Short Communication

The Complete Amino Acid Sequence of Lectin-C from the Roots of Pokeweed (Phytolacca americana)

Ken-ichi Yamaguchi, Ayumi Mori, and Gunki Funatsu

Laboratory of Protein Chemistry & Engineering, Faculty of Agriculture, Kyushu University, Fukuoka 812, Japan
Received March 10, 1995

The complete amino acid sequence of pokeweed lectin-C (PL-C) consisting of 126 residues has been determined. PL-C is an acidic simple protein with molecular mass of 13,747 Da and consists of three cysteine-rich domains with 51-63% homology. PL-C shows homology to chitin-binding proteins such as wheat germ agglutinin, and all eight cysteine residues in the three domains of PL-C are completely conserved in all other chitin-binding domains.

Pokeweed lectin is a lectin specific for N-acetylglucosamine-containing saccharides and stimulates peripheral lymphocytes to undergo mitosis by binding to their cell surfaces. Waxdal\textsuperscript{11} isolated five mitogenic lectins (Pa-1–Pa-5) from the roots of pokeweed (Phytolacca americana) and demonstrated that they have distinct physiological properties and different biological activities; the predominant mitogens, Pa-2 and Pa-4, bind to the same receptors on erythrocytes, but Pa-2 is a hemagglutinin while Pa-4 is not. In a preceding paper,\textsuperscript{21} we isolated three mitogenic lectins (PL-A, PL-B, and PL-C) from young pokeweed roots by a different procedure and showed that they are homologous proteins, of which PL-A and PL-C correspond to Pa-2 and Pa-4, respectively, and PL-B is a novel mitogenic lectin. However, their molecular masses are unclear, and the sugar-content of PL-C did not coincide with that of Pa-4.

In this paper, we describe the analysis of amino acid sequence of PL-C, present most abundantly in the pokeweed roots throughout the year.

Before the amino acid sequence analysis, the molecular mass of PL-C was estimated to be approximately 15 kDa by tricine SDS–PAGE in the presence of 2-mercaptoethanol (2-ME) according to the method of Schägger and Jagow.\textsuperscript{31}

On the basis of this molecular mass, the amino acid composition of PL-C obtained by chemical analysis was calculated to be Asx\textsubscript{11}, S(12), Thr\textsubscript{5}, S(4), Ser\textsubscript{6}, S(8), Glx\textsubscript{17}, 2(17), Pro\textsubscript{3}, 2(9), Gly\textsubscript{22}, 1(22), Ala\textsubscript{2}, 3(2), Val\textsubscript{3}, 0(3), Met\textsubscript{0}, 0(1), Leu\textsubscript{4}, 2(14), Tyr\textsubscript{2}, 2(6), Phe\textsubscript{2}, 1(2), Lys\textsubscript{4}, 2(4), His\textsubscript{3}, 0(7), Arg\textsubscript{5}, 2(5), Trp\textsubscript{1}, 2(3), and Cys\textsubscript{2}, 2(2). The N-terminal amino acid sequence of PL-C was -Asp-Pro-Thr by carboxypeptidase Y (Oriental Yeast Co., Ltd.)-digestion.

The amino acid sequence of PL-C was determined by sequencing the peptides obtained by lysylendopeptidase- and thermolysin-digestions of the reduced and S-pyridylethylated (RPe)-PL-C, prepared by the method of Friedman et al.\textsuperscript{44} Digestions of RPe-PL-C with lysylendopeptidase (Wako Pure Chem. In.) and thermolysin (Sigma Chemical Co.) were done using 1/100 (w/w) enzymes in 4 M urea-50 mM Tris–HCl buffer, pH 9.0, and in 8 M urea-100 mM ammonium bicarbonate solution containing 2.5 mM CaCl\text sub{2}, pH 7.8, respectively, at 37°C for 3 h. Fragmentation of the large peptides obtained above was done using 1/50 (w/w) chymotrypsin (Sigma Chemical Co.) in 100 mM ammonium bicarbonate solution, pH 7.8, at 37°C for 2 h. The peptides were separated by reverse-phase (RP)-HPLC with a YMC-GEL C4 column (4.6 × 250 mm or 4.6 × 150 mm) by a linear gradient of acetonitrile from 0 to 30% either in 0.1% trifluoroacetic acid solution or in 5 mM potassium phosphate buffer, pH 6.0. Amino acid analysis of peptide was done with a PICO-TAG amino acid analyzer (Waters Co.) after hydrolysis with constant-boiling HCl containing 0.05% 2-ME in vacuo at 110°C for 24 h. Amino acid sequence analysis was done by either manual Edman degradation using DABITC/PITC double coupling method\textsuperscript{21} or automated Shimadzu gas-phase protein sequencer PSQ-1.

Five peptides (L1–L5) were isolated from the lysylendopeptidase-digest of RPe-PL-C by RP-HPLC. By direct sequencing, peptides L2, L3, and L4 were sequenced completely, and peptides L1 and L5 up to the 16th and 22th residue, respectively. Peptides L1 and L5 were digested with chymotrypsin and the resulting peptides were separated by RP-HPLC, yielding five (L1C1–L1C5) and three peptides (L5C1–L5C3) from peptides L1 and L5, respectively. By sequencing them, the sequences of L1 and L5 were completed. Thus, the sequence of five peptides (L1–L5) was established. From their sequences, peptides L1 and L5 were found to be derived from the N- and C-terminal of PL-C, respectively.

RPe-PL-C was digested with thermolysin and the resulting peptides were isolated by RP-HPLC, yielding 10 peptides (Th1–Th10). Sequencing of these peptides except Th3 were directly completed and peptide Th3 was sequenced up to 13th residue. Peptide Th3 was digested with chymotrypsin, and the resulting peptides (Th3C1–Th3C4) were separated by RP-HPLC and sequenced. Thus, the sequences of all ten thermolytic peptides were established (Fig. 1).

From the overlappings of the peptides obtained by lysylendopeptidase- and thermolysin-digestions of RPe-PL-C, the complete amino acid sequence of PL-C was established. Amino acid sequence analysis of PL-C is summarized in Fig. 1. The amino acid composition of PL-C calculated from the sequence is Asp\textsubscript{10}, Asn\textsubscript{2}, Thr\textsubscript{5}, Ser\textsubscript{6}, Glu\textsubscript{8}, Gln\textsubscript{3}, Pro\textsubscript{3}, Gly\textsubscript{2}, Ala\textsubscript{2}, Val\textsubscript{3}, Met\textsubscript{1}, Leu\textsubscript{4}, Tyr\textsubscript{2}, Phe\textsubscript{2}, Lys\textsubscript{4}, His\textsubscript{3}, Arg\textsubscript{5}, Trp\textsubscript{1}, and Cys\textsubscript{2} which agrees fairly well with that found by chemical analysis of PL-C. PL-C was titrated with 5,5'-dithio-bis(2-nitrobenzoic acid) in the presence of 6 M guanidine–HCl by the method of Ellman.\textsuperscript{61} However, free sulfhydryl group was not detected, indicating that twenty-four cysteine residues are linked by disulfide bonds.

PL-C consists of 126 amino acid residues and its molecular mass was calculated to be 13,747 Da. The occurrence
Fig. 1. Summary of Amino Acid Sequence Analysis of Pokeweed Lectin-C

1 and Th indicate the peptides derived from digests of RPe-PLC with lysylendopeptidase and thermolysin, respectively. L1C, LSC, and Th3C indicate the peptides derived from chymotryptic digests of peptides L1, L5, and Th3, respectively. Sequence data on individual peptides are indicated as follows: ——, sequenced by the DABITC/PTC method; ———, sequenced by gas-phase sequencer; ——, sequenced by carboxypeptidase Y-digestion; ——, not identified by sequencing.

Fig. 2. Comparison of Amino Acid Sequence of Pokeweed Lectin-C with Those of Wheat Germ Agglutinin (WGA-B), Hevein, and Chitin-Binding Domains of Class I Chitinases from Tobacco, Potato, and Rye (RSC-α).

*stands for the domains of PLC and WGA. Identical residues in the three domains of PLC are bold-typed and those in all eleven domains are boxed. Several gaps (—) are inserted to obtain the maximal homology among these domains.

of 18 acidic and 9 basic residues is compatible with the isoelectric point of 4.35 of this protein. The sequences of residues 1-42, residues 43-83, and residues 84-126 of PLC were found to possess intermolecular homology (Fig. 2), indicating that PLC consists of three homologous domains. They are designated domain A, domain B, and domain C in the sequential order, respectively. The number of the identical amino acid residues between domains A and B, domains A and C, and domains B and C are 21 (51% homology), 21 (51%), and 26 (63%), respectively. Eight cysteine residues in the three domains are completely conserved, and five glycine, three glutamine, and two serine residues are identical.

As shown in Fig. 2, the amino acid sequence of PLC shows high similarity to chitin-binding proteins or domains such as wheat germ agglutinin and class I chitinases, and fourteen amino acid residues including eight cysteine residues (Cys4, Gly5, Gly10, Cys13, Cys18, Cys19, Ser20, Gly23, Cys25, Gly26, Cys32, Cys36, Gin37, and Cys40 in the domain A) are absolutely conserved in these chitin-binding domains. Hemagglutinating activity of PLC was remarkably low, but PLC was strongly adsorbed on a chitin column, clearly indicating that PLC is a chitin-binding protein.

These results suggest that PLC apparently arose through fusion and/or duplication of chitin-binding domains and has biological functions related to defence against chitin-containing pathogens and predators in the roots of pokeweed.

Details of the sequencing of PLC will be presented in our following paper.

References