

Complete nucleotide sequence of the gene encoding a glycinin A₂B_{1a} subunit precursor of soybean

Chikafusa Fukazawa, Takayuki Momma, Wakako Higuchi and Kyoko Uda¹

Genetic Engineering Laboratory, National Food Research Institute, Ministry of Agriculture, Forestry and Fisheries, Yatabe, Tsukuba, Ibaraki 305 and ¹Tokyo Kasei University, Itabashi-ku, Tokyo, Japan
Submitted July 22, 1987 Accession no. Y00398

We have determined the complete nucleotide sequence of a A₂B_{1a} gene of glycinin, the most abundant storage protein of soybean, using the cDNA as a probe (1). The gene spans 2864- base pairs from the capsite (▼) to the 3'-end (↘), and is interrupted by three intervening sequences occurring in the mRNA sequence coding for Gly-75 (between the first letter and the second letter of that codon) and Arg-159 of A₂ subunit region, and for Lys-338 of B_{1a} subunit region. The TATA sequence (-31, -23) and the AATAAA polyadenylation signal (+2843 - +2848) are underlined. Detailed analysis of the 5'-flanking region is presented elsewhere (2).

```

tttgcaattgcatgcaatacaaacacttggattttgtacataatgttgatgtgcaactgtgcaagccactcacacccatgaacttaatgaggtgtaacacacaaggctt -121
ccatagcagtcgcatctgaagaatgttctcaagctcagcaccctcctgtgagctgtccctcaccctctctctcttccctataataaccacgctcaggttctccgctt -1
▼ (exon 1) MetAlaLysLeuValLeuSerLeuCysPheLeuLeuPheSerGlyCysPheAlaLeuArgGluGlnAla
CACAAACACAAACATTCTCTCCATTGTCTTTGAACACTCATCAACGCAAGCTGTTCTTCCCTTTGTTCCCTCTTTTCAGTGGCTGCTTCGGCTTGAGAGACGAGCA 113
(+) cap site 20 30 40
GlnGlnAsnGluCysGlnIleGlnLysLeuAsnAlaLeuLysProAspAsnArgIleGluSerGluGlyGlyPheIleGluThrTrpAsnProAsnAsnLysProPheGlnCys
TTGGCATGATATTCCCGGGTTGTCTAGCACTTATCAAGAGCGCCGAAGTCTCAGCAACGAGGACGAGGCCAGAGGCCCAAGACCTCAACAAAGCCATTCCAGTGT 227
AlaGlyValAlaLeuSerArgCysThrLeuAsnArgAsnAlaLeuArgProSerTyrThrAsnGlyProGlnGluIleTyrIleGlnGlnG
GCCGGTGTGGCCTCTCTCGCTGCACCTTAAACCGCAATGCCCTCTGAGACCTCTCTACACCAACGCTCCCAAGAAATCTACATACACAAAGTccataatctgtttaatt 342
ttttttatcatagatttaattgtagttttttgtagtaaaattgatgaaactgtggcttttttcttcttaaccgtccaaacctctctatatatacaatataataataac 456
lyAsnGlyIle
ttgtttcaatctctcagttttgtgaaccaatgcattttgaaagaaatgggttatgctaattgttttttttcttctgtgtgtgtaatatatgaaagGTAATGGTATT 570
(exon 2) 90 100 110
PheGlyMetIlePheProGlyCysProSerThrTyrGlnGluProGlnGluSerGlnGlnArgGlyArgSerGlnArgProGlnAspArgHisGlnLysValHisArgPheArg
TTTGGCATGATATTCCCGGGTTGTCTAGCACTTATCAAGAGCGCCGAAGTCTCAGCAACGAGGACGAGGCCAGAGGCCCAAGACCTCAACAAAGCCATTCCAGTGT 684
GluGlyAspLeuIleAlaValProThrGlyValAlaTrpTrpMetTyrAsnAsnGluAspThrProValValAlaValSerIleIleAspThrAsnSerLeuGluAsnGlnLeu
GAGGGTGATTGTGTCAGTGCCTACTGGTGTGTCAGTGTGATGTACAAACATGAAGACACTCCTGTTGTTGCCGTTTCTATTATTGACACCAACAGCTTGAGAACCAAGCTC 798
AspGlnMetProArg
GACCAGATGCTTACGGTgagaccacatagcaatattagatatattataattctttaaaggtttaaatatcatttttagttcgtggagttgcactttctaatttagtacctatagatt 912
aaaatagccaattgtagctccttatagttgtgtttttttatccaattggttctgttcttgaatacataggacaattatgtatgataaaaaaagaaactggagcacatctgta 1026
160 (exon 3) 170
ArgPheTyrLeuAlaGlyAsnGlnGluGlnGlu
acgttaagattagaattcttaagtctctaactactagctggttacggattgacaactattgttttgacaattcttggcagAGATTCTATCTTGCTGGGAACCAAGAGCAAGAG 1138
PheLeuLysTyrGlnGlnGlnGlnGlnGlySerGlnSerGlnLysGlyLysGlnGlnGluGluGluAsnGluGlySerAsnIleLeuSerGlyPheAlaProGluPheLeu
TTTCTAAATATCAGCAGCAGCAGCAAGAGGTTCCCAAAAGCCAGCAAGAAAGCAACAGCAAGAAAGAAACGAAGGAACACATATTGAGTGGCTTCGCCCTCGAATCTTTG 1252
LysGluAlaPheGlyValAsnMetGlnIleValArgAsnGluGlnGluAsnGluGluGluAspSerGlyAlaIleValThrValLysGlyLeuArgValThrAlaPro
AAGAAGAGGCTTCGGCTGAACATGCAGATAGTGAGAACTCAAGGTGAGAACGAGAGGAGATAGTGGAGCATTGTGACATGTGAAGGAGGCTTGAAGAGTCACAGCTCCA 1366
250 260 270 280
AlaMetArgLysProGlnGlnGluGluAspAspAspGluGluGluGlnProGlnCysValGluThrAspLysGlyCysGlnArgGlnSerLysArgSerArgAsnGlyIle
GCCATGAGGAGGCCACAGCAAGAGAGATGATGATGATGAGGAGAGAGCAGCCACAGTGGCTGGAGCAGACAGCAAGGTTGCCAACGCCAAAGCAAGGAGCAGAAATGGCAT 1480
AspGluThrIleCysThrMetArgLeuArgGlnAsnIleGlyGlnAsnSerSerProAspIleTyrAsnProGlnAlaGlySerIleThrThrAlaThrSerLeuAspPhePro
GATGAGACCATTTGCACAAATGAGACTTCGCCAAACATTGGTCAGAATTCTCATCCTGACATCTACAACCTCAAGCTGGTAGCATCACACCGCCACCGCTTGACTTCCCA 1594
AlaLeuTrpLeuLysLeuSerAlaGlnTyrGlySerLeuArgLys
GCCCTCTGGCTCTCAAACTAGTGGCCAGATGATGATCCTCCGCAAGgtacgtacatcatcatcatgtttataaataaagctcaacaaatgttaaaaaatggaatagttatatat 1708
agttaataaaagtgtgtgacgacaaagttatcttagcacttgcatatgcatattatcatgtgtctttttttatatatagaataatgccgaagtattaaagttagtcacagtttcttg 1822
tgtcgattttcatcaatcaataatctctgcatgagatgtaattatgcatgtgtttattattttttgtacatgtgatacaattagtagataaattttgtccttttagatatatgac 1936
tcattgagtagtaaaattttttcaacctgtaaaattgttttgacttttaacaaaaaatactttgattaaaaaaagaaaaatattttgtgcaaaaaatattctgattata 2050
caaaaaatattttttctatgagactatattgacattgataaaaatcctgataaatttccataggattaggaagaaaaaggtataagatagtgagataatttaattatc 2164
(exon 4)
atgatgaaagataaaagaaaaataataattataaanaatttgatgataaactgtgtgtgataacgtacacttgcatggaattatgtatgtgtgcgaatgctATGCTATGTT 2278
ValProHisTyrThrLeuAsnAlaSerIleIleTyrAlaLeuAsnGlyArgAlaLeuValGlnValValAsnGlyGluLeuGlnGlu
GTGCGACATACACCTCGAAGCGCAACAGCATAATACGCAATGTAATGGCGGCTTGTGACAGTGGTAAATTCGAATGTCGAGAGAGCTTTGATGAGAGAGCTCGACAG 2392
GlyGlyValLeuIleValProGlnAsnPheSerGlnSerAspAsnPheGluTyrValSerPheLysThrAsnAspArgProSerIleClyAsnLeuAla
GAGGGTGGCTGATGCTCCCAAACTTTGCGTGGCTGCAAAATCCAGAGCGATAACTTTCAGTGTGTGTCATTCAAGCAACATATAGACCCCTCGATCGGAACCTTGCA 2506
GlyAlaAsnSerLeuLeuAsnAlaLeuSerGluGluIleThrPheAsnLeuLysThrPheAsnLeuLysThrPheAsnLeuLysThrPheAsnLeuLysThrPheLeuVal
GGGGCAACTCAATTGTTGAACGCAATGGCAGAGGAAGTATGACGACACTTTAACTAAAGAGCCAGCAGCCAGCAGGTGAAGAACACACCCCTTCAGCTTCTCTGTT 2620
ProGlnGluSerGlnArgAlaValAla**
CCACCTCAGAGCTCTCAGAGAGAGCTGTGGCTTAGAGAGGCCCTTTTGCATGTGCTACACACTTTTGCCCTTTGGGCTTAGCTATGCTGCTGAGCAGTAATAAGATGAAT 2734
AAAAGAACAAAGCGCTTAGATGCCCTTTGTTAAGTCGAGAATAACGGGTATAGTAACATATGTACCCCTGTGTATGAGTCACTCAAGTACCTATGCGGGAAGACATAAATA 2848
poly (A) signal
AGAAGAAATTCCAATaattttctgccaactctcttttttttttttttttttttttttttttttttttttccttcttcgagaacaaaataaggaaattttgtgataac..... 2948
```

References: (1) Momma, T. et al. (1985) FEBS Lett. **188**, 117-122. (2) Fukazawa, C. et al. submitted.