Volume 17 Number 11 1989 Nucleic Acids Research

The glycinin *Gy₁* gene from soybean

Thomas L.Sims and Robert B.Goldberg

Department of Biology, University of California, Los Angeles, CA 90024-1606, USA

Submitted April 20, 1989 EMBL accession no. X15121

The nucleotide sequence of the Gy_1 glycinin storage protein gene is shown below. It was obtained by sequencing both strands of the gene contained in lambda DA28-30, a clone isolated from a genomic library prepared using leaf DNA from the soybean variety Dare. The transcribed region is shown in bold type and the first base in this region is considered +1. Potential regulatory sequences are shown in bold underlined type. These include the CACA (-428) and CATGCAT elements (-109 and -253), legumin box (-90 to -117), TATA box (-32) and polyadenylation signals (2498 and 2612). This sequence is discussed in Nielsen *et al.* (1989) Plant Cell 1:313-328, and is accession number X15121 in the EMBL database.

```
 -638  TAGCCTAAGT ACGTACTCAA AATGCCAACA AATAAAAAAA AAGTTGCTTT AATAATGCCA AAACAAATTA ATAAAACACT TACAACACCG GATTTTTTTT
 -538  AATTAAAATG TGCCATTTAG GATAAATAGT TAATATTTTT AATAATTATT TAAAAAGCCG TATCTACTAA AATGATTTTT ATTTGGTTGA AAATATTAAT
 -438  ATGTTTAAAT CAACACAATC TATCAAAATT AAACTAAAAA AAAAATAAGT GTACGTGGTT AACATTAGTA CAGTAATATA AGAGGAAAAT GAGAAATTAA
 -338  GAAATTGAAA GCGAGTCTAA TTTTTAAATT ATGAACCTGC ATATATAAAA GGAAAGAAAG AATCCAGGAA GAAAAGAAAT GAAACCATGC ATGGTCCCCT
 -238  CGTCATCACG AGTTTCTGCC ATTTGCAATA GAAACACTGA AACACCTTTC TCTTTGTCAC TTAATTGAGA TGCCGAAGCC ACCTCACACC ATGAACTTCA
 -138  TGAGGTGTAG CACCCAAGGC TTCCATAGCC ATGCATACTG AAGAATGTCT CAAGCTCAGC ACCCTACTTC TGTGACGTTG TCCCTCATTC ACCTTCCTCT
                                                                                                          MetAlaLysL
  -38  CTTCCCTATA AATAACCACG CCTCAGGTTC TCCGCTTCAC AACTCAAACA TTCTCCTCCA TTGGTCCTTA AACACTCATC AGTCATCACC ATGGCCAAGC
       euValPheSe rLeuCysPhe LeuLeuPheS erGlyCysCy sPheAlaPhe SerSerArgG luGlnProGl nGlnAsnGlu CysGlnIleG lnLysLeuAs
   63  TAGTTTTTTC CCTTTGTTTT CTGCTTTTCA GTGGCTGCTG CTTCGCTTTC AGTTCCAGAG AGCAGCCTCA GCAAAACGAG TGCCAGATCC AAAAACTCAA
       nAlaLeuLys ProAspAsnA rgIleGluSe rGluGlyGly LeuIleGluT hrTrpAsnPr oAsnAsnLys ProPheGlnC ysAlaGlyVa lAlaLeuSer
  163  TGCCCTCAAA CCGGATAACC GTATAGAGTC AGAAGGAGGG CTCATTGAGA CATGGAACCC TAACAACAAG CCATTCCAGT GTGCCGGTGT TGCCCTCTCT
       ArgCysThrL euAsnArgAs nAlaLeuArg ArgProSerT yrThrAsnGl yProGlnGlu IleTyrIleG lnGlnG
  263  CGCTGCACCC TCAACCGCAA CGCCCTTCGT AGACCTTCCT ACACCAACGG TCCCCAGGAA ATCTACATCC AACAAGGTCC ATCTTGTCCA AACTTCACAT
  363  ATAAATATAT AATAGACTTA AATATGTTTA AGGGTTTGAT AAATGAGGGA ATTTTATTTT AGATTTTTAA TAATTTTTTT TGTTTTGAGT TTTTATATAT
  463  TAAAATTTTT GTTTTGATTT CTTCCATATG ACGTAACATA ATCATATCAT TGATAATGTT GGGTTCCTAA TTTTTGTTTG TTTGTTGTTT TGTAATATGA
           lyLysG lyIlePheGl yMetIleTyr ProGlyCysP roSerThrPh eGluGluPro GlnGlnProG lnGlnArgGl yGlnSerSer ArgProGlnA
  563  ATAGGTAAGG GTATTTTTGG CATGATATAC CCGGGTTGTC CTAGCACATT TGAAGAGCCT CAACAACCTC AACAAAGAGG ACAAAGCAGC AGACCACAAG
       spArgHisGl nLysIleTyr AsnPheArgG luGlyAspLe uIleAlaVal ProThrGlyV alAlaTrpTr pMetTyrAsn AsnGluAspT hrProValVa
  663  ACCGTCACCA GAAGATCTAT AACTTCAGAG AGGGTGATTT GATCGCAGTG CCTACTGGTG TTGCATGGTG GATGTACAAC AATGAAGACA CTCCTGTTGT
       lAlaValSer IleIleAspT hrAsnSerLe uGluAsnGln LeuAspGlnM etProArg
  763  TGCCGTTTCT ATTATTGACA CCAACAGCTT GGAGAACCAG CTCGACCAGA TGCCTAGGGT GAGCCACATA GCAATATTAG ATATTATAAT TCTTTAAAGG
  863  TTTAAATATC ATTTTAGTTC GTGGAGTTGC ACTTTCTAAT TTAGTACCTA TAGATTAAAA TATGCCAATT GAATCCTTAT AGTTGTGTTT TTTTATCCAA
  963  TTTGGTTCTT GTCTTGAAAT AAATGGACAA TATTGTAGCT GATAAAAAAA GGAAACTGGA CTACATTGTA ACGTTAAGAT TAGAATTCTT AAGTTCTAAT
                                                            A rgPheTyrLe uAlaGlyAsn GlnGluGlnG luPheLeuLy sTyrGlnGln
 1063  ACTAGCTGGT TACAGATTGA CAACTATTTG TTTTGACAAT TCTTGGCAGA GATTCTATCT TGCTGGGAAC CAAGAGCAAG AGTTTCTAAA ATATCAGCAA
       GluGlnGlyG lyHisGlnSe rGlnLysGly LysHisGlnG lnGluGluGl uAsnGluGly GlySerIleL euSerGlyPh eThrLeuGlu PheLeuGluH
 1163  GAGCAAGGAG GTCATCAAAG CCAGAAAGGA AAGCATCAGC AAGAAGAAGA AAACGAAGGA GGCAGCATAT TGAGTGGCTT CACCCTGGAA TTCTTGGAAC
       isAlaPheSe rValAspLys GlnIleAlaL ysAsnLeuGl nGlyGluAsn GluGlyGluA spLysGlyAl aIleValThr ValLysGlyG lyLeuSerVa
 1263  ATGCATTCAG CGTGGACAAG CAGATAGCGA AAAACCTACA AGGAGAGAAC GAAGGGGAAG ACAAGGGAGC CATTGTGACA GTGAAAGGAG GTCTGAGCGT
       lIleLysPro ProThrAspG luGlnGlnGl nArgProGln GluGluGluG luGluGluGl uAspGluLys ProGlnCysL ysGlyLysAs pLysHisCys
 1363  GATAAAACCA CCCACGGACG AGCAGCAACA AAGACCCCAG GAAGAGGAAG AAGAAGAAGA GGATGAGAAG CCACAGTGCA AGGGTAAAGA CAAACACTGC
       GlnArgProA rgGlySerGl nSerLysSer ArgArgAsnG lyIleAspGl uThrIleCys ThrMetArgL euArgHisAs nIleGlyGln ThrSerSerP
 1463  CAACGCCCCC GAGGAAGCCA AAGCAAAAGC AGAAGAAATG GCATTGACGA GACCATATGC ACCATGAGAC TTCGCCACAA CATTGGCCAG ACTTCATCAC
       roAspIleTy rAsnProGln AlaGlySerV alThrThrAl aThrSerLeu AspPheProA laLeuSerTr pLeuArgLeu SerAlaGluP heGlySerLe
 1563  CTGACATCTA CAACCCTCAA GCCGGTAGCG TCACAACCGC CACCAGCCTT GACTTCCCAG CCCTCTCGTG GCTCAGACTC AGTGCTGAGT TTGGATCTCT
       uArgLys
 1663  CCGCAAGGTA CGTACATCAT TCATCAAAGA TCAACATACA TTTATACATT AAACTAATAT TTGTTGCCAA ATATTTATTA ATTTTATTGA TAATTAATTT
 1763  TTTTAGAAAA TTTGTTTGAT CACTTTTAAT GGAGTCTTTC ATCTTAATTA CATTATTTAT ACTTAGACTA ATGATTTATT GATTAATAAT AATCTTAGAT
 1863  ACACTATAAA ATGTGTGACG GAGTTATCTT AACACTTGCA TGGATTCTAT CTTTTCTGTC TTTATATATA GAAATAGAGA GAAAAAAAAA GAAAAGATTG
                                                                                                       As nAlaMetPhe
 1963  ATGAAAAAAG CAAAACAAAA AATAGTATTA TTATAAAAAT ATTGGATGAA TTTGTTGTGA CTCTTGCATG CATTGATGTA CGATGCAGAA TGCAATGTTC
       ValProHisT yrAsnLeuAs nAlaAsnSer IleIleTyrA laLeuAsnGl yArgAlaLeu IleGlnValV alAsnCysAs nGlyGluArg ValPheAspG
 2063  GTGCCACACT ACAACCTGAA CGCGAACAGC ATAATATACG CATTGAATGG ACGGGCATTG ATACAAGTGG TGAATTGCAA CGGTGAGAGA GTGTTTGATG
       lyGluLeuGl nGluGlyArg ValLeuIleV alProGlnAs nPheValVal AlaAlaArgS erGlnSerAs pAsnPheGlu TyrValSerP heLysThrAs
 2163  GAGAGCTGCA AGAGGGACGG GTGCTGATCG TGCCACAAAA CTTTGTGGTG GCTGCAAGAT CACAGAGTGA CAACTTCGAG TATGTGTCAT TCAAGACCAA
       nAspThrPro MetIleGlyT hrLeuAlaGl yAlaAsnSer LeuLeuAsnA laLeuProGl uGluValIle GlnHisThrP heAsnLeuLy sSerGlnGln
 2263  TGATACACCC ATGATCGGCA CTCTTGCAGG GGCAAACTCA TTGTTGAACG CATTACCAGA GGAAGTGATT CAGCACACTT TCAACCTAAA AAGCCAGCAG
       AlaArgGlnI leLysAsnAs nAsnProPhe LysPheLeuV alProProGl nGluSerGln LysArgAlaV alAla *
 2363  GCCAGGCAGA TAAAGAACAA CAACCCTTTC AAGTTCCTGG TTCCACCTCA GGAGTCTCAG AAGAGAGCTG TGGCTTAGAG CCCTTTTTGT ATGTGCTACC
 2463  CCACTTTTGT CTTTTTGGCA ATAGTGCTAG CAACCAATAA ATAATAATAA TAATAATGAA TAAGAAAACA AAGGCTTTAG CTTGCCTTTT GTTCACTGTA
 2563  AAATAATAAT GTAAGTACTC TCTATAATGA GTCACGAAAC TTTTGCGGGA ATAAAAGGAG AAATTCCAAT GAGTTTTCTG TCAAATCTTC TTTTGTCTCT
 2663  CTCTCTCTCT CTTTTTTTTT TCTTTCTTCT GAGCTTCTTG CAAAACAAAA GGCAAACAAT AACGATTGGT CCAATGATAG TTAGCTTGAT CGATGATATC
 2763  TTTAGGAAGT GTTGGCAGGA CAGGACATGA TGTAGAAGAC TAAAATTGAA AGTATTGCAG ACCCAATAGT TGAAGATTAA CTTTAAGAAT GAAGACGTCT
 2863  TATCAGGTTC TTCATGACTT GGAGCTC
```