Drosophila virilis Cu–Zn superoxide dismutase gene sequence

Jan Kwiatowski and Francisco J.Ayala

Department of Ecology and Evolutionary Biology, University of California, Irvine, CA 92717, USA
Submitted January 23, 1989 EMBL accession no. X13831

The sequence of 2103 bp given below, isolated from a lambda EMBL3 genomic library of Drosophila virilis, codes for Cu-Zn superoxide dismutase (SOD). An open reading frame and a termination codon are identifiable starting at positions 154 and 1163, respectively. The putative TATA box that starts at position 15 corresponds to one of three putatively identified in the D. melanogaster SOD gene (1). There are three putative polyadenylation signals (at 1311, 1359 and 1377); the middle one corresponds to the one identified in D. melanogaster. The coding region consists of two exons; the 549 bp intron separates the codons for amino acids 22 and 23, similarly as for D. melanogaster and for the first intron of the human gene (2). The inferred sequence of 153 amino acids has the same length as in D. melanogaster, but differs at 20 sites; the DNA coding sequences exhibit 81.4% identity. Two stretches in noncoding regions are also very similar in both species. One is 26 bp long, starting 60 nucleotides downstream from the 5' end of the intron, with 88% identity, but 1 bp is missing in D. virilis; it might be involved in splicing. The other, extending from positions 1329 to 1384 in D. virilis (1369 to 1430 in D. melanogaster), includes the putative polyadenylation signal in the middle, and exhibits 6 different and 47 identical (89%) bp, but gaps of 3 bp in D. melanogaster and 9 bp in D. virilis.

```
   1 AAGCTTACAAACGATATTTCTATCGATAGCTTTTGTGCCCTACCAGTTAGTGACATTAAGCATTCCGTAGCTTGCTCTTCGCACGACAAG
  91 TTCAGTGACTAAAATTGTTTAATTATTTTTCTATCTACGTTTCCGAATAAGCAAAACACAAAAATGGTGGTCAAAGCTGTATGCGTTATC
                                                                    MetValValLysAlaValCysValIle
 181 AATGGCGATGCCAAGGGCACAGTTTTCTTCGAACAGGAGGTAAGAGACTGTATATTATCTGCTTTCAACTTGAAACAAATATGGCAGCCC
      AsnGlyAspAlaLysGlyThrValPhePheGluGlnGlu
 271 ACAAAAACTAAGGTCAATTAGCCGGCCACACCGCACGACACCGAGCTAATAAATGTGTATGCTTGTGTATGTATGTAATTGTGTGTGTAT
 361 ATGTGCATACATATGTCATTGTGGACTGTTGCGGCGCTACTTGAATTTCAATTGCGTGTAGTTTGTGCTGTGTCATCTGATAAATGCATA
 451 ATTATATAAGCACAACAACATTGCACACATGCATGCCAGAAATGCGCAAGTATGTGTGTACGGTTAAAGGCAAACGGCGAAAATTCAATT
 541 CGCCTCTTATACCTCTCTTTAACCGTATACACATTTACTCTTGGCTTAGGTATTGAATTGTTTTCTCGAUCGGATTAGAATATGACCTTA
 631 ATTTCAACCTACCTTTGCCCAGCTAAAATGTTTCTGTCCTCATGGTTAGTTGATTCAGCTTATCAGCTTGTTGTGGCATTTTAATTGTTC
 721 CTATCTCGTTTTAACCCTATGCGTATCTGTTTTCGATCCACTTCAACAGGGCGAGGGTTGCCCCGTGAAGGTTACCGGCGAGGTAACCGG
                                                                    GlyGluGlyCysProValLysValThrGlyGluValThrGly
 811 CTTGGCAAAGGGTCAGCATGGCTTCCATGTGCATGAGTTCGGCGACAACACCAATGGCTGCATGTCGTCGGGACCGCACTTCAATCCCTA
      LeuAlaLysGlyGlnHisGlyPheHisValHisGluPheGlyAspAsnThrAsnGlyCysMetSerSerGlyProHisPheAsnProTyr
 901 CCAGAAGGAGCATGGCGCACCGACCGACGAGAATCGCCATCTGGGCGATTTGGGCAACATCATCGCCAATGGAGACGGTCCCACTCCCGT
      GlnLysGluHisGlyAlaProThrAspGluAsnArgHisLeuGlyAspLeuGlyAsnIleIleAlaAsnGlyAspGlyProThrProVal
 991 GAACATCTGCGATTGCAAGATCACATTGCTCGGCGCCAATAGCATTATTGGACGCACCGTTGTGGTGCACGCCGATCCAGATGACCTGGG
      AsnIleCysAspCysLysIleThrLeuLeuGlyAlaAsnSerIleIleGlyArgThrValValValHisAlaAspProAspAspLeuGly
1081 CAAGGGCGGACACGAGCTGAGCAAGACAACTGGCAACGCGGGAGCTCGCATTGGCTGCGGTGTCATCGGCATTGCCAAAATCTAAACAAC
      LysGlyGlyHisGluLeuSerLysThrThrGlyAsnAlaGlyAlaArgIleGlyCysGlyValIleGlyIleAlaLysIle***
1171 TCTAGAGCCTAAAAGCATTTCAAGCAAATGCATGCATAATAACCCACAAATATATATATATAAATATTGTATATGGATTTAATTTAGTGT
1261 ATCCACGGCGCGTAAATACGGTTTAAACTTTCGATTGGATATATACGTACATTAAAGAGATTTTCCTTTGTTAGTTGTTCAAAAGCAACC
1351 TTTAAGGCAATAAATGGTGTTATCAAATTAAAACTTGCTCTCACTAGGAGCCAGAATTGCTATGAGCAACAATTTTACGTACTAGTCCAA
1441 ACATTTTATTGGTTTCCAAAATACGTATATAAATTAGTAATATTTACAATTCTTGAAATCTAATTAGTCAAAGCTATGACATATGACTAT
1531 GAAACAATATCAACGGGTTTAGAAACAAAAACAGGATTTCTAATTTAGAATATGGGTTTGGTGGCGAGTGCTTCTGTTGGCTTCCATTAT
1621 GTCTTGTTGAAGTAGGACTGATAGACGATGCCCAGTGAAGAAGTGGCAAAGAGCGTAAAGTAGCTCCACCAGGTGCAAAGCCCAGAAAGC
1711 CCTTGTACACTTCCGTGGAGAACACCTGAAAGGTACACATAGCTCAGTTAAAAAAGAACTTTTGATATGAAAACTCTTACATTTTCAA
1801 ACTAGCAAATGAGGCCTGCAATGATTGCGCAGATTCCAGGATATTAAACTCGGCCTCCTCAACATTGGCGTCTTTGGGTGTTGTGGCCAA
1891 TTCTCCATCGGCGGCACATTCATGCTGCAGCACAGCCGTCTCTTTAAAGATGCCCGTCCACGTAAAGCCGCCCAGGCTCTTGTACATGAG
1981 CGTCAGATAAAGACACAATATCAGCGGTGCAGCAAATTGCAACGTAACAACACAGAGATAATAGAAAATTGACGAGATCTGCAAGGCAAA
2071 GACAATATATCAGTCGTTAATTATCCAAAGCTT
```

Acknowledgements: D. virilis library obtained from D. Maier, constructed by Dr. J. Flach in Dr. S. Artavanis-Tsakonas laboratory (Yale Univ.). Research supported by U.S. DOE.

REFERENCES:

1. Seto, N.O.L., Hayashi, S. and Tener, G.M. (1987) Nucl. Acids Res. 15, 10601.
2. Levanon, D., Lieman-Hurwitz, J., Dafni, N., Wigderson, M., Sherman, L., Bernstein, Y., Laver-Rudich, Z., Danciger, E., Stein, O. and Groner, Y. (1985) EMBO J. 4, 77-84.