

Sequence polymorphism of the group 1 allergen of Bermuda grass pollen

Z.-N. CHANG*†, H.-J. PENG‡, W.-C. LEE†, T.-S. CHEN*, K.-Y. CHUA§, L.-C. TSAI‡, C.-W. CHI‡ and S.-H. HAN†¶

*Institute of Biotechnology in Medicine, †Centre for Immunology, National Yang-Ming University, ‡Department of Medical Research and Education, Veterans General Hospital-Taipei, §Department of Paediatrics, National University of Singapore, and ¶Institute of Microbiology and Immunology, National Yang-Ming University, Taipei, Taiwan, Republic of China

Summary

Background Cyn d 1, the major allergen of Bermuda grass pollen, consists of a number of isoforms.

Objective To examine the extent of sequence variation of Cyn d 1 isoforms at the molecular level.

Methods A Bermuda grass pollen λ ZAP II cDNA expression library was immunoscreened with anti-Cyn d 1 monoclonal antibodies. The reactive clones were isolated, subcloned into *Escherichia coli*, and sequenced. Some of them were expressed in the yeast *Pichia pastoris* to obtain recombinant Cyn d 1 proteins.

Results Ten cDNA clones were obtained, all these clones encode the full length of Cyn d 1 protein. Their deduced mature proteins can be grouped into: the long ones with 246 amino acids, and the short ones with 244 amino acids. The last two amino acids (AG) of the long Cyn d 1 are deleted in the short Cyn d 1. The remaining amino acid sequences share more than 98% identity; a total of nine amino acid variations were observed. Two recombinant Cyn d 1 proteins (rCyn d 3–2 and rCyn d 5–4) with three amino acid substitutions showed differential IgE-binding profiles.

Conclusion The present study extended our understanding of the primary structure of isoforms of Cyn d 1.

Keywords: Bermuda grass pollen, group 1 allergen, isoform, polymorphism, cDNA clone

Clinical and Experimental Allergy, Vol. 29, pp. 488–496. Submitted 26 January 1998; revised 18 August 1998; accepted 29 September 1998.

Introduction

Bermuda grass (*Cynodon dactylon*) pollen (BGP) is an allergen of worldwide importance, and Cyn d 1 is its major allergenic component [1–4]. The group 1 allergen of BGP has molecular weights (M_r) of 32–34 kDa and isoelectric focusing points (pIs) of 5.6–7.3 [5]. Using two-dimensional gel electrophoresis, 11 isoforms (nine acidic ones, a neutral one, and a basic one) were observed [5]. Previously, these isoforms had been characterized at the protein level [5]. The acidic Cyn d 1 has higher M_r than that of the neutral and the basic Cyn d 1. All the isoforms are allergenic since they react with human IgE in sera of

BGP-allergic patients, but the neutral and the basic Cyn d 1 have higher IgE-binding capacity than that of the acidic Cyn d 1 as demonstrated by an inhibition of IgE binding assay. With the use of monoclonal antibodies (MoAbs) [5], it was revealed that all isoforms of Cyn d 1 share a common antigenic determinant whereas the neutral and the basic Cyn d 1 have their unique ones. Data obtained from N-terminal sequencing analysis disclosed that these isoforms have similar but not identical amino acid sequences.

The primary structure of Cyn d 1 has recently been reported by Smith *et al.* [6] who isolated and characterized several complementary DNA (cDNA) clones encoding Cyn d 1. None of their clones contained the known N-terminal amino acid sequence of Cyn d 1, and the 5' ends of these proteins were subsequently cloned by the anchor polymerase chain reaction (PCR).

Correspondence: Z.-N. Chang, Associate Professor, Faculty of Medical Technology, National Yang-Ming University, Taipei, Taiwan 112, Republic of China.

In addition to elucidating the primary structure, the study on isoforms of an allergen is important because amino acid substitutions of the primary structure of a protein may change its antigenicity. The present study extends our knowledge on isoforms of Cyn d 1 at the molecular level. A total of 10 cDNA clones containing the full length of the Cyn d 1 were evaluated. A comparison was made among the nucleotide sequences as well as the deduced amino acid sequences and between those reported by Smith *et al.* [6].

Materials and methods

Immunoscreening of the cDNA library

BGP obtained from the Allergon AB (Carthage, MI, USA) and International Biologicals (Piedmont, OK, USA) were pooled for the mRNA preparation. Poly (A)⁺ mRNA was isolated as previously described, and the cDNA library was commercially obtained from Stratagene (La Jolla, CA, USA), prepared from BGP mRNA in the Lambda ZAP II expression vector via EcoRI site. The immunoscreening was performed as described by Shen *et al.* [7]. Briefly, a primary library of 1×10^6 recombinant phages was amplified to generate a higher titre of 10^9 plaque forming units. About 1×10^5 phages were used to infect the *Escherichia coli* strain XL-1 blue, plated on LB-medium plus 10 mM MgSO₄ and incubated for 4 h at 42 °C. Nitrocellulose filters presoaked with 10 mM isopropyl thiogalactoside were used to induce the synthesis of fusion proteins. The proteins were allowed to bind to the membranes by a further incubation at 37 °C for 3 h. After washing, the filters were blocked with 1% (v/v) bovine serum albumin in TRIS buffer (20 mM, pH 7.5, and 150 mM NaCl). The filters were incubated with anti-Cyn d 1 MoAbs at 4 °C overnight. The filters were then washed with Tris buffer and incubated with 1 : 5000 diluted (v/v) alkaline phosphatase conjugated goat anti-mouse IgG (Jackson Lab) at room temperature for 1 h. After washing with TRIS buffer containing 0.05% Tween 20 (v/v), the positive clones were detected by immersing the filters in the specific substrate colour development solution.

Sequencing of the cDNA inserts

Positive plaques were picked with a Pasteur pipette and purified by repeating the above procedure. In vivo excision of the pBluescript phagemid from the Lambda ZAP II vector was performed as described by the manufacturer (Stratagene). Inserts containing pBluescript SK(−) phagemids were excised from the bacteriophage lambda ZAP II and transformed to *E. coli* XL 1-Blue MRF' cells. SORL cells were used to plate the rescued phagemid. DNA for sequencing was purified by Wizard Minipreps DNA purification System (Promega Corp., Madison, WI, USA).

The inserts were sequenced by dideoxy chain termination method using the walking primer strategy [8]. The commercially available T3 and T7 primers (Stratagene) were used.

Amino acid sequence analysis and pI prediction

The sequence was deduced by using of the PC/GENE sequence analysis program (IntelliGenetics, Sweden).

Expression of Cyn d 1 in the yeast *Pichia pastoris* system

The test was performed with an EasySelectTM *Pichia* expression kit (Invitrogen Corp., San Diego, CA, USA). The coding region of two long Cyn d 1 cDNA clones (3–2 and 5–4) was amplified by PCR with EcoRI site at the 5' end and XbaI site at the 3' end. Oligonucleotide primers used were 5'TTAGAATTGCGCCATCGGCGACAAGCCA3' and 5'TTGTCTAGATCAGGCTCCGAAGTGGAG3'. The PCR was performed with an initial denaturation at 94 °C for 4 min followed by a successive 25 thermal cycles of denaturation at 94 °C for 30 s, annealing at 50 °C for 70 s and extension at 72 °C for 30 s, and finally elongated at 72 °C for 4 min. The PCR products were purified, and were directly inserted into the multiple cloning region of pPICZα *Pichia* expression vector. The resulting Cyn d 1-pPICZα constructs were arranged in a linear alignment with Pme I, and were used to transform into *P. pastoris* strain KM71. The transformants were then grown for 2 days at 30 °C in buffered minimal glycerol-complex medium. The expression of the recombinant Cyn d 1 proteins was induced with methanol (final volume 0.5%). Three days after methanol induction, the cells were pelleted, and the supernatants were collected and analysed with SDS-PAGE and Western blotting as previously described [5]. An aliquot of 10 µl (20 µg) of the purified protein was loaded onto wells of 11% SDS-PAGE. Western blots were probed with mouse anti-Cyn d 1 MoAbs and sera from patients allergic to Bermuda grass pollen, followed by horseradish peroxidase labelled antimouse Ig, or antihuman IgE [5].

Results

Ten cDNA clones were obtained. The coding region of the clone 3–2 with 1269 base pairs was completely sequenced (Fig. 1). After the deduced amino acids were compared with the known N-terminal amino acid sequence (AIGDK) [5], the first amino acid of the mature protein was identified to be the 123rd deduced residue. The methionine encoded by nucleotides 289–291 appears to be the putative initiation site of the Cyn d 1 protein since the nucleotides around it (AGAAATGGC) are 78% identical to the plant's initiation

of translation consensus sequences (AACAAATGGC) [9]. For the other nine clones, the nucleotides encoding Cyn d 1, and the signal peptides were sequenced.

All the 10 clones contain the full length of Cyn d 1. The nucleotide sequences encoding these 10 full length proteins are shown in Fig. 2. They can be separated into two groups. Six clones (3-2, 5-4, 8-13, 8-8, 33, and 6-6) contain a long Cyn d 1 with 738 nucleotides whereas the other four (clones 2, 4-8, 9-4, and 9-9) contain a short Cyn d 1 with 732 nucleotides. These two groups of cDNA clones differ

mainly at the 3' end with six nucleotides (GGAGCC) missing in the short Cyn d 1. In addition, their nucleotide sequences were similar to each other, only a total of 27 variations were identified.

Figure 3 shows the deduced amino acid sequences of these clones. For comparison, three partial Cyn d 1 proteins reported by Smith *et al.* (CD1, 23B, and 3B) [6] are also shown. Besides the two additional residues (GA), only nine variations were identified among all the 10 deduced proteins. They were at positions 16 (G or N), 54 (V or G), 89 (V

```

1/1          31/11
GTC GAG GGT CCA CCG CGG TGG CGG CGC TCT AGA ATA GTG GAT CCC CCG GGC TGC AGG AAT
V E G P P R W R R S R I V D P P G C R N
61/21        91/31
TCG GCA CGA GCA TTG CTC CAT CTC GAC CGC GCT GTC GTT TCC TCC GAC TGA TAA TTC GTT
S A R A L L H L D R A V V S S D * * F V
121/41       151/51
CCG TTA ACT AGG TGT GTC TGT GTA TCA TTC TCT TGT GTT ATT AAA GTG TTG ATG ATT CAA
P L T R C V C V S F S C V I K V L M I Q
181/61       211/71
ATG ACG ACT TCC AAT CGA CAA TTG AGT AGT AGT ACC ACT TGT GTT TGT GTT TCC TCC CTT
M T T S N R Q L S S S T T C V C V S S L
241/81       271/91
GAT TCA GAA GAA ACC TCG TGC CGA ATT CGG CAC GAG CAG AAG CAA GAA ATG GCG CAG ACG
D S E E T S C R I R H E Q K Q E M A Q T
301/101      331/111
ACG ATG AAT CAG AAA CTG GCG CTG GTT GCG GTG CTT GTG GCT ACG ATG GTG GCC GGC GGG
T M N Q K L A L V A V L V A T M V A G G
361/121      391/131
TCG TGC GCC ATC GGC GAC AAG CCA GGG CCC AAC ATC ACG GCG ACC TAC GGC AGC AAG TGG
S C A I G D K P G P N I T A T Y G S K W
421/141      451/151
CTG GAG GCC AAG GCC ACC TTC TAC GGC AGC AAC CCG CGC GGT GCC GCC CCC GAT GAC CAC
L E A K A T F Y G S N P R G A A P D D H
481/161      511/171
GGC GGC GCT TGC GGG TAC AAG GAC GTC GAC AAG CCT CCC TTC GAC GTC ATG ACC GGC TGC
G G A C G Y K D V D K P P F D V M T A C
541/181      571/191
GGC AAC GAG CCC ATC TTC AAG GAC GGC CTC GGC TGC GGC GCA TGC TAC GAG ATC AAG TGC
G N E F I F K D G L G C G A C Y E I K C
601/201      631/211
AAG GAA CCC GTC GAG TGC TCC GGC GAG CCC GTC CTC GTC AAG ATC ACC GAC AAG AAC TAC
K E P V E C S G E P V L V K I T D K N Y
661/221      691/231
GAG CAC ATC GCC GGC TAC CAC TTC GAC CTC TCC GGC AAG GCC TTC GGC GCC ATG GCC AAG
E H I A A Y H F D L S G K A F G A M A K
721/241      751/251
AAG GGC CAG GAA GAC AAG CTG CGC AAG GCC GGT GAG CTG ACT CTG CAG TTC CGC CGC GTC
K G Q E D K L R K A G E L T L Q F R R V
781/261      811/271
AAG TGC AAG TAC CCC TCC GGC ACC AAG ATC ACC TTC CAC ATC GAG AAG GGA TCC AAC GAC
K C K Y F S G T K I T F H I E K G S N D
841/281      871/291
CAT TAC CTG GCG CTC CTC GTC AAG TAC GCC GCC GGC GAT GGC AAC ATT GTC GCC GTC GAC
H Y L A L L V K Y A A G D G N I V A V D
901/301      931/311
ATC AAG CCC AAG GAC TCC GAC GAG TTC ATT CCC ATG AAG TCG TCC TGG GGC GCC ATC TGG
I K P K D S D E F I P M K S S W G A I W
961/321      991/331
AGG ATC GAC CCC AAG AAG CCG CTC AAG GGC CCC TTC TCC ATC CGC CTC ACC TCC GAG GGC
R I D P K K P L K G P F S I R L T S E G
1021/341     1051/351
GGC GCC CAT CTC GTC CAG GAC GAC GTC ATC CCA GCC AAC TGG AAG CCA GAC ACC GTC TAC
G A H L V Q D D V I P A N W K P D T V Y
1081/361     1111/371
ACC TCC AAG CTC CAG TTC GGA GCC TGA GAG ACG ATG ATC CTC CAT GCA TAT CCT CGC CGA
T S K L Q F G A * E T M I L H A Y P R R
1141/381     1171/391
TTG CAA GGG CTC ATA TAT GAC ATG TGC GTG TAC GCA TCT GTC GAA TAA GCA TCC ATA TAT
L Q G L I Y D M C V Y A S V E * A S I Y
1201/401     1231/411
GCA TGA GTT TAA TAT TTC TTT TTA TTT CCC CCC TTC AAT TAT ATG TAC ATC TCA ATG TGG
A * V * Y F F L F P P F N Y M Y I S M W
1261/421
AGA GTT ATG
R V M

```

Fig. 1. The nucleotide sequence and deduced amino acid sequence of cDNA clone 3-2. The amino acids identical with the N-terminal amino acid sequence previously reported [5] were underlined starting from the first residue.

```

3-2  GCCATCGGCGACAAGCCAGGGCCCAACATCACGGCGACCTACGGCAGCAAGTGGCTGGAGGCCAAGGCCACCTTCTACGGCAGCAACCCGCGCGGTGCCG 100
5-4  -----
8-13 -----
8-8 -----
33 -----
6-6 -----
2 -----
4-8 -----
9-4 -----
9-9 -----

3-2  CCCCCGATGACCACGGCGGCGCTTGGCGGTACAAGGACGTCGACAAGCCTCCCTTCGACGTCATGACCGCCTGCGGGCAACGAGCCCATCTTCAAGGACGG 200
5-4  -----
8-13 -----
8-8 -----
33 -----
6-6 -----
2 -----
4-8 -----
9-4 -----
9-9 -----

3-2  CCTCGGCTGCGGCGCATGCTACGAGATCAAGTGCAAGGAACCCGTCGAGTGCTCCGGCGA GCCCGTCTCGTCAAGATCACCGACAAGAACTACGAGCAC 300
5-4  -----
8-13 -----
8-8 -----
33 -----
6-6 -----
2 -----
4-8 -----
9-4 -----
9-9 -----

3-2  ATCGCCGCCTACCACTTCGACCTCTCCGGCAAGGCCTTCGGCGCCATGGCCAAGAAGGGCCAGGAAGACAAGCTGCGCAAGGCCGGTGAGCTGACTCTGC 400
5-4  -----
8-13 -----
8-8 -----
33 -----
6-6 -----
2 -----
4-8 -----
9-4 -----
9-9 -----

3-2  AGTTCCGCGCGGTCAAGTGCAAGTACCCCTCCGGCACCAAGATCACCTTCCACATCGAGAAGGGATCCAACGACCATTACCTGGCGCTGCTCGTCAAGTA 500
5-4  -----
8-13 -----
8-8 -----
33 -----
6-6 -----
2 -----
4-8 -----
9-4 -----
9-9 -----

3-2  CGCCGCGCGCGATGGCAACATTGTGCGCGTCGACATCAAGCCCAAGGACTCCGACGAGTTTCATTCATGAAGTCGTCTCTGGGGCGCCATCTGGAGGATC 600
5-4  -----
8-13 -----
8-8 -----
33 -----
6-6 -----
2 -----
4-8 -----
9-4 -----
9-9 -----

3-2  GACCCCAAGAAGCCGCTCAAGGGCCCTTCTCCATCCGCTCACCTCCGAGGGCGGCGCCATCTCGTCCAGGACGACGTATCCAGCCAACTGGAAGC 700
5-4  -----
8-13 -----
8-8 -----
33 -----
6-6 -----
2 -----
4-8 -----
9-4 -----
9-9 -----

3-2  CAGACACCGTCTACACCTCCAAGCTCCAGTTCGGAGCCTGA 741
5-4  -----
8-13 -----
8-8 -----
33 -----
6-6 -----
2 -----
4-8 -----
9-4 -----
9-9 -----

```

Fig. 2. The nucleotide sequences of 10 Cyn d 1 cDNA clones. Horizontal bars (—) represent the identical nucleotides.

	10	20	30	40	50	60	70	80
3-2	AIGDKPGPNITATYGSKWLEAKATFYGSNPRGAAPDDHGGACGYKDVKPPFDVMTACGNEPIFKDGLGCGACYRIKCKE							
5-4	-----G-----							
8-13	-----G-----							
8-8	-----G-----							
33	-----G-----							
6-6	-----G-----							
CD1	-----R-----D-----RN-----E-----							
2	-----G-----							
4-8	-----G-----							
9-4	-----N-----G-----							
9-9	-----N-----G-----							
23B	-----G-----							
3B	-----G-----							
	80	100	110	120	130	140	150	160
3-2	PVECSGEPVLVKITDKNYEHIAAYHFDLSGKAFGAMAKKGQEDKLRKAGELTLQFRRVKCKYPSGTKITFHIEKGSNDHY							
5-4	-----K-----							
8-13	-----							
8-8	-----							
33	-----							
6-6	-----P-----							
CD1	-----							
2	-----							
4-8	-----I-----							
9-4	-----							
9-9	-----							
23B	-----							
3B	-----E-----M-----E---D---A---V---PN-							
	170	180	190	200	210	220	230	240
3-2	LALLVKYAAGDGNIVAVDIKPKDSDEFIPMKSSWGAIWRIDPKKPLKGPFSIRLTSEGGAHLVQDDVIPANWKPDVTYTS							
5-4	-----R-----							
8-13	-----A-----							
8-8	-----R-----							
33	-----							
6-6	-----A-----							
CD1	-----							
2	-----K-----							
4-8	-----K-----							
9-4	-----A-----							
9-9	-----A-----							
23B	-----							
3B	-----S---S-G---D-L---Q-----P-----T-----S-G-VE-E---ED-----K-							
	246							
3-2	KLQFGA*							
5-4	-----*							
8-13	-----*							
8-8	-----*							
33	-----*							
6-6	-----*							
CD1	-----*							
2	-----*							
4-8	-----*							
9-4	-----*							
9-9	-----*							
23B	-----*							
3B	-I--*							

Fig. 3. The deduced amino acid sequences of Cyn d 1 cDNA clones. In addition to our 10 cDNA clones, three cDNA clones (CD1, 23B, and 3B) reported by Smith *et al.* [6] are also included. Horizontal bars (–) represent the identical amino acids.

```

3-2  MAQT'TMNQKLALVAVLVATMVAGGSC
5-4  FKKKKKKKT-----
8-13  ----*****-----A-----
8-8   -----
33    -----
6-6   -V-----
2     -----
4-8   GTR-----*---A-----
9-4   ----*****-----A-----
9-9   vector-----A-----
14c1  -----A---RVV

```

Fig. 4. The signal peptides of 10 Cyn d 1s. 14 C1 is the sequence determined by Smith *et al.* [6]. Asterisks (*) indicate deleted residues.

or I), 119 (K or P), 130 (E or K), 180 (K or R), 185 (D or K), 202 (P or R), and 224 (Q or A).

The signal peptide of clone 3-2 consisted of 26 amino acids (Fig. 1). Figure 4 shows the deduced amino acids of the signal peptides of 10 cDNA clones and that of clone 14C1 reported by Smith *et al.* [6]. For clones 5-4, 4-8 and 9-9, part of their 5' signal peptides were different from that of the others. They seem to represent artificial sequences that presumably ligated to these clones during cDNA synthesis and library construction. The remaining seven of our 10 clones contained similar signal peptide. The same

leader sequences were shared by three long Cyn d 1 clones 3-2, 8-8, 33, and a short Cyn d 1 clone 2. In addition, clones 8-13 and 9-4 had deletion of four amino acids (-22th to -19th residues), and clone 4-8 was lacking of the -12th residue.

Recombinant Cyn d 1 proteins

The presence of recombinant Cyn d 1 proteins (encoded by clones 3-2 and 5-4) in the culture medium of the yeast expression vector after induction with methanol is illustrated by SDS-PAGE (Fig. 5A) and immunoblotting (Fig. 5B-D). As shown, the nonrecombinant construct secreted no protein (Lane 4, Fig. 5A), while recombinant constructs 3-2 and 5-4 expressed secretory proteins (designated rCyn d 3-2 and rCyn d 5-4) with M_r of 27 kDa, 34 kDa, and 41-49 kDa (Lanes 2-3, Fig. 5A). A purified native Cyn d 1 was used as positive control (Lane 1), while two anti-Cyn d 1 MoAbs were used in immunoblotting (Fig. 5B,C). MoAb 4-37 reacts with a sequential epitope on all isoforms of Cyn d 1 and MoAb 1-61 reacts with an conformational epitope on the basic and the neutral Cyn d 1 [5]. As shown, MoAb 1-61 reacted with all the three components of recombinant Cyn d 1 proteins (Fig. 5B, Lanes 2-3). On the other hand, MoAb 4-37 reacted with the 41-49 kDa and the 34 kDa components of both recombinant proteins (Fig. 5C, Lanes 2-3), but not with the 27 kDa component which appears to be a degradation product. These two recombinant Cyn d 1 proteins (rCyn d 3-2 and rCyn d 5-4) share a similar reactive profile while reacted with the same MoAbs. However, their reaction profiles with IgE in the pooled serum of BGP-allergic patients were different (Fig. 5D, Lanes 2-3); the degraded

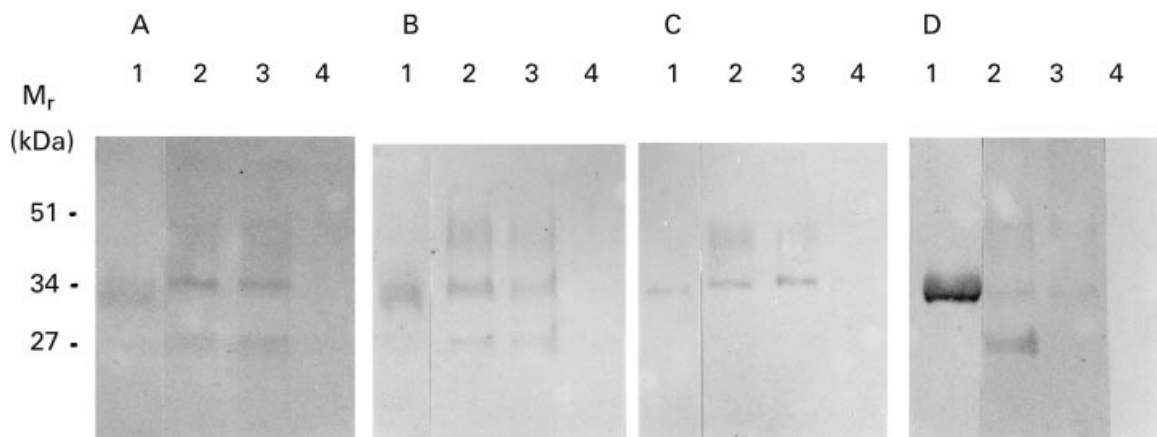


Fig. 5. SDS-PAGE patterns and immunoblotting profiles of recombinant Cyn d 1 proteins. Purified natural Cyn d 1 (lane 1), recombinant proteins rCyn d 5-4 (lane 2) and rCyn d 3-2 (lane 3) secreted into media and secretory proteins from nonrecombinant yeast control (lane 4) were separated by SDS-PAGE (A), transferred onto nitrocellulose membranes for immunoblotting with MoAbs 1-61 (B), 4-37 (C), or IgE in the pooled serum ($n=9$) of BGP-allergic patients (D).

27 kDa component of rCyn d 5–4 was with human IgE-binding capacity while that of rCyn d 3–2 was not. Neither of the recombinant Cyn d 1 proteins reacted with non-specific MoAb or human IgEs of pooled normal serum (data not shown).

Discussion

The present report was designed to evaluate the extent of sequence polymorphism of Cyn d 1. The cDNA library was constructed from BGP obtained via two different resources, and the cDNA clones were screened by immunoassay with a mixture of MoAbs 4–37 and 18–53. MoAb 4–37 is against an epitope common to all isoforms of Cyn d 1 [5], and MoAb 18–53 has recently been found to be against a sequential epitope on the 46–64th residues of Cyn d 1 [10]. All the 10 cDNA clones encode the full length of Cyn d 1 (Fig. 2).

The last six nucleotides (GGAGCC) were deleted in the short Cyn d 1 (Fig. 2), hence it is lack of the last two amino acids A and G (Fig. 3). This is in agreement with report of Smith *et al.* [6]. The signal peptides of the Cyn d 1 protein contain 26 amino acids, and that some clones had deletions of the –22nd to –19th residues (Fig. 4). This had also been observed by Smith *et al.* [6]. Our data revealed that these two kinds of signal peptides (the 26 amino acid and the 22 amino acid) were found in both the short and the long Cyn d 1 clones.

However, the last three amino acids of the signal peptides of all our clones are GSC which agree with the ‘–3, –1’ rule of signal peptide [11] while those identified by Smith *et al.* are RVV which deviates with this rule [6]. The ‘–3, –1’ rule states that residues at these positions are small and neutral amino acids, and those aromatic, charged, and large polar residues are absent from these positions.

The signal peptides of the long Cyn d 1 and the short Cyn d 1 are with 96% to 100% identity (Fig. 4), and the N-terminal 244 amino acids of the long Cyn d 1 share 98.7% to 99.2% identity with the short Cyn d 1 (Fig. 3). These two forms of Cyn d 1 may derive from different alleles of the Cyn d 1 gene. Isoforms of a number of pollen allergens had been studied by cDNA cloning [12–18], they differ from each other from few amino acids to more than one fourth of the complete amino acid sequence. Although the isoforms of some allergens have the same number of amino acids, four isoforms of Pha a 5 were found to consist of 294–320 residues [17]. In addition, the study of the N-terminal 1–20 amino acids of Phl p 6 revealed that one isoform contained two additional amino acids (EG) [19]. It seems that some truncated proteins exist in the isoforms of grass allergens.

The group 1 allergen of many grass pollens has been sequenced. However, it is difficult to assess the extent of the polymorphism since the data were obtained from one or two

cDNA clones [13,16]. Using 10 cDNA clones, the present study identified 9 amino acid variations out of the 246 amino acids of Cyn d 1. In addition, the amino acid sequences of our deduced long Cyn d 1 proteins and that of a partial long Cyn d 1 CD1 determined by Smith *et al.* [6] are with 97% identity, while the deduced partial protein of short Cyn d 1 23B determined by Smith *et al.* [6] is 100% identical to the 46th to 244th amino acids of our clone 2. These data suggest that Cyn d 1 is less polymorphic. Both the group 1 allergen and the group 5 allergen are important major allergens of grasses. The polymorphism study of the group 5 allergen of canary grass pollen showed that four isoforms of Pha a 5 were with 19% to 60% amino acid variations [17].

However, Smith *et al.* had also identified another partial short Cyn d 1 that was deduced from their clone 2B [6]. This partial Cyn d 1 showed only 81.4% identity with the 110th–244th amino acids of our short Cyn d 1; these two clones differ by a total of 25 amino acid residues. Since clone 2B is truncated, it remains unclear whether it encodes Cyn d 1 protein or represents a protein with lower M_r but with similar C-terminal amino acid sequence as that of Cyn d 1. It has been reported that the N-terminal and C-terminal of Phl p 6, a specific allergen of timothy grass pollen, show a 54% identity to an internal part of the group 5 allergens [20]. On the other hand, clone 2B may represent another type of Cyn d 1. The cDNA library used in the present study was constructed from mRNAs of BGPs from two different companies, Allergon AB and ABI. BGP used by Smith *et al.* was obtained from the Greer Laboratories [6]. It is highly possible that our cDNA library did not contain the mRNA sequence for clone 2B. Our previous study revealed that the content of the Cyn d 1's basic isoforms varies in different lots of BGP [5]. The fact that both rCyn d 3–2 and rCyn d 5–4 reacted with MoAb 1–61 (Fig. 5) which is specific against the neutral and the basic Cyn d 1 discloses that the mRNA of these isoforms are present in our cDNA library.

It has been established that the neutral Cyn d 1 and basic Cyn d 1 are more allergenic and antigenic than the acidic Cyn d 1 [5]. The amino acids A and G deleted by the short Cyn d 1 protein are both nonpolar and noncharged, the shortage of them will not influence the pIs of the mature proteins. The pIs of three deduced proteins (2, 4–8, and 5–4) were predicted to be more basic than that of the others, and clone 5–4 was the most basic one since at residues 130 and 202, the amino acids varied from E or P to K.

The recombinant proteins derived from clones 3–2 and 5–4 showed differential IgE-binding profiles (Fig. 5). The fact that the degraded 27 kDa component of rCyn d 5–4 reacted with human IgE while that of rCyn d 3–2 did not suggests that amino acid variations between these two proteins had affected their IgE-binding capacity. rCyn d

3–2 appears to be with less IgE-binding capacity than that of rCyn d 5–4; this was confirmed by a competitive test with a dot-blot immunoassay (data not shown). Most B cell epitopes are conformation-dependent, analysis of the substitution of residues that will cause change of the conformation and lead to the degraded recombinant proteins to be with different IgE-binding capacity is ongoing.

Smith *et al.* had also obtained a recombinant Cyn d 1 protein CD1 from the yeast expression system [6]. The immunoblot profile of their recombinant protein with an anti-Cyn d 1 MoAb is similar to that of our recombinant proteins with MoAb 4–37; two components with MW of 39.8 and 30.2 kDa were found. Interestingly, their data showed that the degraded component of their rCyn d 1 failed to bind human IgE.

The knowledge of the extent of polymorphism of an allergen is not only with academic interest but also with clinical importance. In addition to the IgE binding capacity, different cellular responses had been found to be caused by the amino acid substitution of the isoforms of an allergen. The substitution from tyrosine to histidine at position 50 of Der p 1 influence its reaction with T lymphocytes in about 30% mite allergic patients [21]. In addition, some isoforms of Bet v 1 are with low IgE-binding capacity but high T-cell stimulating ability [22,23]. Therefore, the inclusion of the particular isoform in immunotherapy is essential [24,25]. The various isoform of recombinant Cyn d 1 proteins can serve as useful reagent to study the effect of sequence polymorphism on the T cell epitopes of Cyn d 1.

Acknowledgements

This research was supported in part by a grant from the National Science Council, Republic of China (NSC 86–2314-B010–046).

References

- 1 Ford SA, Baldo BA. Identification of Bermuda grass (*Cynodon dactylon*) pollen allergens by electroblotting. *J Allergy Clin Immunol* 1987; 79:711–20.
- 2 Matthiesen F, Schumacher MJ, Lowenstein H. Characterization of the major allergen of *Cynodon dactylon* pollen, Cyn d 1. *J Allergy Clin Immunol* 1991; 88:763–74.
- 3 Han SH, Chang ZN, Chi CW *et al.* Use of monoclonal antibodies to isolate and characterize Cyn d I, the major allergen of Bermuda grass pollen. *J Allergy Clin Immunol* 1993; 92:549–58.
- 4 Smith PM, Avjioglu A, Ward LR, Simpson RJ, Knox RB, Singh MB. Isolation and characterization of group-I isoallergens from Bermuda grass pollen. *Int Arch Allergy Appl Immunol* 1994; 104:57–64.
- 5 Chang ZN, Liu CC, Tam MF, Peng HJ, Tsai JJ, Han SH. Characterization of the isoforms of the group I allergen of *Cynodon dactylon*. *J Allergy Clin Immunol* 1995; 95:1206–14.
- 6 Smith PM, Suphioglu C, Griffith IJ, Theriault K, Knox B, Singh MB. Cloning and expression in yeast *Pichia pastoris* of a biologically active form of Cyn d 1, the major allergen of Bermuda grass pollen. *J Allergy Clin Immunol* 1996; 98:331–43.
- 7 Shen HD, Choo KB, Lee HH *et al.* The 40-kilodalton allergen of *Candida albicans* is an alcohol dehydrogenase: molecular cloning and immunological analysis using monoclonal antibodies. *Clin Exp Allergy* 1991; 21:675–81.
- 8 Au LC, Lin SB, Chou JS, Teh GW, Chang KJ, Shih CM. Molecular cloning and sequence analysis of the cDNA for ancrod, a thrombin-like enzyme from the venom of *Calloselasma rhodostoma*. *Biochem J* 1993; 294:387–90.
- 9 Lutcke HA, Chow KC, Mickel FS, Moss KHF, Scheele GA. Selection of AUG initiation codon differs in plants and animals. *EMBO J* 1987; 6:43–8.
- 10 Chang ZN, Tam MF, Liu CC, Chi CW, Peng HJ, Han SH. Using monoclonal antibodies to characterize a sequential epitope on the group I allergen of Bermuda grass pollen. *Int Arch Allergy Appl Immunol* 1997; 114:258–64.
- 11 von Heijne G. How signal sequences maintain cleavage specificity. *Mol Biol* 1984; 173:243–51.
- 12 Perez M, Ishioka GY, Walker LE, Chesnut RW. cDNA cloning and immunological characterization of the rye grass allergen Lol, p. I *J Bio Chem* 1990; 265:16210–15.
- 13 Rafnar T, Griffith IJ, Kuo M, Bond JF, Rogers BL, Klapper DG. Cloning of Amb a I (Antigen E), the major allergen family of short ragweed pollen. *J Biol Chem* 1991; 266:1229–36.
- 14 Silvanovich A, Astwood J, Zhang L *et al.* Nucleotide sequence analysis of three cDNAs coding for Poa, p. IX isoallergens of Kentucky bluegrass pollen. *J Bio Chem* 1991; 266:1204–10.
- 15 Griffith IJ, Pollock J, Klapper DG, Rogers BL, Nault AK. Sequence polymorphism of Amb a I and Amb a II, the major allergens in *Ambrosia artemisiifolia* (short ragweed). *Int Arch Allergy Appl Immunol* 1991; 96:296–304.
- 16 Groffith IJ, Smith PM, Pollock J *et al.* Cloning and sequencing of Lol, p. I, the major allergenic protein of rye-grass pollen. *FEBS Lett* 1991; 279:210–5.
- 17 Suphioglu C, Smith MB. Cloning, sequencing and expression in *Escherichia coli* of Pha a 1 and four isoforms of Pha a 5, the major allergens of canary grass pollen. *Clin Exp Allergy* 1995; 25:853–65.
- 18 Duro G, Colombo P, Costa MA *et al.* Isolation and characterization of two cDNA clones coding for isoforms of the *Parietaria judaica* major allergen Par j 1. 0101 *Int Arch Allergy Immunol* 1997; 112:348–55.
- 19 Petersen A, Bufe A, Schlaak M, Recker WM. Characterization of the allergen group VI, in timothy grass pollen (Phl p. 6). I. Immunological and biochemical studies. *Int Arch Allergy Immunol* 1995; 108:49–54.
- 20 Petersen A, Bufe A, Schlaak M, Recker WM. Characterization of the allergen group VI, in timothy grass pollen (Phl p 6). II, cDNA cloning of Phl p 6 and structural comparison to grass group V. *Int Arch Allergy Immunol* 1995; 108:55–9.
- 21 Thomas WR, Smith W, Hales BJ, Carter MD. Functional

- effects of polymorphisms of house dust mite allergens. *Int Arch Allergy Immunol* 1997; 113:96–8.
- 22 Larsen JN. Isoallergens: significance in allergen exposure and response. *ACI News* 1995; 7:141–6.
 - 23 Ferreira F, Hirthenlehner K, Jilek A et al. Dissection of immunoglobulin E and T lymphocyte reactivity of isoforms of the major birch pollen allergen Bet v 1: Potential use of hypoallergenic isoforms for immunotherapy. *J Exp Med* 1996; 183:599–609.
 - 24 Ebner C, Ferreira F. Hypoallergenic isoform variants of allergens – a new concept for specific immunotherapy. *ACI Int* 1997; 9:69–71.
 - 25 Thomas WR, Smith WA, Hales BJ. Allergens: sequence diversity and T-cell recognition. *ACI Int* 1998; 10:16–22.