The glycinin Gy_2 gene from soybean

Vu H.Thanh, Nilgun E.Tumer and Niels C.Nielsen

USDA Agricultural Research Service and Agronomy Department, Purdue University, West Lafayette, IN 47907, USA

Submitted April 20,1989 EMBL accession no. X15122

The nucleotide sequence of the Gy_2 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 (-130 and -199) and CATGCAT elements (-108), legumin box (-89 to -116), TATA box (-27) and polyadenylation signals (2733 and 2844). This sequence is discussed in Nielsen *et al.* (1989) Plant Cell 1:313-328, and is accession number X15122 in the EMBL database.

```
-801  GANTCNTTAC AATAATTTCT TGGATAGAGA AATTATATAT TAGGGACATT TTGTTTTTTG GATTATCCTC GTTTTGATTG TCAATATAAA AAAAAAGTAA
-701  GCGAAAATAA GAAGAAAATG ACTAAGTTTT AAACAATTAT ATTAGTTAAA CAAAAATACT TATATGTTTT AGGTCTATAA TAAATATGAT TAATTTTATT
-601  TTGAGTTCTA ATTTTAATTT TTTTTTCTTT GTATGAAAAT AAGTCCATTT TATTCTCATC CAAGAAATTA TTATATATAC AGTACAAATT ATTCTCTACT
-501  AATCTTTATG ATATGTAAAC CAAAGCATGA TTAATTATTT CCTAATTAAA TTAAAAAAAA TTGAGTAACT AGTATCACAT AAAGATCTGT CAAAACCATG
-401  AATAGTGATA ATTTCATAAC ACCTAATAAT TTCTCTATCC CTAGAGTGAG GTCATTTTTG GACGGCCCTT CTTCTCACTT TGAGGATCCC ATCATTGTCA
-301  CACGGTGTAT TCATATAATT TCATATGTCA TAAACCCGCG AACATGAAAT GAAACCATGG TCCCCCTCTC CACCACCGTT TTCTGGCAAT TGCATGCAAT
-201  ACAAACACAC TTGGTATTTG TCACATAATG TTGATGTCGA GCCACCTCAC ACCCATGAAC TTAATGAGGT GTAACACACA AGGCTTCCAT AGCCATGCAT
-101  ACTGAAGAAT GTCTCAAGCT CAGCACCCCA CTCCTGTGAC GTGTCCCTCA CCCACCTTCC TCTCTTCCCT ATAAATAACC ACGCCTCAGG TTCTCCGCTT
                                                    MetAla LysLeuVall euSerLeuCy sPheLeuLeu PheSerGlyC ysPheAlaLe
  -1  CACAACACAA ACATTCTCTC CATTGTCTTT GAACACTCAT CACCATGGCC AAGCTTGTTC TTTCCCTTTG TTTCCTTCTT TTCAGTGGCT GCTTCGCTCT
      uArgGluGln AlaGlnGlnA snGluCysGl nIleGlnLys LeuAsnAlaL euLysProAs pAsnArgIle GluSerGluG lyGlyPheIl eGluThrTrp
 100  GAGAGAGCAG GCACAGCAAA ATGAGTGCCA GATCCAAAAG CTGAATGCCC TCAAACCGGA TAACCGTATA GAGTCGGAAG GTGGGTTCAT TGAGACATGG
      AsnProAsnA snLysProPh eGlnCysAla GlyValAlaL euSerArgCy sThrLeuAsn ArgAsnAlaL euArgArgPr oSerTyrThr AsnGlyProG
 200  AACCCTAACA ACAAGCCATT CCAGTGTGCC GGTGTTGCCC TCTCTCGCTG CACCCTTAAC CGCAATGCCC TTCGTAGACC TTCCTACACC AACGGTCCCC
      lnGluIleTy rIleGlnGln G
 300  AGGAAATCTA CATACAACAA GGTCCATAAT CTTGTTTAAT TTTTTTTTTA TCAGATGTTA ATGGTTAGTT TTTGTTAGTA AAATTGATTG AAACTGTGGC
 400  TTTTGTTTTT CTTCTTAACC GTCCAACCTA TCTTATATAT TCAAATTAAT AATAACTTGT TTCAATTTCT CAGTTTTGTG AAACCATGCA TTTTGAAAAG
                                                          l yAsnGlyIle PheGlyMetI lePheProGl yCysProSer
 500  AAAATGGTTA TGCTAATGTT TTTTTTTTCT TTGTGTTGTT GTGTAATATA TATGAATAGG TAATGGTATT TTTGGCATGA TATTCCCGGG TTGTCCTAGC
      ThrTyrGlnG luProGlnGl uSerGlnGln ArgGlyArgS erGlnArgPr oGlnAspArg HisGlnLysV alHisArgPh eArgGluGly AspLeuIleA
 600  ACTTATCAAG AGCCGCAAGA ATCTCAGCAA CGAGGACGAA GCCAGAGGCC CCAAGACCGT CACCAAAAGG TACATCGCTT CAGAGAGGGT GATTTGATCG
      laValProTh rGlyValAla TrpTrpMetT yrAsnAsnGl uAspThrPro ValValAlaV alSerIleIl eAspThrAsn SerLeuGluA snGlnLeuAs
 700  CAGTGCCTAC TGGTGTTGCA TGGTGGATGT ACAACAATGA AGACACTCCT GTTGTTGCCG TTTCTATTAT TGACACCAAC AGCTTGGAGA ACCAGCTCGA
      pGlnMetPro Arg
 800  CCAGATGCCT AGGGTGAGAC CACATAGCAA TATTAGATAT TATAATTCTT TAAAGGTTTA AATATCATTT TAGTTCGTGG AGTTGCACTT TCTAATTTAG
 900  TACCTATAGA TTAAAATATG CCAATTGGAT CCTTATAGTT GTGTTTTTTT ATCCAATTTG GTTCTTGTCT TGAAATACAT GGACAATATT GTAGCTGATA
1000  AAAAAAGGAA ACTGGACTAC ATTGTAACGT TAAGATTAGA ATTCTTAAGT TCTAATACTA GCTGGTTACG GATTGACAAC TATTTGTTTT GACAATTCTT
        ArgPh eTyrLeuAla GlyAsnGlnG luGlnGluPh eLeuLysTyr GlnGlnGlnG lnGlnGlyGl ySerGlnSer GlnLysGlyL ysGlnGlnGl
1100  GGCAGAGATT CTATCTTGCT GGGAACCAAG AGCAAGAGTT TCTAAAATAT CAGCAGCAGC AGCAAGGAGG TTCCCAAAGC CAGAAAGGAA AGCAACAAGA
      uGluGluAsn GluGlySerA snIleLeuSe rGlyPheAla ProGluPheL euLysGluAl aPheGlyVal AsnMetGlnI leValArgAs nLeuGlnGly
1200  AGAAGAAAAC GAAGGAAGCA ACATATTGAG TGGCTTCGCC CCTGAATTCT TGAAAGAAGC GTTCGGCGTG AACATGCAGA TAGTGAGAAA CCTACAAGGT
      GluAsnGluG luGluAspSe rGlyAlaIle ValThrValL ysGlyGlyLe uArgValThr AlaProAlaM etArgLysPr oGlnGlnGlu GluAspAspA
1300  GAGAACGAAG AGGAGGATAG TGGAGCCATT GTGACAGTGA AAGGAGGTCT AAGAGTCACA GCTCCAGCCA TGAGGAAGCC ACAGCAAGAA GAAGATGATG
      spAspGluGl uGluGlnPro GlnCysValG luThrAspLy sGlyCysGln ArgGlnSerL ysArgSerAr gAsnGlyIle AspGluThrI leCysThrMe
1400  ATGATGAGGA AGAGCAGCCA CAGTGCGTGG AGACAGACAA AGGTTGCCAA CGCCAAAGCA AAAGGACATT GATGAGACCA TTTGCACAAT
      tArgLeuArg GlnAsnIleG lyGlnAsnSe rSerProAsp IleTyrAsnP roGlnAlaGl ySerIleThr ThrAlaThrS erLeuAspPh eProAlaLeu
1500  GAGACTTCGC CAAACATTG GTCAGAATTC ATCACCTGAC ATCTACAACC CTCAAGCTGG TAGCATCACA ACCGCCACCA GCCTTGACTT CCCAGCCCTC
      TrpLeuLeuL ysLeuSerAl aGlnTyrGly SerLeuArgL ys
1600  TGGCTTCTCA AACTCAGTGC CCAGTATGGA TCACTCCGCA AGGTACGTAC ATCATACATG TTATAAATTA AGCTCAACAA ATGTTAAAAA ATGGTAATAG
1700  TTATATATAG TAATTAAAAT GTGTGACGAC AAAGTTATCT TAGCACTTGC ATATGCATTA TTCATGTGCT TTTTTTTTAT ATATAGAAAT GCCCGAAGTT
1800  ATTAAGTAGT CACAGTTTCT TGTGTCGATT TTACTAATCT AATTAATCTC TGCATGAGAT GTAATTATGC ATGTGTTATT AATTTTTGTT ACATGTGATA
1900  CAATTAGTAG ATAATTTTGT CCTTTAGATA TATGACTCAT GAGTATGAAA TTTTTTTCAA CCTTGAAAAA TTTGTTTTTG ACTTTTAACC AAAAAATACT
2000  TTGATTAAAA AAAAGAAAA ATATTTTTGT GCCAAAAATA TTCTGATTCA CAAAAACATA TTTTTCTATG AGACTATATA TTGACATGTG ATAAAATACC
2100  TGATAATAAT TTCCATAGGA TTAGGAAGAA AAGGTATAAG ATAGTGAGAT AATTAATTAA TATCATGATG AAAAGATAAA AGGAAAATAA TAATATTATA
                                                                    AsnA laMetPheVa lProHisTyr ThrLeuAsnA
2200  AAAATATTGG ATGAATAACA TGTTGTGATT AACGTACACT TGCATGGATA TTGATGTATG GTGCAGAATG CTATGTTCGT GCCACACTAC ACCCTGAACG
      laAsnSerIl eIleTyrAla LeuAsnGlyA rgAlaLeuVa lGlnValVal AsnCysAsnG lyGluArgVa lPheAspGly GluLeuGlnG luGlyGlyVa
2300  CGAACAGCAT AATATACGCA TTGAATGGGC GGGCATTGGT ACAAGTGGTG AATTGCAATG GTGAGAGAGT GTTTGATGGA GAGCTGCAAG AGGGAGGGGT
      lLeuIleVal ProGlnAsnP heAlaValAl aAlaLysSer GlnSerAspA snPheGluTy rValSerPhe LysThrAsnA spArgProSe rIleGlyAsn
2400  GCTGATCGTT CCACAAAACT TTGCGGTGGC TGCAAAATCC CAGAGCGATA ACTTTGAGTA TGTGTCATTC AAGACCAATG ATAGACCCTC GATCGGAAAC
      LeuAlaGlyA laAsnSerLe uLeuAsnAla LeuProGluG luValIleGl nHisThrPhe AsnLeuLysS erGlnGlnAl aArgGlnVal LysAsnAsnA
2500  CTTGCAGGGG CAAACTCATT GTTGAACGCA TTGCCAGAGG AAGTGATTCA GCACACTTTT AACCTAAAGA GCCAGCAGGC CAGGCAGGTG AAGAACAACA
      snProPheSe rPheLeuVal ProProGlnG luSerGlnAr gArgAlaVal Ala *
2600  ACCCTTTCAG CTTCCTTGTT CCACCTCAGG AGTCTCAGAG GAGAGCTGTG GCTTAGAGAG CCCTTTTTGC ATGTGCTACC ACACTTTTGT CCTTTTGGGC
2700  TCTAGCTAGT GCTGGTAGCA GTAAATAAGA ATGAATAAAA GAAACAAAGG CTTAGATGCC CTTTGTTAAG TGTAGAATAA CGGGTATAGT AACTATGTAC
2800  CCTTGTGTAA TGAGTCACTC AAGTACCTAT GCGGGAAAAG CATAAATAAA AGAAGAAATT CCAATAAATT TTCTGTCCAA TCTCTTTTCT TTTCTTTTTA
2900  TTCTTTTCTA TCCTTCTTGC AGAACAAAAT AAGGAAATAT TGTTGATAAC AATTGGTCTA AGAGATAGTT AGCTTGACCG ATGATATCTT TGGGAAGCGA
3000  TTGCAGGACA TGATGAGCAT GAAAGACTAA AATTGAAAGT GTCACACCCC CAAGAAAGAG AGTTTTTTCT TTTTTTTTTT AGATACATAT AACAATATTA
3100  ATAATATTGT TATCATATAT ATATATATAT ATAATAATAA TAATAATAAT TATTATTCAT GCCAATTAAT CGATTTTAAT TTTTCCATAA GGAAATCAAT
3200  ATTTTTTACA TTAAATATTT CCATTATTTT TACATTTTTC TGATTTTACA AATATAGTTA TCATTTTTTT TAATTTAAAA AAGGTAAAAG TAAAATATGT
3300  GTTTTAATTA AAATGTGTTA TTATATTATT TTAAAAAATA GTATAACAAT AGTGTAATAT AAATAGTGTA TACATTTATT TTCTTGTAAA TAAGTCTTCG
3400  TAAAATGCAA TTATACCTGC AGATATTGAT AAATAAATTC ATAAATAGAA ACGCAAAATA TTTGTATAAT AGATACAGAT AGATAAATAA GTGGTTTTTA
3500  TTGTAGAATT TTGTGACATA ACACAAATTT GATCCTCAAT AACTATATAT CGTTTGTTTT TTCTTTTCAC TTAAGCTNAA TTTTACATAT GAGTATGTAA
3600  ACTTAAAAGA AAATTTGTGC TTTTGATATA TATATATATA TATATATATA TATATATATA TATATGAGTT CATGAAAAAC AAAGTGTTCC CAAACCATTT
3700  GCTCTAATTT ATTTGTAAAA ACCTTAATTA ATCGGTTTCA TATGCGTTTG TATACGTAAT TGCTTGTGGT ACTTTTCTTC GGGCTCCATG TCATCGTCTG
3800  CAGCAAAATA TAGATCT
```