The glycinin *Gy₃* gene from soybean

Tae-Ju Cho 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. X15123

The nucleotide sequence of the *Gy₃* glycinin storage protein gene is shown below. It was obtained by sequencing both strands of the gene contained in lambda DA28-6, 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 CATGCA elements (-102 and 249), legumin box (-89 to -117), TATA box (-27) and polyadenylation signals (2922 and 2936). This sequence is discussed in Nielsen *et al.* (1989) Plant Cell 1:313-328, and is accession number X15123 in the EMBL database.

```

-297 TGAAGGAAA GAAAGAGCA GAGGAGAAA AGAAATGAA CATGCATGG TCCCCCCCC AGGACATCAT GGGTTTCTGC CATTTCGAAT ACAACCACTG
-197 AAACACCTTT CTCTTTGTCA CGTAATCGAG ATTCGGAAGC CACCTTAACAC CATTAACTAT ATAGTGTAAG ACAGAGGGGT TCCATAGCCA TGCATCTGTA
-97 AGAATGTCTT AAGCTCAGCA CCCCACTCTC GAGTCGTGTC CCTCACTTCA CCTTCTCTCTC TTCCCTATATA ATAACCACGC CTCAGGTTCCT CCGCTTCACA
      MetAlaL ysLeuValLe uSerLeuCyC sCysPheAla
4 ACACAAACAT TCTCTCCATT TGTCCTTGAA TATAATAC TC AGCATGGCTA AGCTTGTTCT TTCCCTTTGT TTTCTGCTTT TCAGTGGCTG CTGCTTCGCT
      rgGlnGlnPr oGlnGlnAsn GluCysGlnI leGlnArgLe uAsnAlaLeu LysProAspA snArgIleGl uSerGluGly GlyPheIleG
104 TTCAGTTTCA GAGAGCAGCC ACAGACAAAC GAGTGCCAGA TCCAACGCCT CAATGCCCTA AAACCGGATA ACCGTATAGA GTCAAGAAGT GGCCTTCATTG
      nThrTrpAs nProAsnAsn LysProPheG InCysAlaGl yValAlaLeu SerArgCyC hAsnAlaLeu ArgArgProC erIyrThrAs
204 AGACATG6AA CCCTAACACG AACGCTATCC AGTGTCGCGG TGTGCGCTCT TCTGCGCTCA CCTCTCAACG CAACGCGCTT CGACGACACT CCTACGACAA
      nAlaProGln GluIleTyrrI TGcInGlnG leGlnGlnG
304 GAGATCTACA TCCAACTTAT TTTTITATTT TTATCATCAA ATATAGTATT ATAGTTTITTA TTTAAATATAT TATCGAAACT ATATCGAACT AAATCCCACT
      CCGTCCCCAG GAGATCTACA GTTCATTGTT TCCACACTTT ATATACATATA CATACATATA TGTATGTATG TATGATTTAC TATGATTTAC TAACGTATAT
404 TACAACCTTT ATTCAACAGG GTTCATTGTT TTAGTACACG AAATCATTGC ACAAAATGCA AGAAAGAGAA CATCAACGGC TGAACCTTTG TCCAAAACAA
      nThrTrpAs nProAsnAsn LysProPheG InCysAlaGl yValAlaLeu SerLeuIleA spThrAsnSe rPheGlnAsn rPheGlnAsn GluLeuAspG InMetProAr g
504 ACCTATATTT TTTTAGCAGA GGACTGTGAT TTAGTACACG AAATCATTGC ACTTAATGCA CTCTCTATGC GTTTTTTAACT TTTTAAAGTA TTTTAAAGTA TCTTTTACT
604 ATACTTAATT G6AAATAAT AAATAAATAA CTG6AAATG ACTTAATGCA CTCTCTATGC GTTTTTTAACT TTTTAAAGTA TTTTAAAGTA TCTTTTACT
704 CG66GATT7TA TCAACTTACA TATTCCTATG TTATTAGTAT GAATACATTA GCAAATTTATA TCTTTGATATA TTTTAAAGTA TTTTAAAGTA TCTTTTACT
804 GGTAAAAAAA CCATATTAGC AAAACCTCTCA AGAGTAACGT TAAGGACATC GATAATGTTT TGTTTGTATC AGTTTTCAGA ATTTAAACAT TTTGAACGCA TTTGAACGCA
      lySe rGlyIlePhe GlyMetIleP heProGlyCy sProSerThy rPheGluGluP
904 AAATGGTTAT TATAGTAATG TTTTGTGAT TGTTTAATGT GAATAGGTAG TGGTATTTTT GGCATGATAT TCCCGGGTTG TCCTAGCACA TTTGAAGGAG
      roGlnGlnLy sGlyGlnSer SerArgProG InAspArgHi sGlnLysIle TyrHisPheA rgGluGlyAs pLeuIleAla ValProThrG lPheAlaTy
1004 CTCAACAAAA AG6ACAAGAG AGCAGGCCCC AAGACCGTCA CGAGAAGTAC TATCATCTCA GAGAGGGTGA TTTGATTGCA GTGCCAACCG GTTTCGCATA
      rTrpMetTyr AsnAsnGluA spThrProVa lValAlaVal SerLeuIleA spThrAsnSe rPheGlnAsn rPheGlnAsn GluLeuAspG InMetProAr g
1104 CTGGATGTAC AACATAGTAA ACACCTCTGT TGTGGCGGTT TCTCTATTG ACACCAACAG CTTCGAGAAC CAGCTCGACC AGATGCGTAG GGTACGTGAG
1204 CCACATATAT A6CATTAGAT ATTAGAATTC TTTAAAGATT TAAATATCTT TTTGGTTTGT TGTGAGTTTC AGAATTAACA ATGATTTGTT TGTATTAAT
1304 GTCAATTTTA TCTTCATATAT TGTGTTTTT TATGCAATTT GGTCTCTGTT GTGAAATCAC TTTAAATAAC TGGAAAAAAT TGTATTTGTT TGTATTAAT
1404 AATTGGACAA CGTTGTAAAC TGCAATTAAGA TTAGAACCTC TAAGTCTCTAA TACTGGTTAC AGAATTAACA ATGATTTGTT TGTATTAAT TGTATTAAT
      ArgPheT yrLeuAlaGl yAsnGlnGlu GluGlnPheL euGlnTyrrG lProGlnLys GlnGlnGlyG lyThrGlnSe rGlnLysGln LysArgGlnG
1504 CAGAGATTCT ATCTTGCTGG G6ACCAAGAG CAAGAGTTTC TACAGTATCA GCCACAGAG CAGCAAGGAG GTACTCAAGG CCAGAAAGGA AAGCGTCAAG
      InGluGluGl ySerIleLe GlySerIleL euSerGlyPh eAlaProGlu PheLeuGluH isAlaPheVa lValAspArg GlnIleValA rglLysLeuG l
1604 AAGAAGAAGA AAGCAAGGAG GGACGATAT TGAATGGCTT CGCCCCG6AA TCTTGG6AAT ATGCTTCTGT C6TG6AG6AG GAAAGCTACA
      nGlyGluAsn GluIleGluG luLysGlyAl aIleValThr ValLysGlyG lyLeuSerPro ProThrGluG lyThrGlnGln nArgProGlu
1704 AGGTGAGAAG GAAGAGGAGG AAGAGGGTGC CATTGTGACA GT6AAGGAG GTCTCAGCGT GATAAGCCCA CCCCAGGAG AGCAGCAACA AAGACCGGAG
      snIleGlyGl nThrSerSer ProAspIleP heAsnProG l nAlaGlySer lIleThrThrA laThrSerLe uAspPhePro AlaleuSerT rPheLysLe
1804 GAAGAGGAGA AGCCAGATTG TGACGAGAAA GACAAACATT GCCAAGGCCA AAGCAGAAAT GGCATTGAGC AGACCATTTG CACAAATGAG CTTCCGACCA
      uSerAlaGln PheGlySerL euArgLys euArgLys CTTGACATCT TCAACCTCTCA AGCTGGTAGC ATCACACCGC CTACACGCTC GCACTTCCCA GCCCTCTCGT GGCCTCAAAC
1904 ACATTG6CCA GACTTCATCA CTTGACATCT CTAACCTCTCA AGCTGGTAGC ATCACACCGC CTACACGCTC GCACTTCCCA GCCCTCTCGT GGCCTCAAAC
2004 CAGTGCCCAAG TTTGGATCAC TCCGCAAGGT ACTTACATCA TTAATTCATG TTATAAAGC TCAAAAAGTG TTAATAATAGT ACTAATAGTT ATATATAGTA
2104 ATTAATAATG GT6CAACAA CATCTTAAACA CTTGCAATGAT TATGTCATTAT TTTTGTCTCT TTTTGTGTCG AAATATG6GC TATATGTCCTA TATATATATA
2204 TATGTATATG AAAATGTTAA CCAATGTTTT TGGATATTGG TTAAGGAATG AGAAATAATT TTTACTAAGAT TAAGTTTTCAA TATGATCTTC ATATAAACA
2304 TTTTAAAAAT TTTTITTAAT AATGTACTTA TACTACATAT TAACAAAATC ATATATATAT GATTTTTCATA GATTTTTCAA GGAATGTTAGA GAATGTTAGA
      Asn AlaMetPheV alProHisTy rAsnLeuAsn
2404 GATGGAATAA TTACTGTGAT AATAAAAAGA TATGTTATGA TTAACGTACA CTTGATGTAT GGTGCAAGAT GCTATGTTGC TGCCACACTA CAACCTGAAC
      AlaAsnSerI leIleLeuAl eLeuAsnGly ArgAlaLeuF cGlnGlnVala lAsnCysAsn GlyGluArgV aIlePheAspG lGluGluGlnV
2504 GCAACAGACA TAATATG6C CCGGCAATGG ATTAGATTGA PheAlaValA laAlaArgSe rGlnSerAsp AsnPheGluI utyrValSerPh eLysThrAsn AspArgProS erIleGlyAs
2604 TGTAAATGTT GCCACAAAGC TTTGCG6TGG CTGCAAGATC ACAGAGCGAC AACTTCGAGT ATGTTTTCAT CAAGACCAAT GATAGACACT CGATCG6CAA
      nLeuAlaGln AlileAsnSerL euLeuAsnAl aLeuProGlu GluValIleG leGlnGlnAla eAsnLeuGln rAsnGlnGln lLysAsnAsn
2704 CTTGCGAGGT GCACAACTCA TGTGTAACGC ATTGCGGGAG GAAGTGATTC AGCAAACTTT TAACCTAAG6 AG6CAGCAGG AC6GCGAGGT CA6AAGCAAC
      AsnProPheS erPheLeuVa lProProLys GluSerGlnA rgArgValVa lAla *
2804 AACCTTTTCA GCTTCTCTGT TCCACTTAAG GAGTCTCAGA G6AGAGTTGT GGTCTAGAGC CTTTITTGCA TGTGCTACCA CACTTTTGTG CTTTGTGGCA
2904 CTAGTTAATG T6GCTAGCAA TAAATTAAGAA CAGATAAAGC ATCAAAAGGCT TTGCTG6CCTT TTTGTTAAGTA TAAATGTTGA TGAATGTTGA TGAATGTTGA TGAATGTTGA
3004 G6TGAATGAG TCACTCAAGT ACCTATCG6G GAAATTCATA AATTAAGAAA GAAATTTTCAA TGAATTTTCT GTCAATCTCT TTTTCTTTTT ATTTCTTTGT
3104 TTTCTTCTTA TCTTCTTCTG AGGACAAAT AAGGAAAAAT GTATTGAAAA CAATTGATGAC TAAGTATGAT ATGATATATAT AATAGTTTGT
3204 ATTTGAGAA TTTGTCGACT GTTGTGCGCT CTAATTATAT TGAAGAACC TCAATTGTTT TCGTATGCTG TT

```