Plant Physiol. (1991) 97, 839–840
0032-0889/91/97/0839/02/$01.00/0

Received for publication June 3, 1991
Accepted June 8, 1991

Plant Gene Register

Nucleotide Sequence of a Genomic Clone Encoding Arcelin, a Lectin-Like Seed Protein from *Phaseolus vulgaris*[1]

Janice L. Anthony, Raymond A. Vonder Haar[2], and Timothy C. Hall*
Department of Biology, Texas A&M University, College Station, Texas 77843–3258

Arcelin is a novel seed protein first found in several accessions of the wild Mexican bean, *Phaseolus vulgaris* L. var aborigeneus (8). It is closely related to bean lectin, both at the amino acid and nucleotide levels (5). There are five allelic variants of arcelin (4, 6). The Arc 1 (Table I, Fig. 1) protein migrates as two bands on polyacrylamide gels with M_r of 35,000 to 38,000. These two bands are glycosylation variants, because a single band is seen following chemical deglycosylation (7). The presence of arcelin correlates with reduced expression of phaseolin (9), the major seed storage protein in beans. In these seeds, arcelin accounts for about 10% of the total seed protein.

Arcelin has been found to have insecticidal properties on bruchid beetles (bean weevils), a prominent bean pest (5). Naturally occurring seeds containing arcelin and artificial seeds containing purified arcelin were resistant to attack by the Mexican bean weevil *Zabrotes subfasciatus*. Reproduction of the bruchid beetle *Callosobruchus maculatus* was reduced in the presence of black bean phytohemagglutinin, another plant lectin (2).

Two arcelin cDNA sequences have been reported, for Arc 1 (5) and for Arc 2 (3). The sequence reported here is a genomic clone of Arc 1. It is identical with the cDNA for Arc 1 with the exception of a single nucleotide in the 3′-untranslated region.

Table I. *Characteristics of Arcelin from Phaseolus vulgaris*

Organism:
Phaseolus vulgaris L. var aborigeneus
Locus:
Tightly linked to phytohemagglutinin genes (5).
Function:
Encodes a lectin-like seed protein.
Clone Type, Designation:
Genomic sequence; full-length with 3.4 kilobases of 5′-untranslated sequence and 392 bp[a] of 3′-untranslated sequence; clone Z5.
Clone Source:
λZap II size-selected genomic library from leaf DNA.
Method of Identification:
Screened with 309 bp polymerase chain reaction product amplified from same genomic DNA, primers selected from cDNA sequence (5).
Sequencing Strategy:
Deletion subcloning and complete dideoxy sequencing of both strands.
Confirmation:
Identity with cDNA sequence (5).
Regulation and Expression:
Transcript of approximately 1.1 kilobases; message first seen in developing seed 12 DAF, expression peaks 18 DAF; protein first seen on Coomassie-stained gels in seeds harvested 18 DAF.
Features of Gene Structure:
TATA box at 3332 to 3337 bp (1); translation start at 3375 to 3377 bp; translation stop at 4170 to 4172 bp; no introns; varies from cDNA at base 4247 in untranslated 3′ sequence (A in genomic sequence, C in cDNA).
Features of Protein Sequence:
265 amino acids, including 21-amino acid signal sequence as determined by microsequencing of the mature protein; three consensus *N*-linked glycosylation sites.
Antibodies:
Not available; cross-reaction with other plant lectins.
Subcellular location:
Not tested.
GenBank Accession No.:
M68913

[a] bp, base pair.

LITERATURE CITED

1. **Hoffman LM, Donaldson DD** (1985) Characterization of two *Phaseolus vulgaris* phytohemagglutinin genes closely linked on the chromosome. EMBO J **4:** 883–889
2. **Janzen DH, Juster HB, Liener IE** (1976) Insecticidal action of the phytohemagglutinin in black beans on a bruchid beetle. Science **192:** 795–796

[1] Supported in part by a grant from Rhône-Poulenc Agro.
[2] Present address: Ambion Inc., 2130 Woodward Suite 200, Austin, TX 78745.

```
   1 GAATTCTTCAAAGCACTTGATCACTCTTGATCAACATGTTGGTCCTCCTC AAACCAAAAAAGACTTCTTGCAAGTCTTTGATGTGTTGGCCGCACTAGCC
 101 AGATTTGACCACAATGTCGTCCGCGTAGACCTCCACGCTTCGACCGATCA TCCACGCTTCAAATCTGTTTTAGCATTCTGTTTTGATGATATAATTTCGC
 201 TTTGAGTGTTTTAATATTCATTTGTACCTTTAATTGTTGTGTTCCTTCAC ATCTAATGTTGTGTTCAAGTGTATTAAAGTGTCTATGCTTGATTCAACTC
 301 ACATCACTTTCTCTCCAGTAGCAAAGATGACCACCATAAGGTTCCTTCTT TCTCTAGCTTCTATCTATAATTGGGAATTGAAACAATTAGACATAAACAA
 401 TGCCTTTTTACGTGGAGATTTGAAAGAAGATGTATACATGGTTGCTCCTC CTGGAATTGTTAACAAATGCAGGTCTTTTAGCCTGCAAACCTGCACTCAC
 501 TCCTATTGATAATGCATGAAAAATTCTCTTTTACTGGAAGTGTTCCTTTC ACAGATATTCGAGCTTATAGAAGATTGATTGGAAGACTTATGTATCTTAC
 601 TAATACCTGACTTGACATAACATTTTATGTGCAACAACTTTCTCAAATTC TTAATAAGCCTACAGTTGCTGATACTGCAGCATGGCCACTGTTACATGTG
 701 AAATTCAATGGCTAACTTATCTTCTTTAAGATTTCAAAGTTACTTTTGAG AAACCATCTCTTTTATACTATGACAATGACTCCGCAAGATACATTGCAAC
 801 CAATCCAGTGTTTCACGAGCGTACGAAACATATAGAAATATATTGTCACA TTGTTAGGGAAAAATTAAAGAAGAGTCTTATCCAGTTGCTGCCCATTTCC
 901 ACCACAGAGTAGTTGGCTGACATTTATAGCAAAGCTTTAAGTCCTCAATC TTTCAAGAGTATTTGTTCCAAGCTGGATCTCATCAATATTAGTTCACCAG
1001 CTTGCGAGAGATATTAAAGGATATGGATACCAGAGGTATAGAAGAATGGT AATATCGTCATTTTAGATCCTCCTGAAAGATATAAATGGCGGTTATGGTA
1101 AAAAGGCGTCATTTTAGACCCTCTTGAAAGATATAAAACGACGGTTATGG TAAAAGCTGCCATTTGAAATGATAAATTTTTGTTTTAGTTCATTCTGTTT
1201 AATGTTAATTCCATCTAAATTGACTTGTACAATTAAAACTCAACCAACCC AAATACATATATTTAGAGAATCAAATTATATACATTGATAAAAATGTATT
1301 ACAAACATAACATATATATTCCTATCTATCTATCTAATTCCTACACTATT GATTAATTATCCTAGAGTTCTAGAAATTAAGGAAATACGTGGCTCATTTC
1401 TCACATATTCTATGGCGGGAAAAAGATTCACATTAATTAGTGTTAAAGAA TTATAAAAAATAAGTATTATTATCTGCTTCTAATTATGCTCTTACAATGG
1501 TCTCCAAATGCAAATTCTTACAAATCTGACATGTATATGTCCATAGGGTT TTGATATATCACTAATTTATACCTAAAAAATCATGAAAAAATTTAGAATT
1601 TTATAATTAGGGTAAGTATAGGAACTAAAACCTTAAAATACCTGAAAAAC ACCTAAATAATACCGAAAAATAAAAAATTAATAGTAAATGGTGGATTTTT
1701 GTAATTTTTTGGTCCGTCTAAAAAGCACCACTCCCATGACACCCCCTACC AGCATACCCAAATACTTAAAAAGAGTTACCATCACACTGTAATTAAGAAT
1801 CATAGCGAAAGGTTGAATCCTTGTCAAATCAATCCCCAACCCACCAATTC TACCCTTCAAGAAATTCACCTTAAGATCGGAGGATAATTCAAAACAATTT
1901 AGAGTCACCTCTATGTTAAACACACTTTTGGCATTACCTTCACAGACGAA TAAGGTATTGTCTTCATAGTGGAGCATATTAACCTTAACTTTCTTATCCC
2001 CAATTTCCAAACTATCAATCATATTTTTCTCAACGGTCATCCTAGGCGCA CCTGCCAGACCTTCAGCTGCAATCAGAAACAAAAATGGTGTTAACGGATC
2101 ACCTTGACAAAACCCCTTCATTGGAATGAAATCCTTATAGGGACTTCGAT TCACTAACACTGACACCAATGCAAATTCTAAACAAGATTGAATCCATTTA
2201 AATTTTCTATTCAAAATAAATAATTATTAAATTGATTATAATTTAATATC ATAAAATATAAAATTAGAAAATGATCCCATTCTATATTGTTATATATATA
2301 CAATATTAGTTTATTTGAAAATAAAATTAATTTATATTAAAATATGATTT GGTAAAAGATATTATAAAGAGCATCTTGGATGGCCAAAAGTTGTATGACA
2401 CACCGCTTTTGTTAGTACCATATGCACTCTGGCGGCTTTCACACTTCCCA ATGAAACCACTCACCATGCACTCTGCCACCTCAGCTCCCTTCTTCACACA
2501 TGTCTCATGTCACTTTCCACTTTGGCCGCTCGGTCTCCCGTTTCCGAATC CCCTTATTGTTCGTCTTCTTCGTCTTTTTCATCCACTTCCTCCTCCAACA
2601 AAATCCTTATCTGGTTGTAGCTAGAGTAAGAGTCCAGGAAGCTGAATATC TTGTACCCGACCGCACCATCAACAATTTGATCAATGTTAGGCAGGGGTAG
2701 AAGTCCTTTGGATACGCCTTGTTTAAGTCCTTATAGTCCACACACATCCG TCATTTACCATTGGACTTGGTGATCATGACCACATTGACCAACCAGGTTG
2801 TGTATCGGGCTTCTTCAGCGGCCGTCCAAGCGAAGATATAGACATTTTTC TTCAAAGCCTGCTATATGAGCTCGGCGTCAGCGACGGTATAAATTAATTT
2901 TTTAATTTAAATTTTTTATAAATTTTCTAATTGAAATTAGTTTTTTAATT TCAATATTTTATAAATTTTCTAATTTAAATTTTCATGTGTCTTCCTCTCA
3001 CTTCCGAATACATCTTTCACTATGAAACAACTAGCCACCTCAGGTCCTTC CTCTTCCCATGATGACACCACTAGGCATGCATGCCGCCACCTAAGCTGCC
3101 ACCTCTTCTCATTATGGCCATGCACAGTGCCACCTCAGCACCTCCTCTCA ATTCCCATTGCTACCTGCCAAACCGCTTCTCTCCATAAATATCTCTTTAA
3201 ATGTAAACTAATTATTTCATATACTTTTTTGATGACGTGGGTGCATTGGC ATCGTTGTTTAATAATTGTTAATTTGGTGTTTAATAATAAAATGAAAGAA
3301 AAAAGTTGGAAAGATTTTGCATTTGTTGTTCTATAAATAGAGAAGAGAGT GATGGTTAATGCATGAATGCATACATGGCTTCCTCCAACTTACTCACCCT
                                                                                   M  A  S  S  N  L  L  T  L    9
3401 AGCCCTCTTCCTTGTGCTTCTCACCCACGCAAACTCAAGCAACGACGCCT CCTTCAACGTCGAGACGTTCAACAAAACCAACCTCATCCTCCAAGGCGAT
      A  L  F  L  V  L  L  T  H  A  N  S ▼S  N  D  A   S  F  N  V  E  T  F  N  K  T  N  L  I  L  Q  G  D    42
3501 GCCACCGTCTCATCCGAAGGCCACTTACTACTAACCAATGTTAAAGGCAA CGAAGAGGACTCTATGGGCCGCGCCTTCTACTCCGCCCCCATCCAAATCA
      A  T  V  S  S  E  G  H  L  L  L  T  N  V  K  G  N   E  E  D  S  M  G  R  A  F  Y  S  A  P  I  Q  I    75
3601 ATGACAGAACCATCGACAACCTCGCCAGCTTCTCCACCAACTTCACATTC CGTATCAACGCTAAGAACATTGAAAATTCCGCCTATGGCCTTGCCTTTGC
     N  D  R  T  I  D  N  L  A  S  F  S  T  N  F  T  F   R  I  N  A  K  N  I  E  N  S  A  Y  G  L  A  F  A  109
3701 TCTCGTCCCCGTCGGCTCTCGGCCCAAACTTAAAGGCCGTTATCTAGGTC TTTTCAACACAACCAACTATGACCGCGACGCCCATACTGTGGCTGTGGTG
      L  V  P  V  G  S  R  P  K  L  K  G  R  Y  L  G   L  F  N  T  T  N  Y  D  R  D  A  H  T  V  A  V  V   142
3801 TTCGACACCGTCAGCAACCGTATTGAAATCGACGTGAACTCCATCCGGCC TATCGCAACGGAGTCTTGCAATTTCGGCCACAACAACGGAGAAAAGGCCG
      F  D  T  V  S  N  R  I  E  I  D  V  N  S  I  R  P   I  A  T  E  S  C  N  F  G  H  N  N  G  E  K  A   175
3901 AGGTTCGGATCACCTATGACTCCCCCAAGAACGACTTGAGGGTTTCTCTG CTTTACCCTTCTTCGGAAGAAAAGTGCCACGTCTCTGCCACAGTGCCGCT
     E  V  R  I  T  Y  D  S  P  K  N  D  L  R  V  S  L   L  Y  P  S  S  E  E  K  C  H  V  S  A  T  V  P  L  209
4001 GGAGAAAGAAGTTGAGGACTGGGTGAGCGTTGGGTTCTCTGCCACCTCAG GGTCGAAAAAAGAGACCACTGAAACGCACAACGTCCTCTCTTGGTCTTTT
      E  K  E  V  E  D  W  V  S  V  G  F  S  A  T  S   G  S  K  K  E  T  T  E  T  H  N  V  L  S  W  S  F   242
4101 TCTTCCAACTTCATCAATTTTAAGGGCAAAAAATCTGAACGTTCCAACAT CCTCCTCAACAAGATCCTCTAGACTCCCAAAGCCAGCTTCACTGTGACAG
      S  S  N  F  I  N  F  K  G  K  K  S  E  R  S  N  I   L  L  N  K  I  L                                  265
4201 TAAAACCTTCCTTATACGCTAATAATGTTCATCTGTCACACAAACTACAA TAAATAAAATGGGAGCAATAAATAAAATGGGAGCTCATATATTTACACAA
4301 TTTACACTGCCTATTATTCACCATGCCAATTATTACTGCATAATTTCAAA ATTGTCATTTTTTAAAAGTTTATAATAATTAAGAAATATTACTATAAGTT
4401 AAAGTATAACATAGAAAAAAAAAACATTAAATCTTAAGAAATATTACTAT AATTTACCCTTTTTTATCTGAAGAGTCTAATAATTGAGAGATTGACACAA
4501 AATATTTATACCAGCTCCCTCTTTACCAAGAGCTACATTCAGTCTTCGAC CCCACTAAGAATTC
```

Figure 1. Nucleotide and amino acid sequence of arcelin. Nucleotides are numbered on the left; amino acids are numbered on the right. Putative TATA box is underlined, translation start and stop codons are bold and underlined. Arrow indicates signal sequence cleavage site. Single nucleotide deviation from cDNA in the 3′ untranslated region is in bold.

3. **John ME, Long CM** (1990) Sequence analysis of arcelin 2, a lectin-like plant protein. Gene **86:** 171–176
4. **Lioi L, Bollini R** (1989) Identification of a new arcelin variant in wild bean seeds. Bean Improv Coop **32:** 28
5. **Osborn TC, Alexander DC, Sun SSM, Cardona C, Bliss FA** (1988) Insecticidal activity and lectin homology of arcelin seed protein. Science **240:** 207–210
6. **Osborn TC, Blake T, Gepts P, Bliss FA** (1986) Bean arcelin 2: Genetic variation, inheritance and linkage relationships of a novel seed protein of *Phaseolus vulgaris* L. Theor Appl Genet **71:** 847–855
7. **Osborn TC, Burow M, Bliss FA** (1988) Purification and characterization of arcelin seed protein from common bean. Plant Physiol **86:** 399–405
8. **Romero J** (1984) Genetic variability in the seed protein of nondomesticated bean (*Phaseolus vulgaris* L. var aborigeneus) and the inheritance and physiological effects of arcelin, a novel seed protein. PhD thesis. University of Wisconsin, Madison
9. **Romero-Andreas J, Yandell BS, Bliss FA** (1986) Bean arcelin 1. Inheritance of a novel seed protein of *Phaseolus vulgaris* L. and its effect on seed composition. Theor Appl Genet **72:** 123–128