Nucleotide sequence of the chloroplast genes for tRNA^{Gln} and the 4 kD K polypeptide of photosystem II from mustard (*Sinapis alba*)

Heike Neuhaus

Arbeitsgruppe Pflanzliche Zellphysiologie, Ruhr-Universität Bochum, Universitätsstrasse 150, D-4630 Bochum, FRG

Submitted December 1, 1988

Accession no. X13558

The sequence of a mustard chloroplast DNA region containing the genes *trnQ* for tRNA^{Gln} and *psbK* for the 4 kD 'K' protein of photosystem II (1) is presented. The two genes are encoded by opposite DNA strands. They are located much closer to the *rps16* gene for ribosomal protein S16 (H. Neuhaus, A. Scholz & G. Link, submitted) than has been found for the equivalent genes in tobacco (2). By comparison with partial amino acid sequence of the spinach 'K' protein (1), post-translational processing might be inferred to occur between amino acids A and K at position 24 and 25. The degree of sequence conservation for the derived mature protein of 37 amino acids is 97 % between mustard and tobacco (3) and 84 % between mustard and liverwort (4). The conservation between the presumptive N-terminal pre-sequences of 24 residues for the mustard and tobacco proteins is 67 %, while no significant homology was found between the pre-sequences from mustard and the liverwort. The hydropathy profile (not shown) of the derived entire *psbK* polypeptide of 61 amino acids (from the first methionine) reveals one membrane-spanning segment within the 'mature' region.

```

1  TTTATCTCTCTTTTCATTCAATTTTTCCTTAAAAATATGGATTTTAAACAGAAGGAGAAAGATGGTGTACAAAAGCAATAGAA
   rps16 ← -10
91  GATTGATTCGGTAATACTAGACTCTGGACCGGAAGGATTCGAACCTCOGATAGCGGGACCAAAACCCGTTCGCTTACCGCTTGGCTAC
   +1
181  GGGCCATTTTATTTTATTTTCAAGACTACTAAAGAGTATATGCTATGGTGTGTCGTCATTCATTTTAGCCCAATGCAATATAG
   trnQ ← -10 -35
271  ATTACATTTGGTCCATAGTTTTCACAGCTGTAGATAGCAAACTCAACTTACCTTTATTTGATCATTACATAGATTCATTAAGATATGTA
   -10 -35
361  TGAAATATATTATCTTTAATTCATACAGAGATGAAAGGATTTTGGATTGAGTAACTCAACAAAGCTCTTTTAGACTATCTCTCTT
   -10 -35
451  ATTTATTTATTTTTCCTTATATAAAAATATATTAATACTCAATCAAAATTAATTTATCCACAAGACCAATTTTGTATGCTT
   +1
541  AATATATTTAATTGATCTGATTTTGTATTAATTOGGCCCTTTTTCAGCACCTTTTATGTCGCCAAATTGCCAGAGGCTACGCCCTT
   N I F N L I C I C F N S A L F S S T F L V A K L P E A Y A F
631  TTGAATCCAATGCTAGATGTTATGCCGTAATACCTCTTTCTTCTCTCTAGCCCTTGTGTCGCAAGCAGCTGTAAGTTTTCGATGA
   L N P I V D V M P V I P L F F L L L A F V W Q A A V S F R ***
721  AATCTTAATACTGTCTTAGAAAAATTCAGGTAAAGATTCTTCAACAACCTCAAAAGATC

```

Figure legends: Coding regions and putative '-10' and '-35' promoter elements are boxed. Brackets indicate transcription start sites. The triangle marks the putative processing site.

Acknowledgements: I thank Professor G. Link for his interest and encouragement. This work was supported by the DFG (Li 261/7-2) and the Fonds der Chemischen Industrie, FRG

References: 1. Murata et al. (1988) FEBS L. 235, 283-288. 2. Shinozaki et al. (1986) EMBO J. 5, 2043-2049. 3. Deno and Sugiura (1983) Nucleic Acids Res. 11, 8407-8414. 4. Ohya et al. (1988) Nature 322, 572-574