The sequence of the maize plastid encoded *rps3* locus

William E. McLaughlin and Ignacio M. Larrinua*

Lilly Research Laboratories, Lilly Corporate Center, Indianapolis, IN 46285, USA

Submitted May 11, 1987

Accession no. Y00340

The *rps3* locus of the maize plastid genome is found in the large unique region beginning at coordinate 28.93 and ending at coordinate 29.60 on the positive strand (1). This gene like its counterparts in *H. Labacum* (2) and *M. polymorpha* (3) is part of an S10 like operon. It follows the gene for ribosomal protein L22 (4) and precedes that for L16 (5, McLaughlin W E and IM Larrinua, manuscript in preparation). It does not contain a recognizable ribosome binding site. There is a strong bias in the codon usage, the %GC content of the three positions being 44.20, 30.36, and 25.89 respectively. This locus codes for a protein of 224 amino acids with a MW of 25,918 daltons. It has a high degree of homology to the *E. coli* S3 protein, 38.39% of the amino acids are completely conserved while another 12.94% are subject to conservative substitutions. This brings the total homology to 51.33%. The maize plastid S3 protein is 26 amino acids shorter at its carboxy terminus, however it contains internal insertions that add 18 amino acids to its length, thus this protein is only 8 amino acids shorter than its *E. coli* homologue.

```

RPS 3 AAAAAAAAAAAGAAATATAATAGAAAAAT ATG GGA CAA AAA ATA AAT CCA CTC GGT TTC AGA CTT GGT
                                     M G Q K I N P L G F R L G
ACA ACC CAA AAT CAC CAT TCC TTT TGG TTC GCA CAA CCA AAA AAT TAT TCT GAA GGT CTA CAG GAA GAT
T T T Q N H H S F W F A Q P K N Y S E G L Q E D
AAA AAA ATA CGG AAT TGT ATC AAG AAC TAT ATA CAA AAG AAT AGG AAA AAA GGC TCG AAT AGA AAA ATG
K K I R N C I K N Y I Q K N R K K G S N R K M
GAA TCA GAC TCA AGT TCT GAA GTA ATT ACA CAT ATA GAA ATT CAA AAA GAA ATC GAT ACA ATC CAC GTC
E S D S S S E V I T H I E I Q K E I D T I H V
ATA ATC CAT ATT GGA TTC CCC AAC TTA TTA AAG AAA AAG GGA GCA ATC GAA GAA TTA GAG AAA GAT CTC
I I H I G F P N L L K K K G A I E E L E K D L
CAA AAG GAA GTT AAT TCT GTA AAC CAG AGA CTT AAT ATT GCT ATC GAA AAG GTT AAA GAA CCT TAT AGA
Q K E V N S V N Q R L N I A I E K V K E P Y R
CAA CCT AAT ATT CTT GCG GAA TAT ATA GCT TTC CAA TTA AAA AAT AGA GTT TCA TTC CGA AAA GCA ATG
Q P N I L A E Y I A F Q L K N R V S F R K A M
AAA AAA GCT ATT GAA TTA ACT AAA AAA GCG GAT ATA AAG GGA ATA AAA ATC CAA ATT GCG GGT CGT CTA
K K A I E L T K K A D I K G I K I Q I A G R L
GCA GGA AAA GAA ATT GCA CTT GCC GAA TGC ATC AAA AAG GGT AGA CTA CCC CTC CAA ACA ATT CCG GCT
A G K E I A R A E C I K K G R L P L Q T I R A
AAA ATT GAT TAT TGC TGC TAT CCA ATT CGA ACT ATC TAT GGA GTA TTA GGT GTC AAA ATT TGG ATA TTC
K I D Y C C Y P I R T I Y G V L G V K I W I F
GTA GAA GAA GAA TAA CTAACCATGTCTTCCCATTCGCCA6TG
V E E E ?!?

```

*To whom correspondence should be addressed

1. Larrinua, IM et al. 1983 *Pl. Molec. Biol.* 2:129-149.
2. Shinozaki K et al. 1986 *Pl. Molec. Biol. Rep.* 4:110-147
3. Ohyama K et al. 1986 *Pl. Molec. Biol. Rep.* 4:148-147
4. McLaughlin WE and IM Larrinua 1987 *submitted to NAR*
5. Gold, B et al. 1987 *Proc. Natl. Acad. Sci. (USA)* 84:194-198.