THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 270, No. 6, Issue of February 10, pp. 2607–2613, 1995
Printed in U.S.A.

Isoforms of Bet v 1, the Major Birch Pollen Allergen, Analyzed by Liquid Chromatography, Mass Spectrometry, and cDNA Cloning*

(Received for publication, August 15, 1994, and in revised form, October 18, 1994)

Ines Swoboda‡§, Alexander Jilek§¶‖, Fátima Ferreira, Edwin Engel**, Karin Hoffmann-Sommergruber‡, Otto Scheiner‡, Dietrich Kraft‡, Heimo Breiteneder‡, Ernst Pittenauer¶, Erich Schmid¶, Oscar Vicente‡‡, Erwin Heberle-Bors‡‡, Horst Ahorn§§, and Michael Breitenbach**¶¶**

*From the ‡Institut für Allgemeine und Experimentelle Pathologie, Universität Wien, Währinger Gürtel 18–20, A-1090 Vienna, Austria, the **Institut für Genetik und Allgemeine Biologie, Universität Salzburg, Hellbrunnerstrasse 34, A-5020 Salzburg, Austria, the ‡‡Institut für Mikrobiologie und Genetik, Universität Wien, Dr. Bohrgasse 7, A-1030 Vienna, Austria, the ¶Institut für Analytische Chemie, Universität Wien, Währingerstraße 38, A-1090 Vienna, Austria, and the §§Ernst Böhringer Institut für Arzneimittelforschung, A-1120, Vienna, Austria*

Bet v 1, the major allergen of birch pollen, displays a considerable degree of heterogeneity. Several charge variants have been detected by two-dimensional IgE immunoblots and isoelectric focusing techniques. This heterogeneity has been attributed to glycosylation (or other post-translational modifications) or to isogenes coding for Bet v 1 isoforms and/or allelic variants. However, until now, only limited structural data for Bet v 1 have been published. Recently, we described the expression, purification, and immunological properties of recombinant Bet v 1 (rBet v 1) produced in *Escherichia coli* as a non-fusion protein (Ferreira, F. D., Hoffmann-Sommergruber, K., Breiteneder, H., Pettenburger, K., Ebner, C., Sommergruber, W., Steiner, R., Bohle, B., Sperr, W. R., Valent, P., Kungl, A. J., Breitenbach, M., Kraft, D., and Scheiner, O. (1993) *J. Biol. Chem.* 268, 19574–19580). Here, we present a more detailed structural characterization of Bet v 1 by both cDNA cloning and mass spectrometry. Thirteen different cDNA clones coding for Bet v 1 isoforms were obtained by polymerase chain reaction amplification of birch pollen cDNA with a sequence-specific 5′-terminal primer and a nonspecific 3′-terminal primer or by immunological screening of a birch pollen cDNA library. These isoforms are referred to as Bet v 1b to Bet v 1n, whereas the previously isolated Bet v 1 cDNA (Breiteneder, H., Pettenburger, K., Bito, A., Valenta, R., Kraft, D., Rumpold, H., Scheiner, O., and Breitenbach, M. (1989) *EMBO J.* 8, 1935–1938) is now referred to as Bet v 1a. High performance liquid chromatography and plasma desorption mass spectrometry of proteolytic fragments of purified natural Bet v 1 (nBet v 1) and rBet v 1a were used to (i) confirm the primary structure of all Bet v 1 isoforms and (ii) to investigate any possible postsynthetic modifications on rBet v 1a or on the natural mixture of isoallergens obtained from birch pollen. Except for the cleavage of initiating methionine, no postsynthetic modifications were found in either nBet v 1 or rBet v 1a.

In the temperate climate zone of the world, pollen from trees of the order Fagales (*e.g.* birch, alder, hazel, oak, and hornbeam) are a major cause of Type I allergies (Ipsen *et al.*, 1985; Jarolim *et al.*, 1989a). Birch pollen contains a single major allergen with a molecular mass of 17 kDa (Ipsen and Loewenstein, 1983), designated Bet v 1.[1] More than 96% of all tree pollen allergic patients display IgE antibodies to Bet v 1, and 60% react exclusively to this allergen, indicating the importance of this protein in tree pollen allergy (Jarolim *et al.*, 1989a).

Previously, we isolated and sequenced a cDNA clone coding for Bet v 1 (Breiteneder *et al.*, 1989), which shows high sequence similarities to the single major pollen allergen from alder, Aln g 1 (Breiteneder *et al.*, 1992), from hornbeam, Car b 1 (Larsen *et al.*, 1992), and from hazel, Cor a 1 (Breiteneder *et al.*, 1993). This is in good agreement with the observation that patients displaying specific IgE to Bet v 1 also show symptoms during the flowering season of other trees of the order Fagales.

Interestingly, all of these major tree pollen allergens show significant sequence similarities to a family of plant pathogen-activated genes shown to be induced in somatic tissues by infection with fungi and bacteria. They were identified in pea (Fristensky *et al.*, 1988), parsley (Somssich *et al.*, 1988), potato (Matton and Brisson, 1989), bean (Walter *et al.*, 1990), asparagus (Warner *et al.*, 1992), and soybean (Crowl *et al.*, 1992). Although they have been associated with defense response of plants, the precise role of the respective gene products still remains elusive. Computer-aided sequence comparisons do not point to any known biochemical function. Presently, several families of pathogenesis-related proteins and genes are known (Bowles, 1990), but these families show no similarity to Bet v 1 and homologous proteins.

Bet v 1 shows a considerable degree of heterogeneity. Up to 10 charge variants have been observed by two-dimensional IgE immunoblots (Rohac *et al.*, 1991) and isoelectric focusing techniques (Ferreira *et al.*, 1993). Previously, nBet v 1 was described as an acidic glycoprotein (Ipsen and Hansen, 1989; Larsen *et al.*, 1992), and a single consensus site for *N*-glycosylation is present in the Bet v 1 sequence (Breiteneder *et al.*, 1989). Thus, glycosylation (or other post-translational modifications) could be an explanation for the observed heterogeneity. However, until now, conclusive evidence for the presence,

* This work was supported by Grants P10019-MOB (to F. F. and M. B.) and S6002-BIO (to O. S.) from the Austrian Fonds zur Förderung der Wissenschaftlichen Forschung (F.W.F.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession number(s) 77200, 77265–77270, 77271–77273, 77274, 81972, and 82028.

§ The first two authors contributed equally to this paper.

‖ Present address: Inst. f. Genetik u. Allg. Biologie, Universität Salzburg, Hellbrunnerstr. 34, A-5020 Salzburg, Austria.

¶¶ To whom correspondence and reprint requests should be addressed.

[1] The abbreviations used are: Bet v, *Betula verrucosa*; Aln g, *Alnus glutinosa*; Car b, *Carpinus betulus*; Cor a, *Corylus avellana*; HPLC, high performance liquid chromatography; nBet v 1, natural Bet v 1; PCR, polymerase chain reaction; PDMS, plasma desorption mass spectrometry; rBet v 1, recombinant Bet v 1.

nature, and location of structural modifications of the Bet v 1 molecule is not available.

Isogenes coding for Bet v 1 isoforms and/or allelic variants could be another explanation for the heterogeneity observed in two-dimensional IgE immunoblots. Southern blot analysis of birch (Valenta *et al.*, 1991) and differences in the reactivity of two anti-Bet v 1 monoclonal antibodies (Rohac *et al.*, 1991) support this view.

To date, only limited structural analysis of Bet v 1 and related proteins has been published. In a previous paper we described the expression, purification, and immunological properties of rBet v 1 produced in *Escherichia coli* as a non-fusion protein (Ferreira *et al.*, 1993). In the present study, we undertook a more detailed structural characterization of Bet v 1 by both cDNA cloning and mass spectrometry.

EXPERIMENTAL PROCEDURES

Cloning and Sequencing of Bet v 1 Isoforms—Total RNA was extracted from birch pollen (Allergon AB, Engelholm, Sweden) according to Chomczynski and Sacchi (1987). The first cDNA strand was synthesized from poly(A)$^+$-enriched RNA with reverse transcriptase (Amersham), according to the manufacturer's instructions. A synthetic oligodeoxynucleotide, Cora 1, containing an oligo(dT) tract, a spacer sequence, and a *Hin*dIII recognition site (underlined) was used as a primer (5′-GAGAGAGAGAGAAAGCTTT$_{18}$-3′). First-strand cDNA synthesis products were amplified with 2.5 units of Taq DNA polymerase (Boehringer Mannheim) in PCR buffer (Perkin Elmer; 4 mM $MgCl_2$, 0.75 mM of each dNTP, and 0.15 μM each of Cora 2 and Cora 3 primers. The Cora 2 primer was designed to contain an *Eco*RI site (underlined) followed by the first 19 nucleotides of the Bet v 1a coding region (Breiteneder *et al.*, 1989) (5′-GGGAATTCATGGGTGTTTTCAATTACG-3′). The Cora 3 primer was similar to Cora 1, which was used for the first-strand cDNA synthesis, except that it did not contain the oligo(dT) part. Amplification was carried out as described previously (Breiteneder *et al.*, 1993). Amplified PCR products were eluted from an agarose gel, digested with *Eco*RI and *Hin*dIII, and cloned into a pUCBM20 plasmid.

A birch pollen cDNA library was constructed in λZAP (Stratagene, La Jolla, CA) and screened with serum IgE from an allergic individual selected according to typical case history, positive skin prick test, and RAST (radioallergosorbent test) class > 3.5, as described previously (Breiteneder *et al.*, 1989).

Clones coding for Bet v 1 isoforms were isolated, and both strands were sequenced twice according to the dideoxy chain termination method (Sanger *et al.*, 1977) using a T7 Sequencing Kit (Promega).

Purification of Natural Bet v 1 (nBet v 1) and Recombinant Non-fusion Bet v 1a (rBet v 1a)—Protein extract from birch (*Betula verrucosa*) pollen (Allergon AB) was prepared as described previously (Jarolim *et al.*, 1989a). nBet v 1 was purified from pollen extracts by a combination of affinity chromatography on an immobilized anti-Bet v 1 monoclonal antibody (Jarolim *et al.*, 1989b) and reversed phase HPLC on a Hypersil WP300 C_8 column, as described previously (Ferreira *et al.*, 1993).

rBet v 1a was purified from crude *E. coli* lysates by chromatofocusing on a PBE-94 exchanger column followed by reversed phase HPLC (Ferreira *et al.*, 1993). Purified Bet v 1 proteins were analyzed by SDS-polyacrylamide gel electrophoresis according to the method of Laemmli (1970) and visualized by staining with Coomassie Brilliant Blue R-250.

Protein concentration was determined by the micro-Kjeldahl method, using glycine as standard (Jacobs, 1959).

Protease Treatments: Purification and Plasma Desorption Mass Spectrometry (PDMS) of Bet v 1 Proteolytic Fragments—A solution of purified nBet v 1 or rBet v 1a (100 μg in 50 μl of double distilled water) was heated for 20 min at 95 °C and diluted with an equal volume of 0.2 M NH_4HCO_3. One microgram of trypsin (sequencing grade, Boehringer Mannheim) was then added and the mixture incubated at 37 °C for 2 h. Afterwards, trypsin was added again, and incubation was carried out for an additional 4-h period. The reaction was stopped by adding [1/10] vol of trifluoroacetic acid and dried *in vacuo*. Digestion of the Bet v 1 proteins with endoproteinase Glu-C (sequencing grade, Boehringer Mannheim) was done for 4 h at 37 °C using 2 μg of the protease. Peptides resulting from either treatment were separated by reversed phase HPLC on a Waters μBondapak C_{18} column (3.9 × 300 mm). The column was eluted with a linear gradient of acetonitrile (solvent A, 0.1% (v/v) trifluoroacetic acid in water; solvent B, 0.07% (v/v) trifluoroacetic acid in acetonitrile; 0–40% B in 120 min; flow rate, 1 ml/min). UV absorbance was monitored at 214 nm. Fractions were collected and vacuum dried. Peptide solutions were adsorbed on nitrocellulose-layered targets followed by spin drying (Nielsen *et al.* 1988). Spectra were obtained on a Bio-Ion 20 K time-of-flight mass spectrometer (Uppsala, Sweden) using accelerating voltages of 18 kV and −15 kV. After washing the targets three times with 20 μl of 0.1% (v/v) trifluoroacetic acid, data acquisition was repeated. All data are shown as chemical average masses.

Peptide Sequencing by Solid Phase Edman Degradation—This was performed as described previously (Breiteneder *et al.*, 1989).

RESULTS

Cloning and Sequence Analysis of Bet v I Isoforms—Pollen Bet v 1 isoforms all seem to share the same N-terminal amino acid sequence (Ipsen and Hansen, 1989). We therefore assumed that the DNA sequences at the 5′-end of the coding region of the different isoforms are similar, if not identical. Since the full cDNA sequence of one isoform, Bet v 1a, was already known (Breiteneder *et al.*, 1989), it was possible to design a primer for the specific PCR amplification of the coding and 3′-untranslated region of the different Bet v 1 cDNAs after reverse transcription of birch pollen poly(A)$^+$ RNA. A similar strategy has been successfully used to clone several isoforms of Cor a 1, the Bet v 1 homologue from hazel pollen (Breiteneder *et al.*, 1993).

After cloning and sequencing, several PCR-amplified fragments were found to correspond to the original Bet v 1a clone. In addition, 11 different cDNA clones, with lengths ranging from 567 to 756 base pairs, were isolated (Bet v 1b to Bet v11, respectively). Five of these clones showed 3′-untranslated regions of different lengths and contained poly(A) tails, whereas the other six were truncated at the same position in the 3′-noncoding region (approximately 80 nucleotides downstream of the stop codon) because of a single base exchange (G→A) that created a new recognition site for *Hin*dIII, one of the enzymes used for cloning the PCR fragments. Two complete Bet v 1 cDNA clones (Bet v 1 m/n) were isolated by screening a birch pollen cDNA library with human IgE antibodies.

All cDNAs contained open reading frames of 480 nucleotides, coding for putative proteins of 160 amino acids, with calculated molecular masses ranging from 17,450 to 17,573 Da. The deduced amino acid sequences compared with that of Bet v 1a are shown in Fig. 1. In three cases (Bet v 1d/h, Bet v 1f/i, and Bet v 1 m/n) differences in the nucleotide sequences did not result in amino acid changes. Since in all PCR clones the first 19 nucleotides were included in the 5′ PCR primer, additional differences at the DNA level could be possible in this region, but these were not detected.

Therefore, including Bet v 1a, eleven Bet v 1 protein sequence isoforms have been identified altogether, with amino acid identities ranging from 84.4% (because of differences in 25 amino acids) to 99.4% (a single amino acid exchange) for the different pairs.

PDMS Analysis—After reversed phase HPLC, the Bet v 1 preparations (natural and recombinant) migrated as one single band in SDS-polyacrylamide gels (not shown). Purified natural and recombinant Bet v 1a were each treated with trypsin and endoproteinase Glu-C. The resulting peptide mixtures were directly analyzed by PDMS (as an example, see Fig. 2). In addition, the total proteolytic digests were fractionated by reversed phase HPLC (Fig. 3), and the resulting fractions were reanalyzed by PDMS. This procedure had two advantages that allowed greater coverage of the amino acid sequence: (i) signals for peptides not detected by direct analysis of the complex mixture were frequently observed after HPLC fractionation, and (ii) signals for peptides observed in the complex mixture often were much stronger in the partially purified fractions. Table I summarizes the data obtained by PDMS analysis of

```
→ Bet v 1 a   MGVFNYETETTSVIPAARLFKAFILDGDNLFPKVAPQAISSVENIEGNGGPGTIKKISFPEGFPFKYVKDRVDEVDHTNF
  Bet v 1 j   .....A.....
→ Bet v 1 f/i .....I.A.....
→ Bet v 1 d/h .....I.....V.....N.....
→ Bet v 1 g   .....S.....E....I.....N.....
→ Bet v 1 e   .....A.....I.....G.....
  Bet v 1 l   .....A.....M.....K.V.....N.....
  Bet v 1 b   .....E..T.I.....T....S.....E.....A..
  Bet v 1 k   .....S.....E..T.I.....T....S.....E.....A..
→ Bet v 1 c   .....S.....E..T.I.....T....S.....E.....A..
  Bet v 1 m/n .....S.....I.....T....S.....E.....A..

  Bet v 1 a   KYNYSVIEGGPIGDTLEKISNEIKIVATPDGGSILKISNKYHTKGDHEVKAEQVKASKEMGETLLRAVESYLLAHSDAYN
  Bet v 1 j   ..S.....V.....N.....N.....I.....
  Bet v 1 f/i ..S.....V.....N.....N.....I.....
  Bet v 1 d/h .....V.....CV.....N.....
  Bet v 1 g   .....V.....CV.....N.....
  Bet v 1 e   ..S.....V.....N.....N.....I.....
  Bet v 1 l   .....V.....CV.....N.....
  Bet v 1 b   ..S..M....AL.....C.....M...HM..I..K..A.....
  Bet v 1 k   ..S..M....AL.....C.....M...HM..I..K..A.....
  Bet v 1 c   ..S..M....AL.....C.....Q.M...HM..I..K..A.....
  Bet v 1 m/n ..S..M....AL.....C.....M...HM..I..K..A.....
```

FIG. 1. **Deduced amino acid sequence alignment of Bet v 1 isoforms from birch pollen.** *Dots* indicate identical amino acids as in Bet v 1a. *Arrows* mark isoforms identified in the pollen mixture by PDMS analysis. Isoforms b, c, k, and m/n were not individually identified but were confirmed as a group.

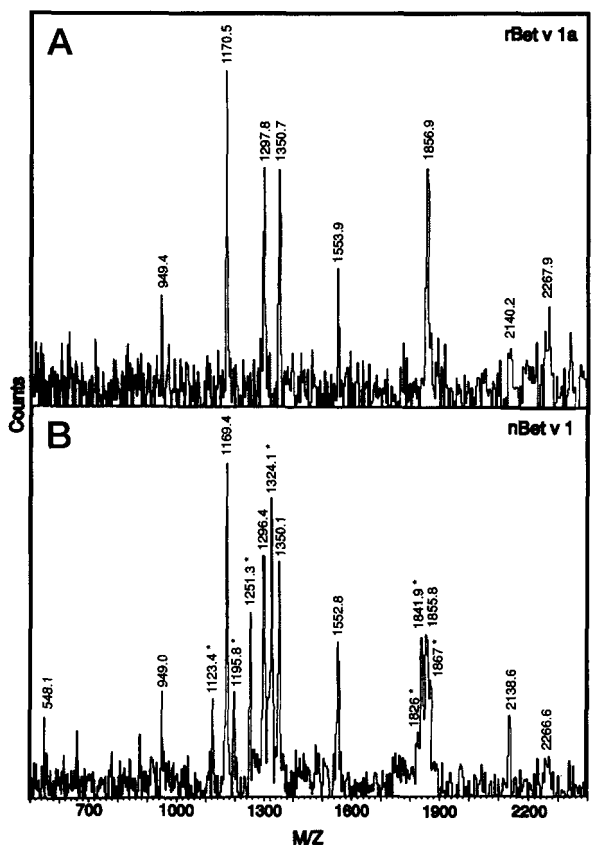

FIG. 2. **Plasma desorption map of tryptic peptides.** Total peptide mixtures of tryptic digests of rBet v 1a (*A*) and nBet v 1 (*B*) applied to PDMS are shown. Additional mass peaks in (*B*) can be explained by newly identified isoforms.

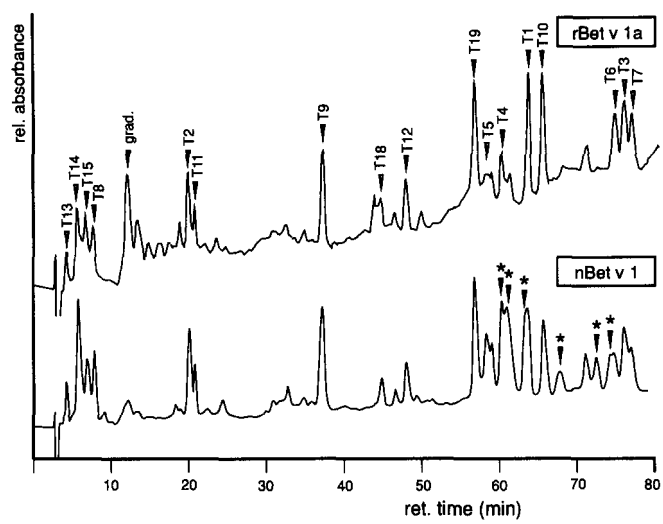

FIG. 3. **Peptide map of Bet v 1 tryptic digests.** Peptides obtained from tryptic cleavages were separated by reversed phase HPLC and identified by PDMS. The *peak* marked (*grad.*) shows the start of linear gradient; *, peaks corresponding to peptides derived from new isoforms.

proteolytic digests of natural and rBet v 1a after HPLC fractionation. The recorded mass signals were mapped onto the cDNA-derived Bet v 1a sequence (Breiteneder *et al.*, 1989) according to their molecular mass and enzyme specificity (Fig. 4).

Forty-three peptides (T peptides) were detected by PDMS in a tryptic digest of nBet v 1, and their molecular weights were determined from the obtained spectra (Table I). Eighteen of the mass signals could be easily matched with the molecular weights of peptides predicted from the amino acid sequence deduced from the published Bet v 1a cDNA sequence (T2–T19). The expected signal at m/z 1987 corresponding to the N-terminal peptide T1 was missing from the spectra. A signal detected at m/z 1856 could be accounted for, assuming that the initial methionine had been removed, leaving glycine as the NH_2-terminal amino acid. This signal could also be matched with the expected T10 fragment (m/z 1856). As two different HPLC fractions of tryptic digests produced an ion at m/z 1856, these fractions were each subdigested with endoproteinase Glu-C, and the resulting peptide mixtures (TE peptides) were analyzed by PDMS (Table I). The two fractions produced $(M+H)^+$ ions at m/z 729/916 and 988, respectively (predicted mass values for endoproteinase Glu-C-subdigested T1 lacking the initial methionine were 729, 249, and 916; for T10 the predicted values were 888 and 987). These fractions were identified as T1 and T10, respectively, thus confirming the above assignment. The removal of initiating methionine was also confirmed by NH_2-terminal sequence analysis of purified nBet v 1 and rBet v1a (data not shown). Therefore, the amino acid sequence defined in the present study corresponds to residues 2–160 of the published cDNA sequence. To be consistent with the NH_2-terminal sequence of the mature Bet v 1 protein, we have renumbered the sequence of Bet v 1a starting with Met^0-Gly^1-Val^2-Phe^3 . . . and applied the same numbering system to the new Bet v 1 cDNA sequences presented here. The signal at m/z 1476 was assigned to peptide T9 (69–80) (Fig. 4) originating from an incomplete trypsin cleavage. As shown in Fig. 4, PDMS analysis and HPLC fractionation of nBet v 1 tryptic digests confirmed 100% of the Bet v 1a sequence.

TABLE I

Mass determinations of proteolytic fragments of nBet v 1 and rBet v 1a after fractionation by reversed phase HPLC

Theoretical m/z values give the exact masses of the peptides plus a proton for formation of the molecular ion, $(M+H)^+$. T, peptides resulting from tryptic digest; E, peptides from endoproteinase Glu-C digest; TE, peptides from digestion of HPLC-purified T1 and T10 peptides with endoproteinase Glu-C. Underlined amino acids mark differences from Bet v 1a sequence. Peptides T19–3/E14–3 and T19–5/E14–5 can be explained by truncation of the COOH terminus by 3 and 5 amino acids, respectively. +, incomplete cleavage; *, methionine sulfoxide; **, sequenced by Edman degradation; °, detected only in peptide mixture.

PEPTIDE	m/z			POSITION	ISOFORM	SEQUENCE
	Theoret.	Observed				
		nBet v 1	rBet v 1a			
T1	1987.2	1855.8	1856.3	0-17	a,b	MGVFNYETETTSVIPAAR
	1973.2	1842.1			c,g,k,m/n	MGVFNYESETTSVIPAAR
	1999.3	1867.0 °			d/h	MGVFNYEIETTSVIPAAR
	1957.2	1826.0 °			e,j,l	MGVFNYETEATSVIPAAR
	1969.3	1837.7			f/i	MGVFNYEIEATSVIPAAR
T2	407.5	407.4	407.7	18-20	a,b,c,d/h,g,e,f/i,j,k,m/n	LFK
	425.6	-			l	MFK
T3	1350.5	1351.0	1350.3	21-32	a,e,f/i,j	AFILDGDNLFPK
	1317.6	1317.1			b,c,k	AFILEGDTLIPK
	1302.5	1301.5			d/h	AFILDGDNLVPK
	1330.6	-			g	AFILEGDNLIPK
	879.0	-			l	AFILDGDK
	456.6	-			l	LVPK
	1316.5	-			m/n	AFILDGDNLIPK
T4	2139.4	2139.7	2140.2 °	33-54	all	VAPQAISSVENIEGNGGPGTIK
T5	2267.5	2270.0	2267.9 °	33-55	all	VAPQAISSVENIEGNGGPGTIKK
T6	1297.5	1297.8	1297.9	55-65	a,f/i,j	KISFPEGFPFK
	1251.5	1251.7			b,c,k,m/n	KITFPEGSPFK
	1324.6	1325.3			d/h,g,l	KINFPEGFPFK
	1263.5	-			e	KISFPEGIPFK
T7	1169.4	1169.7	1169.8	56-65	a,f/i,j	ISFPEGFPFK
	1123.3	1123.4			b,c,k,m/n	ITFPEGSPFK
	1196.4	-			d/h,g,l	INFPEGFPFK
	1135.3	1132.3			e	ISFPEGIPFK
T8	409.5	409.5	409.6	66-68	all	YVK
T9	1475.5	1476.3	1474.7	69-80	a,d/h,g,f/i,j,l	DRVDEVDHTNFK+
	1459.6	1460.3			b,c,k,m/n	ERVDEVDHANFK+
	1417.5	1415.9			e	GRVDEVDHTNFK+
T10	1856.0	1855.7	1856.4	81-97	a	YNYSVIEGGPIGDTLEK
	1835.0	-			b,c,k,m/n	YSYSMIEGGALGDTLEK
	1842.0	1842.5			d/h,g,l	YNYSVIEGGPVGDTLEK
	1815.0	1813.8			e,f/i,j	YSYSVIEGGPVGDTLEK
T11	703.8	704.3	704.3	98-103	a,d/h,g,e,f/i,j,l	ISNEIK
	719.9	719.0			b,c,k,m/n	ICNEIK
T12	1171.4	1171.5	1170.8	104-115	a,b,c,k	IVATPDGGSILK
	1173.4	1174.2			d/h,g,l	IVATPDGGCVLK
	1170.4	-			e,f/i,j	IVATPNGGSILK
T13	461.5	461.9	461.5	116-119	a,b,c,d/h,g,k,l,m/n	ISNK
	488.6	-			e,f/i,j	INNK
T14	548.6	549.1	549.1	120-123	all	YHTK
T15	684.7	684.9	684.9	124-129	a,e,f/i,j	GDHEVK
	716.8	717.1			b,k,m/n	GDHEMK
	707.8	-			c	GDQEMK
	683.7	-			d/h,g,l	GNHEVK
T16	574.6	574.9	-	130-134	a,d/h,g,l	AEQVK
	615.7	616.0			b,c,k,m/n	AEHMK
	588.7	588.8			e,f/i,j	AEQIK
T17	305.3	304.3	-	135-137	a,d/h,g,e,f/i,j,l	ASK
	331.4	331.6			b,c,k,m/n	AIK
T18	949.1	949.4	949.4 °	138-145	a,d/h,g,e,f/i,j,l	EMGETLLR
	916.1	916.3			b,c,k,m/n	EKGEALLR
	658.8	659.0			b,c,k,m/n	GEALLR
T19	1553.7	1552.9	1553.6	146-159	all	AVESYLLAHSDAYN

Treatment of nBet v 1 with endoproteinase Glu-C yielded the 26 peptides (E peptides) shown in Table I. These peptides covered about 98% of the Bet v 1a sequence (Fig. 4). Similarly, as observed in the analysis of tryptic digests, the expected mass signal at m/z 860 (corresponding to the NH_2-terminal peptide E1) was absent in the spectra. The $(M+H)^+$ ion at m/z 729 was unmatched by any expected fragment according to the published cDNA sequence and, thereupon, assigned to peptide E1 lacking the initiating methionine. The $(M+H)^+$ ions at m/z 959 and 1860 were assigned to E1+E2 and E4 peptides, respec-

TABLE I—*continued*

PEPTIDE	m/z			POSITION	ISOFORM	SEQUENCE
	Theoret.	Observed				
		nBet v 1	rBet v 1a			
E1	860.0	729.4	728.6	0-6	all	MGVFNYE
E1+E2	1090.2	959.8	958.9	0-8	a,b,e,j,l	MGVFNYETE*
E2	249.2	-	-	7-8	a,b,e,j,l	TE
	235.2	-			c,g,k,m/n	SE
	261.3	-			d/h,f/i	IE
E3	3618.2	3616.9	3616.2 °	9-42	a	TTSVIPAARLFKAFILDGDNLFPKVAPQAISSVE
	3588.2	-			e,f/i,j	ATSVIPAARLFKAFILDGDNLFPKVAPQAISSVE
	1878.3	1878.9			b,c,g,k	TTSVIPAARLFKAFILE
	1739.0	1735.8			g	GDNLIPKVAPQAISSVE
	1726.0	-			b,c,k	GDTLIPKVAPQAISSVE
	3570.2	-			d/h	TTSVIPAARLFKAFILDGDNLVPKVAPQAISSVE
	3572.2	-			l	ATSVIPAARMFKAFILDGDKLVPKVAPQAISSVE
	3584.2	-			m/n	TTSVIPAARLFKAFILDGDNLIPKVAPQAISSVE
E4	1859.1	1860.0	1861.3	43-60	a,e,f/i,j	NIEGNGGPGTIKKISFPE*
	1873.1	-			b,c,k,m/n	NIEGNGGPGTIKKITFPE*
	1516.7	1516.5			b,c,k,m/n	GNGGPGTIKKITFPE
	1886.1	1889.6			d/h,g,l	NIEGNGGPGTIKKINFPE*
E5	1600.8	1601.6	1601.8	61-73	a,d/h,g,f/i,j,l	GFPFKYVKDRVDE
	1554.7	1554.7 **			b,c,k,m/n	GSPFKYVKERVDE
	1055.2	1055.3			b,c,k,m/n	GSPFKYVKE
	518.5	-			b,c,k,m/n	RVDE
	1508.8	-			e	GIPFKYVKGRVDE
E6	1729.9	1730.4	1729.7	74-87	a,d/h,g,l	VDHTNFKYNYSVIE
	1704.9	1720.3 *			b,c,k,m/n	VDHANFKYSYSMIE
	1702.9	-			e,f/i,j	VDHTNFKYSYSVIE
E7	858.9	858.2	859.1	88-96	a	GGPIGDTLE
	832.9	-			b,c,k,m/n	GGALGDTLE
	844.9	845.2 **			d/h,g,e,f/i,j,l	GGPVGDTLE
E8	590.6	591.0	590.8	97-101	a,d/h,g,e,f/i,j,l	KISNE
	606.7	-			b,c,k,m/n	KICNE
E9	2823.2	2823.6	2823.7	102-127	a,b,k,m/n	IKIVATPDGGSILKISNKYHTKGDHE
	2814.2	-			c	IKIVATPDGGSILKISNKYHTKGDQE
	2824.3	-			d/h,g,l	IKIVATPDGGCVLKISNKYHTKGNHE
	2849.3	-			e,f/i,j	IKIVATPNGGSILKINNKYHTKGDHE
E10	446.5	446.7	446.8	128-131	a,d/h,g,e,f/i,j,l	VKAE
	478.6	-			b,c,k,m/n	MKAE
E11	789.9	790.4	790.0	132-138	a,d/h,g,l	QVKASKE
	857.1	857.3			b,c,k,m/n	HMKAIKE
	803.9	-			e,f/i,j	QIKASKE
E12	336.4	-	-	139-141	a,d/h,g,e,f/i,j,l	MGE
	333.4	333.6			b,c,k,m/n	KGE
E13	802.0	802.0	802.1	142-148	a,d/h,g,e,f/i,j,l	TLLRAVE
	771.9	772.3			b,c,k,m/n	ALLRAVE
E14	1254.3	1254.4	1253.7	149-159	all	SYLLAHSDAYN
T1E1	728.8	729.1		1-6		GVFNYE
T1E2	249.2	-		7-8		TE
T1E3	916.1	916.5		9-17		TTSVIPAAR
T10E1	888.0	-		81-87		YNYSVIE
T10E2	987.1	987.9		88-97		GGPIGDTLEK
E14-3	906.0	906.5		149-156		SYLLAHSD
T19-3	1205.3	1205.0		146-156		AVESYLLAHSD
E14-5	703.8	704.0		149-154		SYLLAH
T19-5	1003.1	1003.2		146-154		AVESYLLAH

tively, originating from incomplete cleavage by endoproteinase Glu-C.

According to the Bet v 1a cDNA sequence, peptides T10 (81–97) and E6 (74–87) should contain the only potential Asn-linked glycosylation site (Asn^{82}). The spectra of nBet v 1 digested with either trypsin (Fig. 2) or endoproteinase Glu-C showed strong signals at m/z 1855 (T10) and 1730 (E6), respectively, demonstrating that the asparagine residue at position 82 was unmodified. Fig. 5 shows the mass spectra recorded on HPLC-purified T10 and E6 peptides. As described above, purified T10 was also subdigested with endoproteinase Glu-C, and mass spectra were recorded on the resulting sample (Table I).

Treatment of rBet v 1a with trypsin yielded the 17 peptides shown in Table I, which covered approximately 95% of the amino acid sequence. The 5% of the rBet v 1a sequence not mapped consisted of small tryptic peptides (one pentapeptide,

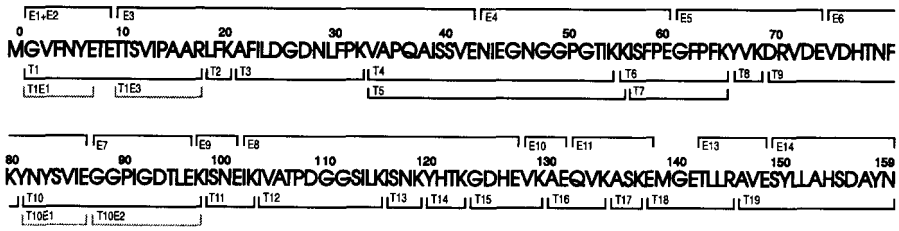

FIG. 4. **Proteolytic peptides of Bet v 1a identified by PDMS.** Peptides were generated by trypsin (*T1–T19*), endoproteinase Glu-C (*E1–E14*), or a combination of the two enzymes (*T1E1–T10E2*).

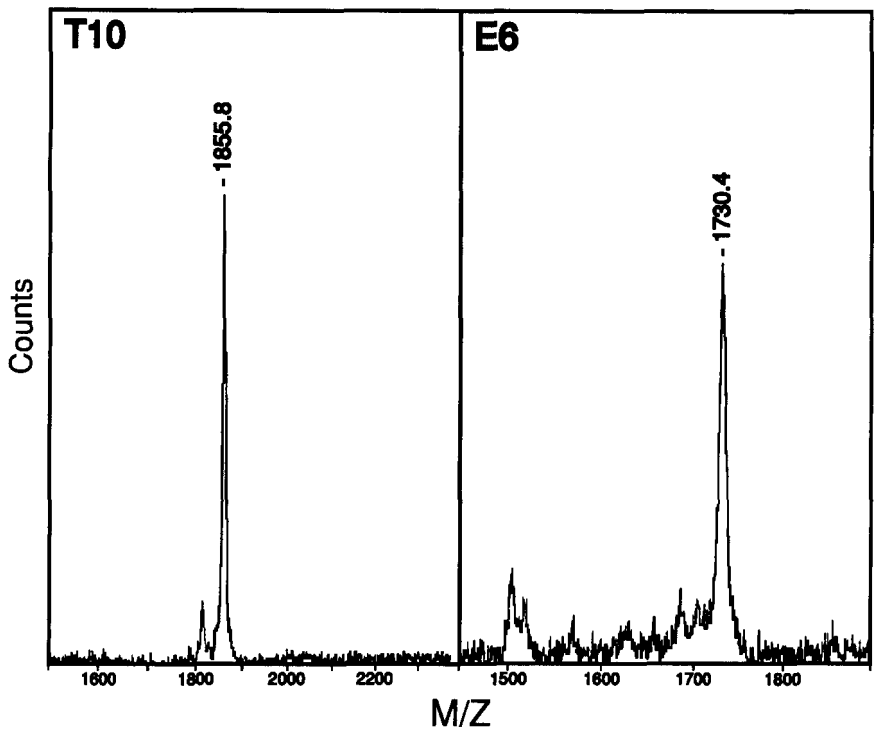

FIG. 5. **PDMS of HPLC-purified T10 and E6 peptides.** Typical mass spectra of peptides T10 and E6, which correspond to trypsin and endoproteinase Glu-C cleavage peptides of nBet v 1 containing the only potential *N*-glycosylation site of the Bet v 1a sequence. Predicted masses for unmodified T10 and E6 were 1856 and 1729, respectively.

T16, and one tripeptide, T17). Additional coverage of the rBet v 1a sequence was achieved by digestion with endoproteinase Glu-C, which produced the 13 peptides listed in Table I. Altogether, 100% of the primary structure of rBet v 1a was confirmed. It should be emphasized that all peptides detected by PDMS of rBet v 1a digests were also detected in nBet v 1 digests.

Next, we attempted to analyze peptides originating from nBet v 1, which could not be assigned to the Bet v 1a sequence by their molecular weight and the specificity of the enzyme. According to their molecular weight, these peptides did not correspond to well characterized autolysis products from trypsin (Vestling *et al.*, 1990). Another possible explanation is that these unmatched mass signals could be caused by peptides carrying postsynthetical modifications. However, the differences between any of the observed mass values and the masses of predicted proteolytic peptides were not consistent with the presence of common post-translational modifications, such as methylation, acetylation, phosphorylation, or *O*-glycosylation. We speculated that they might have originated from isoforms of Bet v 1a. Hence, the data were specifically searched for signals corresponding to proteolytic peptides predicted from the amino acid sequences deduced from the 13 Bet v 1 cDNA sequences obtained in the present study (Bet v 1b-n). In this way, 22 $(M+H)^+$ ions in tryptic digests and 11 in endoproteinase Glu-C digests could be matched with predicted peptides of Bet v 1 isoforms (see Table I). Two of those matched peptides, E5 $(M+H)^+ = 1554$ and E7 $(M+H)^+ = 845$, were sequenced by Edman degradation, confirming these assignments. In total, these matched peptides covered approximately 83–91% of the amino acid exchanges in Bet v 1b, c, k, and m/n; 60–70% for Bet v 1j, f/i, e, and d/h; and 44 and 57% for Bet v 1 g and Bet v 1l, respectively.

Finally, peptides T19–3/E14–3 and T19–5/E14–5 (see Table I) suggested the existence of truncated Bet v 1 isoforms, missing 3 or 5 amino acids at the C terminus, respectively. This region shows 100% sequence identity in all cDNA clones. Interestingly, these shorter peptides were not detected in proteolytic digests of rBet v 1a, but only in preparations of nBet v 1. In this case, we estimated that less than 30% of nBet v 1 consisted of truncated forms (based on relative peak heights of the signals corresponding to truncated *versus* intact peptides), very likely because of proteolysis during the pollen extraction procedure.

DISCUSSION

The aim of the present paper was 2-fold: First, we tried to confirm at the protein level the deduced amino acid sequence of Bet v 1a (formerly referred to as Bet v 1), the major allergen of birch pollen (Breiteneder *et al.*, 1989), and of several other closely related isoallergens that also occur in pollen and whose sequences are presented here. As shown here and by others (Tsarbopoulos *et al.*, 1988; Pedersen *et al.*, 1993), the remarkable detection limit (about 10–100 pmol) and resolving power of PDMS is sufficient to confirm a protein sequence previously determined by cDNA sequencing and to discriminate between closely related isoforms of proteins. Moreover, it is possible to roughly estimate the relative amounts of the isoproteins in the natural mixture by comparing peak areas of isopeptides, which points to the fact that the isoform Bet v 1a represents at least

50% of the total mass of pollen Bet v 1.

Second, it was our aim to investigate any possible postsynthetic modifications on rBet v 1a or on the natural mixture of isoallergens obtained from commercially available birch pollen. The correct postsynthetic modifications of a recombinant protein are of utmost importance if such a protein is to be used for the diagnosis and treatment of human disease, as is the case with recombinant allergens. As shown in several cases, such modifications can strongly influence the immunological properties of proteins (Nilsen *et al.*, 1991; Batanero *et al.*, 1994), and therefore, the present study is closely connected with a previous study in which we investigated the immunological equivalence of rBet v 1a with nBet v 1 by enzyme-linked immunosorbent assay competition experiments (Ferreira *et al.*, 1993). The purified rBet v 1a used here revealed all of the predicted peptides (Fig. 2 and Table I) but no additional peaks. The only postsynthetic modification observed was cleavage of the N-terminal methionine with nearly 100% efficiency, as was expected.

nBet v 1 from birch pollen was purified to electrophoretic homogeneity by immunoaffinity chromatography and HPLC. After purification it still showed the same complex pattern of spots in two-dimensional IgE immunoblots as was seen in the starting material (data not shown). Our interpretation is that no immunoreactive material (and, therefore, no specific isoform) was lost in the purification procedure. All of the mass peaks (with one single exception) obtained from nBet v 1 after proteolytic digestion were either identical with the ones obtained from purified rBet v 1a or could be explained by peptides predicted to arise from the isoform sequences presented here. Isoforms b, c, k, and m/n were so similar that we could only confirm them as a group. Isoform j was so similar to Bet v 1a that it could not be discriminated from it. Isoform l did not lead to any diagnostic peptide discriminating it from all other isoforms, and therefore, we cannot be sure that it really exists at the protein level. It should be considered that some of the sequence differences observed between the Bet v 1 isoforms might have originated from PCR artifacts. However, this seems unlikely since most of the cDNA sequences obtained by PCR were also confirmed individually at the protein level by PDMS analysis (see Table I and Fig. 1).

In the mass spectra of nBet v 1 no peaks were found indicating *N*-glycosylation, *O*-glycosylation, phosphorylation, methylation, or acetylation of the cleavage peptides. This result is in good agreement with earlier work showing that Bet v 1 is a cytoplasmic protein located at or near the place of ER-bound ribosomes in dry pollen (Grote, 1991), since cytoplasmic proteins are frequently unmodified and never have been found to be *N*-glycosylated (Hirschberg and Snider, 1987). Phosphorylation would have explained the charge differences of nBet v 1 in two-dimensional immunoblots (Rohac *et al.*, 1991) but does not seem to occur. It was previously shown that phosphopeptides are well detected by PDMS (Craig *et al.*, 1991). Because of the absence of covalent modifications, the production of rBet v 1 in a form that is immunologically (and conformationally) similar to nBet v 1 in *E. coli* is greatly facilitated.

Finally, comparison of the 14 cDNA sequences showed that three pairs of sequences (f/i, d/h, and m/n) are each coding for the same protein and are different only through silent exchanges. The 3′-noncoding regions within the d/h pair are nearly identical, and therefore, we assume that these sequences represent alleles of the same gene locus. This is conceivable, since the birch pollen used in this study for mRNA and protein extraction was obtained from a variety of different trees. Sequences a and j are probably not allelic, since they show relatively large insertions, deletions, and sequence deviations in their 3′-noncoding regions. However, it is not generally possible to discriminate with certainty between allelic variants and different isoforms by comparison of cDNA sequences. For this, genomic sequences and restriction maps would be needed.

Acknowledgment—We thank Peter Briza for critically reading the manuscript.

REFERENCES

Batanero, E., Villalba, M., and Rodriguez, R. (1994) *Mol. Immunol.* **31,** 31–37
Bowles, D. J. (1990) *Ann. Rev. Biochem.* **59,** 873–907
Breiteneder, H., Pettenburger, K., Bito, A., Valenta, R., Kraft, D., Rumpold, H., Scheiner, O., and Breitenbach, M. (1989) *EMBO J.* **8,** 1935–1938
Breiteneder, H., Ferreira, F., Reikerstorfer, A., Duchene, M., Valenta, R., Hoffmann-Sommergruber, K., Ebner, C., Breitenbach, M., Kraft, D., and Scheiner, O. (1992) *J. Allergy Clin. Immunol.* **90,** 909–917
Breiteneder, H., Ferreira F., Hoffmann-Sommergruber, K., Ebner, C., Breitenbach, M., Rumpold, H., Kraft, D., and Scheiner, O. (1993) *Eur. J. Biochem.* **212,** 355–362
Chomczynski, P., and Sacchi N. (1987) *Anal. Biochem.* **162,** 156–159
Craig, A. G., Engström, A., Bennich, H., Hoffmann-Posorske, E., and Meyer, H. (1991) *Biol. Mass Spectrom.* **20,** 565–574
Crowl, D. N., Maliyakal, E. J., Russel, D., and Amasino, R. M. (1992) *Plant Mol. Biol.* **18,** 459–466
Ferreira, F. D., Hoffmann-Sommergruber, K., Breiteneder, H., Pettenburger, K., Ebner, C., Sommergruber, W., Steiner, R., Bohle, B., Sperr, W. R., Valent, P., Kungl, A. J., Breitenbach, M., Kraft, D., and Scheiner, O. (1993) *J. Biol. Chem.* **268,** 19574–19580
Fristensky, B., Horovitz, D., and Hadwiger, L. A. (1988) *Plant Mol. Biol.* **11,** 713–715
Grote, M. (1991) *J. Histochem. Cytochem.* **39,** 1395–1401
Hirschberg, C. B., and Snider, M. D. (1987) *Ann. Rev. Biochem.* **56,** 63–87
Ipsen, H., and Hansen, O. C. (1989) in *Epitopes of Atopic Allergens* (Sehon, A. H., Kraft, D., and Kunkel, G., eds) pp. 3–8, Union chimique Belge Institute of Allergy, Brussels, Belgium
Ipsen, H., and Loewenstein, H. (1983) *J. Allergy Clin. Immunol.* **72,** 150–159
Jacobs, S. (1959) *Nature* **183,** 262
Jarolim, E., Rumpold, H., Endler, A. T., Ebner, H., Breitenbach, M., Scheiner, O., and Kraft, D. (1989a) *Allergy* **44,** 385–395
Jarolim, E., Tejkl, M., Rohac, M., Schlerka, G., Scheiner, O., Kraft, D., Breitenbach, M., and Rumpold, H. (1989b) *Int. Arch. Allergy Appl. Immunol.* **90,** 54–60
Laemmli, U. K. (1970) *Nature* **227,** 680–685
Larsen, J. N., Ströman, P., and Ipsen, H. (1992) *Mol. Immunol.* **29,** 703–711
Matton, D. P., and Brisson, N. (1989) *Mol. Plant-Microbe Interact.* **2,** 325–333
Nielsen, P. F., Klarskov, K., Hojrup, P., and Roepstorff, P. (1988) *Biomed. Environ. Mass Spectrom.* **17,** 355–362
Nilsen, B. M., Sletten, K., Paulsen, B. S., O'Neill, M., and van Halbeek, H. (1991) *J. Biol. Chem.* **266,** 2660–2668
Pedersen, J., Filimonova, M., Roepstorff, P., and Biedermann, K. (1993) *Biochim. Biophys. Acta* **1202,** 13–21
Rohac, M., Birkner, T., Reimitzer, I., Bohle, B., Steiner, R., Breitenbach, M., Kraft, D., Scheiner, O., Gabl, G., and Rumpold, H. (1991) *Mol. Immunol.* **28,** 897–906
Sanger, F., Nicklen, S., and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467
Somssich, I. E., Schmelzer, E., Kawalleck, P., and Hahlbrock, K. (1988) *Mol. & Gen. Genet.* **213,** 93–98
Tsarbopoulos, A., Becker, G. W., Occolowitz, J. L., and Jardine, I. (1988) *Anal. Biochem.* **171,** 113–123
Valenta, R., Breiteneder, H., Pettenburger, K., Breitenbach, M., Rumpold, H., Kraft, D., and Scheiner, O. (1991) *J. Allergy Clin. Immunol.* **87,** 677–682
Vestling, M. M., Murphy, C. M., and Fenselau, C. (1990) *Anal. Chem.* **62,** 2391–2394
Walter, M. H., Liu, J. W., Grand, C., Lamb, C. J., and Hess, D. (1990) *Mol. & Gen. Genet.* **222,** 353–360
Warner, S. A. J., Scott, R., and Draper, J. (1992) *Plant Mol. Biol.* **19,** 555–561