G.: Molecular studies of human immune response genes for the short ragweed allergen, *Amb a V*. Sequencing of HLA-D second exons in responders and non-responders. *Immunogenetics* 33 (1991) 141–151.

Gene, 101 (1991) 239–246

© 1991 Elsevier Science Publishers B.V. 0378-1119/91/\$03.50

GENE 04011

A bifunctional fusion between β -glucuronidase and neomycin phosphotransferase: a broad-spectrum marker enzyme for plants

(Recombinant DNA; gene cloning; CaMV 35S promoter; tobacco plants; reporter gene; plasmids)

Raju S.S. Datla, Joe K. Hammerlindl, Lawrence E. Pelcher, William L. Crosby and Gopalan Selvaraj

Molecular Genetics Section, Plant Biotechnology Institute, NRC Canada, Saskatoon, SK S7N 0W9 (Canada)

Received by J.L. Slightom: 31 July 1990

Revised: 31 October 1990

Accepted: 11 November 1990

SUMMARY

We have used an in vivo selection approach to isolate a gene encoding a bifunctional fusion peptide between *Escherichia coli* β -glucuronidase (GUS) and neomycin phosphotransferase II (NPT-II) from transposon Tn5 in the N11₂-GUS::NPT-II-COOH configuration. The fused gene is predicted to encode a fusion peptide 885 amino acids long, and was shown in *E. coli* to synthesize a 97-kDa GUS⁺ NPT-II⁺ gene product. Gel-filtration chromatography suggested that, while the native GUS may be active as a dimer and NPT-II as a monomer, the elution profile of the fusion protein is consistent with that of a trimer. The fusion marker has been produced and defined in transgenic *Nicotiana tabacum* plants, where both the chimeric gene and the gene product were stable. The bifunctional gene enabled direct Km^R selection at the callus stage and enzymatic or histochemical assessment of the steady-state production of GUS activity in regenerated plants. In addition to allowing structure-function determination for the GUS and NPT-II domains of the fusion peptide, the *gus::npt-II* gene simplifies vector constructs where both marker domains are desired.

INTRODUCTION

Transcriptional and translational fusions to reporter genes whose products can be easily assayed offer powerful approaches to studying gene structure, expression, regulation, gene product assembly, transport and compart-

mentalization (Rosenberg et al., 1983; Bonnerot et al., 1987; Finnegan et al., 1989). Some of these have also been applied in more extensive studies such as discerning cell lineage during development and purification of gene products (Germino et al., 1983; Silhavy and Beckwith, 1985; Scholtissek and Grosse, 1988). Pioneered with *E. coli lacZ*, gene fusions have been adapted to other organisms including yeast, animal systems and plants. While transcriptional fusions might be construed as more straightforward, translational fusions require reporters that can function despite covalent addition of extraneous polypeptides to their N or C termini. A number of such markers have been identified including NPT-II.

Some markers impart a selective phenotype such as antibiotic resistance while others specify an enzymatic activity for which there is a chromogenic or luminescence assay, making them suitable in combination for selection or facile screening. Of the several available antibiotic resistance genes, *npt-II* is the most widely used marker in

Correspondence to: Dr. W.L. Crosby, Molecular Genetics Section, Plant Biotechnology Institute, NRC Canada, 110 Gymnasium Road, Saskatoon, SK S7N 0W9 (Canada) Tel. (306) 975-4187/5241; Fax (306) 975-4839; Email: bc9944@NRCVM01.bitnet.

Abbreviations: aa, amino acid(s); Ap, ampicillin; bp, base pair(s); CaMV, cauliflower mosaic virus; DTT, dithiothreitol; GUS, β -glucuronidase; *gus*, *E. coli* gene encoding GUS (*uidA*); kb, kilobase(s) or 1000 bp; Km, kanamycin; NPT-II, neomycin phosphotransferase-II; *npt-II*, gene encoding NPT-II; nt, nucleotide(s); oligo, oligodeoxynucleotide; PAGE, polyacrylamide-gel electrophoresis; Polk, Klenow (large) fragment of *E. coli* DNA polymerase I; ^R, resistance/resistant; SDS, sodium dodecyl sulfate; XGluc, 5-bromo-4-chloro-3-indolyl- β -D-glucuronide; ::, novel joint (fusion); [], denotes plasmid-carrier state.