Volume 15 Number 24 1987 Nucleic Acids Research

The sequence of the Cu−Zn superoxide dismutase gene of *Drosophila*

Nina O.L.Seto, Shizu Hayashi and G.M.Tener

Department of Biochemistry, University of British Columbia, Vancouver, BC V6T 1W5, Canada
Submitted October 9, 1987 Accession no.Y00517

The sequence shown below is a 1.8 Kb EcoR1 fragment from a genomic library made from an isogenic stock of *Drosophila melanogaster* (Oregon R) (J. Leung, unpublished). The sequence contains the transcribed region for a Cu-Zn superoxide dismutase (SOD) gene (in capitals), 413 bp of the 5'-untranslated region, and 247 bp of the 3'-untranslated region. The amino acid sequence of the Cu-Zn SOD of *Drosophila* has been reported (1) as well as the sequence of a corresponding cDNA (2). Transcription start and stop sites of the gene are assumed to be the limits of the sequence of the cDNA (2). Two additional triplets in the translated DNA sequence indicate that the initial translation product probably contains not only an N-terminal methionine residue but also a C-terminal valine. Both amino acids would have to be removed to give the amino acid sequence reported for SOD (1). In addition, the gene contains a single 725 bp intron which separates the triplets for amino acids 21 and 22 of SOD. This intron is in the same position as the first intron of the human Cu-Zn SOD gene (3). The transcribed region of the gene has three transitions (overlined) when compared with the sequence of the corresponding cDNA(2) from a different *D. melanogaster* source (4). These changes would not affect the amino acid sequence of the SOD. Possible control sequences are underlined. As with the cDNA, the genomic DNA hybridized to 68A4-9 on *Drosophila* polytene chromosomes.

```
                gaattcctggattcgtttttttatttatacaaaacaaagtaaattgaatagtttcgcgccactgtcattggaataaa -271
tggaaggcttccaagtgaaccacccgttcgttgaacagctcaaaaaattcaacgccatttgtgcagtaaaattgttccgttttcaaattt -181
tgtaatttgtagcttttatccttaaaaatgtaaataaatgtcccaataaaacatgagcttgaaatattacaaagaaaaaatgcttccagt  -91
aacggcataatagtgtgaacgaaacttttgcacaactgaacactaacagtaaaagttccgcatgtatttctaagctgctctgctacggtc   -1
ACACCATAGAAGATACCTGGAAAGTTCTCAACTTTTTTCGTTTTGATAAATTGATTAATTCATTCGAAATGGTGGTTAAAGCTGTCTGCG   90
                                                                M  V  V  K  A  V  C
TAATTAACGGCGATGCCAAGGGCACGGTTTTCTTCGAACAGGAGGTGAGAATCCAAAATCATTTGAACTTCTCTGCTCGGCAAAATGTAC  180
V  I  N  G  D  A  K  G  T  V  F  F  E  Q  E
GAAAAACAGAAGTTCTAAAGGTCAAATAGCCGGCTGCACCCGCGGCCCCCTCTTCCACTTCAATATGCTGCTTTAAATTCTGTCGAGCAT  270
TTTAATTAAGTCCGATTTGAGTTTACGCCTAGTCACCCAGCAAGTGCACCTTTATATTTATATAAGCCGCACCAAAATGCGCATATGTGT  360
GTGCGCTCAAGTGCCTACAGCAAAGGTCACGAAATTAGTACTGGACATAAAAAGGAGTTAAGATATAAAGCTCACTTGTTCGTAAAGTAT  450
CGTTAAATATCAACAAATATTTGTTTTAGAATAAGCATTAGGAATATGGGAATAATTAGAATGATGCTGTTCATAATTAATTTGTACATC  540
AAAGTCAAAGCAGCAATGTCAAGTGTCAAGTAAACGATTATAAACTTGATGATTACAGGTTATGTTTCAGTGCCGAGGAAATTTATGTTT  630
TTAATCTATAAAGATAACCAAATGTTTACTTTGCTGCCTATAAATATTTCCGTTTAACGTGTGTCTATTAACAAATGTTATTTTCTATAA  720
TAACCTATTATCATATGAAGTTGGCCACGCTCGTTATCATAATCAGTGCTTCTGCTCACTATTATACACAACTTGTGTCTTATCAGTATT  810
CGAGTATTATCTGAAGCGTTATAACCCAATCCCTTCATCCCGTCCACAGAGCAGCGGTACGCCCGTGAAGGTCTCCGGTGAGGTGTGCGG  900
                                                S  S  G  T  P  V  K  V  S  G  E  V  C  G
CCTGGCCAAGGGTCTGCACGGATTCCACGTGCACGAGTTCGGTGACAACACCAATGGCTGCATGTCGTCCGGACCGCACTTCAATCCGTA  990
  L  A  K  G  L  H  G  F  H  V  H  E  F  G  D  N  T  N  G  C  M  S  S  G  P  H  F  N  P  Y
TGGCAAGGAGCATGGCGCTCCCGTCGACGAGAATCGTCACCTGGGCGATCTGGGCAACATTGAGGCCACCGGCGACTGCCCCACCAAGGT 1080
  G  K  E  H  G  A  P  V  D  E  N  R  H  L  G  D  L  G  N  I  E  A  T  G  D  C  P  T  K  V
CAACATCACCGACTCCAAGATTACGCTCTTCGGCGCCGACAGCATCATCGGACGCACCGTTGTCGTGCACGCCGATGCCGATGATCTTGG 1170
  N  I  T  D  S  K  I  T  L  F  G  A  D  S  I  I  G  R  T  V  V  V  H  A  D  A  D  D  L  G
CCAGGGTGGACACGAGCTGAGCAAGTCAACGGGCAACGCTGGTGCCCGCATCGGGTGCGGCGTTATTGGCATTGCCAAGGTCTAAGCGAT 1260
  Q  G  G  H  E  L  S  K  S  T  G  N  A  G  A  R  I  G  C  G  V  I  G  I  A  K  V  *
AATCTATTCCGATGTCGGCCACTGTGCTGATCTACTCTATTTAGCACTACCCACTGGAGATATGCAAACGATATACATACTTCTAAACAT 1350
AAATACATAGCCTGTGGTCTGTTAGTTGATACGCAACCTTTGAGGTTCAATAAATTGGTGTTTTGAAATTGCCCCATAAACaaaagttat 1440
agttttcatttgagttgagatggtaagaatgaatatatcacttgttgctcgacgaattc 1499
```

Acknowledgement: This work was supported by the Medical Research Council of Canada.

REFERENCES

(1) Lee, Y.M., Friedman, D.J., and Ayala, F.J. (1985) Arch. Biochem. Biophys. 241, 577-589.
(2) Seto, N.O.L., Hayashi, S., and Tener, G.M. (1987) Nucleic Acids Res. 15, 5483.
(3) Levanon, D., Lieman-Hurwitz, J., Dafni, N., Wigderson, M., Sherman, L., Bernstein, Y., Laver-Rudich, Z., Danciger, E., Stein, O., and Groner, Y. (1985) The EMBO Journal, 4, 77-84.
(4) Poole, S.J., Kauvar, L.M., Drees, B., and Kornberg, T. (1985) Cell, 40, 37-43.