Nucleotide sequence of the *Sesbania rostrata* leghemoglobin (*Srglb3*) gene

Peter Welters, Birgit A.Metz, Jeff Schell and Frans J.de Bruijn

Max-Planck-Institut für Züchtungsforschung, Abteilung Genetische Grundlagen der Pflanzenzüchtung, 5000 Köln 30, FRG
Submitted January 3, 1989   EMBL accession no. X13814

Sesbania rostrata is a tropical leguminous shrub, which can be nodulated on the stem, as well as the root by its nitrogen fixing symbiont *Azorhizobium caulinodans* ORS 571 (1, 2). Using a heterologous *lb* gene probe from soybean (*lbc3*), we isolated a *lb* cDNA and subsequently two genomic *lb* loci from *S. rostrata* (3). The nucleotide and deduced amino acid sequence of one of these *lb* loci (*Srglb3*) is shown in Fig. 1. The first 25 amino acids of GLb3 and GLb2 (3) are identical to those determined for *S. rostrata* LbVI and VII by protein sequencing (4). Inspection of the amino acid differences between GLb3 and 2 and their chromatographic properties (4) suggests that *Srglb2* encodes LbVI and *Srglb3* LbVII.

```
-388  GAATTCACTCTTTATTACTAGAATTAACACAGGATTTCACTAAATAATAGTTTTTTTTTATAAAAAGATATTGCTTCCTTGAAAAAAAGATGTTTGTAAAA  -289
-288  ATTTTTAGAATAATAATTACATTAATTAAAAACTGAAAAAAAATATTAAATGTTTTGAAGTTTTCAATAATTTTTAAATTATTAATTAAATTTTTAAAAAG  -189
-188  ATTATTGTCTCTTAATAATGTCAATGGCCACCCTTCAGGAGCCAAGAGTTTCTTTCAATGACAATCACAAGGAATCAACTCCATCTCTTTGCACTCTCTC   -89
-88   ACCCTCCTATATAAACAAGTATTGGATGGAAAATTATTGCATACATCGGAGAATAAAAATAACAAAAAGGTAAAAAAACAAAAATAA██GGTTTCACA     12
                                                                                                G  F  T
  13  GAGAAACAGGAGGCTTTAGTGAATGCTTCATACGAAGCATTTAAGCAAAACCTTCCTGGCAACAGTGTTCTGTTTTACAGCTT GTAAGTTTTCTCTCTC   111
      E  K  Q  E  A  L  V  N  A  S  Y  E  A  F  K  Q  N  L  P  G  N  S  V  L  F  Y  S  F
 112  AATATATATATGTTATTAATTTGTTTTGTTTTGGAATTAAGAGATGGTGTAATGTAAATGTATCCTTTGGTTTTGGTTTAATTAAAAATAGTATATTGGAGA  211
                                                                                                    I  L  E  K
 212  AAGCACCAGCAGCAAAGGGCATGTTCTCTTTTCTAAAGGACTTTGATGAAGTGCCCCAAAATAATCCCAGTCTCCAGGCCCATGCTGAAAAGGTTTTTGG   311
       A  P  A  A  K  G  M  F  S  F  L  K  D  F  D  E  V  P  Q  N  N  P  S  L  Q  A  H  A  E  K  V  F  G
 312  ACTGGTAAGTGCAACATTTGTAGACCTTTCTATTGCACATCTCTTTAATTCATAAAAAAATGTTAAATTTTCTTCAACAATAAACTTATATTCTAATTAT   411
       L
 412  CTATCGAGCTAGTGGAAAAATAATAAAAATAAAATGATATTTAAGTATTATTTTATTTTAACTAATTAGATAAATTGTTCATAAAACTATTAAGACGATT   511
 512  TAATATTTTTCTAAATATATATTATCATATTCGAGTAATTGTAATAATATCAAGTTGTCTTAATATTTCTTAATTATTACTTTATATCCTTCTTTATCAT   611
 612  AATTCTCATGTCATATTATTATATTTATTGAATTATAGGTGCGTGATTCCGCCGCTCAACTCCGAGCAACCGGAGTAGTGGTATTAGCAGATGCAAGTTT   711
                                                V  R  D  S  A  A  Q  L  R  A  T  G  V  V  L  A  D  A  S  L
 712  AGGTTCTGTCCATGTCCAAAAAGGAGTCCTTGATCCTCACTTTGTGGTATGACATTTAATTGAGTGCTATACGGTGTACTCTTAAATTTTTCATATAGTT  811
       G  S  V  H  V  Q  K  G  V  L  D  P  H  F  V
 812  TAATGATCACTTTTTTTTTTTCTAATTTGGTAATGCATAATAAGTATATATTAACATTATAAATTGCTTATGACACACATATAATAAGACATAAAAAATTTA  911
 912  ATAATATATATAGGCCCTATTTGGATACACTTCTGCAGAAGAACTTCTAGGAGATAAAAATAGAATGAATTAAAGAGGTCAATGAATTTAACTTTCTCCA  1011
1012  TAAGTTGAAATTAGTTACTTATGGAGAAGTTAAATATGAGAGCTTTTGGAGTAGGTGGACTACAAGAACTAATTTCAACTTATGAAGAAGTTAAATGCAT  1111
1112  TTGATCTCTTTGATTTATTCTGTTTTTTTCACCTAGAAGTACTTTTGTAGAAAATTTCAGTCCAAACTGGCCTATAGTTCAGTGACATTAATGCTTGTGACA  1211
1212  TAAATAGCTAGTAAGAGCAGGTCCAACATACTAATGTCATGGATCAATCTTTTGTAGGTGGTTAAAGAAGCATTGTTGAAAACACTAAAGGAAGCCGGAG  1311
                                                                      V  V  K  E  A  L  L  K  T  L  K  E  A  G  G
1312  GGGCCACTTGGAGCGATGAAGTGAGCAATGCTTGGGAAGTAGCCTATGATGAATTGTCAGCTGCAATTAAGAAGGCAATGAGT██AATATTATTCCGAC  1411
       A  T  W  S  D  E  V  S  N  A  W  E  V  A  Y  D  E  L  S  A  A  I  K  K  A  M  S
1412  AAATAAAAAAACTATTATAATAAATACTACTACAAGTTTTAAACTTGTTAAACAAGTTCCTGTAATAAACTGATA                            1485
```

Fig. 1: Nucleotide sequence of the *Srglb3* locus. Start and stop codons are highlighted in black. CAAT- and TATA-boxes are underlined. Major transcriptional start site (3) is indicated by a highlighted *T*. The single base pair gap at the end of the first exon indicates Phe codon (TT T) disruption by the intron. Base pairs highlighted in grey indicate single base pair differences between the *glb2* and 3 loci in the coding regions.

References: 1. Dreyfus. B and Dommergues, Y. (1981), FEMS Micr. Lett. 10: 313-317; 2. Dreyfus, B. et al. (1988), Int. J. Syst. Bact. 38:89-98; 3. Metz, B. et al. (1988), Mol. Gen. Genet., 214: 181-191; 4. Bogusz, D. et al. (1987), Arch. Biochem. Biophys. 254:263-271.