

# Nucleotide sequence of an A-type legumin gene from pea

William G.Rerie<sup>1, 2</sup>, Malcolm I.Whitecross<sup>2</sup> and T.J.V.Higgins<sup>1</sup>

<sup>1</sup>CSIRO Division of Plant Industry, PO Box 1600, Canberra, ACT 2601 and <sup>2</sup>The Australian National University, PO Box 4, Canberra, ACT 2601, Australia

Submitted January 9, 1990

EMBL accession no. X17193

The A-subfamily of legumin genes encodes the predominant 11S-class of seed storage proteins in pea (*Pisum sativum* L.). We have determined the complete nucleotide sequence of one such gene (*LegA2*) located on a 4.2 kb EcoR1 fragment and encoding a mature subunit with a deduced molecular weight of 56,929 Daltons. The coding region spans 1,825 bp (including three short introns) and exhibits a high degree of identity to *LegA* (1), but diverges from this gene in both the immediate 5'- and 3'-flanking regions. Nucleotide positions are numbered relative to the proposed cap-site (+1). Potential regulatory sequences are

underlined and include the TATA (-26 to -32) and the 'legumin' (-100 to -130) boxes in the 5' region and overlapping polyadenylation signals (+1969, +1973) in the 3' region. The proposed signal peptide is delineated by bold-face type preceding the processing site (|). The post-translational processing site between Asn and Gly residues is also indicated.

## REFERENCE

1. Lycett, G. et al. (1984) *Nucl. Acids Res.* **11**, 4493-4506.

```

-500 ATATGACTATATTGTAATTGTAATCAAGCACTTTAATTTGAAGTTGTTATCGTCCACCTATATTTCAACTACACAAAATGTGCTTTACCATTAACTT
-400 TAAAAATTTGTACAGACCATGAACATAAATCCTATCCAACATACATAACAGACATTAAGTGGTGAAGTGAATCAGGTTTCATATATCTAGATATTAC
-300 AAGACAGTAATGATCAAACCTCAGTACATATGTAAGAGAGAAATCAATTATATACATATACATGGTCCCAACACCACCGATTTCAGCTAGCTATCTAAT
-200 TACTCAACTCTCACTGAAGCCACCTCTGCTGTAATGAGACATTTGATTTTATGAGGTGTAACACACAAGGCTTCCATAACCATGCAAGATGAAGAATGT
-100 CGAATGGTCAGCAACTCATGTATCTTCTGGAGCCGATGTGTCCCTCTATACTTTCTTCTATGTTTCACTATAAATCCCTATGCCAGATTAAGGTTCTTCG
      M A T K L L A L S L S F C F L L L G G C F A L R E
1   CGTCACAAACATATATTTCTATCCAACATAGGCTACTAAGCTTCTTGCACCTTCTCTTTCTATTCGTTTTTCTACTTTTGGGGGGCTGTTTTGCTTTGAGAG
101  A A C A G C C A G A G C A A A A T G A G T G C C A G L E N A L P D N R I E S E G G L I E T W N P
      Q P E Q L E R L N A L E P D N R I E S E G G L I E T W N P
      N N K Q F R C A G V A L S R A T L Q H N A L R R P Y Y S N A P Q E
201  C A A C A A A G C A A T T C C G A T T G T G T G T G G C C T C T C T G T A C C C T C A A C A T A A C C C C T T G C A G A C C T T A C T A C C A A T G C T C C C A A G A A
      I F I Q Q G
301  A T T T T C A T C C A A C A A G G T T A C T T A T T T T G A T C T T A T A C C A A C T T C T T T A C T T A C A T T A C A T G C A T A T T A G C A T A T A A T T A G T G T T C T A C T A T A C C A A T
401  A C A G G T A A T G G A T A T T T G G G A T G G T A T T C C C G G T T G C C T G A G A C T T T G A A G A G C C A A G A A T C T G A A C A A G G A G A G G G A C C A G G T A C A G A G C A A
      H Q K V N R F R E G D I I A V P T G I V F W M Y N D O D T P V I A
501  G A C A T A A A A G G T T A A C C G A T T C A G A G A G G G T G A T A T C A T T G C A G T T C C T A C T G G T A T T G T A T T T G G A T G T A C A A C G A C C A A G A C A C T C C A G T T A T T G C
      V S L T D I R S S N N Q L D O M P R
601  C G T C T C T T A C T G A C A T T A G A A G C T C C A A T A A C C A G T T G A T C A A G G G T G A G C T A C T G A G C A T A A T T A A A C T T C C C A T A T A A G A T A A T A T G T
      R F Y L A G N H E Q E F L R Y Q H Q O
701  T G T C C A A A C A G T A A C A T A G A T T C T A T C T A T C T A T G T T T G A C A G A T T C T A T C T T G T G G G A A C C A G A C A A G A G T T T C T A C G A T C A C C A G C A T C A A C A
      G G K Q E Q E N E G N N I F S G A F K R D F L E D A F N V N R H I V
801  A G G A G A A A G C A A G A A C A A G A A A T G A A G C A A C A C A T T T T C A G T G G C T T C A A G A G G A T T T C T T G A A G A T G C T T C A C C G T G A A C A G G C A T A T A G T A
      D R L Q G R N E D E E K G A I V K V K G G L S I I S P P E K Q A R H
901  G A C A G A C T T C A A G C A G G A A T G A A G A C G A A G A G A A G G C C A T T G T C A A A G T G A A G G T G G A C T C A G C A T C A T A A G C C C A C C C G A A G C A A G C G C C
      Q R G S R Q E E D E D E D E E R O P R H Q R G S R Q E E E E D E D
1001  A C C A G A G A G G C A G C A G A C A A G A G A A G A T G A A G A T G A A G A T G A A G A G A G C C G C C C A C C A G A G A G G C A G A C A A G A G A A G A G A A G A T G A A G A
      E E R Q P R H Q R R R G G E E E E E D K K E R R G S O K G K A S R R Q
1101  T G A A G A G A G G C A G C C G C T A T C A A A G G A G A G A G A G A G A A G A A G A C A A A G A G C G C C G C G A A A A G G C A A A A G C A A A G C A G A A G G C A A
      G D M G L E E T V C T A C K L R L N I G P S S P D I Y N P E A G R I
1201  G G A G C A A T G G G C T T G A G A A C A G T T G C A C T G A A C T T C G A T T G A A C A T T G P S C C T T C A T C A C C A G A C A T C A C A C C C T G A A G C T G G T A G A A
      K T V T S L D L P V L R W L K L S A E H G S L H K
1301  T C A A A C T G T T A C C A G C C T G G A C C T C C A G T T C A G G T G G C T C A A A C T A A G T G C T G A G C A T G G A T C T C C A C A A G A T A T G T T T T T C A T A T T T A A T
      N A M F V P P H Y N L N A N
1401  T G T T T T C C A T G A A T C A A T T T C A T G T C G A A C T A T G T G T A A C T A T T A C A A T C T T C A T A C A G A A T G C T A T G T T T G T G C C T A C T A C A C C T G A A T G C A A A
      S I I Y A T A C G C A T T G A A G G A C G T G C A A G G C T A C A A G T G A A C T G C A A T G G C A A C A C C G T G T T G A T G G A G A G C T A G A A G C C G A A G C G G A T T G
1501  C A G T A T A A T A C G C A T T G A A G G A C G T G C A A G G C T A C A A G T G A A C T G C A A T G G C A A C A C C G T G T T G A T G G A G A G C T A G A A G C C G A A G C G G A T T G
      T V P O N Y A V A A K S L S D R F S Y V A F K T N D R A G I A R L A
1601  A C A G T G C C A C A A A C T A T G C T G T G G C T G C A A A G T C A C T A A G C A C A G G T T C T C A T A T G T A G C A T T C A A G A C C A A T G A T A G A G C T G G T A T T G C A A G A C T T G
      G T S T S V I N N L P L D V V A A T F A N L Q R N E A R O L K A S N N P
1701  C A G G C A C T C A T C A G T T A A A T A A C T G C C G T T G D V V A A T F A N L Q R N E A R O L K A S N N P
      F K F L V P A R Q S E N R A S A *
1801  C T T C A A A T T T C A G T T C C A G C T C G T C A G T C T G A G A A C A G A G C T T C G G C T T A G A T T T C G C A C C A A A T C A A T G A A A G T A A T A A T G A A A A G T C T G A A T A A G A A
1901  T A C T A G G C T T A G A T G C C T T T G T A C T T G T G T A A A T A A C T T G A G T C A T G T A C C T T T G G C G G A A C A G A A T A A A T A A A A G G T G A A A T T C C A A T G C T C T A T
2001  G T A T A A G T T A G T A A T A C T T A A T G T T C T A C G G T T G T T C A A T A T C A T C A A A C T C T A A T T G A A A C T T A G A A C C A A A A T C T C A A T C T T T T C T T A A T G A A
    
```