ELSEVIER

Biochimica et Biophysica Acta 1518 (2001) 178–182

BIOCHIMICA ET BIOPHYSICA ACTA
BBA

www.bba-direct.com

Short sequence-paper

Cloning of cDNA encoding a soybean allergen, Gly m Bd 28K[1]

Hideaki Tsuji [a,*], Miki Hiemori [a], Masumi Kimoto [a], Hiromi Yamashita [a], Ryoko Kobatake [a], Motoyasu Adachi [b], Takako Fukuda [b], Noriko Bando [c], Misako Okita [a], Shigeru Utsumi [b]

[a] *Department of Nutritional Science, Faculty of Health and Welfare Science, Okayama Prefectural University, Kuboki 111, Soja 719-1197, Japan*
[b] *Research Institute for Food Science, Kyoto University, Uji, Kyoto 611-0011, Japan*
[c] *Department of Nutrition, School of Medicine, The University of Tokushima, Kuramoto-cho, Tokushima 770-8503, Japan*

Received 4 September 2000; received in revised form 5 December 2000; accepted 12 December 2000

Abstract

A cDNA clone encoding a soybean allergen, Gly m Bd 28K, has been isolated. The clone has a 1567-bp cDNA insert with a 1419-bp open reading frame and a 148-bp 3′-untranslated region, followed by a polyadenylation tail. The open reading frame was shown to encode a polypeptide composed of 473 amino acids. The chemically determined amino acid sequences of the peptides obtained from the allergen, including its N-terminal peptide, were shown to be contained in the N-terminal region of the amino acid sequence deduced from the cDNA, showing that the first half of the cDNA encodes the allergen with a preceding segment of 21 amino acids. The peptide fragment including the allergen was expressed as a fusion protein with glutathione *S*-transferase in *Escherichia coli* and immunoblotted with the sera of soybean-sensitive patients and the monoclonal antibody against the allergen. Furthermore, homology analyses demonstrate that the polypeptide for the cDNA exhibits high homology with the MP27/MP32 proteins in pumpkin seeds and the carrot globulin-like protein. This finding suggests that the polypeptide may consist of a 21-amino acid segment as a part of the signal peptide and the proprotein, which may be converted to two mature proteins, Gly m Bd 28K and a 23-kDa protein, during the development of soybean cotyledons. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Soybean allergen; Gly m Bd 28K; cDNA; Pumpkin MP27/MP32; Carrot globulin-like protein

Soybean is one of the representative allergenic foods in Japan. In a search of soybean allergens, Ogawa et al. found 16 allergens in soybean seeds [1]. Among them, the 68-, 30- and 28-kDa proteins were shown to be major allergens in soybean. The former two allergens have been well characterized [2,3]. Gly m Bd 28K was also purified and some properties of the protein were investigated [4]. However, little information as to its primary structure has been elucidated. For further characterization of the allergenic properties of Gly m Bd 28K, it is necessary to obtain information concerning its complete primary structure. This led us to obtain cDNA clones encoding the allergen. In the present study, we report the cloning and expression of cDNA encoding Gly m Bd 28K. Also, homology analyses of the amino acid sequence deduced from the cDNA are described.

In order to obtain information as to the primary structure of Gly m Bd 28K, the allergen was digested with lysyl endopeptidase, chymotrypsin and BrCN. Five peptides were purified by high performance liquid chromatography on a C_4 column with a linear gradient elution of 0–64% acetonitrile in 0.05% trifluoroacetic acid. The N-terminal amino acid sequences of the allergen and its peptides purified were determined by automated Edman degradation with a Hewlett Packard G1005A automatic sequencer. Total RNA was prepared from developing soybean cotyledons (*Glycine max* L. c.v. Shishioh) and poly(A)$^+$ RNA was purified from the total RNA using a Mini-oligo(dT)-cellulose spin column kit (5′ Prime→3′ Prime Inc., Boulder, CO, USA). The first strand of cDNA encoding the allergen was synthesized from poly(A)$^+$ RNA in the same manner as described in the previous paper [5], except that 5′-ATTGCATTGACAAGCT$_{15}$ was used as a primer and that Moloney murine leukemia virus reverse transcriptase was used instead of avian myeloblastosis virus tran-

Abbreviations: IPTG, isopropyl-1-thio-β-D-galactopyranoside; PCR, polymerase chain reaction

* Corresponding author. Fax: +81-866-94-2143;
E-mail: htsuji@fhw.oka-pu.ac.jp

[1] The nucleotide sequence data reported in this paper have been submitted to the DDBJ Data Bank under the accession number AB046874.

0167-4781/01/$ – see front matter © 2001 Elsevier Science B.V. All rights reserved.
PII: S0167-4781(00)00310-9

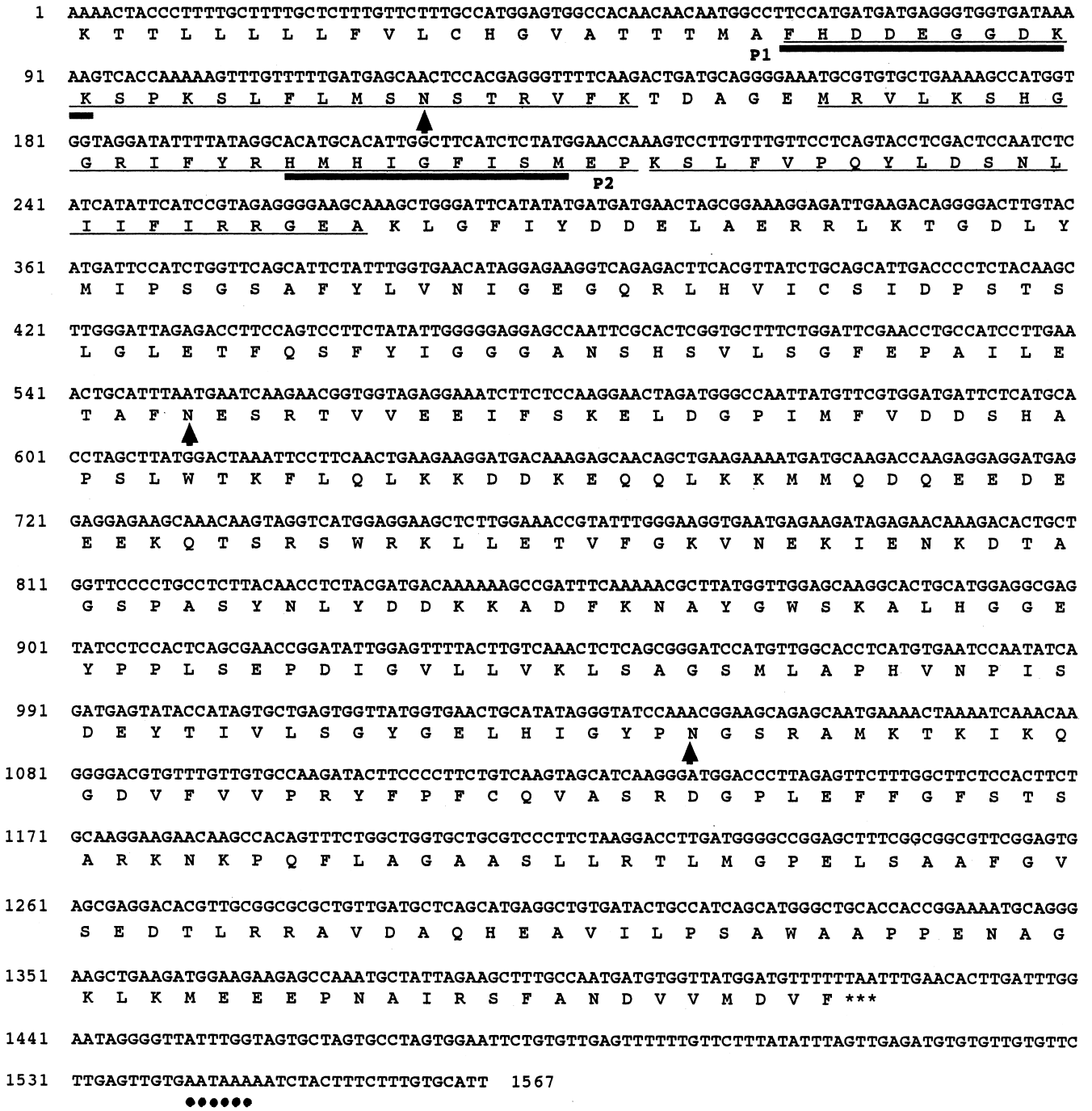

Fig. 1. Nucleotide sequence of cDNA encoding Gly m Bd 28K and its deduced amino acid sequence. The deduced amino acid sequence is shown under the nucleotide sequence of the cDNA. Normal underlines under the amino acid sequence represent the amino acid residues determined by automated Edman degradation. Two bold underlines (P1 and P2) express primers 1 and 2 used in PCR, respectively. Asterisks show the stop codon of the cDNA. Three arrows express the putative asparagine residues to be glycosylated. A signal for polyadenylation of the transcript is shown by a dotted underline.

scriptase. On the basis of the determined amino acid sequences, polymerase chain reaction (PCR) was carried out to obtain a probe suitable for the cloning of cDNAs encoding the allergen. Primers 1 (5′-TT(CT)CA(CT)GA-(CT)GA(CT)GA(AG)GG(AGCT)GG(AGCT)GA(CT)A-AAAA-3′) and 2 (5′-CAT(AGCT)GA(AGT)AT(AG)AA-(AGCT)CC(AGT)AT(AG)TGCAT(AG)TG-3′) were used as sense sequence for FHDDEGGDKK and antisense sequence for HMHIGFISM, respectively, in the PCR. The reaction was run as follows: cycles 1–42, 1 min at 94°C, 1 min at 63°C and 3 min at 72°C; the final cycle, 10 min at 72°C. A PCR product was cloned into a pUC19 plasmid and sequenced by the dideoxy chain terminating method [6]. Since the PCR product was confirmed to encode the above peptide fragments of the allergen, the product was used as a probe in the following experiments. cDNA

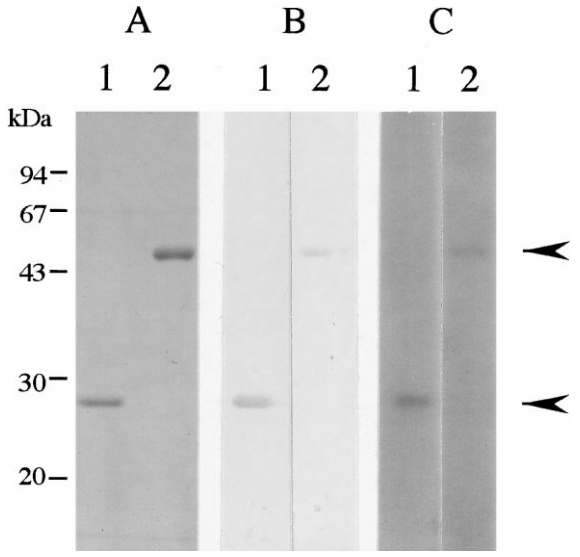

Fig. 2. Immunoblotting patterns of the fusion protein of the protein fragment corresponding to the allergen with glutathione *S*-transferase. *E. coli* JM109 harboring the expression vector for the peptide fragment including the allergen was cultured for 7 h in the presence of 1 mM IPTG. The fusion protein was purified in the same manner as described previously [8]. The purified fusion protein and the allergen purified from soybean were electrophoresed on a 12% polyacrylamide gel and the proteins on a part of the gel were stained with Coomassie brilliant blue R250 (A). The proteins on the other part were electroblotted onto a nitrocellulose membrane. The membrane was divided into several strips. The strips were subjected to staining with the monoclonal antibody C5 (B) or with a serum of a soybean-sensitive patient (C). The upper and lower arrows express the positions of the fusion protein and the purified allergen, respectively. The standard protein markers are shown on the left side of the panel. Lane 1, the allergen purified from soybean; lane 2, the purified fusion protein.

clones encoding the allergen were isolated from the cDNA library in the same manner as described earlier [5], which had been prepared from developing soybean cotyledons (*Glycine max* c.v. Wasesuzunari) using a cloning vector λZAP II in the previous work [7]. Positive phage clones were converted to pBluescript phagemids according to the manufacturer's instruction. The cDNA inserts were sequenced in the same manner as described above. The homology search was performed using FASTA. The expression of the peptide fragment including the allergen was performed essentially in the same manner as described earlier [8]. Primers 3 (5′-CTATATGGATCCCATGATGATGAGGGTGGT-3′) and 4 (5′-TAAGTGGAATTCTTACTTCCCAAATACGGTTTCCAA-3′) were used as sense sequence for FHDDEGG and as antisense sequence for LETVFGK, respectively, in this work. The underlines in the former and latter primers represent the *Bam*H I and *Eco*R I sites, respectively.

The N-terminal amino acid sequence of the allergen was shown to be FHDDEGGDKKSPKSLFLMS-STR. The 20th amino acid residue was confirmed to be an *N*-glycosylated asaparagine residue [9]. By combination of the N-terminal amino acid sequence of the allergen with those of the five purified peptides overlapped, the amino acid sequences of fragments A, B and C were established to be FHDDEGGDKKSPKSLFLMSNSTRVFK, MRVLKSHGGRIFYRHMHIGFISMEP and KSLFVPQYLDSNLIIFIRRGEA, respectively. When PCR was done as described above, a 162-bp DNA fragment was produced. The amino acid sequence deduced from the product was shown to contain those of fragments A and B and to be completely identical with them (Fig. 1). This finding shows that the PCR product codes for the N-terminal region of the allergen. Using this product as a probe, cDNAs encoding the allergen were screened in the above-mentioned cDNA library and four positive phage clones were isolated. Of them, the phage containing the longest cDNA insert was converted to a phagemid and the cDNA in the phagemid was sequenced. The nucleotide length of the cDNA insert, excluding a polyadenylation tail, was 1567 bp. The cDNA was shown to contain an open reading frame with the nucleotide sequence between 1 and 1419, encompassing 473 amino acid residues, and a 148-bp 3′-untranslated region, followed by a polyadenylation tail (Fig. 1). The above-mentioned fragments A, B and C were contained in the amino acid sequence deduced from the nucleotide sequence of the cloned cDNA and their amino acid sequences were completely identical with the corresponding part on the deduced sequence. The allergen was suggested to be composed of 220 amino acids and fragment A was shown to be its N-terminal fragment by the previous study [4]. This means that the product for the cDNA must comprise of a 21-amino acid segment, the whole allergen and other protein(s), which is assumed to correspond to 232 amino acid residues. Possible glycosylation sites were found to be at positions Asn41, Asn184 and Asn349. The former two sites were suggested to be located in the Gly m Bd 28K. The position at Asn41 was proved to be glycosylated [9], but the glycosylation at Asn184 has not been revealed. These findings suggest that the first half of the cDNA cloned encodes the allergen.

In order to confirm that the cDNA encodes the allergen, we attempted to carry out the expression of the allergen with the cloned cDNA. However, the C-terminus of the allergen has never been determined. Therefore, on assumption that the allergen may consist of 220 amino acids, we selected the peptide fragment (sequence Phe22–Lys259), so that it contained the whole allergen. An expression vector of the peptide fragment including the allergen was constructed using pGEX-6P-1. When *Escherichia coli* JM109 harboring the expression vector was cultured in the absence of 1 mM isopropyl-1-thio-β-D-galactopyranoside (IPTG), the 53-kDa fusion protein of the peptide fragment with glutathione *S*-transferase was not detected by sodium dodecyl sulfate–polyacrylamide gel electrophoresis followed by staining with Coomassie brilliant blue R250, but it was shown to be expressed when *E. coli* was cultured in the presence of IPTG. As shown in Fig. 2B, the fusion protein was immunoblotted with monoclonal anti-

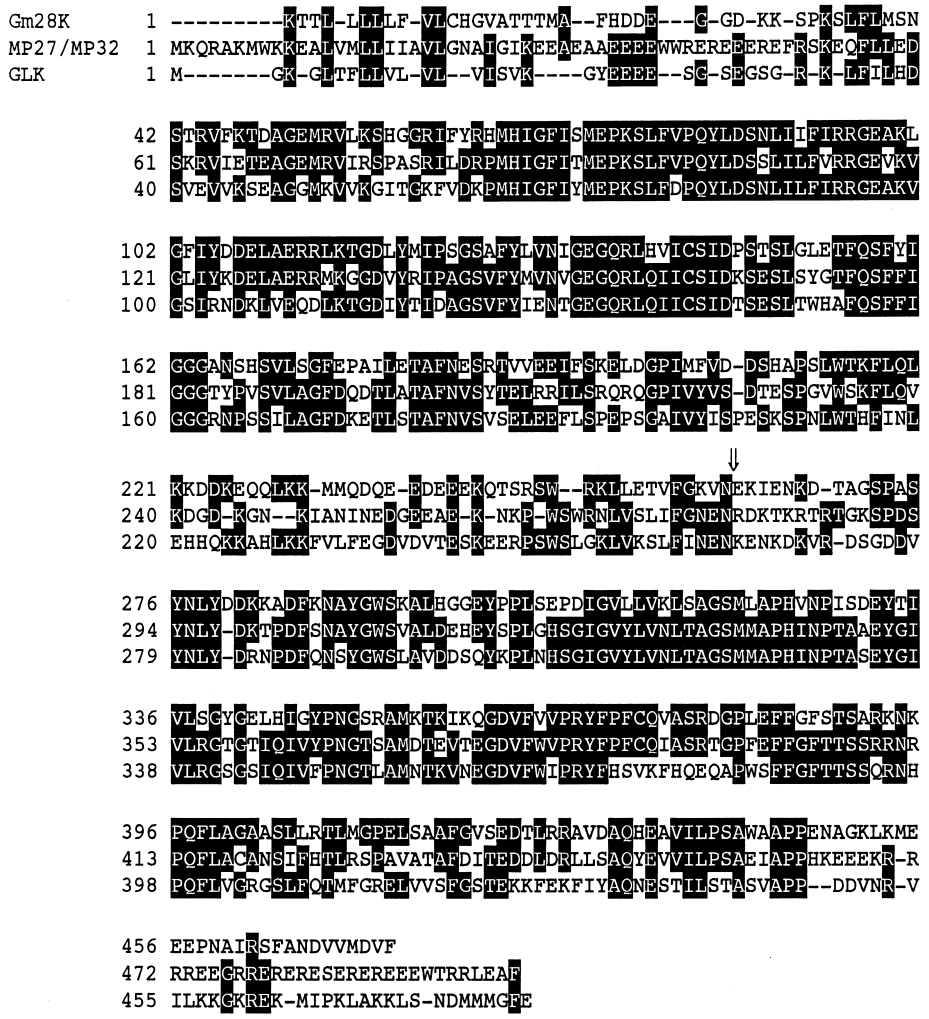

Fig. 3. Alignment of the deduced amino acid sequence of the allergen (Gm28K) with those of the pumpkin MP27/MP32 (MP27/MP32) and the carrot globulin-like protein (GLK). Shaded amino acid residues express the amino acid residues identical between two of the three proteins or among all of the proteins. The open arrow shows a putative cleavage site by a vacuolar processing enzyme, which was suggested in the case of processing of the pumpkin preproprotein to the mature proteins MP27/MP32 [10]. The sequences of MP27/MP32 and GLK were from *Cucurbita* sp. c.v. Kurokawa Amakuri Nankin [10] and *Daucus carota* c.v. Danvers Half-long [11], respectively.

body C5 against the allergen, which had been prepared in the previous work [4]. Furthermore, the fusion protein was observed to be immunostained with a serum of a soybean-sensitive patient, but the fusion protein expressed in *E. coli* was immunoblotted more weakly than the allergen purified from soybean (Fig. 2C). Since the purified allergen contains at least one glycan moiety [9], the difference in the reaction of the purified allergen and the expressed fusion protein with the patient's serum suggests that a considerable amount of IgE antibodies in the patient's serum may bind to the glycan moiety of the allergen as reported in the previous paper [9]. Similar results were obtained with the sera of several other soybean-sensitive patients.

The above-mentioned observations clearly demonstrate that the first half of the cDNA cloned encodes the allergen. We surveyed proteins homologous with the deduced amino acid sequence of the cDNA cloned using FASTA. The deduced protein from the cDNA cloned was shown to exhibit high overall homology with two proteins, the pumpkin proprotein for MP27/MP32 [10] and a carrot globulin-like protein [11] (50.4% identity between Gly m Bd 28K and the pumpkin proteins; 45.9% identity between Gly m Bd 28K and the carrot globulin-like protein; 37.9% identity among the three proteins) (Fig. 3). The 1661-bp cDNA encoding the MP27/MP32 in pumpkin seeds was reported to consist of a 18-bp 5′-untranslated region, a 1497-bp open reading frame, encompassing 499 amino acids, and a 146-bp 3′-untranslated region [10]. The open reading frame codes for a preproprotein, which eliminates a signal peptide of 27 amino acids to form a proprotein. Then, the proprotein was suggested to be cleaved between Asn278 and Arg279 by a vacuolar processing enzyme (asparaginyl endopeptidase) to form the mature proteins, MP27 and MP32 [10]. The fact that Gly m Bd 28K occurs in soybean seeds [4] suggests that the cDNA

cloned from soybean seeds may encode a preproprotein, although the preproprotein may not include the whole, but a part of its signal peptide. The preproprotein might release the signal peptide to form the proprotein and then it would be cleaved by the process enzyme in a similar manner to the processing of the pumpkin proprotein. The alignment of the amino acid sequence of the soybean proprotein with that of pumpkin MP27/MP32 assumes that the soybean proprotein might be cleaved between Asn261 and Glu262, although no information as to the predicted protein with a molecular mass of 23 254 Da (the C-terminal half of the proprotein) has been available.

High homology of Gly m Bd 28K with pumpkin MP27/MP32 and the carrot globulin-like protein suggests that patients sensitive to Gly m Bd 28K might be susceptible to pumpkin MP27 protein and the carrot globulin-like protein, although the cross-reactivity of the sera of Gly m Bd 28K-sensitive patients with the pumpkin and carrot proteins should be elucidated. Recently, many patients with latex allergy have been reported and most of the patients have been shown to exhibit allergic responses to fruits [12]. This phenomenon is known as the latex–fruit syndrome [13]. The allergens identified so far tend to be classified into several groups such as hevein homologs [14] and Bet v 1 homologs [15]. Therefore, it is very interesting that the homology observed among the pumpkin proteins, the carrot protein and Gly m Bd 28K will be well elucidated in connection with the allergenicity of Gly m Bd 28K.

This work was supported in part by a Grant-in-Aid for Scientific Research from the Ministry of Education, Science and Culture of Japan.

References

[1] T. Ogawa, N. Bando, H. Tsuji, K. Okajima, K. Nishikawa, K. Sasaoka, J. Nutr. Sci. Vitaminol. 37 (1991) 555–565.
[2] T. Ogawa, H. Tsuji, N. Bando, Y. Kitamura, Y.-U. Zhu, H. Hirano, K. Nishikawa, Biosci. Biotech. Biochem. 57 (1993) 1030–1033.
[3] T. Ogawa, N. Bando, H. Tsuji, K. Nishikawa, K. Kitamura, Biosci. Biotech. Biochem. 59 (1995) 831–833.
[4] H. Tsuji, N. Bando, M. Hiemori, R. Yamanishi, M. Kimoto, K. Nishikawa, T. Ogawa, Biosci. Biotech. Biochem. 61 (1997) 942–947.
[5] H. Tsuji, T. Oka, M. Kimoto, Y.-M. Hong, Y. Natori, T. Ogawa, Biochim. Biophys. Acta 1309 (1996) 31–36.
[6] F. Sanger, E.F. Nicklen, A.R. Coulson, Proc. Natl. Acad. Sci. USA 74 (1977) 5463–5467.
[7] T. Katsube, M. Adachi, N. Maruyama, K. Ichise, Y. Takenaka, S. Utsumi, Plant Physiol. 109 (1995) 722–723.
[8] H. Tsuji, M. Kimoto, H. Watanabe, T. Sasagawa, T. Oka, H. Yamashita, M. Okita, Biochim. Biophys. Acta 1425 (1998) 628–631.
[9] M. Hiemori, N. Bando, T. Ogawa, H. Shimada, H. Tsuji, R. Yamanishi, J. Terao, Int. Arch. Allergy Immunol. 122 (2000) 238–245.
[10] K. Inoue, A. Motozaki, Y. Takeuchi, M. Nishimura, I. Hara-Nishimura, Plant J. 7 (1995) 235–243.
[11] X. Lin, J.L. Zimmerman, Plant Physiol. 116 (1998) 1603.
[12] A.W. Fränkland, Clin. Exp. Allergy 25 (1995) 580–581.
[13] R. Brehler, U. Theissen, C. Mohr, T. Luger, Allergy 52 (1997) 404–410.
[14] C. Blanco, A. Diaz-Perales, C. Barmen, R. Sanchez-Monge, C. Aragoncillo, R. Castillo, N. Ortega, M. Alvarez, T. Carrilo, G. Salcedo, J. Allergy Clin. Immunol. 103 (1999) 507–513.
[15] H. Breiteneder, Allergy 53 (1998) 31–34.