

The 5' end of the coding region of a U6 RNA gene candidate from tomato starts with GUCC, a phylogenetically highly conserved 5' end sequence of U6 RNA

Attila Szukálek, Tamás Kiss and Ferenc Solymosy

Institute of Plant Physiology, Biological Research Center, Hungarian Academy of Sciences, PO Box 521, H-6701 Szeged, Hungary

Submitted January 26, 1990

EMBL accession no. X51447

In vertebrates, U6 RNA transcribed by RNA polymerase III (1) starts with a G residue, the first nucleotide of the evolutionarily highly conserved sequence GUYC (2). Since the 5' end of broad bean U6 RNA was published by us (3) to start with CUUC and upon RNA sequencing we probably missed the very 5' end of the molecule (2), we reinvestigated the nucleotide sequence of a plant U6 RNA at the genomic level. From a tomato genomic library we isolated and sequenced (4) a U6 genomic locus. Our data (Fig. 1) suggest that the 5' end sequence, GUYC (2), is present also in plant U6 RNA. We note that in our U6 gene candidate both the TATA Box and the upstream sequence element (USE) are richer in T residues and are positioned closer to each other and to the cap site of the putative transcript than in U2 (5) and U5 (6) *bona fide* genes of Arabidopsis. The 3' Box present in these latter is missing from our U6 gene candidate.

ACKNOWLEDGEMENTS

Thanks are due to Dr R.W.Breidenbach for the tomato genomic library and to Dr Y.Ohshima for the mouse U6 DNA clone. This work was supported by grants OKKFT(Tt)/1986, OTKA 564/86 and OTKA 174/88 to F.S. from the Hungarian Academy of Sciences.

REFERENCES

1. Reddy,R. *et al.* (1987) *JBC* **262**, 75–81.
2. Guthrie,C. and Patterson,B. (1988) *Annu. Rev. Genet.* **22**, 387–419.
3. Kiss,T. *et al.* (1987) *Nucl. Acids Res.* **15**, 543–560.
4. Henikoff,S. (1984) *Gene* **28**, 351–359.
5. Vankan,P. and Filipowicz,W. (1988) *EMBO J.* **7**, 791–799.
6. Vankan,P. *et al.* (1988) *Nucl. Acids Res.* **16**, 10425–10440.

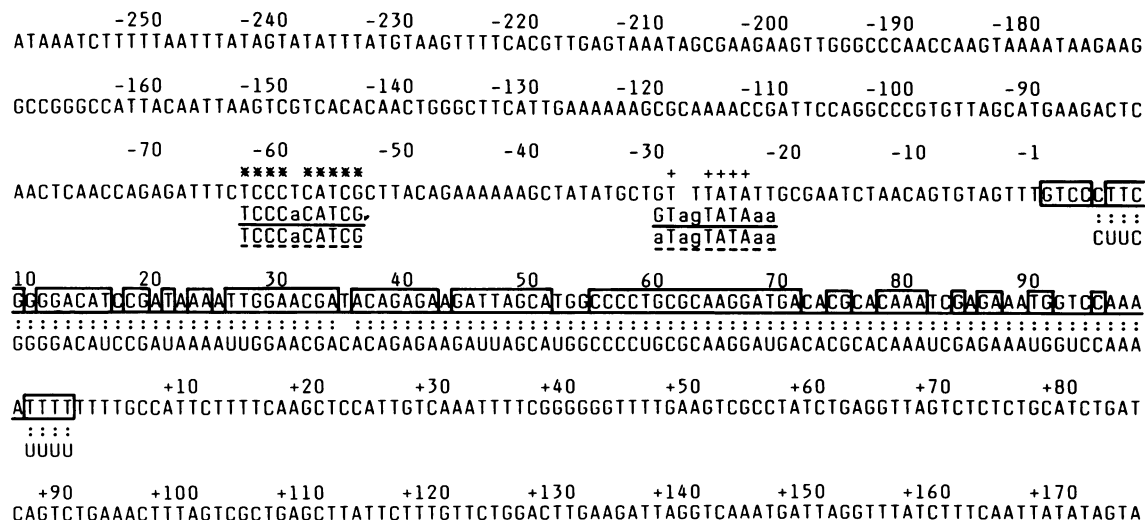


Figure 1. Nucleotide sequence of a U6 RNA genomic locus from tomato. The alleged coding region (103 nts) is underlined and is compared with the nucleotide sequence of broad bean U6 RNA (3) written below. The TATA Box and the USE are marked by crosses and asterisks, respectively, at the positions in which the residues match those in the corresponding regions of Arabidopsis U2.2 (5)/continuous underlining/ and of Arabidopsis U5 (6) /broken underlining/ *bona fide* genes. Mismatches are indicated by lower case letters in the U2.2 and U5 sequences. Evolutionarily highly conserved nucleotide residues (2) in the coding region are boxed.