The families of pathogenesis-related proteins, their activities, and comparative analysis of PR-1 type proteins

L. C. VAN LOON* and E. A. VAN STRIEN

Graduate School Experimental Plant Sciences, Institute of Biology, Section Phytopathology, Utrecht University,
3508 TB Utrecht, The Netherlands

(Accepted for publication April 1999)

PATHOGENESIS-RELATED PROTEINS ASSOCIATED WITH PLANT DEFENSE

In incompatible host-pathogen interactions, damage caused by the pathogen remains restricted as a result of the plant's defensive response. Most effective is the hypersensitive reaction, in which the cells around the infection site rapidly necrose. This response is associated with a coordinated and integrated set of metabolic alterations that are instrumental in impeding further pathogen ingress, as well as in enhancing the capacity of the host to limit subsequent infection by different types of pathogens [27, 77]. Altered ion fluxes across the plant cell membrane, generation of active oxygen species, changes in the phosphorylation state of regulatory proteins and transcriptional activation of plant defense systems culminate in cell death at the site of infection, local accumulation of phytoalexins and cell wall rigidification as a result of callose, lignin and suberin deposition [31, 89]. In addition, various novel proteins are induced which are collectively referred to as "pathogenesis-related proteins" (PRs). These PRs, defined as proteins coded for by the host plant but induced specifically in pathological or related situations [4, 81], do not only accumulate locally in the infected leaf, but are also induced systemically, associated with the development of systemic acquired resistance (SAR) against further infection by fungi, bacteria and viruses. Induction of PRs has been found in many plant species belonging to various families [76], suggestive of a general role for these proteins in adaptation to biotic stress conditions. SAR, likewise, is a generally occurring phenomenon, that engenders an enhancement of the defensive capacity of plants in response to necrotizing infections [76]. Since some of the tobacco PRs were identified as chitinases [45] and β-1,3-glucanases [38] with potential antifungal activity, it has often been suggested that the collective set of PRs may be effective in inhibiting pathogen growth, multiplication and/or spread, and be responsible for the state of SAR [42, 65].

Originally, five main classes of PRs (PR-1–5) were characterized by both biochemical and molecular-biological techniques in tobacco [9, 81]. Thereupon, in 1994 a unifying nomenclature for PRs was proposed based on their grouping into families sharing amino acid sequences, serological relationship, and/or enzymatic or biological activity. By then 11 families (PR-1–11) were recognized and classified for tobacco and tomato [81] (cf. Table 1). Criteria used for the inclusion of new families of PRs were that (i) protein(s) must be induced by a pathogen in tissues that do not normally express the protein(s), and (ii) induced expression must have been shown to occur in at least two different plant-pathogen combinations, or expression in a single plant-pathogen combination must have been confirmed independently in different laboratories.

Individual family members are named by lower case letters in the order in which they are described. In the literature, besides proteins, newly defined mRNAs (cDNAs) are often considered as additional members of the existing families when shown to be induced by pathogens or specific elicitors. However, because PRs are generally defined by their occurrence as protein bands on gels, and classified within each family once the protein has been characterized, cDNA or genomic sequences without information on the corresponding protein cannot be fitted into the adopted nomenclature. Thus, for naming, it is necessary to gather information at both the nucleic acid and the protein level when dealing with a stress-related sequence falling within the definition of PRs. Conversely, homologies at the cDNA or genomic level may be encountered without information on the expression or
characteristics of the encoded protein. Such sequences obviously belong to the PR-type families, but cannot (yet) be considered to correspond to pathogen-induced PRs and named accordingly. In several situations it is difficult to distinguish PRs from related proteins/mRNAs that are present in some organs or appear during specific developmental stages. Homologous proteins/mRNAs in healthy tissues in which no induction by pathogen infection has (yet) been demonstrated, are to be termed PR-like proteins (PRLs). In several situations it is difficult to distinguish PRs from related proteins

<table>
<thead>
<tr>
<th>Family</th>
<th>Type member</th>
<th>Properties</th>
<th>Gene symbols</th>
</tr>
</thead>
<tbody>
<tr>
<td>PR-1</td>
<td>Tobacco PR-1a</td>
<td>unknown</td>
<td>Yp1</td>
</tr>
<tr>
<td>PR-2</td>
<td>Tobacco PR-2</td>
<td>1,3-glucanase</td>
<td>Yp2, [Gn2 ('Glb')]</td>
</tr>
<tr>
<td>PR-3</td>
<td>Tobacco P, Q</td>
<td>chitinase type I, II</td>
<td>Yp3, Chia</td>
</tr>
<tr>
<td>PR-4</td>
<td>Tobacco ‘R’</td>
<td>chitinase type I, II</td>
<td>Yp4, Chib</td>
</tr>
<tr>
<td>PR-5</td>
<td>Tobacco S</td>
<td>thaumatin-like</td>
<td>Yp5</td>
</tr>
<tr>
<td>PR-6</td>
<td>Tomato Inhibitor I</td>
<td>proteinase-inhibitor</td>
<td>Yp6, Pio ('Pin')</td>
</tr>
<tr>
<td>PR-7</td>
<td>Tomato Pao</td>
<td>endoproteinase</td>
<td>Yp7</td>
</tr>
<tr>
<td>PR-8</td>
<td>Cucumber chitinase</td>
<td>chitinase type III</td>
<td>Yp8, Chib</td>
</tr>
<tr>
<td>PR-9</td>
<td>Tobacco ‘lignin-forming peroxidase’</td>
<td>peroxidase</td>
<td>Yp9, Prx</td>
</tr>
<tr>
<td>PR-10</td>
<td>Parsley ‘PR-1’</td>
<td>‘ribonuclease-like’</td>
<td>Yp10</td>
</tr>
<tr>
<td>PR-11</td>
<td>Tobacco class V chitinase</td>
<td>chitinase, type I</td>
<td>Yp11, Chic</td>
</tr>
<tr>
<td>PR-12</td>
<td>Radish Rs-AFP3</td>
<td>defensin</td>
<td>Yp12</td>
</tr>
<tr>
<td>PR-13</td>
<td>Arabidopsis THI2.1</td>
<td>thionin</td>
<td>ypr13, Thi</td>
</tr>
<tr>
<td>PR-14</td>
<td>Barley LTP4</td>
<td>lipid-transfer protein</td>
<td>ypr14, Ltp</td>
</tr>
</tbody>
</table>

a For references to PR-1 to -11 see [87], for PR-12 [74], for PR-13 [26], and for PR-14 [22].

Rhizobacterially-mediated ISR in Arabidopsis shares with SAR the dependence on the functioning of the npr1 gene. The latter, in turn, distinguishes ISR from the JA- and ethylene-dependent inducible defense response pathway effective against Alternaria brassicicola, which is independent of npr1 [20]. Thus, at least three types of microbially-induced resistance, characterized by different signalling pathways, appear to exist in plants.

The level of protection afforded by ISR is usually less than that attainable during SAR [62, 77], in agreement with findings that pathogen-induced hypersensitive necrosis contributes to the level of resistance achieved [14]. The association of PRs with SAR, but not with ISR, has led to the hypothesis that accumulation of PRs is not a prerequisite for the induction of resistance, but that PRs contribute to the protective state [71]. Indeed, the different signalling pathways of SAR and ISR converging at the level of npr1 appear complementary: higher levels of protection are achieved by combining induction by a necrotizing pathogen on the leaves and a non-pathogenic rhizobacterium on the roots, without SAR-associated PR-gene expression being stimulated [85]. SAR is dependent on the accumulation of SA, but not JA or ethylene. ISR requires perception of JA and ethylene but is not associated with significant increases in the levels of these regulatory compounds. In contrast, the JA- and ethylene-dependent pathway induced by, and effective against, A. brassicicola involves increases in SA, JA and ethylene. It appears that only when increases in the levels of any of these signals occur, PRs become detectable in the infected plants. The observations indicate that individual PRs are induced to various extents by these different signals. Consequently, the mixture of signals released or produced upon microbial stimulation appears to determine the magnitude of the plant’s response and its effectiveness to inhibit further infection.
Recent evidence indicates that in Arabidopsis SA-dependent expression of PR-1, PR-2 and PR-5 is required for increased protection against the biotrophic fungus Peronospora parasitica, whereas SA-independent but JA-dependent induction of the plant defensin gene pdf1-2, as well as of PR-3 and PR-4, is associated with the induced resistance against the necrotrophic fungi A. brassicicola [58], Botrytis cinerea [75] and Fusarium oxysporum Lsp. matthioliace (Fom) [1]. Moreover, overexpression of the JA-inducible thionin gene thi2-I increased SA-independent resistance against Fom [21, 22]. These results suggest that the SA- and JA-dependent defense pathways in Arabidopsis contribute to resistance against distinct microbial pathogens. As a result, PRs and similarly induced antimicrobial proteins appear to contribute differentially to the induced resistance against different pathogens. This can explain why in PR-overexpressing transgenic plants only some of these proteins are effective in reducing infection by only some selected fungi and bacteria [77].

**CURRENT STATUS OF THE FAMILIES OF PATHOGENESIS-RELATED PROTEINS**

PDF12 is a member of the group of plant defensins, which together with the thionins, lipid transfer proteins (LPTs), hevein-type, knottin-type and Impatiens antimicrobial peptides are families of peptides with antimicrobial activities that are present in many, and perhaps all, plant species [19]. In a number of plant species, a strong induction of genes encoding either thionins, plant defensins or LPTs has been observed upon infection of the leaves by pathogens. Hence, at least some of these antimicrobial peptides must be considered PRs and can also take part in the inducible defense response of plants. Therefore, following discussion at the 5th International Workshop on Pathogenesis-related Proteins in Plants, held in 1998 at Aussois, France, we now wish to propose inclusion of three additional families of PRs. These comprise the pathogen-induced plant defensins (PR-12), thionins (PR-13) and LTPs (PR-14) (Table 1).

Within each PR-family a type member has been defined, the nucleotide sequence of the mRNA of which may be used in the search for homologues in the same or in different plant species. The type member of PR-12 is Rs-AFP3 from radish, shown to be induced upon infection with A. brassicicola [74]. THI2-I, highly inducible in Arabidopsis seedlings by Fom [20], constitutes the type member of PR-13. The situation for the LTPs is somewhat less unequivocal. Several LTP genes in barley are upregulated in response to infection by Erysiphe graminis or Rhynchosporium secalis [21, 57]. LTPs are present in relatively high concentrations in vascular tissue and in the outer cell layers of the expressed surface of the plant, and share with defensins and thionins their ability to inhibit bacterial and fungal pathogens. However, LTPs appear to be regulated differently in that their expression is reduced upon treatment with MeJA. From barley, four LTPs have been purified to homogeneity and characterized [22]. Of these, LTP4 (cw21) is induced at the mRNA level to the same extent as barley PR-1 upon fungal infection [25]. Although protein levels were not quantified under these conditions, the situation appears sufficiently similar to that of other PR-proteins to include LTPs as PR-14, with the tentative designation of barley LTP4 as the type member. Two members of a novel family from barley, which are similarly induced by E. graminis and have benzothiadiazole-inducible homologs in wheat [15], may qualify for future addition as PR-15 when more details become available.

For the majority of the PR families, activities are known or can be inferred [42, 46]. Thus, the PR-2 family consists of endo-β-1,3-glucanases, and PR-3, -4, -8 and -11 are all classified as endochitinases, even though their specific activities on colloidal chitin vary over 100-fold [12]. A different way of distinguishing these types of chitinases is by class, based on their different specific activities on a range of substrates, with class III (PR-8) basic isoforms possessing substantial lysozyme activity. Substrate preferences of ten tobacco chitinases purified to homogeneity from tobacco leaves reacting hypersensitively to tobacco mosaic virus (TMV) and belonging to five distinct structural classes, were suggested to represent complementary enzymes which may have synergistic effects on their substrates [12]. Several glucanases and chitinases have been shown to have antifungal properties, although these appear to be restricted to certain fungi [42].

PR-6 are proteinase inhibitors implicated in defense against insects and other herbivores, micro-organisms, and nematodes [41, 64]. PR-7 has so far been characterized only in tomato, where it is a major PR and acts as an endoproteinase. Because lysis of fungal cell walls often requires degradation of cell wall proteins in addition to hydrolysis of chitin and glucan [26, 32], it seems reasonable to assume that PR-7 serves as an accessory to antifungal action. The PR-9 family of peroxidases is likely to function in strengthening plant cell walls by catalyzing lignin deposition in reaction to microbial attack. The PR-5 family belongs to the thaumatin-like proteins with homology to permatins, that permeabilize fungal membranes [59]. Some members of this family have been shown to possess antifungal activity, particularly against oomycetes. Recently, a 22 kDa potato PR-5 was shown to bind actin together with a 32 kDa basic chitinase, and it was suggested that the actin-binding complex might be involved in cytoplasmic aggregation, thereby participating in the potato cell’s defense against Phytophthora infestans [73].

The PR-10 family is structurally related to ribonucleases...
and although it is tempting to suppose that these intracellular PRs may be active against viruses, a capability to cleave viral RNA remains to be demonstrated. The PR-12 type defensins, PR-13 type thionins and PR-14 type LTPs all exhibit antifungal and antibacterial activity, exerting their effect at the level of the plasma membrane of the target micro-organism [2, 16, 25]. The only PR family for which no function or relationship is known, consists of the PR-1 proteins. Specific members of the tobacco and tomato PR-1 families have antifungal activity against oomycete fungi, but their mechanism of action is not known. Alexander et al. [1] provided indirect evidence by showing that transgenic tobacco plants which constitutively expressed the PR-1a gene exhibited increased tolerance to Phytophthora parasitica var. nicotianae and Peronospora tabacina. Direct antifungal activity of tomato PR-1 was demonstrated by Niderman et al. [56], both in vitro as an inhibition of germination of P. infestans zoospores, and in vivo as a reduction in the surface area of leaf discs infected with this fungus. Differential activity was found between the acidic tobacco PR-1a and -1b and the basic tomato PR-1c and tobacco PR-1g proteins, with the basic proteins having the highest antifungal activities.

Not all families of PRs have been identified in each plant species examined. In tobacco for example, no PR-7, -10, -12, -13 and -14 are known, suggesting that plant species differ in the types of PRs present or, at least, expressed upon infection. Yet, members of several families of PRs have been described for Arabidopsis, barley, bean, maize, potato, tobacco and tomato [91, and in preparation]. In tobacco, the major PRs are acidic, located extracellularly, and coordinately expressed upon infection, whereas basic vacuolar counterparts with different stress expression patterns are also present in healthy plants where their expression is temporally and spatially controlled in a cell-type and organ-dependent manner. The presence of PR-type proteins in healthy plant tissues, such as glucanases and chitinases in dicotyledons and thionins and LTPs in monocotyledons, appears to be fairly common.

The occurrence of homologous PRs as small multigene families in various species belonging to different plant families, their tissue-specific expression during development and consistent localization in the apoplast as well as in the vacuolar compartment, and their differential induction by endogenous and exogenous signalling compounds, suggest that PRs have important functions extending beyond a role in adaptation to biotic stress conditions. Indeed, the basic tobacco glucanase PR-2d functions developmentally in seed germination by weakening the endosperm, thus allowing the radicle to protrude [82]. Chitinases homologous to PR-3 and PR-4 act as morphogenetic factors in carrot somatic embryogenesis [16, 43]. Several PRs are expressed in cultured cells [6, 2] or upon the transition of plants to flowering [e.g. 55], also suggestive of a developmental role. PR-2-like, PR-3-like and PR-5-like proteins accumulate in the apoplast of winter rye tissues during cold acclimation and exhibit antifreeze activity [3]. Basic PR-5 proteins (osmotin) are induced in e.g. tobacco and tomato in response to osmotic stress [61]. PR-10 proteins are homologous to a large family of food and tree pollen allergens from both monocotyledonous and dicotyledonous plant species [96].

Some PRLs contain a PR-type sequence as part of a longer protein. Thus, the tobacco stigma- and style-specific glucanases Sp41a and Sp41b [57] and a 39 kDa thaumatin-like protein that is secreted from stigmas [44] both contain a C-terminal glycosylated extension with affinity for concanavalin A, suggestive of a role in carbohydrate-mediated cell signalling. From Arabidopsis, a gene encoding a PR-type receptor protein kinase (PR5K) was isolated comprising an extracellular domain related to the PR-5 proteins, a central transmembrane-spanning domain, and a functional intracellular serine/threonine protein kinase domain. Infection with turnip crinkle virus or treatment with SA induced PR-5 but not PR5K. It was suggested that the PR-5 domain might bind a polypeptide ligand, raising the possibility that authentic PR-5 proteins may also interact with polypeptides [87]. A similar receptor-like kinase (CHRK1) containing a class V chitinase-related domain in its N-terminus, was identified in a cDNA library from tobacco flowers. CHRK1 mRNA was induced by TMV but the protein lacked chitinase activity, suggesting that it functions as a chitin-binding receptor [39]. Such observations raise the question whether PR genes evolved primarily to limit damage inflicted by invading pathogens, or were adapted from other functions to serve an accessory protective role.

**COMPARATIVE ANALYSIS OF PR-1 TYPE PROTEINS**

PR-1 is a dominant group of PRs induced by pathogens or SA, and is commonly used as a marker for SAR. Since their discovery in 1970, numerous researchers have attempted to assess the function of PR-1 proteins in plants, but without much success [15]. Their limited antifungal activity suggests a function in plant defense, but its mode of action or relationship to other proteins is unknown. In tobacco and tomato, PR-1 proteins belong to small multigene families. Properties of the known PR-1 proteins from these species are listed in Table 2. The acidic N. tabacum PR-1a was the first to be purified and characterized [4], and has been taken as the type member of the PR-1 proteins [81]. The basic tomato PR-1a and -1b are more similar to the basic PR-1g of tobacco than to the acidic tobacco PR-1a and -1b. The structure of tomato PR-1b (P14a) was solved recently by nuclear
magnetic resonance and found to represent a unique molecular architecture [29]. The protein contains four \( \alpha \)-helices (I–IV) and four \( \beta \)-strands (A–D) arranged anti-parallel between helices I, III and IV and II, respectively (Fig. 1). The tight packing of the \( \alpha \)-helices on both sides of the central \( \beta \)-sheet results in a compact, bipartite molecular core, which is stabilized by hydrophobic interactions and multiple hydrogen bonds. This compact structure reflects the high stability of PR-1 proteins and their insensitivity to several proteases [79].

Because the function of the PR-1 family is still unclear, these considerations prompted us to search the databases for homologues of PR-1 proteins. Homology searches were performed using the NCBI Blast [2] network service, on the Genbank, PDB, Swiss-Prot, PRF and PRD databases. Protein sequences were aligned with the Pileup program from UWGCG 8.0 [18] and adapted where required. Figure 2 shows the aligned amino acid sequences of 36 PR-1 type proteins from different plant species, including several PRLs, a probably partial sequence of a protein found in a tumourous tobacco hybrid (N-tum) and the derived sequence of a putative pseudogene from tobacco (NtPR1-Y). Not included are PR-1a, -1b and -1c sequences from different tobacco (Nicotiana tabacum) cultivars that differ in only one or a few nucleotides (resulting in at most, a single amino acid difference) from the sequences presented. The primary translation products contain a hydrophobic signal sequence, which is cleaved off upon entry in the endoplasmic reticulum. The mature proteins are mostly about 135 amino acids long, contain six conserved cysteine residues forming disulphide bridges, and show a high level of sequence conservation throughout different plant families, including both mono- and dicotyledons, with 31 \% sequence identity among all PR-1 type proteins (Table 3) and pairwise sequence identity up to 96 \%. Percentage similarity to the type member, tobacco PR-1a, ranges from 97 \% to 46 \% (Table 3). No consistent amino acid sequence differences are apparent between acidic and basic isoforms. The structural elements comprising the four \( \alpha \)-helices and \( \beta \)-strands appear to be highly conserved (Fig. 2), suggesting that all proteins share the same \( \alpha \)-\( \beta \)-\( \alpha \) sandwich structure with tomato PR-1a. The PR-1 type proteins exhibiting short C-terminal extensions have mostly been characterized as basic proteins, and these sequence additions reflect their slightly higher molecular weights.

Related sequences have been found in proteins from yeasts (Fig. 3), insects (Fig. 4) and vertebrates (Fig. 5). Figure 3 compares the sequence of tobacco PR-1a with five partial sequences from the yeasts. At the protein level, the yeast sequences show 25–39 \% identity and up to 51 \% similarity with the type member (Table 3). No consistent amino acid sequence differences are apparent between acidic and basic isoforms. The structural elements comprising the four \( \alpha \)-helices and \( \beta \)-strands appear to be highly conserved (Fig. 2), suggesting that all proteins share the same \( \alpha \)-\( \beta \)-\( \alpha \) sandwich structure with tomato PR-1a. The PR-1 type proteins exhibiting short C-terminal extensions have mostly been characterized as basic proteins, and these sequence additions reflect their slightly higher molecular weights.

**Table 2. Properties of PR-1 proteins from tobacco and tomato**

<table>
<thead>
<tr>
<th>Member</th>
<th>Molecular weight (kDa)</th>
<th>Isoelectric point</th>
<th>Percentage homology at protein level</th>
<th>Serogroup</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nt.a</td>
<td>15</td>
<td>4.0</td>
<td>100</td>
<td>I</td>
</tr>
<tr>
<td>Nt.b</td>
<td>15</td>
<td>4.4</td>
<td>94</td>
<td>II</td>
</tr>
<tr>
<td>Nt.c</td>
<td>15</td>
<td>4.5</td>
<td>91</td>
<td>III</td>
</tr>
<tr>
<td>Nt.d</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ns.e</td>
<td>15</td>
<td></td>
<td></td>
<td>I</td>
</tr>
<tr>
<td>Ng.f</td>
<td>14</td>
<td></td>
<td></td>
<td>II</td>
</tr>
<tr>
<td>Nt.g</td>
<td>17</td>
<td>10.7</td>
<td>67</td>
<td>IV</td>
</tr>
<tr>
<td>Nl.h</td>
<td>15</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Le.a</td>
<td>14</td>
<td>10.7</td>
<td>60</td>
<td>IV</td>
</tr>
<tr>
<td>Le.b</td>
<td>14</td>
<td>10.9</td>
<td></td>
<td>IV</td>
</tr>
<tr>
<td>Le.c</td>
<td>14</td>
<td></td>
<td></td>
<td>IV</td>
</tr>
</tbody>
</table>

*a For references see [91]. For N.t see also [36]. Nl.h is the less acidic PR-1 from Nicotiana langsdorffii [39]. Le.a = tomato PR-1a = P14 [77] = c2 [36] = PR-1b2 [76] = P14b [36]; Le.b = tomato PR-1b = P6 [17] = c3 [29] = PR-1b1 [76] = P14a [56]; Le.c = tomato PR-1c = P14c [50].

**Families of pathogenesis-related proteins**

---

**Fig. 1. Ribbon drawing of the three-dimensional structure of tomato PR-1b.** The \( \alpha \)-helices I to IV are shown in red and yellow, the \( \beta \)-strands A to D in green, other polypeptide segments in gray, and the polypeptide termini are marked V and C. From Fernández et al. [29].
Fig. 2. For caption see facing page.
Fig. 2. Aligned amino acid sequences of 36 PR-1 type proteins from plants, including signal sequences. Identical amino acid residues are highlighted in black if present in all sequences and in gray if present in at least 50% of all sequences. A "consensus sequence" is depicted above the sequence of the PR-1 type member tobacco PR-1a (NtPR1a). The cleavage site between the signal peptide and mature protein is indicated by the arrowhead. The positions of the three disulfide bonds (C-C), the four helices αI to αIV and the four parallel strands βA to βD are marked. Residue numbers are indicated to the right. AtPR1–1: Arabidopsis thaliana PR-1 type (pr1–1) (Accession number M59196); AtPR1–2: A. thaliana PR-1 type (pr1–2) (X69600); AtPR1–3: A. thaliana PR-1 type (pr1–3) (X69600); AtPR1–4: A. thaliana PR-1 (pr1–4) (M90508); BnPR1–1: Brassica napus PR-1a (pr1–1) (U21849/U70666); BnPR1–2: B. napus PR-1 (pr1–2) (U64806); CaPR1: Capsicum annuum basic PR-1 (AF053343); CsPRL1: Camellia sinensis PRL-1 (AB015047); HvPR1–1: Hordeum Šulgare PR-1 type (pr1–1) (Z21494); HvPR1–2: H. Šulgare PR-1 (pr1–2) (Z48728); HvPR1a: H. Šulgare PR-1a (X74939); HvPRB12: H. vulgaris basic PR-1 type (pr1–2) (Z26333); LeP14–1: Lycopersicon esculentum PR-1 type (A22635); LeP14–2: L. esculentum PR-1 type (A22636); LePR1A1: L. esculentum, acidic PR-1a (pr1A1) (X71592); LePR1A2: L. esculentum, acidic PR-1b (Y08844); LePR1A3: L. esculentum acidic PR-1 (Y08844); LePR1A4: L. esculentum acidic PR-1 (Y08844); LePR1B2: L. esculentum acidic PR-1 (Y08844); LePR1B3: L. esculentum acidic PR-1 (Y08844); LePR1B4: L. esculentum acidic PR-1 (Y08844); LePR1C1: L. esculentum acidic PR-1 (Y08844); LePR1C2: L. esculentum acidic PR-1 (Y08844); LePR1C3: L. esculentum acidic PR-1 (Y08844); LePR1C4: L. esculentum acidic PR-1 (Y08844); LePR1D1: L. esculentum acidic PR-1 (Y08844); LePR1D2: L. esculentum acidic PR-1 (Y08844); LePR1D3: L. esculentum acidic PR-1 (Y08844); LePR1D4: L. esculentum acidic PR-1 (Y08844); N-tum: partial PR-1 type clone from F1 hybrid between Nicotiana glauca and Nicotiana langsdorffii (D94356); NgPR1: Nicotiana glutinosa PR-1f (U49341); NtPR1a: Nicotiana tabacum PR-1a (A015047); NtPR1b: N. tabacum PR-1b (D90197); NtPR1c: N. tabacum PR-1c (S07580); NtPR1d: N. tabacum PR-1d (S07580); NtPR1e: N. tabacum PR-1e (S07580); NtPR1f: N. tabacum PR-1f (S07580); NtPR1g: N. tabacum PR-1g (X10465); NtPR1h: N. tabacum basic PR-1 (X66942); OatPR1: Oryza sativa PR-1 type (U89895); TaPR1: Triticum aestivum PR-1 (pr1–1) (A015047); TaPR1b: T. aestivum PR-1 (pr1–2) (A015047); ZmPR1: Zea mays PR-1 (S07580).
### Table 3. Homology of PR-1 type proteins from plants and other organisms in percentage amino acid identity (upper) and similarity (lower; in bold) with tobacco PR-1a (NtPR1a), excluding the signal sequence. For abbreviations see legends to Figs 2–5

<table>
<thead>
<tr>
<th>Protein</th>
<th>NtPR1b</th>
<th>NtPR1c</th>
<th>NtPR1-Y</th>
<th>NgPR1</th>
<th>HvPR-1-1</th>
<th>HvPRB13</th>
<th>HvPRB12</th>
<th>HvPR1a</th>
<th>ZmPR1-2</th>
</tr>
</thead>
<tbody>
<tr>
<td>NtPR1a</td>
<td>94%</td>
<td>94%</td>
<td>83%</td>
<td>92%</td>
<td>56%</td>
<td>57%</td>
<td>56%</td>
<td>59%</td>
<td>65%</td>
</tr>
<tr>
<td>ZmPR1</td>
<td>96%</td>
<td>97%</td>
<td>85%</td>
<td>95%</td>
<td>68%</td>
<td>68%</td>
<td>69%</td>
<td>70%</td>
<td>79%</td>
</tr>
<tr>
<td>OaPR1</td>
<td>53%</td>
<td>59%</td>
<td>57%</td>
<td>54%</td>
<td>55%</td>
<td>55%</td>
<td>59%</td>
<td>56%</td>
<td>52%</td>
</tr>
<tr>
<td>TaPR1-1</td>
<td>68%</td>
<td>72%</td>
<td>68%</td>
<td>72%</td>
<td>71%</td>
<td>67%</td>
<td>71%</td>
<td>71%</td>
<td>64%</td>
</tr>
<tr>
<td>TaPR1-2</td>
<td>63%</td>
<td>58%</td>
<td>60%</td>
<td>54%</td>
<td>58%</td>
<td>53%</td>
<td>61%</td>
<td>60%</td>
<td>63%</td>
</tr>
<tr>
<td>HvPR1-1A</td>
<td>79%</td>
<td>73%</td>
<td>74%</td>
<td>67%</td>
<td>73%</td>
<td>68%</td>
<td>76%</td>
<td>73%</td>
<td>76%</td>
</tr>
<tr>
<td>HvPR1b</td>
<td>60%</td>
<td>55%</td>
<td>50%</td>
<td>41%</td>
<td>52%</td>
<td>46%</td>
<td>31%</td>
<td>34%</td>
<td>50%</td>
</tr>
<tr>
<td>HvPR1c</td>
<td>72%</td>
<td>67%</td>
<td>62%</td>
<td>48%</td>
<td>67%</td>
<td>62%</td>
<td>46%</td>
<td>50%</td>
<td></td>
</tr>
<tr>
<td>HvPR1d</td>
<td>25%</td>
<td>24%</td>
<td>36%</td>
<td>35%</td>
<td>39%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HvPR1e</td>
<td>38%</td>
<td>36%</td>
<td>49%</td>
<td>47%</td>
<td>51%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hs-GLIO</td>
<td>18%</td>
<td>18%</td>
<td>13%</td>
<td>16%</td>
<td>30%</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hs-RTVP</td>
<td>30%</td>
<td>30%</td>
<td>25%</td>
<td>28%</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hs-GPepi</td>
<td>46%</td>
<td>46%</td>
<td>45%</td>
<td>45%</td>
<td>40%</td>
<td>40%</td>
<td>43%</td>
<td>46%</td>
<td>33%</td>
</tr>
<tr>
<td>Hs-TryInh</td>
<td>31%</td>
<td>32%</td>
<td>29%</td>
<td>29%</td>
<td>25%</td>
<td>22%</td>
<td>28%</td>
<td>31%</td>
<td>22%</td>
</tr>
<tr>
<td>Hh-Helo</td>
<td>46%</td>
<td>46%</td>
<td>45%</td>
<td>45%</td>
<td>40%</td>
<td>40%</td>
<td>43%</td>
<td>46%</td>
<td>33%</td>
</tr>
</tbody>
</table>

cDNA cloning of the major allergen (Dol-Ag5) in the venom of the white-face hornet wasp Dolichovespula maculata revealed that the primary structures of the two isoforms present have sequence similarity with tobacco PR-1a: in a 130-residue overlap of these proteins, 35–39 residues were identical [23] (Fig. 4). Similar sequences were also revealed in venom proteins of other wasps [36, 47] and fire ants [35], as well as in the fruit fly Drosophila melanogaster (Dm-Ag5) [66], being up to 30% homologous with tobacco PR-1a (Table 3). The vespid and ant venom proteins are all major allergens and have been suggested to have evolved for use against other insects or invertebrates [36], but no further physiological function is known so far.
Fig. 4. Aligned amino acid sequences of tobacco PR-1a (NtPR1a), including signal sequence, and corresponding parts of PR-1 type proteins from insect venom allergens. Identical amino acid residues are highlighted in black if present in all sequences and in gray if present in at least 50% of the sequences. A “consensus sequence” is depicted below the sequences. The cleavage site between the signal peptide and mature protein is indicated by the arrowhead. The positions of the conserved cysteine residues forming disulphide bonds in tobacco PR-1a are marked. Residue numbers are indicated to the right. Dm-Ag5: *Drosophila melanogaster* antigen 5 type (partial sequence) (Accession number: L49036); Dm-VAg5: *Dolichoaspis espusula maculata* antigen 5 (J03601); Si-VAg5: *Solenopsis invicta* allergen III (B37330); Vp-VAg5: *Vespula pensylvania* antigen 5 (C44522/C44583).

Fig. 5. Aligned amino acid sequences of tobacco PR-1a (NtPR1a), including signal sequence, and corresponding parts of PR-1 type proteins from vertebrates. Identical amino acid residues are highlighted in black if present in all sequences and in gray if present in at least 50% of the sequences. A “consensus sequence” is depicted below the sequences. The positions of the conserved cysteine residues forming disulphide bonds in tobacco PR-1a are marked. Residue numbers are indicated to the right. Hh-Helo: *Heloderma horridum* helothermine (Accession number: U13619); Hs-CRS3: *Homo sapiens* CRISP-3 (X95240); Hs-GLIO: *H. sapiens* GliPR (U16307); Hs-GPepi: *H. sapiens* acidic epididymal glycoprotein homolog (S80310); Hs-RTVP: *H. sapiens* RTVP-1 (X91911); Hs-SGP28: *H. sapiens* SGP28 (X94325); Hs-TrInh: *H. sapiens* P25TI (D45027); Mm-SCP1: *Mus musculus* CRISP-1 (L05559); Tm-Crvp: *Trismeresurus mucrosquamatus* venom protein (U59447).

A related protein family in vertebrates (Fig. 5) has been described as cysteine-rich secretory proteins (CRISPs). The homology of the CRISPs with plant PR-1 proteins is restricted to their N-terminal part, with 22–31% identity and up to 46% similarity with tobacco PR-1a (Table 3). The amino acid sequence GHYTQVVW is a particularly well-conserved region in the two groups of proteins, suggestive of an important functional role of this domain. Mostly, two of the conserved PR-1 Cys residues involved in the formation of the disulphide bridges are absent from the vertebrate proteins. On the other hand, CRISPs contain a conserved spacing of up to 16 Cys residues in the
C-terminal half, which most probably forms a discrete, compact domain [15].

It has been speculated that CRISP may encode lytic enzymatic activities, which would be consistent with the observed association of the mammalian androgen-regulated acidic epididymal glycoprotein (AEG, or CRISP-1) with the sperm surface [1i, 34], as well as with the presence of the guinea pig counterpart (AA1) [33] of the mouse testis-specific gene-encoded protein TPX-1 (CRISP-2) [37] in the acrosome, where these proteins could assist in the maturation of the sperm and be involved in degrading egg structures during fertilization. In contrast, the mouse CRISP-3 protein has been shown to be expressed in the male salivary gland [30] and in developing B-cells, where a possible lytic activity might be related to antimicrobial activity in saliva and in the blood or lymph [29]. Helothermine, a toxin with hypothermic effects originating from the salivary secretion of the Mexican beaded lizard, Heloderma horridum, has been found to be another member of the CRISP family [16, 57].

Based on the finding that helothermine blocks the ryanodine receptor in striated muscle, the possibility was suggested that CRISP-like proteins may endogenously regulate ryanodine receptors in mammalian tissues.

The CRISP sequences show some stretches of complete identity and an overall 30% identity to two further groups of mammalian proteins, encompassing proteins from human tumor cells of glial origin (GliPR, RTVP-1) and macrophages [54, 63], and specific granule protein 28 (SGP28) from human neutrophils [49]. Based on its intracellular location within the specific granules, SGP28 has been hypothesized to function either as an antimicrobial B-cell or as a type of matrix protease. Another related 25 kDa protein (P25TI), frequently expressed on human neuroblastoma and glioblastoma cell lines and also present at low levels in the brain, placenta and lymphocytes, was characterized as exhibiting weak trypsin-inhibiting activity, but its sequence has no homology to other proteinase inhibitors [85]. PR-1 proteins from plants have been tested for proteinase-inhibitory activity, but none has been reported.

These observations that PR-1 proteins form a specific family within the plant kingdom and show homologies and structural motifs in common with proteins from fungi, invertebrate and vertebrate animals and humans, make the PR-1 family a distinct and highly conserved group of proteins. Their widespread occurrence suggests that these proteins share an evolutionary origin and possess activity essential to the functioning of living organisms. A close structural similarity of the human glima protein GliPR and tomato PR-1b has been interpreted as evidence for an origin of these proteins from a common ancestor that has evolved into a large "PR-protein superfamily" [11]. It was even inferred that human GliPR and plant PR-1 proteins operate according to the same molecular mechanism, establishing a possible functional link between the human immune system and defense-related activity in plants. It is striking that the proteins from insects and vertebrates function as major antigens in venoms or are implicated otherwise in defense. However, most plant PR-1 proteins are not very antigenic and the functions of all PR-1 type proteins remain essentially unknown. The weak trypsin-inhibitory activity of P25TI provides a first indication of some biochemical activity of a PR-1 type protein, but it must be realized that this protein is substantially larger than any of the plant PR-1 proteins. Although it was suggested that all proteins of the PR-protein superfamily could, on the basis of present knowledge of the molecular structure, have similar functions, the most convincing homology is observed with the proteins from yeast, which have been implicated in morphogenesis. It must be concluded that the high extent of sequence conservation of the plant PR-1 proteins from different plant families is remarkable, but so far does not offer any clues concerning their mode of action. To define their function(s) in plants, new strategies need to be devised.

This research was supported in part by the EU/Bioinformatics program under contract number B104 CT950247.

REFERENCES


Families of pathogenesis-related proteins


