Note

Structural Classification of Plant Chitinases: Two Subclasses in Class I and Class II Chitinases

Tomohiro Araki* and Takao Torikata

Department of Biochemistry, School of Agriculture, Kyushu Tokai University, Aso, Kumamoto 869-14, Japan

Received August 17, 1994

For the classification of plant chitinases, phylogenetic relationships were analyzed besides the established classification of domain structure and C-terminal extension sequences. Two genetically different subclasses (high and low molecular weight) were found for the main structure of class I and class II chitinases in their phylogenetic trees. The genetic distance of these subclasses showed that the high molecular weight subclass may be an ancestral molecule.

Plant chitinases are considered host-synthesized proteins induced in response to pathogen attack and to contribute to self-defence systems.1-5 Chitinases are classified into three classes: I, II, and III.6 Class I chitinases were originally found as basic chitinases, have a molecular mass of about 33 kDa, and are composed of a N-terminal domain, a hinge region, and a main structure. Class II chitinases have a molecular mass of about 28 kDa, mainly have acidic pKα, and are composed of only the main structure that is homologous to that of class I chitinases. Class III chitinases have the same molecular mass of class II chitinases, but are structurally different from former two classes.

The N-terminal (cysteine-rich) domain of class I chitinases is similar to the wheat germ agglutinin (WGA) domain.7) The single domain protein corresponding to one of WGA domains was found as hevin. This finding suggests that the class I chitinase is the product of a gene incorporating the hevin gene.8)

We have reported the complete amino acid sequence of yam acidic class I chitinase.9,10) However, the sequence similarities between this enzyme and other class I chitinases were 40-50%.

In this study, we analyzed the phylogenetic relationship of plant chitinases and described a new structural classification of chitinases by protein genealogies of chitinases.

The protein sequence data were obtained from Protein Sequence Database of the Protein Identification Resource (PIR) at National Biochemical Research Foundation, U.S.A. and the nucleotide sequence data were from DNA Data Bank of Japan, National Institute of Genetics, Japan; EMBL Nucleotide Sequence Data Library, European Molecular Biology Laboratory, Germany; Genetic sequence Data Bank Gen Bank, U.S.A. data bank systems.

Phylogenetical analysis was performed using the "ODEN" program of the Molecular Evolutionary Analysis System for DNA and Amino Acid Sequences at DNA Research Center, National Institute of Genetics, Japan.

Complete amino acid sequences and nucleotide sequences of the chitinase moiety were collected from the database. The nucleotide sequences were translated into amino acids for phylogenetic analysis.

Three trees were constructed for plant chitinases: 1, main structures of class I chitinases and class II chitinases; 2, N-terminal domains of class I chitinases; 3, class III chitinases. As the main structure of class I chitinase is homologous to class II chitinase, this region was analyzed with class II chitinases. The hinge regions of class I chitinases, which have inconsistent sizes and sequences, were not analyzed in this study. The amino acid sequences of each group were aligned, and the protein genealogies were constructed by the neighbor-joining method11) using the mutational distance obtained by Kimura's method.12) The distance was counted for

* To whom correspondence should be addressed.

Fig. 1. Phylogenetic Tree of Thirty-five Main Structure Sequences from Class I and Class II Chitinases.

Class I-L (a) and class II-L (b) subclasses are boxed in. Class II chitinases are indicated by solid squares. The accession numbers of sequence database (see the text) for the chitinase sequences are as follows. Zea mays 2, M84165; Zea mays 3, M84164; Phaseolus vulgaris 2, X57187; Brassica napus 2, X61488; Diococereus japonicus, A44309; Populus trichocarpa 1, M23337; Urtica dioica, M23332; Populus trichocarpa 2, M23336; Petunia hybrida, X51427; Lycopersicon esculentum 4, Z15139; Nicotiana tabacum 5, X51425; Nicotiana tabacum 6, X51426; Lycopersicon esculentum 3, Z15141; Piaum sativum, X63899; Solanum tuberosum 1, X14133; Lycopersicon esculentum 1, Z15140; Solanum tuberosum 2, X07130, X15494; Nicotiana tabacum 4, X51599, X64519, M15173; Nicotiana tabacum 1, S44869; Nicotiana tabacum 2, X16938, X16939; Nicotiana tabacum 3, X64518; Lycopersicon esculentum 2, Z15138; Solanum tuberosum 3, X67693; Phaseolus vulgaris 1, M13968, S43926; Arabidopsis thaliana, M38240; Brassica napus 1, M95835; Secale cereale 1, JC2071; Hordeum vulgare, M62094; Secale cereale 2, JN0884; Allium sativum 1, M94105; Allium sativum 2, M94106; Zea mays 1, L00973; Oryza sativa 3, X5467; Oryza sativa 2, X56063; Oryza sativa 1, S15997.
Fig. 2. Phylogenetic Tree of Twenty-seven N-Terminal Domain Sequences from Class I Chitinases.

Class I-L subclass is boxed in. For the accession numbers of chitinase sequences, see Fig. 1. Duplicated N-terminal domains of *Urtica dioica* (nettle) lectin are designated -a and -b from the N-terminus of polypeptide chain.

each position that contained no deletion in all chitinases.

Two phyletic lineages were found in the protein genealogy of class I and class II chitinases (Fig. 1); one was a high molecular weight type chitinase and the other was a low molecular weight type chitinase that contained deletions in the molecule. High molecular weight subclasses of class I chitinases (class I-H) and class II chitinases (class II-H) showed no genetic distance. However, low molecular subclasses (class I-L and class II-L) showed independent lineages. This shows the genes of low molecular weight subclasses evolved from the high molecular weight subclass gene. As *P. trichocarpa* chitinases and *U. dioica* lectin have no deletion in the polypeptide chain, the molecular type is considered H-type. However, these chitinases did not belong to class I-H subclass. The genetic distance found in *P. trichocarpa* chitinases may be responsible for the difference of species. The critical classification of arboresous chitinases should be done when the sequence data of these chitinases is accumulated. The genetic distance in *U. dioica* lectin is considered to reflect the differential pathway of gene duplication and gene recombination.

To estimate the relation of the N-terminal domain and the main structure of class I chitinase, the protein genealogy of the N-terminal domain was constructed (Fig. 2). The tree obtained showed the same relationship of two lineages found in the tree of the main structure. That is, the N-terminal domains from class I-L formed a divergence from those of class I-H. Further, the N-terminal domains of class I-H had no deletions in the molecule. This strongly suggests that the class I-L gene evolved after the incorporation of the N-terminal domain gene into the class II-H gene.

The protein genealogy of class III chitinases showed no definitive evidence for evolutionary lineage (data not shown). In spite of the low sequence similarity (40-50%), no subclasses were found in this class.

The classification of plant chitinases is shown in Fig. 3. Class I comprises class IH and class IL subclasses. Further, Class IH comprises class IA (contain C-terminal extension sequence) and class Ib. Class II comprises class III and class II. There is no genetic difference between class IIH and the main structure of class IH. The reported class IV chitinase corresponds to class IL in this classification.

The possible relationship of these classes and related proteins is shown in Fig. 4. An ancestral single domain gene evolved into a double domain gene by gene duplication and evolved into class the II-H gene. The class II-H gene evolved into the class I-H gene by incorporating the N-terminal domain gene. This pathway is supported by the presence of direct repeats at the 5' and 3' flanking regions of the class I chitinase gene. As the trees of N-terminal domain and the main structure of class I chitinases are similar, there is no genetic distance before and after the incorporation of N-terminal domain gene. Class I-H and class II-H genes then evolved into class I-L and class II-L genes, respectively, by another pathway. The *U. dioica* gene may be formed by an independent pathway from class II-H gene incorporating a double N-terminal domain gene. The *U. dioica* gene should be classified into another class when the sequence data is accumulated. Class III chitinase gene may have evolved from a different ancestral gene by convergent evolution.

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

References