

Genomic characterization of members of the Bet v 1 family: genes coding for allergens and pathogenesis-related proteins share intron positions

Karin Hoffmann-Sommergruber ^{a,*}, Monika Vanek-Krebitz ^a, Christian Radauer ^a, Jun Wen ^b,
Fatima Ferreira ^c, Otto Scheiner ^a, Heimo Breiteneder ^a

^a Institute of General and Experimental Pathology, AKH-EBO-3Q, University of Vienna, Währinger Gürtel 18–20,
A-1090 Vienna, Austria

^b Department of Biology, Colorado State University, Fort Collins, Colorado, USA

^c Institute of Genetics and General Biology, University of Salzburg, Salzburg, Austria

Received 27 December 1996; accepted 4 April 1997

Abstract

Bet v 1, the major birch pollen allergen, is a member of a multigene family; a number of isoforms and homologous proteins from closely related species (alder, hazel and hornbeam) has been isolated and their cDNAs cloned and characterized. Genomic clones coding for Bet v 1 and homologues from apple and hazel were isolated and sequenced. Some of these clones contained intervening sequences. The exon–intron formation is highly conserved throughout this family of pathogenesis-related proteins in dicot plants and is also found in Aoprl (*Asparagus officinalis*), a monocot species. Phylogenetic analysis suggested a possible common origin of the intron position in these homologous proteins at codon 62 in various families of flowering plants, including Fagaceae, Rosaceae and Apiaceae. This conserved ‘proto-splice site’ may point to a structure/function relationship. A conserved sequence motif (P-loop) was also found in all members of this protein family. Moreover, there is a certain degree of sequence similarity among the proteins derived from various species throughout the dicots and the only monocot examined. This fact is reflected by cross-reactivity from monoclonal and polyclonal antibodies raised against Bet v 1. © 1997 Elsevier Science B.V.

Keywords: Allergen; Bet v 1; Cor a 1; Mal d 1; Cross-reactivity; Genomic clone; Intracellular pathogenesis-related protein; Intron

1. Introduction

In the past two decades, there has been an ongoing debate on the origin, evolution and significance of spliceosomal introns. The issue became polarized into two opposing theories. The ‘introns-early’ theory or ‘exon theory of genes’, postulates the presence of many introns in the genome of the common ancestor of life, followed by (sometimes incomplete) loss of introns in many independent lineages. There, introns were thought to direct the recombinatorial shuffling of short exons of

15–20 amino acid units of protein sequences. A crucial prediction of this theory was the definition of functional units within proteins that are delimited by intron insertion and redefined as a supersecondary structure of compact modules (Palmer and Logsdon, 1991; Gilbert and Glynias, 1993; Mattick, 1994).

The contradictory theory—intron-late—proposes that introns were not present in ancestral proteins, but arose during eukaryotic evolution.

A number of ancient and well characterized proteins

* Corresponding author. Fax: +43 1 404005130;
e-mail: Karin.Hoffmann@akh-wien.ac.at

Abbreviations: Aoprl, wound-induced protein from asparagus; Bet v 1, major birch pollen allergen; Bet v 1-Sc1–3, Bet v 1-encoding cDNAs expressed in cell culture upon pathogen exposure; BIP1, murine monoclonal anti-Bet v 1 IgG; bp, base pair(s); *bvgc*, genomic clones encoding Bet v 1-sequences; *cagc*, genomic clones encoding Cor a 1-sequences; cDNA, DNA complementary to RNA; Cor a 1, major

hazel pollen allergen; dNTP(s), desoxynucleosidetriphosphate(s); IgE, immunoglobulin E; IgG, immunoglobulin G; IPR, intracellular pathogenesis-related protein; kDa, kilodalton; Mal d 1, major apple allergen; *mdgc*, genomic clones encoding Mal d 1-sequences; mRNA, messenger RNA; NHS, normal human serum; PAGE, polyacrylamide-gel electrophoresis; PcPR1, pathogen-inducible protein from parsley; PCR, polymerase chain reaction; PI49, pathogen-inducible protein from pea; SDS, sodium dodecylsulfate; STH2, pathogen- and stress-inducible proteins from potato; *Taq* polymerase, polymerase from *Thermus aquaticus*.

like triose phosphate isomerase, glyceraldehyde-3-phosphate dehydrogenase and alcohol dehydrogenase has been used as models for the analysis of intron positions providing inconclusive data or data of inconsistent interpretation (Stone and Schwartz, 1990; Gilbert and Glynias, 1993; Kersanach et al., 1994; Logsdon et al., 1995).

One experimental approach could be the analyses of intron positions in homologous genes from different species and even phyla in the eukaryotes on the hypothesis that a common position reflects the situation in the progenote, if present there. However, one could also argue that matching intron positions would account for preferential sites for intron insertion (proto-splice site). In this context, the introns-early and introns-late theory may appear unlinked (Mattick, 1994).

Bet v 1, the major birch pollen allergen, a 17-kDa protein, belongs to the class of intracellular pathogenesis-related proteins (IPRs) on the basis of sequence similarity (Breiteneder et al., 1989). It was the first published cDNA-derived sequence from a group of proteins with high sequence similarities from species of the order Fagales (alder, hazel and hornbeam (Breiteneder et al., 1992; Larsen et al., 1992; Breiteneder et al., 1993)). Moreover, Bet v 1-homologous proteins were identified in a broad range of plant derived food, such as apple and celery, providing molecular evidence for cross-reactivity of tree pollen allergy and allergy to food of plant origin (Breiteneder et al., 1995; Ebner et al., 1995; Vanek-Krebitz et al., 1995).

A subset of sequences from the Bet v 1 family, isolated from a cDNA library, different from those that are constitutively expressed in mature pollen, encodes sequences that are transcriptionally activated by microbial pathogens (Swoboda et al., 1994). They were designated Bet v 1-Sc1, Sc2 and Sc3 and revealed sequence identities (at the amino acid level) of 70–89% to the different pollen-derived Bet v 1 isoforms (Swoboda et al., 1995a,b).

Studies of the expression of the corresponding sequences in response to microbial attack revealed that Bet v 1-Sc-mRNA was rapidly induced in a model system of cultured birch cells upon interaction of cells with bacteria or fungi (Swoboda et al., 1994).

The observation that the expression of Bet v 1 genes in pollen grains was regulated differently to the inducible forms in tissue cultures implies that different functions of proteins might exist within one family. Therefore, data from the genomic organization of this gene family within one species and from distantly related species could contribute to conclusions about the diversity or conservation of functions.

Here, we provide data on the exon–intron formation of genomic clones of the Bet v 1 family derived from birch, hazel and apple, compare them to the genomic

clones of the IPR-protein family known from the literature and compare the differential antibody binding patterns in immunoblots.

2. Materials and methods

2.1. Cloning and sequencing of betv1, coral and mald1 genes

Total genomic DNA was extracted from leaves of birch (*Betula verrucosa*), hazel (*Corylus avellana*) and apple (*Malus domestica* cv. Gala; Draper and Scott, 1988). According to the known N-terminal and C-terminal sequences of Bet v 1 (Breiteneder et al., 1989), synthetic oligonucleotides (HB 20, HB 22) of 27-bp length (5'ATG GGT GTT TTC AAT TAC GAA ACT GAG-3') and 31-bp length (5'TTA GTT GTA GGC ATC GGA GTG TGC CAA GAG G-3'), respectively, were designed. Bet v 1-homologous sequences were amplified from genomic DNA with 2.5 units of *Taq* polymerase (Perkin Elmer Cetus, Norwalk, CT, USA); 2.5–4.5 mM MgCl₂, 0.2 mM of each dNTP and with 0.2 μM of each primer. Amplification was carried out as described previously (Breiteneder et al., 1993). Amplified PCR-products were gel purified using a silica matrix (GeneClean II Kit, Bio101, La Jolla, CA, USA) and cloned into the PCR[®]II-plasmid. Competent *E. coli* INVαF' cells (TA cloning[®] system, Invitrogen, San Diego, CA, USA) were transformed with the resulting plasmids.

Clones coding for Bet v 1-homologous sequences were isolated, and both strands were sequenced according to the dideoxy chain termination method using a T7-Sequencing Kit (Promega, Madison, WI, USA).

2.2. Protein extracts from apple, asparagus, parsley, pea and potato

Aqueous protein extracts from apple (*Malus domestica* cv. Gala), asparagus (*Asparagus officinalis*), parsley (*Petroselinum crispum*), pea (*Pisum sativum*) and potato (*Solanum tuberosum*) were prepared as described previously (Ebner et al., 1995).

2.3. SDS-PAGE and immunoblots

Proteins from the extracts were separated by SDS-PAGE in 12% acrylamide gels and either stained with Coomassie Brilliant Blue R-250 or transferred on to nitrocellulose membranes (Schleicher and Schuell, Dassel, Germany). Subsequently, blots were incubated with either mouse monoclonal anti-Bet v 1-IgG (BIP 1. (Jarolim et al., 1989)) or rabbit polyclonal anti-recombinant Bet v 1-IgG, followed by rabbit anti-mouse-IgG and ¹²⁵I-donkey anti-rabbit-IgG. For IgE-immunoblots

a serum pool of birch pollen allergic patients was used. Detection of bound IgE was performed with ^{125}I -rabbit anti-human-IgE (Pharmacia, Uppsala, Sweden). A serum pool of non-allergic individuals and adequate buffer controls was tested in parallel.

2.4. Phylogenetic analysis

Parsimony analyses of the amino acid sequences of BVGC34, MDGC32, CAGC10, PI49, PcPR1, STH2 and Aopr1 were performed with PAUP (Phylogenetic Analysis Using Parsimony version 3.1, (Swofford, 1993)) to reconstruct a gene tree. This gene tree was used to examine the evolution of intron positions. The parsimony analyses employed a branch and bound search, and the tree was rooted using the sequence from the only monocot species (Aopr1 from *Asparagus*). Bootstrap analysis (Felsenstein, 1985) was performed with 1000 replicates to show the relative support of the monophyletic groups.

3. Results

3.1. Analysis of cloned betv1, mald1 and coral genes

3.1.1. betv1 genes

Three PCR-products differing in length (483 bp, 572–584 bp and 636 bp) were obtained from genomic DNA. Sequencing revealed four different clones *bvgc11*, *bvgc181*, *bvgc31* and *bvgc63* (GenBank/EMBL Databank accession Nos. Z72429, Z72437, Z72432, Z72435) of 483 bp in length with an open reading frame of 160 amino acid residues, five clones *bvgc21*, *bvgc25*, *bvgc45*, *bvgc70* and *bvgc681* (accession Nos. Z72430, Z72431, Z72434, Z72436, Z72438) with one intron of 89–101 bp length and one clone *bvgc34* (accession No. Z72433) containing an intron of 153 bp in length (Fig. 1). All these introns were positioned at amino acid 62, being inserted between nucleotides 1 and 2. All sequences at the 5' and 3' splice sites were in accordance with the known motifs for plant introns 5' AG:GT–AG:GC 3', and the intron sequences displayed a high AU content (Fig. 1).

With the exception of a single nucleotide exchange, clones *bvgc11* (at position 31 cytosine instead of adenine) and *bvgc31* (position 28 guanine instead of adenine) were identical to the pollen-derived cDNA-clone Bet v 1a (99.8% sequence identity (Breiteneder et al., 1989)). Amino acid sequences encoded by clones *bvgc25*, *bvgc45* and *bvgc34* showed high similarity to isoforms Bet v 1b, c, k and m (98% sequence identity (Swoboda et al., 1995a)). Clone *bvgc21* revealed sequence similarity to isoform Bet v 1j [95% identity (Swoboda et al., 1995a)],

whereas clones *bvgc18* and *bvgc63* resembled the pathogen inducible isoform Bet v 1-Sc1 isolated from suspension cultured birch cells (90% and 91% sequence identity, respectively (Swoboda et al., 1995b)).

3.1.2. mald1 genes

PCR products of 480 bp and 594 bp were obtained. Sequencing identified one clone (*mdgc10*, accession No. Z72425) of 480 bp and deduced 159 amino acid residues and three clones of 594 bp in length and deduced 160 amino acid residues (*mdgc32*, accession No. Z72426; *mdgc36*, accession No. Z72427; *mdgc38*, accession No. Z72428). These three clones contained a single intron positioned at amino acid residue 62 inserted in between nucleotides 1 and 2 (AG:GT and AG:GC) of 111 bp in length (Fig. 1) with the known plant splice sites. Sequence comparison of the coding regions of *mdgc10*, *mdgc36*, *mdgc38* and *mdgc32* revealed 75–79% identity to MDAP (Atkinson et al., 1996) and 73–76% identity to Mal d 1 (Vanek-Krebitz et al., 1995).

3.1.3. coral genes

PCR products of 564 bp and 577 bp in length were obtained when performing PCR with primers HB20 and HB22. Sequencing revealed that each coded for 160 amino acid residues plus one intron positioned at codon 62, inserted between nucleotides 1 and 2, of 81 bp in length (clone *cagc10*, accession No. Z72439) and of 94 bp in length (clone *cagc11*, accession No. Z72440). All introns were delimited by the 5' and 3' splice site of AG:GT and AG:GC. Clone *cagc10* showed 86% sequence identity at the DNA-level to the known cDNA-derived isoform Cor a 1/11 (Breiteneder et al., 1993) and 87.4% identity to the pathogen inducible isoform Bet v 1-Sc1 (Swoboda et al., 1995b), whereas clone *cagc11* revealed 76% identity to Cor a 1/11, and 78.3% identity to Bet v1-Sc3.

3.1.4. Genomic analysis of intron position and length of members of the IPR-family

Genomic clones obtained by PCR from DNA of apple, hazel and birch coding for Bet v 1 and Bet v 1-related sequences displayed a coding region of 160 amino acid residues for birch, hazel, and apple (exception: *mdgc10*, 159 amino acid residues; Fig. 1). Introns were at identical positions at codon 62 inserted between nucleotides 1 and 2 (phase 1, Fig. 1B). All intervening sequences showed the reported 5' and 3' splice sites (Simpson et al., 1993) and high AU content ranging from 69% (*cagc10*) to 79% (*bvgc21*), which is characteristic of plant introns (Goodall and Filipowicz, 1989; Fig. 1A,B).

Sequences coding for IPR-proteins reported in the literature harboured intervening sequences inserted at a

A	
	1 50
<i>mdgc32</i>	GTAAATAAC CCTCATTACT CTCCTCCTCT GTTCGTAT..
<i>cagc11</i>	GTAA.TTTC TACCACAATA ATCTCCAAGG AGTTGTAC..
<i>bvvc21</i>	GTAA.TTAA TTACTTTCTT AATT....AA AAGCA.....
<i>bvvc681</i>	GTAA.TTAA TTACTTTCTT AATT....AA AAGCA.....
<i>bvvc70</i>	GTAA.TTAA TTACTTT.TC AATT....AA AAGCA.....
<i>bvvc25</i>	GTTAG.TTGA TTACTTTCTT AATTAATTAA AAGCAC.....
<i>bvvc45</i>	GTTAG.TTAA TTACTTTCTT AATTAATTAA AAGCAC.....
<i>bvvc34</i>	GTTAG.TTAA TTACTTTCTT AATTAATTAA AAGCACACAC ACACACACAC
<i>cagc10</i>	GTAAATTTAA TTACTTTCTA TCCTATAT..
	51 100
<i>mdgc32</i>	GTTTTACAC TTGAATGAAC
<i>cagc11</i>	GTTTTGTGT TTCATAGCAT
<i>bvvc21</i>	TC ATATTTATTT
<i>bvvc681</i>	TC ATATATATTT
<i>bvvc70</i>	TC ATATATATTT
<i>bvvc25</i>	AT ATATATATGT
<i>bvvc45</i>	AT ATATATATGT
<i>bvvc34</i>	ACACACACAC ACACACATAT ATATATATAT ATATATATAT ATATATATGT
<i>cagc10</i>	TCTCC
	101 150
<i>mdgc32</i>	TCAAAATTAT TAATGACTTT TTTTGTACTA ATTTATGTTT TAATTCATTT
<i>cagc11</i>	GCA..... .T TATATAGATA CATTATTTTG TCAATATTTT
<i>bvvc21</i>	ATTTATGTAG GTATATATTG ATCATGTGCT TTTTTTGTGA ...TACCTT
<i>bvvc681</i>	ATATATGTAG GTATATATTG ATCATGTGCT TTTTTTGTGA ...TACCCT
<i>bvvc70</i>	ATATATGTAG GAATATATTG ATCATGTGC. .TTTTTGTGA ...TTCCCT
<i>bvvc25</i>	ATATATGTAG GTATATATTG ATCATGTGCT TAATTTTGTGA TACGTACCCT
<i>bvvc45</i>	ATGTATGTAG GTATATATTG ATCATGTGCT TAATTTTGTGA TACGTACCCT
<i>bvvc34</i>	ATGTATGTAG GTATATATTG ATCATGTGCT TAATTTTGTGA TACGTACCCT
<i>cagc10</i>	AACGCATAAT TCATATAACT TACTTGTGCT ATTTTCTATA T.....CCT
	151
<i>mdgc32</i>	GCAG
<i>cagc11</i>	GCAG
<i>bvvc21</i>	TTAG
<i>bvvc681</i>	TCAG
<i>bvvc70</i>	TCAG
<i>bvvc25</i>	TCAG
<i>bvvc45</i>	TCAG
<i>bvvc34</i>	TCAG
<i>cagc10</i>	ACAG

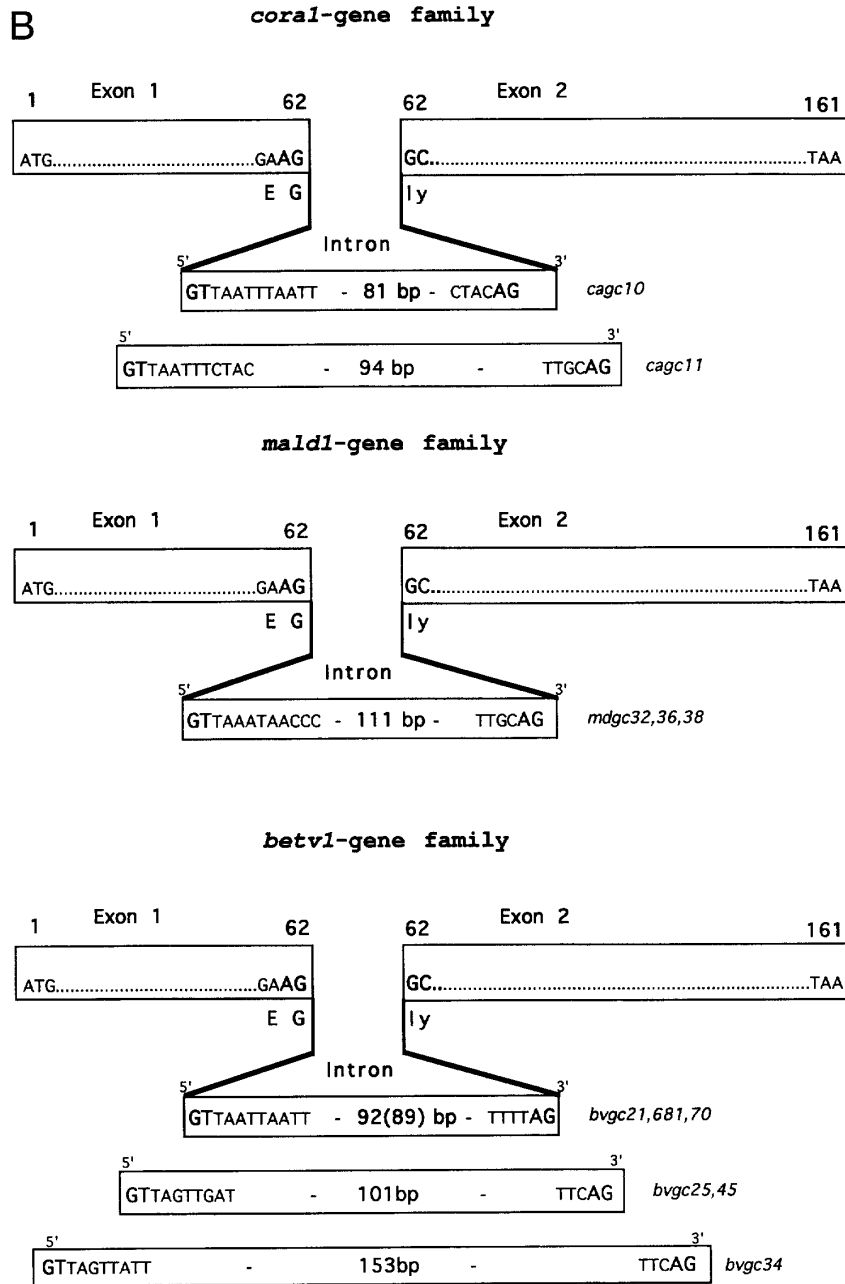
Fig. 1. Intron alignment. (A) Intron sequences from birch (*bvvc21*, 25, 34, 45, 70, 681), apple (*mdgc32*, 36, 38) and hazel (*cagc10*, 11) are aligned. gaps are indicated by dots (intron sequences of *mdgc32*, *mdgc36*, and *mdgc38* were identical). (B) Intron position in the *coral*, *mald1* and *betv1* gene family.

highly conserved position (Fig. 2). It is codon 62 for parsley (Ala; PcPR1 (van de Löcht et al., 1990)), and for hazel, birch and apple (Gly; this paper), codon 61 for pea (Asp; PI-49 (Chiang and Hadwiger, 1990)) and codon 57 for potato (Gly; STH-2 (Matton et al., 1993)). Even the monocot asparagus displays an intron positioned at codon 63 (Ala; AoPR1 (Warner et al., 1993)).

There is a striking conserved sequence motif GXGXXG at amino acid residues 47–52 throughout the family of IPR-proteins (Fig. 2); the exception is STH-2, where it is reduced to GXG (amino acid posi-

tions 45–47). This motif is flanked by conserved lysine residues 14 amino acid residues upstream, as well as 18 amino acid residues down-stream (exception PcPR1 upstream and PI49 downstream; Fig. 2). This motif is known as 'P-loop' (phosphate binding loop) and is frequently found with variations in protein kinases as well as in nucleotide binding proteins (Saraste et al., 1990).

The Bet v 1a sequence was analyzed using a dot plot program, but no evidence of a duplication of the sequence was found (data not shown).



3.1.5. Phylogenetic analysis of *Bet v 1*-homologous sequences and sequences of *IPR*-proteins

Parsimony analysis generated two maximally parsimonious trees with a length of 359 steps, a consistency index (CI) of 0.903 (0.784 excluding uninformative characters), retention index (RI) of 0.514 and rescaled consistency index of 0.464 (Fig. 3). The two trees differed only in the

relative position of STH-2 (*Solanum*) and PcPR1 (*Petroselinum*): STH-2 basal or PcPR1 basal among the dicot species. A larger parsimony analysis using all the available *Bet v 1*-homologous sequences (67 sequences from eight different plant families) revealed a basal position of STH-2 (Wen et al., 1997). The tree with STH-2 basal was therefore preferred (Fig. 3).

3.1.6. SDS-PAGE analysis of protein extracts and immunoblots

Immunoblots displayed IgE-binding from birch pollen allergic patients to Bet v 1 (Fig. 4A: lane 6) and its dimer (34 kDa) as well as to homologous proteins from parsley (lane 5), potato (lane 3) and apple (lane 1). A different binding profile revealed the Western blot performed with the monoclonal anti-Bet v 1 antibody (panel B) recognizing Bet v 1 (lane 6), the homologous protein from parsley with reduced binding intensity (lane 5) and Mal d 1 (lane 1). The polyclonal anti-Bet v 1 antibody (panel C) showed binding to Bet v 1 and the respective homologous proteins from parsley (lane 5), asparagus (lane 4), potato (lane 3), pea (lane 2) and apple (lane 1).

4. Discussion

Since their discovery in 1977 (Williamson, 1977), introns have been classified into three groups according to the mechanisms of excision and intron features. Nuclear pre-mRNA introns are understood to have evolved from group II introns (Jaquier, 1990; Palmer

and Logsdon, 1991). Nevertheless, the understanding of sequences not coding for proteins is still incomplete.

There is a close correlation between intron density and developmental complexity, according to Palmer and Logsdon (1991), ranging from one intron per kb of coding sequence in simpler eukaryotes, rising to three or four introns per kb in plants and fungi, and up to six introns per kb in animals.

So far, introns have been assumed to display higher sequence variations between species than their flanking exons (Mattick, 1994). However, there are reports of unexpected high sequence similarities of introns. For example, the single intron of the oligodendrocyte-myelin glycoprotein from mouse and human displays 75% sequence identity overall and, in one region, over 90% (Mikol et al., 1993); sequence cassettes from the third intron of the γ -actin gene are highly conserved between human and *Xenopus* (Erba et al., 1988). Whether these data represent isolated cases or are part of a molecular principle remains to be investigated. Nevertheless, they may contribute to a new understanding of the complexity and functionality of introns. It seems that introns contain information, but how this information is encoded and transduced is unclear (Mattick, 1994).

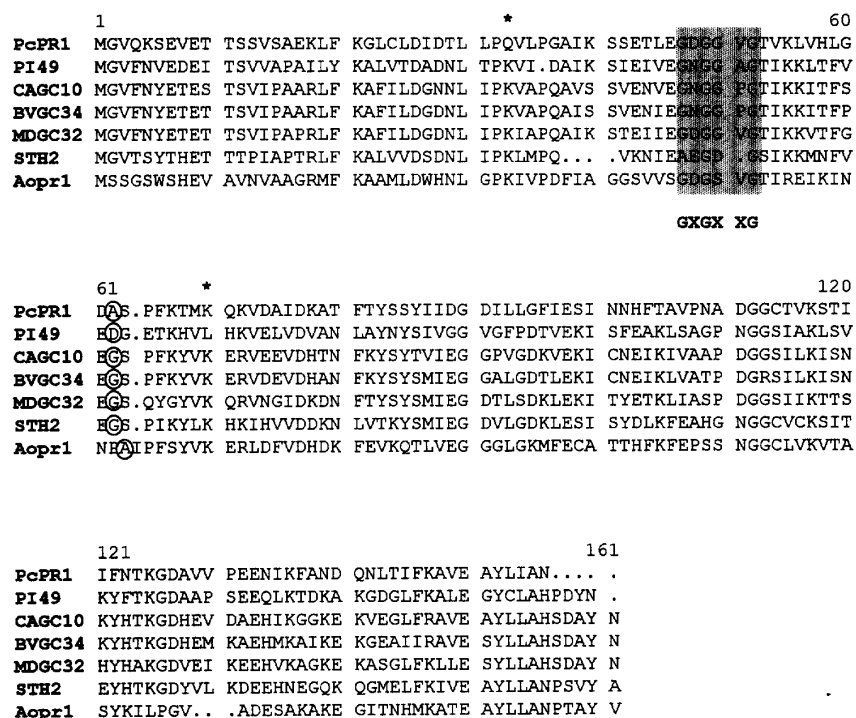


Fig. 2. Alignment of amino acid sequences from genes of the PR-family containing introns compared to BVGC, MDGC and CAGC clones. Intron positions are indicated as ellipses where they interrupt codons (phase 1). A conserved motif for phosphate binding loop (P-loop; shaded region) and the flanking lysine residues are marked (*). This motif is reduced to GXGX in the case of *Solanum*. PCPR1: *Petroselinum*; PI49: *Pisum*; CAGC10: *Corylus*; BVGC34: *Betula*; MDGC32: *Malus*; STH2: *Solanum*; Aopr1: *Asparagus*.

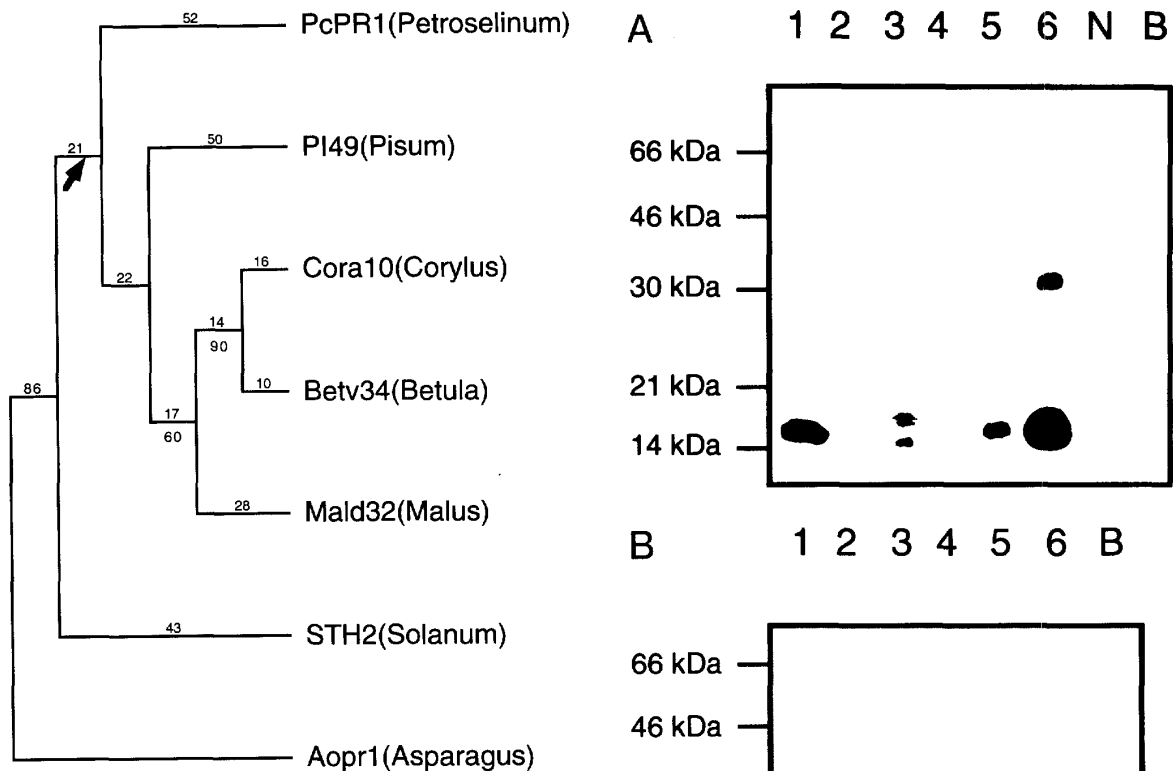
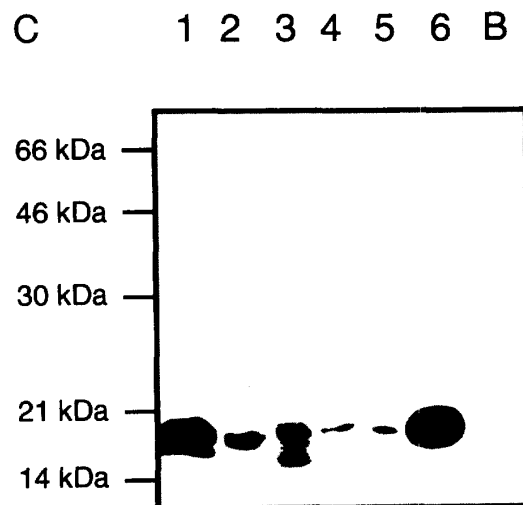


Fig. 3. One of the two maximally parsimonious trees from the analysis using the amino acid sequences from Bet v 1-homologues. The common origin of the intron position 62 in various plant groups as the most parsimonious hypothesis is marked by an arrow. The numbers above the lines indicate the branch length, and the numbers below the lines are the bootstrap values with 1000 replicates.

Possible implications are, on the one hand, the development of new molecules based on RNA-processing, like a set of related proteins originating from a single gene that can be produced via alternative splicing. On the other hand, it is feasible that introns bear sequence motifs and signatures of information leading to a model of informational RNA in parallel to mRNA. Thus, communication at the RNA level (RNA–RNA, RNA–DNA, and RNA–protein interaction) via RNA signals can be envisaged (Mattick, 1994).

Fig. 4. Western Blot. Aqueous protein extracts from apple (*Malus domestica*, lane 1) pea (*Pisum sativum*, lane 2), potato (*Solanum tuberosum*, lane 3), asparagus (*Asparagus officinalis*, lane 4), parsley (*Petroselinum crispum*, lane 5) and birch pollen (*Betula verrucosa*, lane 6) were separated by 12% SDS–PAGE and transferred to nitrocellulose sheets. Subsequently, binding of IgE from birch pollen allergic patients' sera (A), binding of monoclonal murine anti-Bet v 1 IgG (BIP 1, B) and binding of polyclonal rabbit anti-Bet v 1 IgG (C) were tested. Detection was performed by ^{125} I-labelled anti-IgE, and anti-IgG, respectively (for controls, buffer B and normal human serum NHS were included).



Another aspect concerns the differential rate of gene expression controlled by introns. Earlier observations on plants indicated that introns are not required to achieve normal expression levels, e.g. the seed storage protein phaseolin is expressed to similar amounts with or without its introns (Chee et al., 1986). Yet, there is an increasing number of reports, in particular in heterologous gene expression assays, showing that gene expression is enhanced by intron sequences (Callis et al., 1987; Luehrsen and Walbot, 1990). The mechanisms of this enhancement are not well understood. Introns may stimulate initiation of transcription, increase stability of pre-mRNA in the nucleus, or raise the rate of mRNA export to the cytoplasm (Luehrsen and Walbot, 1990).

Investigating the *betv1*-gene family for their exon–intron formation, genomic clones (*bvgc*) without intervening sequences and clones harboring one intron were found. *bvgc* clones (*bvgc21*, 25, 34, 45, 681 and 70) harboring one intron showed high sequence similarities to the isoforms constitutively expressed in mature birch pollen (95%–99% identity), whereas *bvgc* clones with no introns (*bvgc11* and *bvgc31*; 99.8% sequence identity) showed close similarity to Bet v 1 pollen isoforms as well as homology to Bet v 1-Sc sequences, the pathogen-inducible proteins (*bvgc181*, *bvgc63*; 90–91% identity to Bet v 1-Sc1 (Swoboda et al., 1995b)). These findings point to an individual pattern of gene expression for each member of this gene family.

Comparing the positions where introns were inserted, a strikingly conserved pattern can be observed. Intervening sequences were inserted at amino acid residue 62 in Bet v 1-sequences (Betulaceae), Cor a 1-sequences (Corylaceae), Mal d 1-sequences (Rosaceae) and PcPR1 (Apiaceae), amino acid residue 61 in PI49 (Fabaceae), amino acid residue 57 in STH2 (Solanaceae) and amino acid residue 63 in Aopr1 (Liliaceae). Whether it is a parallel insertion at a proto-splice site or this is due to common ancestry remains unknown. The most parsimonious explanation based on sequences (Fig. 3) is of a single origin of intron position at codon 62 (a common ancestry). This would also be the original position of the intron in the case of PI49, but changed to position 61 due to a secondary loss of one amino acid residue at position 36.

There are some concepts of introns delineating functional units from proteins called 'compact modules'. In the case of Bet v 1, this conserved conformation of the proteins is recognized by either monoclonal-anti Bet v 1 antibodies or by polyclonal rabbit anti-Bet v 1 serum or human IgE with overlapping, yet distinct binding specificities. By performing cross-inhibition assays with two monoclonal antibodies, different epitopes could be defined, distinct from B-cell-epitopes. Nevertheless, correct three-dimensional folding of the proteins was still mandatory for IgE-binding. Only full-length clones coding for Bet v 1 were picked from λ -ZAP libraries,

when screening with allergic patients' sera. The conformation seemed to be conserved in even less homologous sequences since cross-reactivity can be observed with IgE and monoclonal antibodies among distantly related species.

When sequences from different species are compared for their differences and/or similarities, conclusions about their phylogenetic distance can be drawn. *bvgc* clones and *cagc* clones are more closely related to each other and to *mdgc* clones than to clones from parsley, pea or potato.

Although an increasing number of members of this gene family has already been isolated and sequenced, only little evidence is available on their function. Expression upon pathogen induction has been shown for a few sequences (Swoboda et al., 1994). Aopr1 may be involved in the phenylpropanoid pathway, again linked to a possible reactivity upon pathogen attack (Warner et al., 1994).

A conserved motif, present in cDNA-clones as well as in genomic clones can be detected. A GXGXXG sequence at aa residues 47–52 is flanked by lysine residues 14 amino acid residues upstream, as well as 18 (15) residues downstream. This motif is common throughout all the allergen-coding sequences from the Bet v 1 family as well as in the class of intracellular pathogenesis-related proteins found so far. This sequence is characteristic for protein kinases (P-loop, Saraste et al., 1990), as well as for other nucleotide binding proteins, facilitating binding of the ribose into the turn of the glycine-rich sequence, whereas phosphate is bound to lysine. Yet, there are no experimental data corroborating this assumption. First, there is no overall pattern, common for all ATP- or GTP-binding protein P-loops. Second, homology based exclusively on this sequence motif does not necessarily mean that the candidate protein must have a kinase activity. An enzymatic function as a protein kinase has not yet been shown for a member of the Bet v 1-family. Most of the known protein kinases contain a catalytic domain of 260 amino acids with conserved motifs spread across the whole sequence. The glycine-rich sequence may be a nucleotide binding motif.

All members of this ubiquitous class of proteins are very similar in amino acid composition, protein size (approx. 17 kDa) and total charge, which suggests the homology (a common ancestry) of these sequences. A preliminary phylogenetic analysis of various members of these pathogenesis related proteins (Wen et al., 1997) shows the potential of Bet v 1-homologues as phylogenetic markers. The phylogenetic relationships revealed by Bet v 1-homologues are congruent with those from morphology (Cronquist, 1988). Comparing isoform sequences isolated from one species, it seems that they originated from one sequence, diverging from each other via duplication events—yet forming a monophyletic

group. These single amino acid exchanges result in quite different capacities in binding human IgE antibodies as shown for Bet v 1 isoforms and Cor a 1 isoforms (Breiteneder et al., 1993; Ferreira et al., 1996; Weiss et al., 1996). If single amino acid substitutions affect the three-dimensional folding of the protein, IgE-binding intensities can vary greatly. Furthermore, these changes in conformation might be one explanation why highly ubiquitous Bet v 1 proteins from species such as *P. crispum*, *S. tuberosum*, and *B. verrucosa* are able to elicit an allergic response in atopic humans, whereas the homologous proteins present in species like *A. officinalis* or *P. sativum* display rather low allergenic characteristics and cross-reactivities. This might also explain why a number of fruits and vegetables (like *D. carota*, *A. graveolens*, *P. crispum*, *M. domestica*, and *C. avellana*) induce oral allergic symptoms—whereas a range of other vegetables (like *L. esculentum*, *P. sativum*, *S. tuberosum*, *F. vesca* and *A. officinalis*) evoke little or no allergic reaction in Bet v 1-prensensitized patients.

Acknowledgement

This work was supported by grant P11410-MOB from the FWF Austria and a Career Enhancement Grant from the Colorado State University (to J.W.).

References

- Atkinson, R.G., Perry, J., Matsui, T., Ross, G.S., Macrae, E.A., 1996. A stress-, pathogenesis-, and allergen-related cDNA in apple fruit is also ripening-related. *NZ J. Crop Hortic. Sci.* 24, 103–107.
- Breiteneder, H., Pettenburger, K., Bito, A., Valenta, R., Kraft, D., Rumpold, H., Scheiner, O., Breitenbach, M., 1989. The gene coding for the major birch pollen allergen Bet v 1, is highly homologous to a pea disease resistance response gene. *EMBO J.* 8, 1935–1938.
- Breiteneder, H., Ferreira, F., Reikerstorfer, A., Duchêne, M., Valenta, R., Hoffmann-Sommergruber, K., Ebner, C., Breitenbach, M., Kraft, D., Scheiner, O., 1992. Complementary cloning and expression in *Escherichia coli* of Aln g 1, the major allergen in pollen of alder (*Alnus glutinosa*). *J. Allergy Clin. Immunol.* 90, 909–917.
- Breiteneder, H., Ferreira, F., Hoffmann-Sommergruber, K., Ebner, C., Breitenbach, M., Rumpold, H., Kraft, D., Scheiner, O., 1993. Four recombinant isoforms of Cor a 1, the major allergen of hazel pollen, show different IgE-binding properties. *Eur. J. Biochem.* 212, 355–362.
- Breiteneder, H., Hoffmann-Sommergruber, K., O'Riordain, G., Susani, M., Ahorn, H., Ebner, C., Kraft, D., Scheiner, O., 1995. Molecular characterization of Api g 1, the major allergen of celery (*Apium graveolens*), and its immunological and structural relationships to a group of 17-kDa tree pollen allergens. *Eur. J. Biochem.* 233, 484–489.
- Callis, J., Fromm, M., Walbot, V., 1987. Introns increase gene expression in cultured maize cells. *Genes Dev.* 1, 1183–1200.
- Chee, P.P., Klassy, C., Slightom, J., 1986. Expression of a bean storage protein 'phascolin minigene' in foreign plant tissues. *Gene* 41, 47–57.
- Chiang, C.C., Hadwiger, L.A., 1990. Cloning and characterization of a disease resistance response gene in pea inducible by *Fusarium solani*. *Mol. Plant-Microbe Interact.* 3, 78–85.
- Cronquist, A., 1988. The Evolution and Classification of Flowering Plants, 2nd ed. The New York Botanical Garden, Bronx, NY.
- Draper, J., Scott, R., 1988. The isolation of plant nucleic acids. In: Draper, J., Scott, R., Armitage, P., Walden, R. (Eds.) *Plant Genetic Transformation and Gene Expression: A Laboratory Manual*. Blackstone Scientific Publications, Oxford, pp. 199–236.
- Ebner, C., Hirschwehr, R., Bauer, L., Breiteneder, H., Valenta, R., Ebner, H., Kraft, D., Scheiner, O., 1995. Identification of allergens in fruits and vegetables. IgE-crossreactivities with important birch pollen allergens Bet v 1 and Bet v 2 (birch profilin). *J. Allergy Clin. Immunol.* 95, 962–969.
- Erba, H.P., Eddy, R., Shows, T., Kedes, L., Gunning, P., 1988. Structure, chromosomal location, and expression of the human γ -actin gene: differential evolution, location, and expression of the cytoskeletal β - and γ -actin genes. *Mol. Cell Biol.* 8, 1775–1779.
- Felsenstein, J., 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39, 783–791.
- Ferreira, F., Hirtenlehner, K., Jilek, A., Godnik-Cvar, J., Breiteneder, H., Grimm, R., Hoffmann-Sommergruber, K., Scheiner, O., Kraft, D., Breitenbach, M., Rheinberger, H.J., Ebner, C., 1996. 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.* 183, 1–11.
- Gilbert, W., Glynias, M., 1993. On the ancient nature of introns. *Gene* 135, 137–144.
- Goodall, G.J., Filipowicz, W., 1989. The AU-rich sequences present in the introns of plant nuclear pre-mRNAs are required for splicing. *Cell* 58, 473–483.
- Jaquier, A., 1990. Self-splicing group II and nuclear pre-mRNA introns: How similar are they? *Trends Biochem. Sci.* 15, 351–354.
- Jarolim, E., Tejkl, M., Rohac, M., Schlerka, G., Scheiner, O., Kraft, D., Rumpold, H., 1989. Monoclonal antibodies against birch pollen allergens: Characterization by immunoblotting and use for single-step affinity purification of the major allergen Bet v 1. *Int. Arch. Allergy Appl. Immunol.* 90, 54–60.
- Kersanach, R., Brinkmann, H., Liaud, M.-F., Zhang, D.X., Martin, W., Cerff, R., 1994. Five identical intron positions in ancient duplicated genes of eubacterial origin. *Nature* 367, 387–389.
- Larsen, J.N., Stroman, P., Ipsen, H., 1992. PCR based cloning and sequencing of isogenes encoding the tree pollen major allergen Car b 1 from *Carpinus betulus*, hornbeam. *Mol. Immunol.* 29, 703–711.
- van de Löcht, U., Meier, I., Hahlbrock, K., Somssich, I.E., 1990. A 125 pb promoter fragment is sufficient for strong elicitor-mediated gene activation in parsley. *EMBO J.* 9, 2945–2950.
- Logsdon, J.M., Tyshenko, M.G., Dixon, C., Jafari, J.D., Walker, V.K., Palmer, J.D., 1995. Seven newly discovered intron positions in the triose-phosphate isomerase gene: Evidence for the introns-late theory. *Proc Natl. Acad. Sci. USA* 92, 8507–8511.
- Luehrsen, K.R., Walbot, V., 1990. Intron enhancement of gene expression and the splicing efficiency of introns in maize cells. *Mol. Gen. Genet.* 225, 81–93.
- Mattick, J., 1994. Introns: evolution and function. *Curr. Opin. Genet. Dev.* 4, 823–831.
- Matton, D.P., Prexcott, G., Bertrand, C., Camirand, A., Brisson, N., 1993. Identification of *cis*-acting elements involved in the regulation of the pathogenesis-related gene STH-2 in potato. *Plant Mol. Biol.* 22, 279–291.
- Mikol, D.D., Rognoparut, P., Allwardt, B.A., Marton, L.S., Stefansson, K., 1993. The oligodendrocyte-myelin glycoprotein of mouse: primary structure and gene structure. *Genomics* 17, 604–610.
- Palmer, J.D., Logsdon, J.M., 1991. The recent origin of introns. *Curr. Opin. Genet. Dev.* 1, 470–477.
- Saraste, M., Sibbald, P.R., Wittinghofer, A., 1990. The P-loop—a common motif in ATP- and GTP-binding proteins. *Trends in Biochem. Sci.* 19, 430–434.

- Simpson, C.G., Leader, D.J., Brown, J.W.S., 1993. Characteristics of plant pre-mRNA introns and transposable elements. In: Croy, R.R.D. (Ed.), *Plant Molecular Biology*. Labfax Blackwell Scientific Publications, Durham, pp. 183–251.
- Stone, E.M., Schwartz, R.J. (Eds.), 1990. *Intervening Sequences in Evolution and Development*. Oxford University Press, New York, NY.
- Swoboda, I., Scheiner, O., Kraft, D., Breitenbach, M., Heberle-Bors, E., Vicente, O., 1994. A birch gene family encoding pollen allergens and pathogenesis-related proteins. *Biochim. Biophys. Acta* 1219, 457–464.
- Swoboda, I., Jilek, A., Ferreira, F., Engel, E., Hoffmann-Sommergruber, K., Scheiner, O., Kraft, D., Breiteneder, H., Pittenauer, E., Schmid, E., Vicente, O., Heberle-Bors, E., Ahorn, H., Breitenbach, M., 1995. Isoforms of Bet v 1, the major birch pollen allergen, analyzed by liquid chromatography, mass spectrometry and cDNA cloning. *J. Biol. Chem.* 270, 2607–2613.
- Swoboda, I., Scheiner, O., Heberle-Bors, E., Vicente, O., 1995. cDNA cloning and characterization of three genes in the Bet v 1 gene family that encode pathogenesis-related proteins. *Plant Cell Envir.* 18, 865–874.
- Swofford, D.L., 1993. *PAUP: Phylogenetic Analysis Using Parsimony version 3.1*. Distributed by the Illinois Natural History Survey, Champaign, IL.
- Vanek-Krebitz, M., Hoffmann-Sommergruber, K., Laimer daCamara Machado, M., Susani, M., Ebner, C., Kraft, D., Scheiner, O., Breiteneder, H., 1995. Cloning and sequencing of Mal d 1, the major allergen from apple (*Malus domestica*), and its immunological relationship to Bet v 1, the major birch pollen allergen. *Biochem. Biophys. Res. Commun.* 214, 538–551.
- Warner, S.A.J., Scott, R., Draper, J., 1993. Isolation of an asparagus intracellular PR gene (AoPR1) wound responsive promoter by the inverse polymerase chain reaction and its characterization in transgenic tobacco. *Plant J.* 3, 191–201.
- Warner, S.A.J., Gill, A., Draper, J., 1994. The developmental expression of the asparagus intracellular PR protein (AoPR1) gene correlates with sites of phenylpropanoid biosynthesis. *Plant J.* 6, 31–43.
- Weiss, C., Kramer, B., Ebner, C., Susani, M., Briza, P., Hoffmann-Sommergruber, K., Breiteneder, H., Kraft, D., Scheiner, O., Breitenbach, M., Ferreira, F., 1996. High-level expression of tree pollen isoallergens in *Escherichia coli*. *Int. Arch. Appl. Immunol.* 110, 282–287.
- Wen, J., Vanek-Krebitz, M., Hoffmann-Sommergruber, K., Scheiner, O., Breiteneder, H., 1997. The potential of Bet v 1 homologues, a nuclear multigene family, as phylogenetic markers in flowering plants. *J. Mol. Evol.*, in press.
- Williamson, B., 1977. DNA insertions and gene structure. *Nature* 270, 295–297.