Isolation and Characterization of cDNA Clones Encoding Jacalin Isolectins

Hui Yang‡ and Thomas H. Czapla

From the Department of Biotechnology Plant Breeding Division, Pioneer Hi-Bred International, Johnston, Iowa 50131

Four jacalin cDNA clones (pSKcJA1, pSKcJA3, pSKcJA15, and pSKcJA17) have been obtained from an Artocarpus integrifolia (jackfruit) seed cDNA library. These clones share over 94% sequence homology, and their deduced polypeptide sequences confirm the existence of multiple jacalin isolectins in jackfruit seeds. The deduced amino acid sequences show that jacalin appears to be initially synthesized as a prepropeptide with the following structure: N-signal (21 residues) → propeptide (39 residues) → β-peptide (20 residues) → linker region (4 residues) → α-peptide (133 residues). These observations are supported by Western blot analysis of jackfruit seed extract and by immunoprecipitation of in vitro translated products of both pSKcJA3 transcript and jackfruit seed poly(A)+ RNA. Sequence analysis of the 39-residue propeptide reveals that it has the potential to facilitate proper folding of jacalin protein. The unusual primary structure of jacalin prepropeptide suggests a quite interesting processing of this lectin precursor into mature α- and β-subunits.

Lectins are carbohydrate-specific, cell-agglutinating proteins found widely in plants, animals, and microorganisms (for review, see Eizler, 1985a). Biological roles in defense (Christeens and Raikhel, 1991), in cell-cell recognition (Knox et al., 1976), or as catalytic enzymes (Hangins et al., 1980) have been proposed. Because of the variety of carbohydrate specificities found among them, lectins are useful reagents for studying carbohydrate moieties of glycoproteins and distribution and function of carbohydrates on cell surfaces (Lis and Sharon, 1986). Lectins from plants within taxonomic groups can be distinguished from lectins of less closely related plants by their sequences (for review, see Eizler, 1985a). Plant lectins from Leguminosae or Graminaceae have been most extensively studied. Jacalin is a lectin from Artocarpus integrifolia Lamark (jackfruit), a member of the Moraceae family. Jacalin binds to Calb1-3GalNac on tumor-associated T-antigen (Sasety et al., 1986), to human IgA and IgD (Aucouturier et al., 1987), and has potent mitogenic effects on T- and B-cells (Bunn- Moreno and Campos-Neto, 1981). Jacalin also has an insecticidal activity against Southern corn rootworm (Diabrotica undecimpunctata howardi), a pest of corn (Young et al., 1989, 1991). Sequence analysis revealed that both α- and β-subunits have variations in each chain, indicating the existence of multiple jacalin isoforms (Young et al., 1989, 1991). It has been reported that jacalin is a αβ22 tetramer of Mr = 43,000-54,000, which may be glycosylated (Vijayakumar and Forrester, 1986; Aucouturier et al., 1987). Although x-ray crystallography studies on jacalin structure are underway (Banejee et al., 1991), little else is known about the molecular structures of jacalin isoforms.

In this study, we report the isolation and characterization of four cDNA clones from A. integrifolia seeds, which encode different isoforms of jacalin lectin. We present evidence that jacalin lectin is synthesized in vivo as an unusual prepropeptide, which probably undergoes both co- and post-translational processing. The possible functions of the 39-residue propeptide region are also discussed.

MATERIALS AND METHODS

Amplification and Subcloning—Primer 5 and primer 3 (synthesized by the Nucleic Acid Facility at Iowa State University) were added at a final concentration of 0.4 µM to a mixture (40 µl) containing 67 mM Tris, 1.4 mM dNTPs (Pharmacia LKB Biotechnology Inc.), 10.7 mM (NH₄)₂SO₄, 6.6 mM MgCl₂, 1 unit of Taq DNA polymerase (Perkin-Elmer/Cetus), and 50 ng of A. integrifolia genomic DNA. This mixture was initially denatured at 94 °C for 4 min and then incubated in an ERICOMP thermocycler (San Diego, CA) for 40 cycles: 1 min at 94 °C, 1 min at 37 °C, and 2 min at 72 °C, followed by a 10-min extension at 72 °C added to the final cycle. The reaction products were resolved by agarose gel electrophoresis (2% NuSieve agarose gel, 1% NuSieve agarose, and 1% agarose), and the band of interest was eluted from the gel, blunted-ended with T4 polymerase, digested by EcoRI and Sall, and subcloned between the EcoRI and SalI sites of pBluescript II KS(+) DNA. The resulting plasmid (pKSG53) was used as a hybridization probe.

RNA Isolation and Northern Blot Analysis—Immature fruits and fully developed leaves were collected from jackfruit trees in Hawaii. Seeds (about 18 mm in diameter) were obtained from the fruits and were used for further studies. Total RNA was isolated from the seeds and leaves by the method of Roberts and Lord (1981). Poly(A)+ RNA was purified by the PolyTract mRNA Isolation System (Promega, Madison, WI) following the supplier's protocol. For Northern blot analysis, 1 µg of poly(A)+ RNA was fractionated by 1% agarose gel electrophoresis (Liu and Chou, 1990), transferred to Amersham N-Hybrid membrane (Amersham Corp.) as described (Sambrook et al., 1989). The membrane was baked for 1 h, UV irradiated for 45 s, and hybridized with 32P-labeled EcoRI Sall fragment of pKSG53. Hybridizations were performed overnight at 65 °C in 5% SSC (1× SSC = 0.15 M NaCl, 0.15 M sodium citrate) containing 0.015 M sodium phosphate, pH 7.5, 2.5 × Denhardt's solution (1× Denhardt's solution = 0.2 g/liter Ficoll, 0.2 g/liter bovine serum albumin, and 0.2 g/liter polyvinylpyrrolidone), 100 mg/ml sheared salmon sperm DNA, and 32P-labeled probe (5 × 10⁶ cpm/ml), after prehybridization in the solution described above without probe for 3 h at 65 °C. The probe was labeled with [α-32P]dATP by the random prime method (Feinberg and Vogelstein, 1983) and purified using a Quick spin G-25 column (Boehringer Mannheim). Following hybridization, membranes were washed in 0.2 × SSC, 0.1% SDS for 20 min at room temperature, and then for 2 min at 65 °C. Blots were then exposed to Kodak XR films with intensifying screens.
Cloning and Sequencing Jacalin cDNA—Poly(A)" RNA was examined for the presence of jacalin message by Northern blot analysis before being used for cDNA synthesis (see "Results and Discussion"). The double-stranded cDNA containing both EcoRI and NotI restriction sites was made from jackfruit seed mRNA using the Capture Clone Magnetic cDNA Synthesis System (Promega) according to the manufacturer's protocol. The cDNA was ligated to dephosphorylated EcoRI-NotI XZAPII arms (Stratagene). After in vitro packaging with Gigapack II packaging extract (Stratagene), 5 × 10^6 plaque-forming units were screened in Escherichia coli SURE cells (Stratagene). Plaque lifts on nitrocellulose filters (Schleicher & Schuell) were prepared with hybridized "P-labeled EcoRI-SalI fragment of pKSg53 at 65 °C overnight under the same conditions as described above for Northern blot analysis. Filters were washed in 0.2 × SSC, 0.1% SDS at room temperature for 20 min, and then for 2 min at 65 °C. Plates giving the strongest hybridization signal were isolated and rescreened under the same condition. Phagemids were obtained from purified positive plaques through in vivo excision (Stratagene's protocol), and inserts were sequenced using a model 373A DNA sequencing system (Applied Biosystems, Foster City, CA) at the Iowa State University Nucleic Acid Facility. Sequencing was carried out with a Tag Dye Deoxy terminator cycle sequencing kit (Applied Biosystems), using M13-20 primer, reverse primer, and a primer complementary to an internal segment of the cDNA clones, respectively. All sequences were determined three times in each direction.

Southern Blot Analysis—Genomic DNA isolated from A. integrifolia was cleaved with various restriction enzymes for 24 h, resolved and rescreened under the same condition. Phagemids were obtained from purified positive plaques through in vivo excision (Stratagene's protocol), and inserts were sequenced using a model 373A DNA sequencing system (Applied Biosystems, Foster City, CA) at the Iowa State University Nucleic Acid Facility. Sequencing was carried out with a Tag Dye Deoxy terminator cycle sequencing kit (Applied Biosystems), using M13-20 primer, reverse primer, and a primer complementary to an internal segment of the cDNA clones, respectively. All sequences were determined three times in each direction.

In Vitro Translation and Immunoprecipitation—pSKcJA3 was linearized with NotI for sense RNA transcripts, and "caps" transcript was generated using the Riboprobe II core system (Promega). The manufacturer's protocol was followed with the modification of adding "7GpppG at a final concentration of 0.3 mM to the reaction. One μg of capped sense transcript or poly(A)" RNA from jackfruit seeds was then translated in a wheat germ extract (Promega) using 60 μCi of [35S]methionine (Amersham Corp., 1250 Ci/mmol) per 50-μl reaction. The in vitro translation products were immunoprecipitated with anti-Maclura promotiera antibody (Bethyl Laboratories, Montgomery, TX; Namjuntra et al., 1989) by the method of Firestone and Winguth (1990). Immunoprecipitated products were analyzed by SDS-polyacrylamide gel electrophoresis on 15% acrylamide gel and visualized by autoradiography.

Analysis of Jacalin Lectins Synthesized in Vivo—Three g of seeds (about 18 mm in diameter) of immature jackfruits were ground in 40 ml of phosphate-buffered saline (0.01 M sodium phosphate, pH 6.8, and 137 mM NaCl). The crude extract was incubated at 4 °C overnight with constant stirring. Insoluble material was removed by centrifugation and proteins in the supernatant were precipitated by addition of (NH4)2SO4 to a final concentration of 60%. The pellet was collected, dissolved in phosphate-buffered saline, and analyzed on SDS-polyacrylamide gel electrophoresis. The proteins were then transferred to a polyvinylidene difluoride membrane (Bio-Rad) in electrophoresis buffer (14.4 g/liter glycine, 6 g/liter Tris base, and 10% methanol) and analyzed on SDS-polyacrylamide gel electrophoresis on 15% acrylamide gel and visualized by autoradiography.

RESULTS AND DISCUSSION

Amplification of a Jacalin-specific Sequence—A probe for a coding sequence of jacalin was generated using polymerase chain reaction (PCR, 1 Saiki et al., 1985). Two degenerate oligonucleotides were derived from jacalin α-peptide (Young et al., 1989); primer 5 (a mixture of 16 sequences) corresponds to the sequence GKAFFDG, and primer 3 (a mixture of eight sequences) corresponds to the sequence GDFQVVY. The restriction site added to each primer was included to facilitate subcloning (Fig. 1A). One fragment of about 110 bases long was generated by PCR using jackfruit genomic DNA as the amplification template (Fig. 1B, compare lane 2 with lane 3). This PCR fragment was subcloned into pBluescript II ks plus (Stratagene), and the resulting clone (pKSg53) was sequenced. The data confirmed that the 110-base pair fragment could encode a portion of the jacalin sequence (sequence not shown).

Northern Blot Analysis—to examine the distribution of jacalin lectin message in seed and leaf, poly(A)" RNAs from both tissues were hybridized with the insert of pKSg53. An abundant message of about 1 kilobase was detected in seed mRNA (Fig. 2, lane 1), but no message was seen in mRNA from the leaves (Fig. 2, lane 2). The absence of jacalin transcript in our collected leaves was not surprising since most lectins are primarily stored in seed tissues (Etzler, 1985b).

Isolation, Sequence, and Characterization of cDNA Clones—About 0.3% of the plaques in a cDNA library derived from A. integrifolia seed gave a positive signal. Twenty putative jacalin
clones were randomly selected. Restriction map analysis indicated that all these cDNA clones were most probably derived from a single species of mRNA with differences in size arising from incomplete cDNA synthesis. The complete nucleotide sequence of one of the clones, pSkcJA3, and its deduced amino acid sequence are displayed in Fig. 3. An open reading frame of 651 nucleotides encodes a putative polypeptide of 217-amino-acid residues with a calculated molecular weight of 23,517. The sequence TGCAGAEGCT around the ATG frame of 651 nucleotides encodes a putative polypeptide of residues 1-217 of a protein with a calculated molecular weight of 241 CCATGGGGAGCCAAAGTAAGCACTAGCTCCAATGGTARAGCTTTTGATGAGTCTATTAAAAAAAAAA. The putative signal sequence is from residues 1-21 of the protein sequence, and the polyadenylation sites are between residues 218 and 217.

**Fig. 2.** RNA blot analysis of jacalin mRNA levels. Poly(A)⁺ RNA (1 µg), isolated from jackfruit seed (lane 1) or leaf (lane 2), was separated on a 1% agarose gel. After transferring the RNA to a nylon membrane, the filter was hybridized with a '*'P-labeled EcoRI-Sal1 sequence of pSKcJA3.

Translation is assumed to start at the first initiation codon at nucleotides 22-24. The putative signal sequence matches the sequences of the reported jacalin subunits, and the polyadenylation sites are also indicated. The 3' untranslated region contains two in-frame termination codons and two potential polyadenylation signals and ends in a polyadenylic acid tail.

By using the rule for protein processing of von Heijne (1986), the most probable cleavage site in the deduced polypeptide is between residues 21 and 22, suggesting that amino acid residues 21-22 correspond well with the consensus sequence of the translational initiation context TAAA-CATGGCCT (9 out of 12 nucleotides identical, Joshi, 1987). The untranslated region contains two in-frame termination codons and two potential polyadenylation signals and ends in a polyadenylic acid tail.

**Fig. 3.** Nucleotide sequence and deduced amino acid sequence of pSkcJA3. Translation is assumed to start at the first ATG (nucleotides 22-24) and continues until the first termination codon, TAA (nucleotides 673-675). The putative signal sequence is marked with dashed underlines. Residues 61-80 and residues 218-217 match the sequences of the α- and β-subunits of jacalin, respectively, and are shown by bold letters and bold underlines. Two translational termination codons (dotted underlines) and the polyadenylation site (double underlines) are also indicated.

**Fig. 4.** Alignment of the deduced amino acid sequences of cDNA clones pSkcJA1, pSkcJA3, pSkcJA15, and pSkcJA17, and the published protein sequences of jacalin lectins. Complete amino acid residues deduced from clone pSkcJA3 are shown. — denotes residues identical to that of pSkcJA3. Individual amino acid residues are displayed by letters at positions where there is a difference. Bold letters represent differences in the 3' untranslated region.
examine whether the variations in amino acid sequence arose from the genetic variants of their mRNAs, we decided to sequence more putative cDNA clones. The deduced amino acid sequences of clones pSKcJA1, pSKcJA15, and pSKcJA17 are displayed in Fig. 4, along with the amino acid sequence deduced from clone pSKcJA3 and the published protein sequences of jacalin lectins. pSKcJA1, pSKcJA15, and pSKcJA17 showed over 94% homology to pSKcJA3 both in the nucleotide sequence (data not shown) and in the deduced protein sequence (Fig. 4), indicating that cDNA clones pSKcJA1, pSKcJA15, and pSKcJA17 also encode jacalin lectins. Fig. 4 also reveals that 10 out of 13 variant amino acid residues found in jacalin lectins are represented among the amino acid sequences deduced from the cDNA clones. In fact, the deduced β-chain sequences from clones pSKcJA1, pSKcJA15, and pSKcJA17 are almost identical to the published amino acid sequences of p2-, Dl-, and @3-subunits (Young et al., 1989), respectively, with the exception of the second residue (lysine) encoded by pSKcJA1. These data argue against the possibility that the differences in the nucleotide sequence of these clones are the result of cloning artifact or sequencing error. It is thus quite clear that the isoforms of jacalin lectin are genetically determined. We conclude that the amino acid variations found in jacalin lectin represent the sequence differences among various jacalin isoforms, each of which is encoded by its own mRNA. The difference between our cDNA derived amino acid sequence(s) and the published protein sequence of Young et al. (1989, 1991) may be attributed to a variation within the jackfruit population.

The calculated pl (isoelectric point) values for the deduced β-chain sequence of clones pSKcJA1, pSKcJA3, pSKcJA15, and pSKcJA17 are 10.80, 6.45, 9.51, and 4.24, respectively. These very different pl values most likely arise from the different numbers of charged residues among various β-subunit isoforms (Fig. 4). The deduced α-chain sequence of clone pSKcJA1 has a calculated pl of 7.52, whereas the pl values of the other three deduced α-chains have an identical value of 5.86. Two potential N-linked glycosylation sites were found at Asn-101 and Asn-158 for pSKcJA3, pSKcJA15, and pSKcJA17, whereas only one such site (Asn-101) was found in pSKcJA1.

Southern Blot Analysis—Different isoelectrons could be encoded either by different genes at different locations (i.e. at different loci), or by different genes at the same location (i.e. by different alleles). Genomic DNA isolated from A. integrifolia seeds was digested by EcoRV, EcoRI, HindIII, HindII, BbsI, and XbaI, respectively, separated on a 0.8% agarose gel, and probed with 32P-labeled EcoRI-SalI fragment of pKSg53. Positions of DNA size markers (HindIII digest of λDNA) are shown on the left.

extracts (lane 3), in addition to the mature forms of the 14–17-kDa jacalin proteins (lanes 2 and 3) previously seen by others (Aucouturier et al., 1987; Hagiwara et al., 1988). The presence of a ~25-kDa protein supports the speculation from the cDNA sequence that jacalin lectins are synthesized initially as a molecule larger than the mature protein. Immunoprecipitation of in vitro translated products revealed two bands of ~24 and ~17 kDa, respectively, for both pSKcJA3 transcript (Fig. 6B, lane 1) and jackfruit poly(A)+ RNA (Fig. 6B, lane 2). This ~24-kDa immunoreactive product corresponds to the size of jacalin precursor as predicted from the cDNA sequence. Translation from an in-frame internal methionine codon (Met-56) would produce a 17,735-dalton protein (Fig. 3 above) which is similar in size to the ~17-kDa product (Fig. 6B). These data strengthen the contention that clone pSKcJA3 (and, very likely, clones pSKcJA1, pSKcJA15, and pSKcJA17) encodes jacalin lectin and that jacalin protein is synthesized as a ~25-kDa preproprotein in vivo.

The Possible Function of the 39-residue Propeptide—A number of well studied lectin precursors have the structure of NH2–signal peptide → subunit I sequence → several amino acid residues → subunit II sequence–COOH (Hemperly et al., 1982; Butterworth and Lord, 1983; Higgins et al., 1983; Car- rington et al., 1985; Wilkins and Raikhel, 1989). The primary structure of jacalin protein deduced from the cDNA clones is quite unusual because it bears a 39-amino-acid propeptide region (residues 22–60) near the NH2 terminus of the precursor. Lectin precursors from pea (Higgins et al., 1983), wheat germ (Raikhel and Wilkins, 1987), barley (Lerner and Raikhel, 1989), rice (Wilkins and Raikhel, 1989), and garlic (Van Damme, 1991) have proprotein regions in addition to the leader sequence. In all cases, the propeptides are located at the COOH termini as extensions. The function of these COOH-terminal propeptides remains largely unclear, although it has been shown that the COOH-terminal sequence of barley lectin is necessary for the efficient sorting of the protein to the plant cell vacuoles (Bednarek et al., 1990). Sporamin, a storage protein, has a 16-residue propeptide region adjacent to the NH2-terminal signal sequence (Hattori et al., 1985). In a study with transformed tobacco cells, deletion of this pro-domain results in secretion of mature sporamin out of cells rather than translocation to the vacuoles (Matsuoka and Nakamura, 1991). Since jacalin lectin belongs to the category of storage proteins, it may be possible to
assume that at least a portion of the 39-residue propeptide segment is required for translocating jacalin lectin to an intracellular destination, perhaps the vacuole. Two vacuolar proteins in yeast, carboxypeptidase Y (Hasilik and Tanner, 1978) and protease A (Ammerer et al., 1986), are synthesized as a preproprotein that has a propeptide between the signal sequence and the mature polypeptide. These propeptides have been shown to bear a vacuolar targeting signal (Valls et al., 1987; Johnson et al., 1987; Klionsky et al., 1988), which is later cleaved after the arrival of the protein in the vacuole (Woolford et al., 1986).

Protein precursors to carboxypeptidase Y (Winther and Sørensen, 1991), subtilisin (Zhu et al., 1989), and α-lytic protease (Silen et al., 1989; Silen and Agard, 1989) have a propeptide sequence that enables proper folding of the mature protein, presumably either by stabilizing the transition state for a correct folding pathway or by destabilizing incorrect folding intermediates (Silen et al., 1989). Two structural features of peptide sequences important for protein folding include a relatively high frequency of charged amino acids and a relatively higher percentage of amino acid residues favoring α-helical secondary structure (Sadis et al., 1990). As shown in Table I, jacalin propeptide has a predominant percentage of both charged amino acids and α-helix formation elements, as compared to the mature subunits. These features are quite comparable to those chaperone-type of propeptides for carboxypeptidase Y, subtilisin, and α-lytic protease (Table I). Thus, the 39-residue pro-domain of jacalin precursor might also be involved in the correct folding of α- and β-chain. The COOH-terminal propeptide sequence for pea, wheat germ, barley, rice, or garlic lectin precursor has a significantly higher percentage of helical structure than the corresponding mature lectin, but lacks a predominant percentage of charged residues (data not shown).

A Speculation on Processing of Jacalin Precursor into Mature Subunits—The primary structure of jacalin precursor indicates that at least three endoproteolytic cleavages should take place before mature jacalin protein is formed: the cleavage of the signal sequence, the removal of the 39-residue propeptide segment, and the processing between α- and β-subunits (Fig. 7). Because most proteins destined either for export from the cell or to certain intracellular locations are cotranslationally processed (Blobel, 1980; reviewed by Chrispeels, 1991), the signal peptide of jacalin preproprotein might be removed similarly, that is, at the time when jacalin precursor is being directed into the endoplasmic reticulum. The signal peptide is presumably both necessary and sufficient for directing the protein into the secretory pathway via the endoplasmic reticulum (see review, Giersch, 1988), although it may not need to be proteolytically removed in order to function (see review, Verner and Schatz, 1988). Within the lumen of the ER, the protein may undergo further maturation through proteolytic processing, proper folding, and oligomer assembly.

The proteolytic cleavage between the propeptide and α-chain and the cleavage between α- and β-chains may occur post-translationally because of the detection of the large ~25-kDa protein product in vivo (Fig. 6A, lane 3). It is not clear which of these two post-translational cleavages comes first, nor do we know whether they employ different proteolytic enzymes or whether they occur at different intracellular locations. However, a controlling mechanism in these processes seems necessary to ensure correct folding and assembly of jacalin for its exit from the ER and subsequent transport (Hurtley and Helenius, 1989; Pelham, 1989). One advantage for α- and β-chains being on one signal molecule is, perhaps,  

<table>
<thead>
<tr>
<th>Protein</th>
<th>% helical structures</th>
<th>% charged residues</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jacalin</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Carboxypeptidase Y (Valls et al., 1987)</td>
<td>64.1</td>
<td>12.7</td>
</tr>
<tr>
<td>α-Lytic protease (Silen et al., 1988)</td>
<td>47.3</td>
<td>6.0</td>
</tr>
<tr>
<td>Subtilisin (Vasantha et al., 1984)</td>
<td>75.2</td>
<td>18.1</td>
</tr>
</tbody>
</table>
FIG. 7. A speculation on the processing of jacalin prepropeptide. Leader peptide (residues 1–21), pro-peptide (residues 22–60), β-peptide (residues 61–80), linker peptide (residues 81–144), and α-peptide (residues 85–217) of jacalin prepropeptide are indicated. Three small upward arrows indicate three possible processing events required for generating jacalin polypeptide. The removal of the 4-residue linker peptide may occur post-translationally.

REFERENCES


