Characterization and Comparison of Arcelin Seed Protein Variants from Common Bean

Lynn M. Hartweck, Robert D. Vogelzang, and Thomas C. Osborn*
Department of Agronomy, University of Wisconsin, Madison, Wisconsin 53706

ABSTRACT

Four variants of arcelin, an insecticidal seed storage protein of bean, Phaseolus vulgaris L., were investigated. Each variant (arcelin-1, -2, -3, and -4) was purified, and solubilities and Ms were determined. For arcelins-1, -2, and -4, the isoelectric points, hemagglutinating activities, immunological cross-reactivities, and N-terminal amino acid sequences were determined. On the basis of native and denatured Ms, the variants were classified as being composed of dimer protein (arcelin-2), tetramer protein (arcelins-3 and -4), or both dimer and tetramer proteins (arcelin-1). Although the dimer proteins (arcelins-1d and -2) could be distinguished by Ms and isoelectric points, they were identical for their first 37 N-terminal amino acids and had similar immunological cross-reactions, and bean lines containing these variants had a DNA restriction fragment in common. The tetramer proteins arcelin-1t and arcelin-4 also could be distinguished from each other based on Ms and isoelectric points; however, they had similar immunological cross-reactions and they were 77 to 93% identical for N-terminal amino acid composition. The similarities among arcelin variants, phytohemagglutinin, and a bean a-amylase inhibitor suggest that they are all encoded by related members of a lectin gene family.

Arcelin-1 is the most thoroughly characterized of the protein variants. Using artificial seeds reconstituted from bean flour with the addition of purified arcelin protein, arcelin-1 was shown to be the component of the wild lines that confers resistance to Z. subfasciatus (18). Arcelin and PHA isolated from an arcelin-1-containing line had similar solubility properties, subunit Ms, deglycosylated Ms, and amino acid compositions; however, they had very different isoelectric points, and PHA agglutinated native erythrocytes but arcelin agglutinated only Pronase-treated cells (20). These similarities and differences might be explained by the observation that the cDNA nucleotide-derived amino acid sequence of arcelin-1 is approximately 60% identical to PHA-E and PHA-L (18). Arcelin has one methionine (PHA-E and PHA-L have no methionine) and additional attached oligosaccharide residues (18, 20) that may affect the chemistry and conformation of the mature protein.

Another difference between arcelin-1 and PHA concerns their effects on seed-feeding insects. Janzen et al. (11) proposed that phytohemagglutinins are involved in the specificity of host-insect interactions between Phaseolus species and seed-feeding predators. Although phytohemagglutinin from P. vulgaris did not provide protection to P. vulgaris predators, they did provide protection against seed-feeding pests of other Phaseolus species. Osborn et al. (18) showed that arcelin-1 protected seeds against P. vulgaris pests and recently the four arcelin variants were found to be associated with different levels and types of resistance to the Mexican bean weevil or the common bean weevil (3, 8). Characterization of arcelin variants and comparisons to each other as well as to PHA may aid in understanding specific host-pest interactions, thereby enabling plant breeders to plan an effective strategy to control bruchid pests.

In this paper, the solubilities and mol wt of the four arcelin variants are reported. The isoelectric points, hemagglutinating activities, immunological cross-reactivities, and N-terminal amino acid sequences of arcelin-1, -2, -4 also have been reported. These properties are compared among the arcelin variants and to those of PHA proteins. Lines containing these proteins also are compared for genomic DNA restriction fragment patterns of arcelin and PHA genes.

MATERIALS AND METHODS

Plant Materials

Four backcross lines of bean, Phaseolus vulgaris L. cv Sanilac, each containing one of the four allelic variants of arcelin, were used in this study. These lines were derived as
reported earlier for SARCI (18) except the wild bean lines
G12866, G12891, and G12949 (CIAT accession numbers; 19)
were used as donor lines for arcelins-2, -3, and -4, respect-
ively, and the backcross-derived lines were designated as
SARC2, SARC3, and SARC4, respectively. Seeds from field-
grown plants were harvested, dried, and cleaned for storage.

Protein Purification

Dry bean flour was stirred (10%, w/v) in 10 mM NaCl (pH
2.4) for 1 h at room temperature and centrifuged at 16,500g
for 20 min at 4°C. The supernatant was extensively dialyzed
for 24 to 48 h at 4°C against four to six changes of ddH2O
(30X volumes each), centrifuged to pellet the precipitate (glob-
ulin), and both the supernatant (albumin) and globulin were
lyophilized. The globulin pellet was finely pulverized and
reextracted (2%, w/v) in 10 mM NaCl (pH 2.4) by stirring for
30 min at room temperature, and the remaining insolubles
were removed by centrifugation. The supernatant (redissolved
globulin) was removed and adjusted to 150 mM NaCl by
adding 0.1 volume of 10X PBS (pH 7.4).

PHA was removed from the globulin or albumin fractions
by recirculation 6 to 8 times through in-line 47 mm filter
holders, each containing 10 pieces of activated affinity chro-
matography matrix (Nalge Co., No. 755–5013) coupled accord-
ing to the manufacturer's directions with fetuin (Sigma,
No. F2379) as the ligand. The membranes were washed with
200 mL PBS; the first 50 mL contained unbound proteins
washed from the affinity membrane and was added to the
original protein elution solution, and the remainder was dis-
carded. PHA was eluted from the membranes with 80 to 100
mL of 0.5 mM NaCl (pH 3.0), containing 50 mM glycine, and
the membranes were regenerated with four alternating washes
of 1.0 mM NaCl containing 0.1 mM sodium acetate (pH 4.0) or
0.1 M boric acid (pH 8.0), followed by PBS reequilibration.
The globulin or albumin solution containing unbound pro-
teins was tested for hemagglutinating activity, and if residual
PHA was detected the affinity purification was repeated. The
PHA-free albumin and globulin protein solutions and the
eluted PHA solutions were dialyzed against ddH2O and
lyophilized.

Arcelin was further purified from the PHA-free protein
solutions by dissolving the lyophilized powder to 25 mg/mL
in PBS and passing it through a gel filtration column (Seph-
acryl S-300, 1.5 x 120 cm) with an eluent (PBS) flow rate of
6.45 mL/h. Elution fractions between beginning and ending
half-peak heights of each arcelin peak were collected and
pooled, then dialyzed and lyophilized. Each lyophilized pro-
tein peak was redissolved at 25 mg/mL, and gel filtration
and collection were repeated. The native Ms of the arcelin
peaks were determined by comparison to standard proteins as
described previously (20). All proteins were dissolved in PBS
for electrophoresis, hemagglutination assays, immunoblotting,
isoelectric focusing, and sequencing.

Electrophoresis

Protein samples were prepared and separated by SDS-
PAGE as described previously (19, 20) except that each pu-
ified protein was first dissolved at 1 mg/mL in PBS, then an
aliquot was mixed with an equal volume of sample loading
buffer. Proteins were chemically deglycosylated as described
previously (20) using trifluoromethane sulfonic acid (4). The
Ms of denatured proteins were determined by comparing the
mobility of proteins to low mol wt standard proteins (Sigma).
Isoelectric points were determined (20) and IEF/SDS-PAGE
was performed as described previously (19).

Hemagglutination

Purified proteins were tested for agglutination of human
erythrocytes using the procedures described previously (2, 18)
with the following modifications: for Pronase treatment of
erthrocytes, the washed erythrocytes were resuspended to
their original volume in PBS containing 1% (w/v) Pronase
(CalBiochem), incubated at 30°C for 30 min, then washed
two to three times with PBS, and finally resuspended to 3% (v/v)
in PBS. Purified proteins were dissolved in PBS (2 mg/mL),
serially diluted 1:1 with PBS, and assayed for agglutinating
activity on native and Pronase-treated erythrocytes (20).
Protein concentrations of solutions were determined by using
the bicinchoninic acid colorimetric assay (23). Bovine serum
albumin (Sigma) was used as the standard. Agglutinating
activity of purified proteins is reported as the minimum
protein concentration (µg/mL) required to agglutinate
erthrocytes.

Immunoblotting

Electrophoresed gels were equilibrated in transfer buffer (10
mM Caps [pH 11.0], 10% [v/v] methanol) containing 0.2% (w/v)
SDS, for 15 to 20 min before transferring at 500 mA
for 30 min onto Immobilon-P PVDF membrane (Millipore
Corp.), according to the procedure of Matsudaira (15). Buffer
temperature was maintained at 10°C by submerging the trans-
fer cell in an ice-water bath. The blotted membranes were
incubated overnight at 4°C in block solution containing 20
mM Tris (pH 7.5), 150 mM NaCl, 1% (w/v) BSA, and 5% (w/v)
nonfat dry milk solids (Carnation). Antisera containing
antibodies against each of the purified arcelins or against
deglycosylated PHA were obtained as described previously
(20). Immunoblots were incubated, washed, and developed as
described previously (20) except that the primary antisera
were diluted 1:1000 in block solution.

Protein Sequencing

Proteins were prepared for sequencing as either purified
proteins from gel filtration (arcelin-1 dimeric fraction, arcelin-
1 tetrameric fraction, and arcelin-2) or by transferring purified
protein separated by SDS-PAGE to Immobilon-P PVDF
membrane, as described for immunoblotting, and dissecting
the membrane piece containing the major polypeptide (arce-
lin-4). Protein sequencing was performed at the University of
Wisconsin Biotechnology Center Protein Sequencing Facility
with either an Applied Biosystems gas phase 470A protein
sequencer or a Pulsed-Liquid 477A protein sequencer using
standard cycles (5). Phenylthiohydantoin amino acid analysis
was performed on-line using an Applied Biosystems model
120A analyzer. All chemicals used in sequencing were from
Applied Biosystems.
**Restriction Fragment Analysis**

Bean genomic DNA was purified and analyzed by Southern blot hybridization as described previously (17). DNA was quantified by fluorometry (Hoefer Scientific Instruments), digested with EcoRI, and 10 µg was loaded onto 0.8% agarose gels for electrophoretic separation of fragments. DNA was Southern blotted onto Nytran membranes (Schleicher and Schuell) according to manufacturer’s directions and hybridized with nick-translation probes. Additionally, 10⁻²⁵ g of λ-DNA that had been digested with HindIII was included in the nick translation mix for hybridization to mol wt markers of HindIII digested λ. Southern blots containing DNAs from SARC lines and Sanilac were probed first with pAR1–11, a recombinant plasmid containing a 1-kb, full-length cDNA sequence of arcelin-1 (18). After removing the arcelin probe (20 min in 0.2 M NaOH followed by 20 min in 0.5 M Tris [pH 7.5], 0.1X SSC and 0.1% SDS), the blots were checked for any remaining radioactivity by autoradiography and then hybridized with pSC1, a recombinant plasmid containing a 1.7 to 1.8 kb cDNA sequence of PHA (24). Blots were washed as previously described (17) except the final wash was in 0.1X SSC and 0.1% SDS at 65°C.

**RESULTS**

**Protein Purification**

Based on comparisons to previously reported electrophoretic patterns of arcelins (19), arcelins appeared to be the major proteins in the crude extracts (10 mM NaCl [pH 2.4]) of bean flour from SARC lines (Fig. 1, lanes 1, 4, and 6). Although arcelin-1 and arcelin-2 were soluble in both the globulin and albumin fractions derived from the crude extracts, the globulin fractions were more enriched for these arcelin proteins (data not shown) and were used to further purify arcelins-1 and -2 by removal of PHA and gel filtration. Arcelin-2 eluted as a single peak from gel filtration and migrated as a single band by SDS-PAGE (Fig. 1, lane 5), whereas arcelin-1 eluted as one major and one minor peak that had slightly different SDS-PAGE patterns (Fig. 1, lanes 2 and 3, respectively). Arcelin-3 and -4 appeared predominantly in the albumin fraction (data not shown), so PHA was removed from the albumin fractions by affinity chromatography before final purification by gel filtration. Arcelin-3 and -4 eluted as single peaks, and, because of their similar electrophoretic patterns (19), only the SDS-PAGE pattern of arcelin-4 is shown (Fig. 1, lane 7). PHA protein was separated from arcelins by affinity chromatography. The PHAs from SARC1 and SARC4 had similar SDS-PAGE patterns, whereas PHA from SARC2 had a distinct pattern that lacked a lower band common to the PHAs from SARC1 and SARC4 (Fig. 1, lanes 8, 9, and 10) and contained an additional higher M₄ band which was barely visible in the SARC1 or SARC4 PHAs.

**Mol Wt and Isoelectric Points**

The Mₛ of native, denatured, and deglycosylated arcelin and PHA proteins and IEF points of the denatured proteins are reported in Table I. Native and denatured Mₛ of the arcelin protein variants were compared in order to infer the quaternary conformations of these proteins under the conditions used to determine native Mₛ. Arcelin-2 appeared to exist as a dimer of polypeptides while polypeptides of arcelin-3 and arcelin-4 formed only tetramers. Arcelin-1 protein was separated into a major fraction in which polypeptides associated as dimers (arcelin-1d) and a minor tetrameric fraction (arcelin-1t). In comparing the dimer proteins, arcelin-1d had slightly greater native and denatured Mₛ and a more basic isoelectric point than arcelin-2 (Table I). These properties were confirmed by comparing two-dimensional electrophoretic patterns of the dimer proteins (Fig. 2, a–c). In comparing the arcelin-1t and arcelin-4 tetramer proteins, arcelin-1t had a slightly lower denatured Mₛ and more basic isoelectric point than arcelin-4 (Table I) and these properties also were apparent in two-dimensional gels of the proteins (Fig. 2, d–f). The PHA from SARC1, a tetramer protein, had a more acidic isoelectric point than the arcelin proteins (Table I). After deglycosylation, all arcelins and PHA had lower Mₛ than the denatured forms, indicating that all these proteins are glycosylated.
Hemagglutination

The hemagglutinating activities of the purified arcelin variants and the PHA fraction from SARC1 were assayed (Table I). None of the purified arcelin variants agglutinated native erythrocytes, whereas PHA agglutinated native erythrocytes at a minimum concentration of 8.2 μg/mL. The arcelin variants agglutinated erythrocytes only after Pronase treatment. This activity could be due to small amounts of contaminating PHA in the arcelin preparation; however, this is unlikely based on the calculations of the minimum amount of PHA contamination required to cause a positive hemagglutination reaction. According to the reaction with Pronase-treated erythrocytes, a contamination of the arcelin-1t preparation with PHA at 0.95% by weight would have caused arcelin-1t to agglutinate Pronase-treated erythrocytes at a concentration of 41.9 μg/mL. However, the lack of native erythrocyte agglutination with arcelin-1t suggests that the arcelin sample would have contained less than 0.66% PHA contamination. Therefore, the positive agglutinating activity of arcelin-1t with Pronase-treated erythrocytes is unlikely due to PHA contamination. The same argument can be used for arcelin-4. Because relatively high concentrations of arcelin-1d and arcelin-2 were required to agglutinate Pronase-treated cells, we cannot completely rule out the possibility that the observed agglutination is due to PHA contamination. However, since the native Mr of arcelin-1d and arcelin-2 were considerably lower than those of PHA, separation by gel filtration was probably effective and it is not likely that the collected arcelin peaks were contaminated with PHA.

Immunoblotting

Immunoblots containing seed protein extracts from each SARC line and corresponding purified arcelin proteins were incubated with antibodies made against one of the three purified arcelins or against deglycosylated PHA (Fig. 3, a–d). The arcelin-1 antibody (raised against protein containing arcelin-1d and -1t) identified dimer and tetramer proteins, as
well as PHA, in both the crude and purified preparations (Fig. 3a) and its reactions to the dimer proteins appeared to be stronger. Since the arcelin-1 preparation used to produce the antibodies had more dimer protein than tetramer (20), the antisera may have contained more antibodies with specificity for dimer proteins.

The arcelin-2 antibody had the strongest reaction with dimer proteins while the arcelin-4 antibody reacted more strongly with tetramer proteins (Fig. 3, b and c). The arcelin-4 antibody also bound some phaseolin in the crude extracts. This may represent nonspecific binding or possible antigenic similarity between arcelin-4 and phaseolin since no phaseolin contamination was observed in the purified arcelin-4 preparation used to make antibodies. All the arcelin antibodies bound to purified PHA but the arcelin-2 antibody bound less strongly than the others (Fig. 3, a–c). The deglycosylated PHA antibody had an extremely slight affinity for arcelin-1d and no detectable affinity for the other purified arcelins (Fig. 3d).

This antibody did react with PHA protein in the crude extracts of SARC1, SARC2, and SARC4 and to purified PHA from SARC1.

Protein Sequencing

The first 37 to 39 N-terminal amino acids of arcelin-1d and -1t, arcelin-2, and the major polypeptide of arcelin-4 (see Fig. 1, lane 7) were determined and compared (Fig. 4), along with the nucleotide-derived amino acid sequences of the PHA-E and PHA-L proteins (9) and the lectin-like protein (10), which has been identified as an α-amylase inhibitor (16). Two different sequences of arcelin-1t were obtained. One sequence (arcelin-1t1) contains an insertion between the 14th and 15th positions with respect to the dimer sequences. At position 21, two residues were indicated (alanine and aspartic acid); however, it could not be determined which tetramer sequence

![Figure 3. Crude protein extracts and purified arcelin and PHA fractions separated by SDS-PAGE, Western blotted, and reacted to antibodies against arcelin-1 (panel a), arcelin-2 (panel b), arcelin-4 (panel c), or deglycosylated PHA (panel d). Shown are: crude SARC1 (lane 1); pure arcelin-1d (lane 2); pure arcelin-1t (lane 3); crude SARC2 (lane 4); pure arcelin-2 (lane 5); crude SARC4 (lane 6); pure arcelin-4 (lane 7); pure PHA from SARC1 (lane 8).](image)

![Figure 4. Amino acid sequence of arcelin-2 (A-2, top sequence) is compared with the amino acid sequences of arcelin-1d (A-1d), arcelin-1t1 (A-1t1), arcelin-1t2, (A-1t2), arcelin-4 (A-4), the deduced amino acid sequences of PHA-E, PHA-L (9), and the α-amylase inhibitor protein (16) encoded by a lectin-like sequence (10)(αAl). (●), An amino acid identical to the arcelin-2 sequence; (x), an undetermined amino acid; (——), a space added to align the sequences for greatest similarity. Shaded areas show residues which have been conserved among many lectin proteins (6). Unshaded boxes were drawn to highlight residues common to some or all of the proteins. See "Results" for an explanation of "D/A" at position 21.](image)
contained which amino acid. The consensus sequence asparagine-g-threonine (where γ may be any amino acid and threonine may be replaced by serine) indicates a potential glycosylation site at the asparagine residue (14). The amino acid sequences of the dimer proteins indicated an "x" at position 12, but the previously published nucleotide sequences indicate that an asparagine residue occupies this position (arcelin-1d, ref. 18; arcelin-2, ref. 12). If position 12 was an asparagine, it may have been glycosylated, preventing its determination by amino acid sequencing. Although arcelin-1t1 and -1t2 did not contain this site, they may have a potential glycosylation site at position 31 and arcelin-4 may have glycosylation sites at 7 and 31.

The percentages of identity between arcelin, PHA (9) and α-amylase inhibitor (10) protein sequences are reported in Figure 5. The two arcelin dimer proteins were 100% identical while the three tetramer proteins were 77 to 93% identical. Although the two tetramer protein sequences from SARC1 were only 82% identical, the arcelin-1t1 protein was 93% identical to the arcelin-4 protein. The homologies observed between arcelin quaternary groups were lower than those within groups: the dimers were 53 to 55% identical to the tetramers and 58 to 67% identical to the PHA and α-amylase inhibitor proteins. The tetramers were 45 to 64% identical to the PHA and α-amylase inhibitor proteins.

Although the nucleotide-derived PHA sequences contained glutamine at positions 3 and 9 of PHA-E and position 9 of PHA-L, these glutamines were indicated as glutamic acid by Edman degradation of the proteins (6). Such an error changes the identity values between the proteins, but the identities were calculated with the derived amino acids as given.

**Restriction Fragment Analysis**

When the PHA probe (pSC1) was hybridized to Southern blots containing bean genomic DNAs (Fig. 6a), a common band was observed in all SARC lines and in Sanilac (12.3 kb), an additional band was observed only in SARC1, SARC3 and SARC4 (23.1), and two unique bands were found in SARC2 (15.8 and 4.4 kb). A dark 8.1 kb band in SARC4 also was faintly visible in SARC1 and SARC3. There were faint bands in SARC3 and SARC4 (5–6 kb) that were hybridized more strongly to the arcelin probe and were assumed to be more homologous to the arcelin probe than the PHA probe.

When a cloned arcelin sequence (pAR1-11) was used to probe the blots (Fig. 6b), the same bands were observed with some smaller sized bands. The slight hybridization to Sanilac DNA showed that the arcelin probe was somewhat homologous to sequences in Sanilac that could be PHA and/or sequences encoding the α-amylase inhibitor. SARC1 was found to have a unique band (5.9 kb), a band in common to SARC3 and SARC4 (5.2 kb) and a different band in common to SARC2 (2.4 kb). There were no unique bands in SARC2, although the 2.4 kb band was extremely intense in SARC2 and faint in SARC1. Although SARC2 has no polypeptides in common with SARC3 and SARC4 (19), they shared a 4.4 kb fragment which also hybridized in SARC2 to the PHA probe. A smaller band at 3.4 kb was shared by only SARC3 and SARC4. The arcelin variant lines, SARC3 and SARC4, were identical except that the 8.1 kb band in SARC4 hybridized more strongly to the PHA cDNA clone.

**DISCUSSION**

The arcelin-1 protein was shown previously to consist of dimer and tetramer fractions, with the dimer fraction being more abundant (20). This observation was confirmed in the present study and sufficient quantities of these fractions were isolated to allow additional characterization. The arcelin-1d preparation was very similar to the one reported previously, although it appeared to have none or very little of the lower molecular mass (35.8 kD) band observed previously (20). The

<table>
<thead>
<tr>
<th></th>
<th>A-1t1</th>
<th>A-1t2</th>
<th>A-4</th>
<th>PHA-E</th>
<th>PHA-L</th>
<th>α AI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dimer</td>
<td>53</td>
<td>55</td>
<td>53</td>
<td>64</td>
<td>67</td>
<td>58</td>
</tr>
<tr>
<td>A-1t1</td>
<td>82</td>
<td>77</td>
<td>61</td>
<td>54</td>
<td>45</td>
<td></td>
</tr>
<tr>
<td>A-1t2</td>
<td>93</td>
<td>59</td>
<td>56</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td>A-4</td>
<td></td>
<td>64</td>
<td>52</td>
<td>55</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHA-E</td>
<td></td>
<td></td>
<td>76</td>
<td>59</td>
<td></td>
<td></td>
</tr>
<tr>
<td>PHA-L</td>
<td></td>
<td></td>
<td></td>
<td>47</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Figure 5.** Percent identity of arcelin proteins, PHA-E, PHA-L, and α-amylase inhibitor, based on the first 37 N-terminal amino acids. The proteins designated are the same as in Figure 4. Arcelin-1d and arcelin-2 were 100% identical and are listed together as "dimer".

**Figure 6.** Autoradiograph from Southern blot of SARC1, SARC2, SARC3, SARC4, and Sanilac (lanes 1, 2, 3, 4, and 5, respectively) DNAs cut with EcoRI and probed with pSC1 (24) (panel a) or pAR1-11 (18) (panel b). Lane M contains a HindIII digest of λ-DNA which was used to calculate restriction fragment sizes in lanes 1 to 5 (sizes in kb on the right).
arcelen-1t preparation also appeared to be similar to the one reported previously, except for a slightly higher mobility in SDS gels, which may have been due to loading a much higher quantity of protein in this study. The three other arcelen variants, which had not been characterized previously, were purified and classified as being composed of dimer protein (arcelin-2) or tetramer protein (arcelin-3 and arcelin-4). Recently, a fifth arcelin variant has been reported (13) with a subunit molecular mass of 31 to 32 kD, which is quite distinct from the variants characterized here (34–37 kD), and the quaternary composition was not reported.

Although all arcelen variants could be distinguished on the basis of Mr, and isoelectric point, variants within the dimer and tetramer groups had some very similar characteristics. The arcelen dimer proteins had similar immunological cross-reactions and they shared an exact amino acid sequence for their first 37 N-terminal amino acids. The amino acid sequence of the arcelin-1d fraction corresponded exactly with the predicted amino acid sequence of the cDNA clone of arcelin-1, pAR1-11 (18), and thus this cDNA codes for the arcelin-1d protein. Recently, a cDNA clone of arcelin-2 was sequenced, and based on a nucleotide-derived amino acid sequence comparison there were only four amino acid differences between arcelin-2 and arcelin-1d (12). An important amino acid difference occurred at position 130, where one of the three potential glycosylation sites present in arcelin-1d was absent in arcelin-2. If these sites were glycosylated, then the predicted differences in glycosylation would agree with our observation of a slightly lower mol wt for arcelin-2 compared to arcelin-1d and they could explain slight differences in isoelectric points between the dimers. Since the arcelin-1d and arcelin-2 cDNA sequences are 99.3% identical (12), there should be a highly conserved region in the DNAs of SARC1 and SARC2 lines which includes these coding sequences. The 2.4 kb restriction fragment (Fig. 6b) which was common to both SARC1 and SARC2 when hybridized to the arcelin-1d cDNA clone may contain this coding region.

The arcelin-3 and arcelin-4 tetramer proteins, which were shown previously to have polypeptides in common (19), had very similar solubility properties and mol wt, and DNAs from the parent lines probed with the arcelin-1d cDNA clone had almost identical restriction fragment patterns. Arcelin-1t had a lower mol wt and more basic isoelectric point than arcelin-4, but appeared to be very similar immunologically and in N-terminal amino acid sequence. The two arcelin-1t sequences which were identified probably represent two different polypeptides. These polypeptides may complicate as the major band observed on SDS-PAGE gels, or one of these polypeptides may be the minor, higher Mr band (Fig. 1, lane 3). The two arcelin-1t sequences differed by an insertion in arcelin-1t after position 14 of the reference sequence. A comparison of lectin proteins from many species showed that additions and deletions within the region from the phenylalanine at position 11 and the leucine at position 16 were very common (6). Curiously, the two arcelin-1t sequences were not as similar to each other (82%) as arcelin-1 was to the major polypeptide of arcelin-4 (93%) for their first 37 amino acids. It is possible that the minor, higher Mr polypeptide of the arcelin-4 variant (Fig. 1, lane 7) is very similar to the arcelin-1t sequence.

The restriction fragment analysis seemed to corroborate the tetramer amino acid sequence results. Two restriction fragments were associated with DNAs from the tetramer-containing lines: a 5.9 kb fragment, which was unique to arcelin-1 and may contain the coding sequence of one of the two arcelin-1t proteins, and a 5.2 kb fragment observed in the SARC1, SARC3, and SARC4 lines, which may contain the coding sequence for a slightly different tetramer protein.

Although we do not know how arcelen, PHA, and α-amylase inhibitor proteins compare for their entire length, the estimated identity based on the first 37 amino acids of the dimer proteins and PHAs was similar to the calculated degree of identity between the entire sequences (9, 18). Foriers et al. (7) also concluded that the identity between the N-terminal amino acids of lectin proteins was a useful predictor of identity between the entire sequence of lectin proteins.

In comparing the N-terminal sequences of arcelin, PHA, and α-amylase inhibitor protein, there were some regions of very high amino acid conservation; in particular, the region from position 16 to 22 was 92% identical. This region has been identified as functioning to target PHA protein within the cell to the vacuole (25). Foriers et al. (6), found that these amino acids, as well as those at positions 5, 6, and 11, are highly conserved in the lectin sequences from many species. These positions were also conserved in all arcelin proteins analyzed in this study.

Although arcelins and PHA had similar N-terminal amino acid sequences, they appeared to be quite different for hemagglutinating and immunological reactivities. None of the arcelin variants agglutinated native erythrocytes. Pronase treatment, which removes or denatures some cell surface proteins exposing binding sites, resulted in positive agglutination reactions; however, the concentration of arcelin required for agglutination was 100 times that of PHA. The arcelin antibodies recognized all arcelin types and PHA to some degree; however, the arcelin-2 antibody was most specific for arcelin dimers, and the arcelin-4 antibody recognized arcelin tetramers and PHA more strongly than arcelin dimers. In agreement with a previous study (20), the antibody to deglycosylated PHA reacted very weakly with arcelin-1d. These data suggest that antibody recognition can be complex and might be affected by many factors including: the amino acid sequence, the quaternary structure of the protein, and the attachment of glycosyl moieties.

The similarities among the different arcelin variants and between arcelin and PHA, together with the tight linkage of genes encoding these proteins (19), suggests that they are all encoded by related members of a lectin gene family. The SARC1 line appears to contain a family consisting of genes which encode PHA, the arcelin dimer, and arcelin tetramer proteins, whereas other arcelin-containing lines consist of genes encoding PHA and either arcelin dimer or arcelin tetramer proteins. Since most bean lines do not contain arcelin (19), genes encoding arcelin have probably evolved from PHA genes by some mechanism of gene duplication and divergence (18). The SARC1 gene family may have evolved from a rare recombination in a hybrid between lines containing either the dimer or tetramer genes.

Arcelin-1 has been shown to be an antibiosis factor for bruchid pests (18) and bean lines containing the other arcelin variants also showed varying levels of antibiosis with different
species of bruchids (3, 8). Although PHA at its naturally occurring low level did not exhibit this quality (18), Janzen et al. (11) reported that PHA was toxic to cowpea weevils (Callosobruchus maculatus Fabricius). The differences in antibiosis of these proteins to various pests is probably due in part to their biochemical properties; however, the relative quantities of these proteins in the seed also may be an important factor. Osborn et al. (18) showed that artificial seeds which contained higher levels of arcelin-1 exhibited higher levels of resistance to bruchids. Harmsen (8) postulated that differences in the percentage of emergence and days to emergence of bruchids were due to differences in arcelin type as well as arcelin concentration in the seed. Additional insect feeding trials with purified arcelins and PHA are being conducted to determine the relative effects of protein type and concentration on bruchids. Research also is underway to investigate whether genetically increasing arcelin or PHA concentration within the seeds of bean lines results in increased levels of resistance to bruchids.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the technical assistance of Bradley Hodgeman and David Austin, and Cynthia Wadsworth and Ronald Niece at the University of Wisconsin Biotechnology Center for amino acid sequencing. We thank Maarten Chrispeels for providing the PHA cDNA clone, pSC1, and Leslie Hoffman for providing antibody to deglycosylated PHA.

LITERATURE CITED