 © *1990 Oxford University Press*

Nucleotide sequence of the mouse preprosomatostatin gene

Guy Fuhrmann, Roland Heilig[1], Jules Kempf[1] and Alix Ebel*
Centre de Neurochimie du CNRS LP 6511, 5 rue Blaise Pascal 67084 Strasbourg Cedex and [1]Unité de Biologie Moléculaire et de Génie Génétique de l'INSERM U 184, 11 rue Humann 67085 Strasbourg Cedex, France

Submitted January 31,1990 EMBL accession no. X51468

A mouse brain genomic library cloned in lambda 2001 was screened with a rat somatostatin cDNA probe (1). A 22.5 Kb fragment was isolated and subcloned in pEMBL 18^+ plasmid. A 2564 bp insert, sequenced by the dideoxy method of Sanger (2), contains a 116 amino acids preprosomatostatin coding region. The mouse preprosomatostatin gene (Fig. 1) consists of two exons of 238 and 361 bp, separated by a single intron of 665 bp; its 5′- and 3′-flanking regions show short phylogenetically conserved sequences which are described to act as regulatory elements (3).

REFERENCES

1. Goodman,R.H., Jacobs,J.W., Dee,P.C. and Habener,J.F. (1982) *J. Biol. Chem.* **257**, 1156–1159.
2. Sanger,F.S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
3. Fuhrmann,G., Heilig,R., Kempf,J. and Ebel,A. (manuscript in preparation.

```
-795 · TCTAGATGCCTCTGCTCTTCCAACTGTACAGAGAAATGATCTGCCAGTCCCTTAAGCTTTCTTGGTTCAAGTTCTGAGGC
-715 · TTGTTAGATGATTTCTGAGACTGATTCCCAGGGCTGTTTCCAGGTGCCAAATGTAGGCTTTCTTTCTCCCCCTCCCTCCT
-635 · GTGTGTGTGTGTGTGTGTCTGTCTGTCTGTCTATCTGTCTCTCCAGTGGTTTTCTTTTGTTAAAATATAAAGATAGGCCG
-555 · CTTGGACAAAGTGAGGTTCCTTTACAGCTCAATTTCATCCCTTTTTTCCTACAAGGCTTTAAGAGATGGAGGGAGAGAAT
-475 · ATAGTTCAGTCCTCTTAATTGCAAATTCATTCTGAGATTGTTTCCTAGACAGATCGCTCTAAGTCTCACTCGCCATACAA
-395 · AAAGTTAAAGGTGAATGCAAGTCCAGTAATCTGGGTACATTGACAGGTACCCAACTGAGTGTGATGATGTATTGCTAACC
-315 · AAGGACTGAGTGATCTCTGTGTAATTAAGTGTGCTCCTATGTGGCTGAAATATGGGAGCGGCATGTCAGCACTGAGTGAA
-235 · GGTAAGATTGTTTGGTCTCTGTGGCATGGAGAATTTCATGTGCCTGCGTGGGTGCAGGCTTTCTTTTTCTTTTTTTAAA
-155 · AAAATAAACCACTTTAGATCGTGTCGCCTCCCCTCACTTCTGTGATTGATTTTGCGAGGCTAATGGTGCGTAAAAGCACT
- 75 · GGTGAGATCTGGGGGCGCCTCCTTGGCTGACGTCAGAGAGAGAGTTTAAAAAGGGGAGACCGTGGAGAGCTCCAT ÅGC
  + 4 · GGCTGAAGGAGACGCTACCGAAGCCGTCGCTGCTGCCTGAGGACCTGCGACTAGACTGACCCACCGCGC
  +73 · TCCAGCTTGGCTGCCTGAGGCAAGGAAG ATG CTG TCC TGC CGT CTC CAG TGC GCC  CTG
                                     M   L   S   C   R   L   Q   C   A   L  ·10
+131 · GCT GCG CTC TGC ATC GTC CTG GCT TTG GGC GGT GTC ACC GGC GCG CCC TCG
        A   A   L   C   I   V   L   A   L   G   G   V   T   G   A   P   S  ·27
+182 · GAC CCC AGA CTC CGT CAG TTT CTG CAG AAG TCT CTG GCG GCT GCC ACC GGG
        D   P   R   L   R   Q   F   L   Q   K   S   L   A   A   A   T   G  ·44
+233 · AAA CAG GTAAGGACTGTCCCCCTTACGAATCCCCCAGCCTTCCTGCTGCAGCCCCTGCGACAGGTGTTTAGCGGGCG
        K   Q                                                              ·46
+310 · CTTCTCAGAGTCGCTCAGCCCCTAAGCTCCCAGGGAAACTTTTGAAGTTAGGGTCCCCTCTTACTCTTTTCAGAATTGAT
+390 · CAGCGCAGGTGGTCACCTTGCAGGTAAGTTCCTCCCTGGCTTTCAAGAAAATTCAGAAAGCAGGTAGGAGAGACTGGGCT
+470 · CTATCCCTGGTGCTGGCAAGAGGGTTGCTGACCCCGGTGCTGAAACGCAATGTTTGTTCACAAAACGATTCTGGGGCAAG
+550 · AGGCTCTTCAGTTTATGGGAAGGTTTGCTTTTACCTTTCAAAAAATCTGGCAAGAACTCAGCTCCACTTCATATCTCATC
+630 · ACAGGAAAATGGACGTGCGCAAACTATTGGCATTCTATGTAGTTATAAACGACTTCTTAATCATCTGTGATTTCTCTGTG
+710 · TTTTAAGGCAGAGCACTTTCTGAAAGACTTGCGTTTGGGGAAGCTTTTTTTCTCCTGCATAATCTAATGAATACATACAG
+790 · CAATCACCCATATGATTGTGAAAACTGGGTTTTGAGTGATTAAATCTTACTTTCAAACCACATTTTCCCCCCTCTCCCAT
+870 · TCCTCCCTTTTGTTCTTCTCACTGCCCTATCCAG GAA CTG GCC AAG TAC TTC TTG GCA GAG CTG
                                          E   L   A   K   Y   F   L   A   E   L  ·56
+934 · CTG TCC GAG CCC AAC CAG ACA GAG AAT GAT GCC CTG GAG CCC GAG GAT TTG
        L   S   E   P   N   Q   T   E   N   D   A   L   E   P   E   D   L  ·73
+985 · CCC CAG GCA GCT GAG CAG GAC GAG ATG AGG CTG GAG CTG CAG AGG TCT GCC
        P   Q   A   A   E   Q   D   E   M   R   L   E   L   Q   R   S   A  ·90
+1036 · AAC TCG AAC CCA GCA ATG GCA CCC CGG GAA CGC AAA GCT GGC TGC AAG AAC
        N   S   N   P   A   M   A   P   R   E   R   K   A   G   C   K   N  ·107
+1087 · TTC TTC TGG AAG ACA TTC ACA TCC TGT TAGCTTTAATATTGTTGTCCTAGCCAGACCTCT
        F   F   W   K   T   F   T   S   C  STOP                            ·116
+1147 · GATCCCTCTCCCCCAAACCCCATATCTCTTCCTTAACTCCTGGCCCCCGATGCTCAACTTGACCCTGCA
+1216 · TTAGAAATTGAAGACTGTAAATACAAAATAAAATTATGGTGAGATTATG AAGAGCAAGCGTTTGACTTCTT
+1287 · ATTGAGTAAATTGTTTTGTTCAGAAATACATGATGAGCATGGGTTGTGGTTTCCGTAAACATATTTTAGGTGTACATACC
+1367 · CACATACACAAAGGTGTATTGAAAATTCCCTGCGGTAATAAAAGTTAGACTTTAAAAATGACTGTTCCTTTAAAACTTGA
+1447 · ATAACTTTGAATGTCTATAAATTGAACCATGTATCCGAATGATGGCTATTGGGAACAGCTGCCATAAACATCCTTTGCAT
+1527 · GTTAAAGACATTTGGGAATAGCTGTAGGATTACCAAAATGTGTTTTAGAATTTATGGGCAGTCAGGCTGAAGCGCTTAAC
+1607 · ACTGTACTATGCTGATTCACAAAACATAAAACATATTTAAAATGGAAAGAAAATAATCAAAAAAGTGGGAAAGAACTTGG
+1687 · TTTAAAGAGGTGAAACAAACTTTCAGCTTTTAAATTTTGTTTTGGTTTGGTTTTTTGTTTGTTTGTGAACATTTACCTCT
+1767 · AGA
```

Figure 1. Mouse preprosomatostatin gene. Exons are in high capital letters; intron and flanking regions are in small capital letters. TATA box and poly(A) addition site are underlined.

* To whom correspondence should be addressed