

CLONING AND SEQUENCING OF cDNA CODING FOR *Cry j I*,
A MAJOR ALLERGEN OF JAPANESE CEDAR POLLEN*

Toshio Sone*, Naoki Komiyama, Kimiko Shimizu, Takahiro Kusakabe¹,
Keiko Morikubo, and Kohsuke Kino

Meiji Institute of Health Science, 540 Naruda, Odawara, Kanagawa 250, Japan

Received January 24, 1994

cDNA clones coding for *Cry j I*, a major allergen of the Japanese cedar (*Cryptomeria japonica*) pollen, have been isolated. Two of the clones were sequenced and found to code for a putative 21-residue signal peptide and a 353-residue mature protein with a derived molecular weight of 38.5 KDa. Five possible N-linked glycosylation sites were found in the sequence. Comparison of the nucleotide sequences of the two clones revealed sixteen nucleotide differences and these led to five amino acid exchanges in the mature allergen, indicating that an isoform of the *Cry j I* molecule exists. The deduced amino acid sequence of the *Cry j I* shows 46 - 48 % identities with those of the *Amb a I* family and *Amb a II*, the major allergens of short ragweed. These findings should facilitate study of the structure-function relationship between the allergen and the immunocompetent cells.

© 1994 Academic Press, Inc.

Pollen from Japanese cedar (*Cryptomeria japonica*) is a potent seasonal aeroallergen in Japan with more than 10 % of the population being affected by cedar pollinosis. The crude extracts of Japanese cedar pollen available for clinical use contain active allergens and large amounts of sugar-like materials which are allergenically inactive. Purification of the allergens from the pollen should improve the accuracy of clinical diagnosis and the efficacy of immunotherapy. Along a line, two major allergens from Japanese cedar pollen were purified: the first one designated as Sugi basic protein (SBP) (1) or *Cry j I* (2), and the second as *Cry j II* (3). *Cry j I* is a basic glycoprotein with molecular weight of 41 - 45 KDa and pI of 8.9 - 9.2. The allergen was separated into four fractions based on charge and size differences (1). Taniai *et al.* (4) proposed that the difference might be due in part to the existence of N-linked carbohydrate chains. A 27-residue N-terminal

* The nucleotide sequence data reported in this paper will appear in the GSDB, DDBJ, EMBL and NCBI nucleotide sequence databases with the following accession numbers: D26544 and D26545.

*To whom correspondence should be addressed.

¹ Present address: Department of Biochemistry, Saga Medical School, Nabeshima, Saga 816, Japan.

sequence and a 28-residue internal sequence of the *Cry j I* molecule have been determined (4-6), however, the complete amino acid sequence has yet to be reported.

Molecular cloning is a powerful tool to characterize the allergens precisely. In the present study, we have isolated cDNA clones coding for the *Cry j I* genes and have sequenced two of these cDNA clones. Both clones consist of a 1.3-Kbp cDNA encoding a typical signal peptide and a mature *Cry j I* protein. The deduced amino acid sequence of the *Cry j I* reveals that an isoform of the *Cry j I* allergen is present. Cloning of the genes coding for *Cry j I* should facilitate the reproducible preparation of a recombinant allergen. The information of the complete amino acid sequence of *Cry j I* may lead to the determination of T- and B-cell epitopes.

MATERIALS AND METHODS

Plant materials. Pollen from *Cryptomeria japonica* was collected from mature male flowers of Japanese cedar trees located in Kanagawa and Shizuoka Prefectures and was stored at -70 °C prior to use.

Preparation, enzyme digestion, and sequencing of *Cry j I*. *Cry j I* protein was purified by the method described by Yasueda *et al.* (1). Enzymatic digestion of *Cry j I* by lysylendopeptidase (EC 3.4.21.50) in 8 M Urea was performed by the method described elsewhere (7, 8). The peptide mixture was applied to a reverse-phase HPLC using a Protein C4 column (Vydac, CA, U.S.A.), or a C18 column (Nihon Millipore, Tokyo) and each peptide fraction was collected. The amino acid sequence of the peptides was determined by using a 470A Protein Sequencer (Applied Biosystems Inc., CA, U.S.A.).

Isolation of RNA. Isolation of RNA was performed with a modified method described by Briteneder *et al.* (9). Frozen pollen (1 g) was homogenized with a Teflon pestle at a maximum speed for 30 strokes in a mixture of 15 ml homogenizing buffer (100 mM LiCl, 10 mM EDTA, 1 % SDS, 20 % 2-mercaptoethanol, and 100 mM Tris-HCl, pH 9.5) and 15 ml phenol/chloroform/isoamylalcohol (24:24:1). After phase separation by centrifugation, the aqueous phase was extracted again with phenol and then with CHCl₃ alone. The pH of the solution was adjusted to 7.5 by the addition of 3 M sodium acetate. An equal volume of 4 M LiCl was added to the solution and the mixture kept at -20 °C overnight. The frozen materials were allowed to thaw at room temperature and were then centrifuged. Total RNA were obtained as precipitates (~ 0.8 mg).

Amplification of screening probe by reverse-transcriptase polymerase chain reaction (RT-PCR). Single-stranded cDNA was synthesized from 1 µg total RNA by MoLV-reverse transcriptase supplied by GeneAmp RNA PCR Kit (Perkin-Elmer Cetus Instruments, CT, U.S.A.). *Taq* DNA polymerase, a forward primer and a reverse primer (1 µM) listed in Fig. 1, and the buffers were added to the solution containing single-stranded cDNA and the mixture was kept at 95 °C for 5 min to denature the reverse transcriptase. PCR amplification was performed for 20 cycles with each cycle consisting of 1 min at 95 °C for denaturation, 1 min at 43 °C to anneal primers and 1 min at 72 °C for polymerization. The PCR products were loaded onto a 1.5 % agarose gel and then were electrophoresed. An amplified cDNA band (named CryPCR4-6) was eluted from the gel and then subcloned into the pCR II vector (Invitrogen Co., CA, U.S.A.). The pCR II vector containing CryPCR4-6 was selected and amplified according to the protocol attached to the kit. The amplified vector was digested with *EcoRI* and the reaction mixture loaded onto 1.5 % agarose gel. After electrophoresis, CryPCR4-6 was eluted from the gel and used as a probe for screening.

cDNA library and cloning of *Cry j I* cDNA. cDNA was synthesized from 1 µg poly (A⁺) RNA using the cDNA Synthesis System Plus Kit (Amersham International Ltd., Buckinghamshire, England) and was cloned into the *EcoRI* site of the lambda-gt10 vector.

An aliquot of cDNA library was plated into 15 cm dishes and the plaques appeared were transferred to nitrocellulose membranes. Positive plaques were selected by hybridization screening using [³²P]-labeled CryPCR4-6 as a probe. The plaques were picked off, and amplified. A secondary screening was then performed.

Nucleotide sequencing. The cDNA insert obtained by *Eco*RI digestion of lambda-gt10 vector was subcloned into the *Eco*RI site of the pUC19 vector. The DNA sequence was determined by fluorescence-based sequencing of templates amplified by asymmetric PCR using 370A DNA Sequencing System (Applied Biosystems Inc., CA, U.S.A.) (10).

RESULTS AND DISCUSSION

The peptide fragments after lysylendopeptidase digestion of *Cry j* I molecule were separated into about 28 peaks by reverse-phase HPLC using either a Protein C4 column or a C18 column (data not shown). Nine sharp peaks appeared, their fractions selected and their amino acid sequences determined. Seven peptide fractions of which one corresponded to an *N*-terminal sequence showed reproducible amino acid sequences, respectively (shown in Fig. 1A). The amino acid sequences of the *N*-terminus and the No.5 internal peptide were identical to those of the *N*-terminus and an internal peptide reported previously by Taniai *et al.* (4).

Two peptide sequences, one with NWAQ located within the *N*-terminal sequence and another with FNQFGP within the No.3 internal peptide, were selected to synthesize primers for PCR. Four forward primers (FP) and two reverse primers (RP) were synthesized (shown in Fig. 1B). The inosine residue was inserted at highly redundant positions. A nucleotide was placed in the 17-mer primers at positions 13 - 17, since *Taq* DNA polymerase might be expected to drive well. A cDNA (CryPCR4-6) having 660 bp was amplified by PCR when FP-4 and RP-6 were used for primers.

The lambda-gt10 cDNA library of 2 x 10⁴ plaques was screened with [³²P]-labeled CryPCR4-6 and 22 positive clones (≈ 0.11 %) were obtained. Two of the cDNA clones (named pCCI-2-2 and pCCI-15), which were the longest among the clones, were sequenced (shown in Fig. 2). Both clones consisted of 1,280 and 1,271 bp nucleotides. The

A		B	
<i>N</i> -terminal Sequence		Amino Acid	Asn Trp Ala Gln Asn Arg
		Genetic Code	AAY TGG GCN CAR AAY MGN
DNPIDSCWRGDSNWAQNRMK		Forward Primer	
-LADCAVGFGSST		FP-1	AAY TGG GCI CAR AAT CG-3'
		FP-2	AAY TGG GCI CAR AAT AG-3'
Internal Sequence		FP-3	AAY TGG GCI CAR AAC CG-3'
		FP-4	AAY TGG GCI CAR AAC AG-3'
1. GGDLYTVTNSDDDPVNPAPGTLRY	Amino Acid	Phe Asn Gln Phe Gly Pro	
2. VMLLGHDDAYSDDK	Genetic Code	TTY AAY CAR TTY GGN CCN	
3. XXVAFNQFGPNXGQRMPRARYGL	Reverse Primer		
4. QVTIRIGXK	RP-5	AAA TTR GTY AAR CCI GG-5'	
5. EAFNVENGNA TPQLTK	RP-6	AAG TTR GTY AAR CCI GG-5'	
6. NAGVLTCSLSK			

Figure 1. The amino acid sequences of an *N*-terminus and six internal peptides from *Cry j* I (A) and the oligonucleotide primers synthesized for PCR amplification (B). X in (A) indicates an undetermined amino acid residue. Underlined amino acid sequences were chosen to synthesize the primers for PCR. Letters M and I in (B) mean A or C, and inosine, respectively.

A pCCI-2-2

TGCTCATAATCATAGCATAGCCGTATAGAA 30
 AGAAATCTACACTCTGCTACCAAAAAATGGATTCCCTTGGTTA 75
 M D S P C L -16
 GTAGCATTACTGGTTCTCTCTTTTGAATTTGGATCTTGGCTTTCT 120
 V A L L V L S F V I G S C F S -1
 GATAATCCCATAGATAGCTGCTGGAGGAGGACTCAAACTGGGCA 165
 D N P I D S C W R G D S N W A 15
 CAAAACAGAATGAAGCTCGCAGATTGTGCAGTGGGCTTCGGAAGC 210
 Q N R M K L A D C A V G F G S 30
 TCCACCATGGGAGGCAAGGAGGAGATCTTTATACAGTCAACAAC 255
 S T M G G K G G D L Y T V T N 45
 TCAGATGACGATCTGTGAATCTGCACTGAGGAACCTGCGCTAT 300
 S D D D P V N P A P G T L R Y 60
 GGAGCAACCCGAGATAGGCCCTGTGGATAATTTTCAGTGGGAAT 345
 G A T R D R P L W I I F S G N 75
 ATGAATATAAAGCTCAAAATGCCATGTACATTTGCTGGGTATAAG 390
 M N I K L K M P M Y I A G Y K 90
 ACTTTTGATGGCAGGGAGCACAAGTTTATATGGCAATGGCGGT 435
 T F D G R G A Q V Y I G N G S 105
 CCTGTGTGTTTATCAAGAGAGTGAGCAATGTATCATACACGGT 480
 P C V F I K R V S N V I I H G 120
 TTGCACTGTACGGCTGTAGTACTAGTGTTTTGGGGAATGTTTG 525
 L H L Y G C S T S V L G N V L 135
 ATAAACGAGAGTTTGGGGTGGAGCCTGTTCATCTCAGGATGGC 570
 I N E S F G V E P V H P Q D G 150
 GATGCTCTTACTCTGGCAGCTGCTACAATATTGGATTGATCAT 615
 D A L T L R T A T N I W I D H 165
 AATTCTTTCTCCAACTCTCTGATGCTGGTTCGATGTCACCTCT 660
 N S F S N S S D G L V D V T L 180
 TCTTCGACTGGAGTTACTATTTCACAACATCTTTTTTCAACCAT 705
 S S T G V T I S N N L F F N H 195
 CATAAAGTGATGTTGTAGGGCATGATGATCATATAGTATGAC 750
 H K V M L L G H D D A Y S D D 210
 AAATCCATGAAGGTGACAGTGGCGTTCAATCAATTTGGACCTAAC 795
 K S M K V T V A F N Q F G P N 225
 TGTGGACAAAGAATGCCAGGGCAGCATATGGACTTGTACATGTT 840
 C G Q R M P R A R Y G L V H V 240
 GCAACAATAATTATGACCCATGGACTATATATGCAATTTGGTGGG 885
 A N N N Y D P W T I Y A I G G 255
 AGTTCAATCCCAACCTCTAAGTGAAGGGAATAGTTTCACTGCA 930
 S S N P T I L S E G N S F T A 270
 CCAATGAGAGCTACAAGAAGCAAGTAACCATACGTATTTGGATGC 975
 P N E S Y K K Q V T I R I G C 285
 AAAACATCATCATCTTGTTCAAATTTGGGTGTGGCAATCTACACAA 1020
 K T S S S C S N W V W Q S T Q 300
 GATGTTTTTATAATGGAGCTTATTTTGTATCATCAGGGAATAT 1065
 D V F Y N G A Y F V S S G K Y 315
 GAAGGGGTATATATACACAAGAAAGAGCTTTCAATTTTGAG 1110
 E G G N I Y T K K E A F N V E 330
 AATGGGAATGCAACTCTCAATTTGACAAAAATGCTGGGGTTTTA 1155
 N G N A T P L T K N A G V L 345
 ACATGCTCTCTCTAAACGTTGTTGATGATGCATATATCTAGC 1200
 T C S L S K R C * 353
 ATGTTGTACTATCTAAATTAACATCAACAAGAAATATATCATGAT 1245
 GTATATTGTTGTAATAAAAAAAAAAAAAAAAAA 1280

B pCCI-15

CTCATAATCATAGCATAGCCGTATAGAA 29
 AGAAATCTACACTCTGCTACCAAAAAATGGATTCCCTTGGTTA 73
 M D S P C L -16
 GTAGCATTACTGGTTCTCTCTTTTGAATTTGGATCTTGGCTTTCT 118
 V A L L V L S F V I G S C F S -1
 GATAATCCCATAGATAGCTGCTGGAGGAGGACTCAAACTGGGCA 163
 D N P I D S C W R G D S N W A 15
 CAAAACAGAATGAAGCTCGCAGATTGTGCAGTGGGCTTCGGAAGC 218
 Q N R M K L A D C A V G F G S 30
 TCCACCATGGGAGGCAAGGAGGAGATCTTTATACAGTCAACAAC 253
 S T M G G K G G D L Y T V T N 45
 TCAGATGACGATCTGTGAATCTGCACTGAGGAACCTGCGCTAT 298
 S D D D P V N P A P G T L R Y 60
 GGAGCAACCCGAGATAGGCCCTGTGGATAATTTTCAGTGGGAAT 348
 G A T R D R P L W I I F S G N 75
 ATGAATATAAAGCTCAAAATGCCATGTACATTTGCTGGGTATAAG 388
 M N I K L K M P M Y I A G Y K 90
 ACTTTTGATGGCAGGGAGCACAAGTTTATATGGCAATGGCGGT 433
 T F D G R G A Q V Y I G N G S 105
 CCTGTGTGTTTATCAAGAGAGTGAGCAATGTATCATACACGGT 478
 P C V F I K R V S N V I I H G 120
 TTGCACTGTACGGCTGTAGTACTAGTGTTTTGGGGAATGTTTG 523
 L H L Y G C S T S V L G N V L 135
 ATAAACGAGAGTTTGGGGTGGAGCCTGTTCATCTCAGGATGGC 568
 I N E S F G V E P V H P Q D G 150
 GATGCTCTTACTCTGGCAGCTGCTACAATATTGGATTGATCAT 613
 D A L T L R T A T N I W I D H 165
 AATTCTTTCTCCAACTCTCTGATGCTGGTTCGATGTCACCTCT 658
 N S F S N S S D G L V D V T L 180
 TCTTCGACTGGAGTTACTATTTCACAACATCTTTTTTCAACCAT 703
 S S T G V T I S N N L F F N H 195
 CATAAAGTGATGTTGTAGGGCATGATGATCATATAGTATGAC 748
 H K V M L L G H D D A Y S D D 210
 AAATCCATGAAGGTGACAGTGGCGTTCAATCAATTTGGACCTAAC 793
 K S M K V T V A F N Q F G P N 225
 TGTGGACAAAGAATGCCAGGGCAGCATATGGACTTGTACATGTT 838
 C G Q R M P R A R Y G L V H V 240
 GCAACAATAATTATGACCCATGGACTATATATGCAATTTGGTGGG 883
 A N N N Y D P W T I Y A I G G 255
 AGTTCAATCCCAACCTCTAAGTGAAGGGAATAGTTTCACTGCA 928
 S S N P T I L S E G N S F T A 270
 CCAATGAGAGCTACAAGAAGCAAGTAACCATACGTATTTGGATGC 973
 P N E S Y K K Q V T I R I G C 285
 AAAACATCATCATCTTGTTCAAATTTGGGTGTGGCAATCTACACAA 1018
 K T S S S C S N W V W Q S T Q 300
 GATGTTTTTATAATGGAGCTTATTTTGTATCATCAGGGAATAT 1063
 D V F Y N G A Y F V S S G K Y 315
 GAAGGGGTATATATACACAAGAAAGAGCTTTCAATTTTGAG 1108
 E G G N I Y T K K E A F N V E 330
 AATGGGAATGCAACTCTCAATTTGACAAAAATGCTGGGGTTTTA 1153
 N G N A T P L T K N A G V L 345
 ACATGCTCTCTCTAAACGTTGTTGATGATGCATATATCTAGC 1198
 T C S L S K R C * 353
 ATGTTGTACTATCTAAATTAACATCAACAAGAAATATATCATGAT 1243
 GTATATTGTTGTAATAAAAAAAAAAAAAAAAAA 1271

Figure 2. Nucleotide and predicted amino acid sequences of the two clones, pCCI-2-2 (A) and pCCI-15 (B), coding for *Cry j 1*. An ATG initiation codon and a TGA stop codon present bold types and an asterisk, respectively. The putative signal peptide is indicated by negative numbers. Underlined amino acids were identified by amino acid sequencing (shown in Fig. 1A). Dotted underlines indicate the position of PCR primers. Five possible N-linked glycosylation sites are shown by boxes. The nucleotide differences and the amino acid exchanges between pCCI-2-2 and pCCI-15 indicate inverted letters in (B).

sequence contained an open reading frame (ORF) of 1,122 bp starting with an ATG initiation codon and terminating with a TGA stop codon at position 1,180 (pCCI-2-2). The deduced amino acid sequence of the *Cry j 1* consisted of 374-residue polypeptide and

it contained a 21-residue signal peptide and a 353-residue mature protein. This sequence contained an *N*-terminal sequence and six internal peptides determined biochemically after lysylendopeptidase digestion of native *Cry j* I molecule (Fig. 1A). Taken together, the cDNA clones isolated are believed to encode the genes of the major allergen *Cry j* I.

Molecular weight of mature *Cry j* I based on the sequence was calculated to 38.5 KDa. The molecular weight of the native *Cry j* I was estimated to be 41 - 46 KDa (1) and 45 - 50 KDa (4) by SDS-PAGE, respectively. Five possible *N*-linked glycosylation sites having the characteristic Asn-X-Ser/Thr motif were found in the *Cry j* I sequence (Fig. 2A). Then, the difference of the molecular weight between native protein and predicted amino acid sequence of *Cry j* I is probably due to the glycosylation of the molecule. Actually, Taniai *et al.* (6) have reported the existence of an *N*-linked glycosylation site which corresponded to an Asn at position 333 in our sequence, therefore, at least one of Asn-residues in the sequence may be glycosylated.

Comparison of the nucleotide sequences of pCCI-2-2 with pCCI-15 revealed sixteen nucleotide differences (1.2 %) (shown in Fig. 2B). These nucleotide differences resulted in six amino acid exchanges at positions -10 (Leu to Phe), 122 (His to Tyr), 181 (Ser to Thr), 200 (Leu to Ser), 337 (Gln to His), and 340 (Lys to Gln). Similarly, amino acid exchanges were found at position 24 (Cys to Lys) and at position 327 (Phe to Pro) by comparison with the complete sequence and partial sequences of the peptides from native *Cry j* I reported previously (5, 6). Furthermore, Griffith *et al.* (11) have pointed out more amino acid exchanges at positions 38 (Gly to Glu), 51 (Val to Ala), 60 (Tyr to His) and 74 (Gly to Glu). These observations indicate that the genes coding for *Cry j* I are polymorphic and may belong to related genes of a small multigene family to the extent of 1 to 2 % nucleotide differences.

The nucleotide sequences and the deduced amino acid sequence of the *Cry j* I were compared to the sequences in the data bases of EMBL-GDB (Rel. 33.0, Dec. 1992) and NBRF-PDB (Rel. 35.00, Dec. 1992) and no proteins showing above 50 % sequence identity with *Cry j* I were found. However, the amino acid sequence of *Cry j* I showed 45.5 - 46.7, 48.0, 44.5, 43.4, and 45.1 % identities with those of *Amb a* I family (*Amb a* I.1 - I.4, isoforms), *Amb a* II (both are allergens of short ragweed) (12 - 14) (shown in Fig. 3), LAT56, LAT59 (pollen protein in tomato) (15), and 9612 (pistil-expressed gene in tomato) (16), respectively. Short ragweed (*Ambrosia artemisiifolia*) pollen is an important airborne allergen in North America and the *Amb a* I family and the *Amb a* II are the major allergens in patients with pollinosis. The similarity of the amino acid sequence of *Cry j* I with those of the *Amb a* I family and *Amb a* II is interesting in relation to the mechanism of pollinosis. *Cry j* I was found to share sequence identity in limited regions with pectate lyase A and E (38.5 % of 117 amino acids and 36.7 % of 109 amino acids, respectively) of *Erwinia chrysanthemi* (a bacterial plant pathogen), and pectate lyase A and B (24.3 % of 148 amino acids and 28.3 % of 145 amino acids, respectively) of *Erwinia carotovora* suggesting that *Cry j* I and allergen/proteins homologous to *Cry j* I mentioned above might have an enzyme activity similar to a pectate lyase (13, 15, 17).

<i>Cry j I</i> (pCCI-2-2)	DNPIDSCWRG DSNWAQNRMK LADCAVGFGS STMGGKGGDL YTVTNS*DDD	49
<i>Cry j I</i> (pCCI-15)	-----	49
<i>Amb a I.1</i>	Y-I--G---- KAD--E--KA -----Q--K G-V--D--I ----SEL---	50
<i>Amb a II</i>	H-I--K---C KPD--E--QA -GN--Q--K A-H--W--I -M--SDQ---	50
<i>Cry j I</i> (pCCI-2-2)	PVNPAFGTLR YGATRDRLW IIFSGNMNIK LKMPMYIAGY KTFDGRGAQV	99
<i>Cry j I</i> (pCCI-15)	-----	99
<i>Amb a I.1</i>	VA--KE---- F--AQN-----ERD-V-R -DKE-VVNSD --I-----K-	100
<i>Amb a II</i>	V---KE---- F---Q-----QRD-I-Y -QQE-VVTSD --I-----K-	100
<i>Cry j I</i> (pCCI-2-2)	YIGRGGPCVF IKRVSNVLIH GLHLYGCSTS VLGNVLINES FGVEPVHPQD	149
<i>Cry j I</i> (pCCI-15)	----- --Y-----	149
<i>Amb a I.1</i>	E-I-A--*FT LNG-K-----NINMHDVKVN P*-GLIKSND GPAA-RAGS-	147
<i>Amb a II</i>	ELVY--*IT LMN-K-----NIDIHDVRLV PG-RIKS-GG PAI*-R-QS-	147
<i>Cry j I</i> (pCCI-2-2)	GDALTIRTAT NIWIDHNSFS NSSDGLVDVT LSSTGVTISN NLFFNHHKVM	199
<i>Cry j I</i> (pCCI-15)	----- -T-----	199
<i>Amb a I.1</i>	---ISISGSS Q-----C-L- K-V-----AK -GT-RL-V-- S--TQ-QF-L	197
<i>Amb a II</i>	---IHTGSS D-----CTL- K-F-----N WG----- CK-TH-E-AV	197
<i>Cry j I</i> (pCCI-2-2)	LLGHDDAYS DSKMKVTVA F NQFGPNCQQR MPRARYGLVH VANNNYDPWT	249
<i>Cry j I</i> (pCCI-15)	S-----	249
<i>Amb a I.1</i>	-F-AG-ENIE -RG-LA----- -T-TD-VD-- --C-H-FFQ -V-----K-G	247
<i>Amb a II</i>	---AS-THFQ -LK-H--L-Y -I-TNIVHE- --C-F-FFQ IV--F--R-D	247
<i>Cry j I</i> (pCCI-2-2)	IYAIGGSSNP TILSEGNSFT APNESYKQV TIRIGCKTSS SCSNWWQST	299
<i>Cry j I</i> (pCCI-15)	-----	299
<i>Amb a I.1</i>	S-----AS- ----Q--R-C --D-RS--N- LG-H-*EAAA ESMK-N-RTN	296
<i>Amb a II</i>	K----- --Q--K-V --DFI--N- CL-T-AQEPE W*MT-N-RTQ	296
<i>Cry j I</i> (pCCI-2-2)	QDVFYNGAYF VSSGKYEGGN IYTKKEAFNV ENGNA TPQLT KNAGVLTCSL	349
<i>Cry j I</i> (pCCI-15)	----- --H-- Q-----	349
<i>Amb a I.1</i>	K--LE---I- -A--VDPVLT PEQSAGMIPA -P-ESALS-- SS----S-QP	346
<i>Amb a II</i>	N--LE---I- -A--SDPVLT AEQNAGMMQA -P-DMV---- M-----P	346
<i>Cry j I</i> (pCCI-2-2)	SKRC	353
<i>Cry j I</i> (pCCI-15)	----	353
<i>Amb a I.1</i>	GAP-	350
<i>Amb a II</i>	GAP-	350

Figure 3. Sequence comparison among *Cry j I* (pCCI-2-2 and pCCI-15), *Amb a I.1*, and *Amb a II*. *Amb a I.1* (a representative of four isoforms of *Amb a I* family) and *Amb a II* are the pollen allergens of short ragweed. Horizontal bars mean identical amino acids. The asterisk indicates a blank that has been inserted to maximize identity.

As for the sequence similarity mentioned above, the tree pollen allergens such as *Cor a I* (hazel), *Bet v I* (birch), *Aln g I* (alder) and *Car b I* (hornbeam) have similar sequences (18) and the grass pollen allergens such as *Poa p IX* (Kentucky bluegrass) (19) and *Lol p Ib* (rye grass) (20) are also related. It appears that allergens of pollen from trees and grasses may in future be classified into several groups.

In the present study, we have reported the complete amino acid sequence of *Cry j I*, a major allergen of Japanese cedar pollen, by using molecular cloning techniques. We also clearly demonstrated that an isoform of the *Cry j I* molecule exists. The recombinant *Cry j I* molecule may be available in future and it may contribute to the standardization of the allergens used for the diagnosis and the immunotherapy. The information of the complete amino acid sequence of *Cry j I* may facilitate experimental determination of the

T- and B-cell epitopes of the allergen by use of synthetic overlapping peptides based on its sequence.

ACKNOWLEDGMENT

The authors would like to thank Dr. H. Tsunoo of Meiji Institute of Health Science for encouraging us to pursue this study.

REFERENCES

1. Yasueda, H., Yui, Y., Shimizu, T. and Shida, T. (1983) *J. Allergy Clin. Immunol.* 71, 77-85.
2. Marsh, D. G., Goodfriend, L., King, T. P., Lowenstein, H. and Platts-Mills, T. A. E. (1986) *Bull. WHO* 64, 767-770.
3. Sakaguchi, M., Inouye, S., Taniai, M., Ando, S., Usui, M. and Matuhasi, T. (1990) *Allergy* 45, 309-312.
4. Taniai, M., Ando, S., Usui, M., Kurimoto, M., Sakaguchi, M., Inouye, S. and Matuhasi, T. (1988) *FEBS Lett.* 239, 329-332.
5. Kawashima, T., Taniai, M., Usui, M., Ando, S., Kurimoto, M. and Matuhasi, T. (1992) *Int. Arch. Allergy Immunol.* 98, 118-126.
6. Taniai, M., Kayano, T., Takakura, R., Yamamoto, S., Usui, M., Ando, S., Kurimoto, M., Panzani, R. and Matuhasi, T. (1993) *Mol. Immunol.* 30, 187-189.
7. Masaki, T., Tanabe, M., Nakamura, K. and Soejima, M. (1981) *Biochim. Biophys. Acta.* 660, 44-50.
8. Tsunasawa, S., Masaki, T., Hirose, M., Soejima, M. and Sakiyama, F. (1989) *J. Biol. Chem.* 264, 3832-3839.
9. Breiteneder, H., Hassfeld, W., Pettenburger, K., Jarolim, E., Breitenbach, M., Rumpold, H., Kraft, D. and Scheiner, O. (1988) *Int. Arch. Appl. Immunol.* 87, 19-24.
10. Gibbs, R. A., Nguyen, P.-N., McBride, L. J., Koepf, S. M. and Caskey, C. T. (1989) *Proc. Natl. Acad. Sci. U.S.A.* 86, 1919-1923.
11. Griffith, I. J., Lussier, A., Garman, R., Koury, R., Yeung, H. and Pollock, J. (1993) *J. Allergy Clin. Immunol.* 91, 339. (Abstract).
12. Rafnar, T., Griffith, I. J., Kuo, M.-C., Bond, J. F., Rogers, B. L. and Klapper, D. G. (1991) *J. Biol. Chem.* 266, 1229-1236.
13. Griffith, I. J., Pollock, J., Klapper, D. G., Rogers, B. L. and Nault, A. K. (1991) *Int. Arch. Allergy Appl. Immunol.* 96, 296-304.
14. Rogers, B. L., Morgenstern, J. P., Griffith, I. J., Yu, X.-B., Counsell, C. M., Brauer, A. W., King, T. P., Garman, R. D. and Kuo, M.-C. (1991) *J. Immunol.* 147, 2547-2552.
15. Wing, R. A., Yamaguchi, J., Larabell, S. K., Ursin, V. M. and McCormick, S. (1989) *Plant Mol. Biol.* 14, 17-28.
16. Budelier, K. A., Smith, A. G. and Gasser, C. S. (1990) *Mol. Gen. Genet.* 224, 183-192.
17. McCormick, S. (1991) *Trends Genet.* 7, 298-303.
18. Breiteneder, H., Ferreira, R., Hoffmann-Sommergruber, K., Ebner, C., Breitenbach, M., Rumpold, H., Kraft, D. and Scheiner, O. (1993) *Eur. J. Biochem.* 212, 353-362.
19. Silvanovich, A., Astwood, J., Zhang, L., Olsen, E., Kisil, F., Schon, A., Monaputra, S. and Hill, R. (1991) *J. Biol. Chem.* 266, 1204-1210.
20. Singh, M. B., Hough, T., Thecrakulpisut, P., Avlioglu, A., Davies, S., Smith, P. M., Taylor, P., Simpson, R. J., Ward, L. D., McCluskey, J., Puy, R. and Knox, R. B. (1991) *Proc. Natl. Acad. Sci. U.S.A.* 88, 1334-1338.