

Short communication

Sequential and structural homology between intracellular pathogenesis-related proteins and a group of latex proteins

Peter Osmark*, Brian Boyle and Normand Brisson

Department of Biochemistry, Université de Montréal, Montréal, Canada H3C 3J7 (*author for correspondence)

Received 18 March 1998; accepted 1 June 1998

Key words: Bet v 1, homology, latex proteins, pathogenesis-related proteins, PR-10, sequence analysis

Abstract

The intracellular pathogenesis-related proteins have been identified in a broad range of flowering plants. Some display quite different patterns of expression, in many cases unrelated to the pathogenic response. Nevertheless, these proteins are all very similar and in most cases share more than 35% sequence identity. In this report we investigate the significance of a rather weak similarity between the intracellular pathogenesis-related (IPR or PR-10) proteins and a group of proteins identified in the latex of opium poppy and in *Arabidopsis*, among others. A sequence analysis held together with the recently published three-dimensional structure of Bet v 1, an IPR protein from birch pollen, strongly suggests sequential and structural homology between the two protein families.

The intracellular pathogenesis-related (IPR or PR-10 [13, 12]) proteins are a family of small homologous, primarily acidic proteins which have been identified in a variety of angiosperms, monocots as well as dicots. Their biological function is basically unknown, although expression patterns for many of the molecules have been thoroughly studied and a few members have been reported to possess RNase activity [1, 5]. Usually, these proteins accumulate in connection with fungal and bacterial infection, but several are regulated by other stress factors, such as drought, wounding, and darkness, or they are developmentally regulated in certain tissues. A number of the IPR proteins are of immunological interest as they possess strong allergenic properties. These include the major allergens from birch and hornbeam pollen, apple, carrot, and celery.

A weak sequence similarity between these proteins and a family of proteins first identified in the latex of opium poppy (*Papaver somniferum*) [6] led us to undertake a thorough examination of their mutual relationship. These latter proteins have been designated major latex proteins, or MLPs [7], and have later been found in *Arabidopsis*, bell pepper, melon, strawberry, and tobacco. Also in this case the function of the pro-

teins in the plant is unknown, but a few show patterns of expression similar to some of the IPR proteins. Thus, the bell pepper proteins are known to be induced by wounding [8] as are IPR proteins in potato [4], soybean [2] and asparagus [13]. Furthermore, all of the known MLPs, like the IPR proteins, have a theoretical pI in the acidic range (4.75–6.65). Curiously, no plant is so far known to contain both MLP and IPR proteins. No IPR protein has, for example, ever been identified in the extensively studied tobacco plant, although several are present in potato, which is a close relative.

The sequence similarity between MLP and IPR proteins has been noticed earlier [8], but the extent of identity is so low, less than 25%, that it can not alone be used to establish whether the two families are homologous. There is, however, compelling evidence that this is in fact the case when the sequence alignment is analyzed in detail and compared to the known three-dimensional structure of one of the IPR proteins, the birch pollen allergen Bet v 1 [3] (PDB entry 1bv1).

Figure 1 shows the alignment of 12 IPR- and 6 MLP sequences (for clarity, only a representative subset of the known sequences has been included). The alignment, which is based on 94 IPR and 13 MLP sequences, was initially performed with the pro-

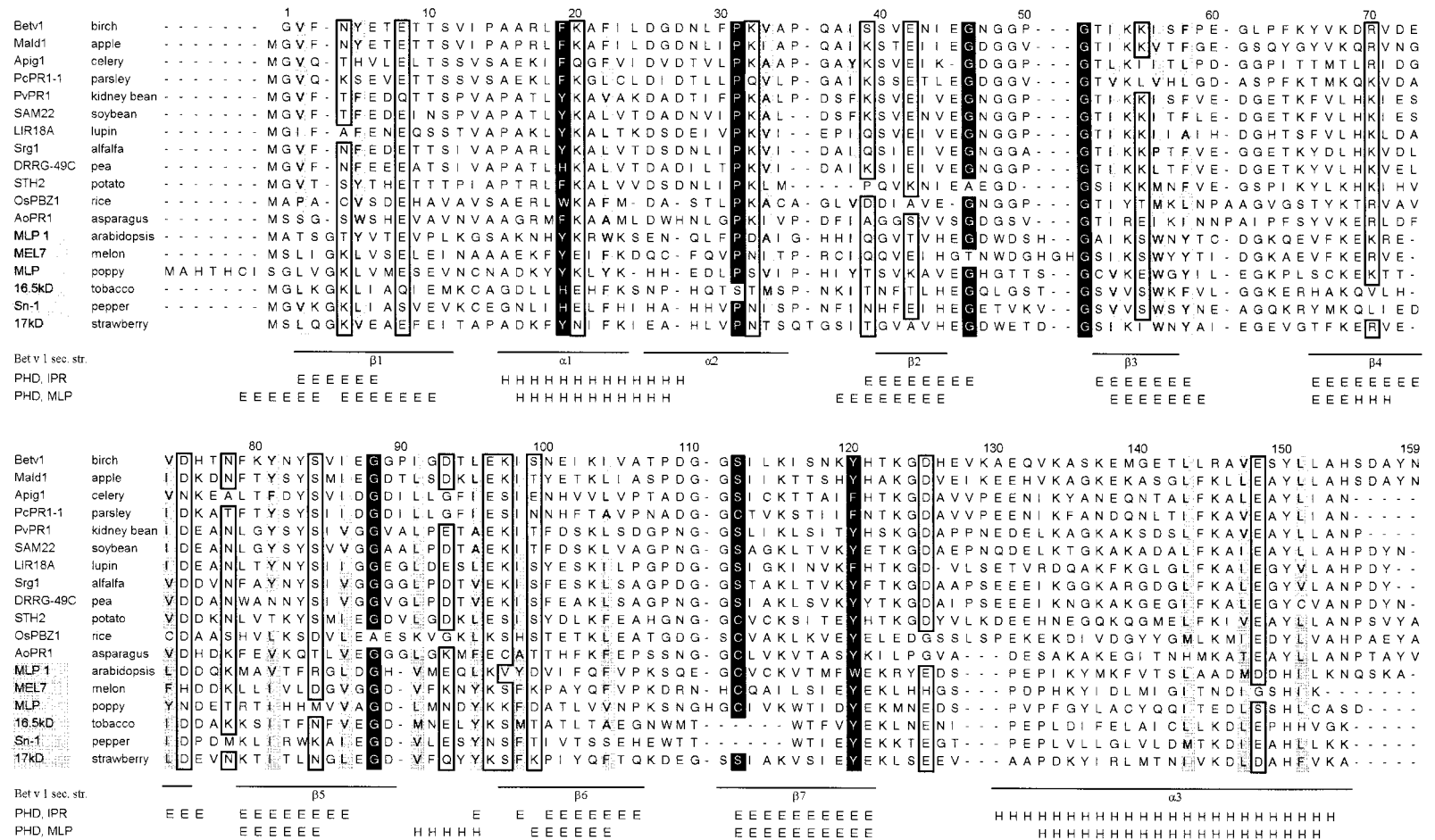


Figure 1. Alignment of 12 IPR and 6 MLP sequences, the latter indicated by gray shading. Although only a subset is shown, the alignment is based on 94 IPR and 13 MLP sequences. Conserved (>90%) residues are coloured black; positions with conserved non-polar side chains are marked with gray, polar are boxed. Below the alignment the location of secondary structure elements in the X-ray structure of Bet v 1 is indicated with horizontal bars. In addition is shown the result of a secondary structure prediction of the aligned IPR sequences and the aligned MLP sequences, respectively, performed using the PHD programme [9, 10]; 'H' refers to a predicted helical conformation, 'E' to β -strand conformation. The sequence numbering is that of Bet v 1. The plant source and NCBI accession code of the sequences shown is as follows: Bet v 1 (*Betula verrucosa*), gi1542869; Mal d 1 (*Malus domestica*), gi1313967; Api g 1 (*Apium graveolens*), gi763063; PcPR1-1 (*Petroselinum crispum*), gi130822; PvPR1 (*Phaseolus vulgaris*), gi21043; SAM22 (*Glycine max*), gi18743; L1R 18 A (*Lupinus luteus*), gi1039333; Srg1 (*Medicago sativa*, gi1136332; DRRG-49C (*Pisum sativum*), gi118934; STH2 (*Solanum tuberosum*), gi169550; PBZ1 (*Oryza sativa*), gi1215811; AoPR1 (*Asparagus officinalis*), gi16078; MLP 1 (*Arabidopsis thaliana*), gi2129641; MEL7 (*Cucumis melo*), gi1843442; MLP (*Papaver somniferum*, gi294060; 16.5 kD (*Nicotiana tabacum*), gi82170; Sn-1 (*Capsicum annuum*, gi2129926; 17 kD (*Fragaria vesca*), gi2465015.

gram CLUSTAL W [11] and then modified taking into account the known structure of Bet v 1.

Immediately apparent from the alignment is the overall similarity in the extent of the polypeptide chains, which are from 145 to 162 amino acids in length. In fact, 102 out of the 107 protein sequences seem to begin at the position corresponding to Gly₁ in Bet v 1 (in the cases where it has been investigated the N-terminal methionine is cleaved off posttranslationally). This is in accordance with the structure of Bet v 1, where the amino group of Gly₁ is involved in two hydrogen bonds to that the N-terminus cannot be shortened or extended without disruption of the local structure. The five exceptions include a *PR-10* gene from sorghum (*S. bicolor*), which begins one codon earlier than the vast majority, and the four *MLP* sequences from opium poppy, which are 7 or 8 codons longer at the 5' end, perhaps reflecting a difference in the structure at the protein N-termini in these cases.

Only a few residues are conserved in both protein families, these being His/Phe/Trp/Tyr₁₉, Pro₃₁, Gly₄₆, Gly₅₁, Gly₈₈, Gly₁₁₁, Cys/Ser₁₁₂, and Phe/Trp/Tyr₁₂₀ (numbering according to Bet v 1). However, if only the distribution of polar (C, D, E, K, N, Q, R, S, T) and apolar residues (A, F, I, L, M, P, V, W, Y) is considered, the overall pattern is quite consistent. An exception is the stretch of sequence corresponding to the first half of the long C-terminal helix, α 3, of Bet v 1, a part of the primary structure which is little conserved even among the IPR proteins alone. Most notably, side chain hydrophobicity is conserved at sequence positions which in Bet v 1 are occupied by residues forming the three important hydrophobic clusters [Ile₉₁, Leu₉₅, Ile₉₈, Tyr₁₂₀], [Ile₃₈, Ile₅₆, Phe₅₈], and [Phe₁₉, Phe₂₂, Ile₁₀₄]. An usually large internal cavity is present in the molecule, which therefore does not have a central hydrophobic core.

Shown below the sequence alignment is the secondary structure of the X-ray structure of Bet v 1 along with the secondary structure prediction of the IPR family and the MLP family, respectively. The prediction was performed using the PHD method [9, 10], which reportedly has an accuracy of 72% on average. The programme predicts the conformation of each amino acid in the sequence as being either extended (E), helical (H), or neither. For Bet v 1 the programme predicts the structure with an accuracy of 72.3% (not shown). Improved accuracy can be obtained by performing a PHD prediction on alignments of homologous sequences; Figure 1 shows the results of a PHD prediction of 25 aligned IPR sequences (only

one from each species of many almost identical isoforms of, for example, the tree pollen allergens was included) and of the 13 MLP sequences, respectively. For the IPR proteins the prediction is 83.6% in agreement with the structure of Bet v 1, and in the case of the MLP sequences this number is 71.2%. One of the larger discrepancies in both cases is at the position of the α 2 helix with the unusual conserved proline in the middle, but apart from that the location of the secondary structural elements is in excellent agreement throughout the peptide chain. Not only does this strongly suggest that the proteins are indeed of common origin, but also that the overall structure is remarkably conserved. This is sustained by the few and short gaps present in the alignment, a fact that indicates that the extent of structural elements is very much the same in the proteins with short connecting loops, as it is the case in Bet v 1. It is, however, not unexpected that the agreement is better among the IPR proteins than between the MLP's and Bet v 1 which only share between 14% and 24% identity and are liable to show local differences in some parts of the structure, e.g. at the N-terminus of α 3.

In conclusion, we find that there is compelling evidence that the MLP and the IPR proteins are indeed related. Consequently, they are likely to play similar roles in the different plants, which should be kept in mind in further studies of members of both families. The exact biological function of the molecules remains to be revealed as does the reason why they exist in two different subfamilies.

Acknowledgements

This research was supported by a grant from the Natural Sciences and Engineering Research Council of Canada. P.O. was supported by a grant from the Carlsberg Foundation.

References

1. Bufe A, Spangfort MD, Kahlert H, Schlaak M, Becker W-M: The major birch pollen allergen, Bet v 1, shows ribonuclease activity. *Planta* 199: 413–415 (1996).
2. Crowell DN, John ME, Russell D, Amasino RM: Characterization of a stress-induced, developmentally regulated gene family from soybean. *Plant Mol Biol* 18: 459–466 (1992).
3. Gajhede M, Osmark P, Poulsen FM, Ipsen H, Larsen JN, van Neerven RJJ, Schou C, Lowenstein H, Spangfort MD: X-Ray and NMR structure of Bet v 1, the origin of birch pollen allergy. *Nature Struct Biol* 3: 1040–1045 (1996).

4. Matton DP, Brisson N: Cloning, expression, and sequence conservation of pathogenesis-related gene transcripts of potato. *Mol Plant-Microbe Interact* 2: 325–331 (1989).
5. Moiseyev GP, Fedoreyeva LI, Zhuravlev YN, Yastnetskaya E, Jekel PA, Beintema JJ: Primary structures of two ribonucleases from ginseng calluses; new members of the PR-10 family of intracellular pathogenesis-related plant proteins. *FEBS Lett* 407: 207–210 (1997).
6. Nessler CL, Burnett RJ: Organization of the major latex protein gene family in opium poppy. *Plant Mol Biol* 20: 749–752 (1992).
7. Nessler CL, Kurz WGW, Pelcher LE: Isolation and analysis of the major latex protein genes of opium poppy. *Plant Mol Biol* 15: 951–953 (1990).
8. Pozueta-Romero J, Klein M, Houlné G, Schantz M-L, Meyer B, Schantz R: Characterization of a family of genes encoding a fruit-specific wound-stimulated protein of bell pepper (*Cap-sicum annuum*): identification of a new family of transposable elements. *Plant Mol Biol* 28: 1011–1025 (1995).
9. Rost B, Sander C: Prediction of protein secondary structure at better than 70-percent accuracy. *J Mol Biol* 232: 584–599 (1993).
10. Rost B, Sander C: Combining evolutionary information and neural networks to predict protein secondary structure. *Proteins* 19: 55–72 (1994).
11. Thompson JD, Higgins DG, Gibson TJ: CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl Acids Res* 22: 4673–4680 (1994).
12. van Loon LC, Pierpoint WS, Boller T, Conejero V: Recommendations for naming plant pathogenesis-related proteins. *Plant Mol Biol Rep* 12: 245–264 (1994).
13. Warner SAJ, Scott R, Draper J: Characterisation of a wound-induced transcript from the monocot asparagus that shares similarity with a class of intracellular pathogenesis-related (PR) proteins. *Plant Mol Biol* 19: 555–561 (1992).