Proc. Natl. Acad. Sci. USA
Vol. 88, pp. 4250–4254, May 1991
Biochemistry

Expression of growth/differentiation factor 1 in the nervous system: Conservation of a bicistronic structure

SE-JIN LEE

Carnegie Institution of Washington, Department of Embryology, 115 West University Parkway, Baltimore, MD 21210

Communicated by Donald D. Brown, February 1, 1991

ABSTRACT **Growth/differentiation factor 1 (GDF-1) is a recently described member of the transforming growth factor β superfamily isolated from a day-8.5 mouse embryo cDNA library. Northern (RNA) analysis of embryonic mRNA detected two GDF-1 transcripts [1.4 kilobases (kb) and 3.0 kb in length] displaying distinct temporal patterns of expression. Only the 3.0-kb transcript was detected in adult tissues, where its expression was restricted almost exclusively to the central nervous system. Comparison of murine and human brain cDNA sequences corresponding to the 3.0-kb transcript revealed high conservation of two nonoverlapping open reading frames with poor conservation of the intervening spacer region and the putative 5′ and 3′ untranslated sequences. By immunohistochemical analysis, the protein encoded by the downstream open reading frame (GDF-1) was detected exclusively in the brain, spinal cord, and peripheral nerves in day-14.5 mouse embryos. The upstream open reading frame encodes a protein of unknown function containing multiple putative membrane-spanning domains. These findings raise the possibility that this mRNA may give rise to two different proteins.**

The transforming growth factor β (TGF-β) superfamily encompasses a group of proteins likely to play critical roles in regulating differentiation events during embryonic development. Among the known members, for example, are Mullerian inhibiting substance (MIS), which appears to be required for normal male sex development (1–3); the *Drosophila* decapentaplegic gene product, which is required for normal dorsal–ventral axis formation and for morphogenesis of the imaginal disks (4); the activins, which can induce the formation of mesoderm and anterior structures in *Xenopus* embryos (5–9); and the bone morphogenetic proteins, which are capable of inducing *de novo* cartilage and bone formation (10–12). To identify other extracellular factors that may be essential for mammalian embryonic development, I have been screening mouse embryo cDNA libraries with oligonucleotides corresponding to conserved regions among the known TGF-β family members. I previously reported the isolation of an additional TGF-β family member [designated growth/differentiation factor 1 (GDF-1)] from a day-8.5 mouse embryo cDNA library (13). Here I report the characterization of GDF-1 expression during embryogenesis and in adult tissues.* The data presented here suggest that GDF-1 is expressed specifically in the nervous system, where the GDF-1 mRNA contains an unusual bicistronic structure.

MATERIALS AND METHODS

RNA Isolation, cDNA Libraries, and Hybridizations. All embryonic materials were obtained from random matings of CD-1 mice (Charles River Breeding Laboratories). RNA isolation and Northern (RNA) analysis were done as described (13). Oligo(dT)-primed and random hexanucleotide-primed murine brain cDNA libraries were prepared in the lambda ZAP II vector (Stratagene), according to the instructions provided by Stratagene and Amersham, respectively. Human adult cerebellum and human fetal brain (17 to 18 week abortus) cDNA libraries were purchased from Stratagene. DNA filters were hybridized as described (13), except that filters hybridized with probes from the homologous species were subjected to a final wash in 0.2× standard saline citrate (SSC) (1× SSC is 0.15 M sodium chloride/0.015 M sodium citrate, pH 7) at 68°C. DNA sequencing of both strands was done with the dideoxynucleotide chain-termination method (14), using the exonuclease III/S1 nuclease strategy (15).

Immunohistochemistry. For antibody production, the GDF-1-coding sequence spanning amino acids 13–217 was cloned into the T7-based expression vector pET3a (provided by F. W. Studier, Brookhaven National Laboratory), and the resulting plasmid was transformed into the BL21 (DE3) bacterial strain. Total cell extracts from isopropyl β-D-thiogalactoside-induced cells were electrophoresed on SDS/polyacrylamide gels, and GDF-1 protein was excised, mixed with Freund's adjuvant, and used to immunize rabbits. All immunization procedures were done by Spring Valley Lab, Sykesville, MD. Embryos were fixed in 4% paraformaldehyde and embedded in paraffin. Eight-micron sections were incubated with either immune or preimmune serum (1:500 dilution), and sites of antibody binding were detected by using the Vectastain method, according to the instructions provided by Vector Laboratories.

RESULTS AND DISCUSSION

Expression of GDF-1 mRNA. To begin to characterize the pattern of GDF-1 expression during mouse embryogenesis, Northern analysis was carried out using poly(A)-selected RNA prepared from whole embryos isolated at various gestational ages. The GDF-1 probe detected two mRNA species displaying distinct temporal patterns of expression (Fig. 1*A*). One mRNA species, 1.4 kilobases (kb) in length, was present at days 8.5 and 9.5 but disappeared by day 10.5. The second mRNA species, 3.0 kb in length, appeared at day 9.5 and persisted throughout embryogenesis. The previously reported GDF-1 cDNA sequence is likely to correspond to the 1.4-kb mRNA species because only the 1.4-kb species could be detected in the day-8.5 sample, from which the original cDNA library had been prepared.

To determine the spatial pattern of GDF-1 expression in adult animals, Northern analysis was carried out using poly(A)-selected RNA prepared from a variety of mouse tissues. The GDF-1 probe detected a single 3.0-kb mRNA species expressed almost exclusively in the brain (Fig. 1*B*). On longer exposures of the same blot significantly lower, though detectable, levels were seen in the ovary, oviduct, and adrenal

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: GDF-1, growth/differentiation factor 1; TGF-β, transforming growth factor β; ORF, open reading frame; UOG, upstream of GDF-1; hUOG-1 and mUOG-1, human and murine UOG, respectively; nt, nucleotide(s).

*The sequences reported in this paper have been deposited in the GenBank data base (accession nos. M62301 for the murine GDF-1 sequence and M62302 for the human GDF-1 sequence).

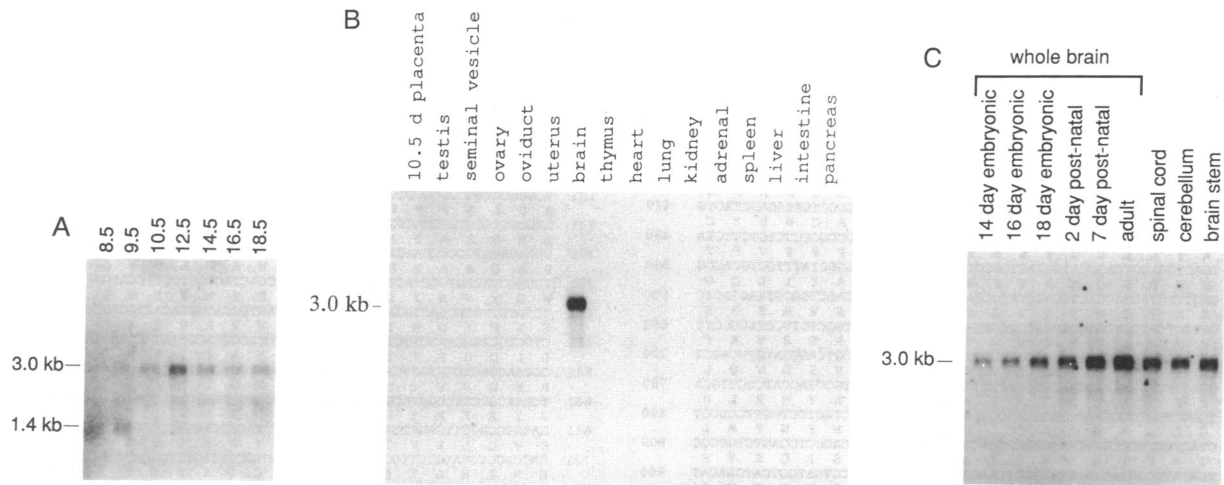

FIG. 1. Expression of GDF-1 mRNA in mouse embryos and tissues. Two micrograms of twice poly(A)-selected mRNA (*A* and *C*) or 5 μg of once poly(A)-selected mRNA (*B*) was electrophoresed, blotted, and probed with the 1.4-kb GDF-1 cDNA. d, Day. In *A*, RNA was prepared from whole CD-1 mouse embryos isolated at the indicated days of gestation. Assignment of the band sizes was based on mobilities of RNA standards.

gland (data not shown). No band corresponding to the 1.4-kb species was detected in any of these tissues. The 3.0-kb species was also detected in adult spinal cord, cerebellum, and brain stem as well as in fetal brains isolated at various developmental stages (Fig. 1*C*). Hence, the expression of the 3.0-kb mRNA appears widespread in the central nervous system.

Nucleotide Sequence of the Larger GDF-1 Transcript. To determine whether the 3.0-kb band represents an alternate transcript derived from the GDF-1-encoding gene or a transcript derived from a different gene homologous to GDF-1, several cDNA libraries [oligo(dT) and random primed] were constructed from poly(A)-selected adult mouse brain mRNA and screened with the 1.4-kb GDF-1 probe. Fourteen hybridizing cDNA clones were isolated, and these could be aligned within a region spanning 2.7 kb based on partial nucleotide sequence analysis of the ends of the clones (Fig. 2*A*). The complete 2.7-kb cDNA sequence, obtained by determining the entire nucleotide sequence of clones mBr-1, mBr-2, and mBr-7, is shown in Fig. 3*A*. Sequence comparison showed that the previously reported 1.4-kb sequence was essentially fully contained within the 2.7-kb sequence [from nucleotides (nt) 1311–2687]. Within this region, the sequence derived

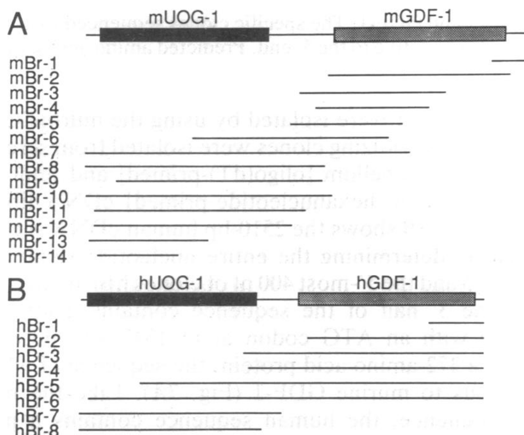

FIG. 2. Schematic representation of clones isolated from murine (*A*) or human (*B*) brain cDNA libraries. Locations of the UOG-1 and GDF-1 open reading frames (ORFs) are shown by the solid and stippled boxes, respectively. All clones were oriented and aligned by determining the sequences at both ends.

from clones mBr-2 and mBr-7 contained three nucleotide differences compared with the sequence derived from a day-8.5 embryo cDNA clone. One of these differences resulted in a change in the predicted amino acid sequence (cysteine to serine at codon 145). Limited nucleotide sequence analysis of multiple cDNA clones revealed that both classes of sequences were present both in day-8.5 embryos, where only the 1.4-kb mRNA species could be detected, and in the adult brain, where only the 3.0-kb mRNA species could be detected. Whether these sequence differences represent different alleles or two different genes is not known. However, Southern analysis with a probe derived from the GDF-1-coding sequence detected a single band at high stringency in three different digests of murine genomic DNA (Fig. 4).

Upstream of the GDF-1-coding region, the 2.7-kb sequence contained an additional 1310 base pairs (bp) not present in the 1.4-kb sequence, leaving a total of 1527 bp upstream of the initiating codon for GDF-1. Northern analysis of the identical RNA samples shown in Fig. 1 *A*, *B*, and *C* using this upstream sequence as a probe detected only a single 3.0-kb mRNA with the same distribution as that detected with the GDF-1-coding sequence probe (data not shown). Unexpectedly, within this upstream region was a second long ORF beginning with a putative initiating methionine codon at nt 74, extending for 350 codons, and terminating 404 nt upstream of the GDF-1-initiating ATG. For simplicity, this second ORF will be hereafter referred to as murine UOG-1 (upstream of GDF-1) (mUOG-1). Because of the multiple stop codons in the region between mUOG-1 and murine GDF-1, at least four frameshifts would be required to translate the two ORFs as a single protein. A search of the National Biomedical Research Foundation and GenBank sequence data bases with the predicted mUOG-1 amino acid sequence and with the entire upstream nucleotide sequence, respectively, revealed no significant homologies with known sequences. However, hydropathicity analysis of the predicted mUOG-1 amino acid sequence revealed multiple clusters of hydrophobic residues, reminiscent of membrane-spanning domains (Fig. 5). Particularly striking is the most distal of these clusters, which is immediately followed by a highly charged C-terminal region. Like certain other proteins with multiple membrane-spanning domains (for example, see ref. 17), mUOG-1 does not contain an obvious N-terminal signal sequence.

Expression of GDF-1 Protein from the Larger Transcript. To determine whether GDF-1 is translated from the 3.0-kb

A

```
   1 GCGCGTGACGCGAGGGCGCGCGGCGACTCGGACCGGTGCAGGCAACAGCGGAGACAGCGG   60
  61 AGAATTGGATAGCATGGCTGCTGCCGCGGCGACCCCCAGGCTCGAGGCGCCAGAGCCCAT  120
                  M  A  A  A  A  A  T  P  R  L  E  A  P  E  P  M
 121 GCCGAGTTATGCGCAGATGTTGCAACGAAGCTGGGCCTCGGCGCTGGCGGCGGCTCAGGG  180
      P  S  Y  A  Q  M  L  Q  R  S  W  A  S  A  L  A  A  A  Q  G
 181 CTGCGGGGACTGCGGCTGGGGACTGGCGCGCCGCGGCCTGGCGGAGCACGCGCACCTGGC  240
      C  G  D  C  G  W  G  L  A  R  R  G  L  A  E  H  A  H  L  A
 241 TGCACCCGAGCTGCTGCTGGCCGTGCTCTGCGCTCTGGGGTGGACAGCGTTGCGCTGGGC  300
      A  P  E  L  L  L  A  V  L  C  A  L  G  W  T  A  L  R  W  A
 301 AGCCACCACACACATCTTTCGGCCCCTGGCCAAGCGGTGTCGTCTGCAGCCTAGAGATGC  360
      A  T  T  H  I  F  R  P  L  A  K  R  C  R  L  Q  P  R  D  A
 361 TGCCAGGTTACCTGAGAGCGCCTGGAAGCTTCTGTTCTACTTGGCCTGTTGGAGCTACTG  420
      A  R  L  P  E  S  A  W  K  L  L  F  Y  L  A  C  W  S  Y  C
 421 CGCTTACCTGCTCCTGGGCACCAGTTATCCTTTCTTCCATGACCCGCCCTCTGTCTTCTA  480
      A  Y  L  L  L  G  T  S  Y  P  F  F  H  D  P  P  S  V  F  Y
 481 TGACTGGAGGTCAGGCATGGCAGTGCCCTGGGACATCGCGGTGGCCTATTTGCTGCAGGG  540
      D  W  R  S  G  M  A  V  P  W  D  I  A  V  A  Y  L  L  Q  G
 541 GAGTTTCTACTGCCACTCCATCTATGCCACCGTGTACATGGACAGCTGGCGTAAGGACTC  600
      S  F  Y  C  H  S  I  Y  A  T  V  Y  M  D  S  W  R  K  D  S
 601 GGTGGTCATGCTGGTGCATCACGTGGTCACCCTGCTCCTCATTGCCTCTTCCTACGCCTT  660
      V  V  M  L  V  H  H  V  V  T  L  L  L  I  A  S  S  Y  A  F
 661 CCGGTACCACAACGTAGGCCTCCTCGTGTTCTTCCTGCATGACGTCAGCGATGTGCAGCT  720
      R  Y  H  N  V  G  L  L  V  F  F  L  H  D  V  S  D  V  Q  L
 721 GGAGTTCACAAAACTCAACATCTACTTTAAGGCTAGGGGTGGTGCCTACCATCGCTTGCA  780
      E  F  T  K  L  N  I  Y  F  K  A  R  G  G  A  Y  H  R  L  H
 781 TGGGCTGGTGGCCAACCTGGGCTGCCTCAGCTTCTGTTTCTGCTGGTTCTGGTTCCGCCT  840
      G  L  V  A  N  L  G  C  L  S  F  C  F  C  W  F  W  F  R  L
 841 CTACTGGTTCCCGCTCAAGGTTCTGTACGCCACTTGCCACTGCAGCCTGCAGTCTGTGCC  900
      Y  W  F  P  L  K  V  L  Y  A  T  C  H  C  S  L  Q  S  V  P
 901 TGACATTCCGTACTACTTCTTCTTCAACATTCTGCTGTTGCTCCTGATGGTCATGAACAT  960
      D  I  P  Y  Y  F  F  F  N  I  L  L  L  L  L  M  V  M  N  I
 961 CTATTGGTTCCTGTACATTGTGGCTTTCGCAGCCAAGGTGCTGACTGGTCAGATGCGTGA 1020
      Y  W  F  L  Y  I  V  A  F  A  A  K  V  L  T  G  Q  M  R  E
1021 ACTGGAAGACTTGAGGGAGTACGACACTCTGGAAGCTCAGACAGCCAAGCCCTGCAAAGC 1080
      L  E  D  L  R  E  Y  D  T  L  E  A  Q  T  A  K  P  C  K  A
1081 CGAGAAGCCACTGAGGAATGGCCTGGTGAAGGACAAGCTCTTCTGAGTCTCTTGTCCTCA 1140
      E  K  P  L  R  N  G  L  V  K  D  K  L  F
1141 ACTTCAGCCATCCAGGACTCTATCCCATCCTACCTGGGATACTGACTCCGCCCCTGGAGA 1200
1201 CTCGACCCAGTCCCTGGAGGTCTGCTCCCACCCCTGGAGGCCCGGTCCCGCCTTTGGCGG 1260
1261 CATGGCCTCGCCCCTAGGACAAATAGCCCCGCCCTAAGATTCAGGATGCTACCCTTCTCCA 1320
1321 GGGACTCTGGCTGCCAGCAGCTCCGCCTTTCAGATCAATTCTCGACCACCCACCTTGGGA 1380
1381 CTGCCGCCCAGTCCTGCCCTCTGGATCAGTGGGGTCCAGACACGCCCCCTCCAGGACCTC 1440
1441 AAAGCACCCCCGACCTAAGGTCACCAGCCCACTGGCCCCAGACGCAGTGGGCTCCGCTGA 1500
1501 CTCTCTTGGACACCTCCTGGGAGGAAAATGCTCCCTGTCTGCCATCGTTTTTGCACCAC 1560
                                  M  L  P  V  C  H  R  F  C  D  H
1561 CTCCTCCTCCTGCTCTTGCTGCCCTCGACGACCCTGGCCCCGCGCCAGCATCCATGGGC 1620
      L  L  L  L  L  L  L  P  S  T  T  L  A  P  A  P  A  S  M  G
1621 CCCGCTGCCGCCCTGCTCCAGGTTCTTGGGCTTCCCGAAGCGCCCCGGAGCGTCCCCACA 1680
      P  A  A  A  L  L  Q  V  L  G  L  P  E  A  P  R  S  V  P  T
1681 CACCGACCTGTGCCTCCTGTCATGTGGCGCCTATTCCGTCGCCGCGACCCCCAGGAGGCC 1740
      H  R  P  V  P  P  V  M  W  R  L  F  R  R  R  D  P  Q  E  A
1741 AGAGTGGGACGCCCTCTGCGGCCATGCCACGTGGAGGAACTAGGGGTCGCCGGAAACATT 1800
      R  V  G  R  P  L  R  P  C  H  V  E  E  L  G  V  A  G  N  I
1801 GTGCGCCACATCCCCGACAGCGGTCTGTCCTCCAGGCCCGCACAACCCGCCAGGACCTCG 1860
      V  R  H  I  P  D  S  G  L  S  S  R  P  A  Q  P  A  R  T  S
1861 GGGCTGTGCCCCGAGTGGACAGTCGTCTTTGACCTGTCGAATGTGGAGCCCACAGAGCGC 1920
      G  L  C  P  E  W  T  V  V  F  D  L  S  N  V  E  P  T  E  R
1921 CCAACACGCGCGCGCTTAGAGTTGCGGCTGGAGGCTGAGAGTGAAGATACAGGGGGGTGG 1980
      P  T  R  A  R  L  E  L  R  L  E  A  E  S  E  D  T  G  G  W
1981 GAGCTAAGCGTGGCACTGTGGGCCGACGCAGAGCATCCAGGGCCTGAGCTGCTGCGCGTG 2040
      E  L  S  V  A  L  W  A  D  A  E  H  P  G  P  E  L  L  R  V
2041 CCGGCGCCACCAGGGGTGCTCCTGCGCGCAGACCTACTGGGGACTGCAGTAGCCGCCAAC 2100
      P  A  P  P  G  V  L  L  R  A  D  L  L  G  T  A  V  A  A  N
2101 GCATCAGTGCCCTGTACTGTGCGCCTGGCGCTGTCACTGCACCCTGGGGCCACTGCAGCC 2160
      A  S  V  P  C  T  V  R  L  A  L  S  L  H  P  G  A  T  A  A
2161 TGTGGGCGCCTGGCTGAGGCCTCCCTGCTGCTGGTGACGCTGGACCCACGCCTGTGTCCC 2220
      C  G  R  L  A  E  A  S  L  L  L  V  T  L  D  P  R  L  C  P
2221 TTGCCGCGATTGCGGCGCCACACGGAGCCCAGGGTAGAAGTTGGTCCAGTGGGCACTTGT 2280
      L  P  R  L  R  R  H  T  E  P  R  V  E  V  G  P  V  G  T  C
2281 CGTACCCGACGGTTGCATGTGAGCTTCCGTGAGGTGGGCTGGCACCGTTGGGTGATCGCG 2340
      R  T  R  R  L  H  V  S  F  R  E  V  G  W  H  R  W  V  I  A
2341 CCGCGTGGCTTCCTAGCCAACTTCTGCCAGGGCACGTGCGCACTACCCGAAACGCTGAGG 2400
      P  R  G  F  L  A  N  F  C  Q  G  T  C  A  L  P  E  T  L  R
2401 GGACCCGGCGGGCCGCCTGCACTCAACCACGCTGTGCTGCGCGCGCTCATGCACGCAGCT 2460
      G  P  G  G  P  P  A  L  N  H  A  V  L  R  A  L  M  H  A  A
2461 GCTCCCACCCCGGGTGCAGGCTCGCCCTGCTGCGTGCCAGAGCGTCTATCACCCATCTCC 2520
      A  P  T  P  G  A  G  S  P  C  C  V  P  E  R  L  S  P  I  S
2521 GTGCTCTTCTTCGACAATAGTGACAACGTGGTCCTGCGACACTACGAAGACATGGTGGTG 2580
      V  L  F  F  D  N  S  D  N  V  V  L  R  H  Y  E  D  M  V  V
2581 GATGAGTGTGGCTGCCGTTGACCACCCGGGACACCCTTTCAGGGACCGCCCCACGCAAAA 2640
      D  E  C  G  C  R
2641 GCAGGGACTGTTTGTTCATGTTTTATTGGTGACAAAAAGCTTAAAACAAATTTGACTAAA 2700
2701 AATTAAGTTCC 2711
```

B

```
   1 GGACACGGCGGGCGAGCGGGCGGTATGGCGGCGGCGGGGCCCGCGGCGGGGCCGACGGGG   60
                               M  A  A  A  G  P  A  A  G  P  T  G
  61 CCCGAGCCCATGCCGAGCTACGCGCAGCTAGTGCAGCGCGGCTGGGGCAGCGCGCTGGCG  120
      P  E  P  M  P  S  Y  A  Q  L  V  Q  R  G  W  G  S  A  L  A
 121 GCGGCGCGGGGCTGCACGGACTGCGGCTGGGGGCTGGCGCGTCGCGGCCTGGCTGAGCAC  180
      A  A  R  G  C  T  D  C  G  W  G  L  A  R  R  G  L  A  E  H
 181 GCGCACCTGGCGCCCGCCGAGCTGCTGCTGCTGGCGCTGGGCGCGCTGGGCTGGACCGCG  240
      A  H  L  A  P  P  E  L  L  L  L  A  L  G  A  L  G  W  T  A
 241 CTGCGCTCCGCGGCCACTGCGCGCCTCTTTCGGCCCCTGGCGAAGCGGTGCTGCCTCCAG  300
      L  R  S  A  A  T  A  R  L  F  R  P  L  A  K  R  C  C  L  Q
 301 CCCAGAGATGCCGCCAAGATGCCCGAGAGCGCTTGGAAGTTTCTCTTCTACCTGGGCAGC  360
      P  R  D  A  A  K  M  P  E  S  A  W  K  F  L  F  Y  L  G  S
 361 TGGAGCTACAGTGCCTACCTGCTGTTTGGCACCGACTACCCCTTCTTCCATGACCCACCA  420
      W  S  Y  S  A  Y  L  L  F  G  T  D  Y  P  F  F  H  D  P  P
 421 TCTGTCTTCTACGACTGGACGCCGGGCATGGCAGTGCCACGGGACATTGCAGCCGCCTAC  480
      S  V  F  Y  D  W  T  P  G  M  A  V  P  R  D  I  A  A  A  Y
 481 CTGCTCCAGGGAAGCTTCTATGGCCACTCCATCTACGCTACGCTATACATGGACACCTGG  540
      L  L  Q  G  S  F  Y  G  H  S  I  Y  A  T  L  Y  M  D  T  W
 541 CGCAAGGACTCGGTGGTCATGCTGCTCCACCACGTGGTCACTCTCATCCTCATCGTCTCC  600
      R  K  D  S  V  V  M  L  L  H  H  V  V  T  L  I  L  I  V  S
 601 TCCTACGCCTTCCGGTACCACAATGTGGGCATCCTTGTGCTCTTCCTGCACGATATCAGT  660
      S  Y  A  F  R  Y  H  N  V  G  I  L  V  L  F  L  H  D  I  S
 661 GACGTGCAGCTTGAGTTCACCAAGCTCAACATTTACTTCAAGTCCCGCGGCGGCTCCTAC  720
      D  V  Q  L  E  F  T  K  L  N  I  Y  F  K  S  R  G  G  S  Y
 721 CATCGGCTGCATGCCTTGGCAGCAGACTTGGGCTGCCTCAGCTTCGGCTTCAGCTGGTTC  780
      H  R  L  H  A  L  A  A  D  L  G  C  L  S  F  G  F  S  W  F
 781 TGGTTCCGCCTCTACTGGTTCCCGCTCAAGGTCCTGTATGCCACCAGTCACTGCAGTCTG  840
      W  F  R  L  Y  W  F  P  L  K  V  L  Y  A  T  S  H  C  S  L
 841 CGCACGGTGCCTGACATCCCCTTCTACTTCTTCTTCAATGCGCTCCTGCTGCTGCTCACC  900
      R  T  V  P  D  I  P  F  Y  F  F  F  N  A  L  L  L  L  L  T
 901 CTTATGAACCTCTACTGGTTCCTGTACATCGTGGCGTTTGCAGCCAAGGTGTTGACAGGC  960
      L  M  N  L  Y  W  F  L  Y  I  V  A  F  A  A  K  V  L  T  G
 961 CAGGTGCACGAGCTGAAGGACCTGCGGGAGTATGACACAGCCGAGGCCCAGAGCCTGAAG 1020
      Q  V  H  E  L  K  D  L  R  E  Y  D  T  A  E  A  Q  S  L  K
1021 CCCAGCAAAGCCGAGAAGCCACTGAGGAACGGCCTGGTGAAGGACAAGCGCTTCTGAACC 1080
      P  S  K  A  E  K  P  L  R  N  G  L  V  K  D  K  R  F
1081 CCTCGGCCCCGCCCCCGTGGACCCGGCCCCACCCCGAATACCCCGGCCACGCTCCCCGTC 1140
1141 CTTGGCCGCCCCTCCACCCCCTCCAACTCTGCTCCTCTAGGGCCGCCGCCACCTCCCCTG 1200
1201 GGACCCCGCCCCCTCATCCTGCCTCCATTTCCCGGCCACGCCCCCCAGGACCCCTGCCCC 1260
1261 TCCGGGGACACCGGCCCCGCCCTCAGCCCACTGGTCCCGGGCCGCCGCGGACCCTGCGCA 1320
1321 CTCTCTGGTCATCGCCTGGGAGGAAGATGCCACCGCCGCAGCAAGGTCCCTGCGGCCACC 1380
                                  M  P  P  P  Q  Q  G  P  C  G  H  H
1381 ACCTCCTCCTCCTCCTGGCCCTGCTGCTGCCCTCGCTGCCCCTGACCCGCGCCCCCGTGC 1440
      L  L  L  L  L  A  L  L  L  P  S  L  P  L  T  R  A  P  V  P
1441 CCCCAGGCCCAGCCGCCGCCCTGCTCCAGGCTCTAGGACTGCGCGATGAGCCCCAGGGTG 1500
      P  G  P  A  A  A  L  L  Q  A  L  G  L  R  D  E  P  Q  G  A
1501 CCCCCAGGCTCCGGCCGGTTCCCCCGGTCATGTGGCGCCTGTTTCGACGCCGGGACCCCC 1560
      P  R  L  R  P  V  P  P  V  M  W  R  L  F  R  R  R  D  P  Q
1561 AGGAGACCAGGTCTGGCTCGCGGCGGACGTCCCCAGGGGTCACCCTGCAACCGTGCCACG 1620
      E  T  R  S  G  S  R  R  T  S  P  G  V  T  L  Q  P  C  H  V
1621 TGGAGGAGCTGGGGGTCGCCGGAAACATCGTGCGCCACATCCCGGACCGCGGTGCGCCCA 1680
      E  E  L  G  V  A  G  N  I  V  R  H  I  P  D  R  G  A  P  T
1681 CCCGGGCCTCGGAGCCTGTCTCGGCCGCGGGGCATTGCCCTGAGTGGACAGTCGTCTTCG 1740
      R  A  S  E  P  V  S  A  A  G  H  C  P  E  W  T  V  V  F  D
1741 ACCTGTCGGCTGTGGAACCCGCTGAGCGCCCGAGCCGGGCCCGCCTGGAGCTGCGTTTCG 1800
      L  S  A  V  E  P  A  E  R  P  S  R  A  R  L  E  L  R  F  A
1801 CGGCGGCGGCGGCGGCAGCCCCGGAGGGCGGCTGGGAGCTGAGCGTGGCGCAAGCGGGCC 1860
      A  A  A  A  A  A  P  E  G  G  W  E  L  S  V  A  Q  A  G  Q
1861 AGGGCGCGGGCGCGGACCCCGGGCCGGTGCTGCTCCGCCAGTTGGTGCCCGCCCTGGGGC 1920
      G  A  G  A  D  P  G  P  V  L  L  R  Q  L  V  P  A  L  G  P
1921 CGCCAGTGCGCGCGGAGCTGCTGGGCGCCGCTTGGGCTCGCAACGCCTCATGGCCGCGCA 1980
      P  V  R  A  E  L  L  G  A  A  W  A  R  N  A  S  W  P  R  S
1981 GCCTCCGCCTGGCGCTGGCGCTACGCCCCCGGGCCCCTGCCGCCTGCGCGCGCCTGGCCG 2040
      L  R  L  A  L  A  L  R  P  R  A  P  A  A  C  A  R  L  A  E
2041 AGGCCTCGCTGCTGCTGGTGACCCTCGACCCGCGCCTGTGCCACCCCCTGGCCCGGCCGC 2100
      A  S  L  L  L  V  T  L  D  P  R  L  C  H  P  L  A  R  P  R
2101 GGCGCGACGCCGAACCCGTGTTGGGCGGCGGCCCCGGGGGCGCTTGTCGCGCGCGGCGGC 2160
      R  D  A  E  P  V  L  G  G  G  P  G  G  A  C  R  A  R  R  L
2161 TGTACGTGAGCTTCCGCGAGGTGGGCTGGCACCGCTGGGTCATCGCGCCGCGCGGCTTCC 2220
      Y  V  S  F  R  E  V  G  W  H  R  W  V  I  A  P  R  G  F  L
2221 TGGCCAACTACTGCCAGGGTCAGTGCGCGCTGCCCGTCGCGCTGTCGGGGTCCGGGGGGC 2280
      A  N  Y  C  Q  G  Q  C  A  L  P  V  A  L  S  G  S  G  G  P
2281 CGCCGGCGCTCAACCACGCTGTGCTGCGCGCGCTCATGCACGCGGCCGCCCCGGGAGCCG 2340
      P  A  L  N  H  A  V  L  R  A  L  M  H  A  A  A  P  G  A  A
2341 CCGACCTGCCCTGCTGCGTGCCCGCGCGCCTGTCGCCCATCTCCGTGCTCTTCTTTGACA 2400
      D  L  P  C  C  V  P  A  R  L  S  P  I  S  V  L  F  F  D  N
2401 ACAGCGACAACGTGGTGCTGCGGCAGTATGAGGACATGGTGGTGGACGAGTGCGGCTGCC 2460
      S  D  N  V  V  L  R  Q  Y  E  D  M  V  V  D  E  C  G  C  R
2461 GCTAACCCGGGGCGGGCAGGGACGCGGGCCCAACAATAAATGCCGCGTGG 2510
```

FIG. 3. Nucleotide sequences of murine (*A*) and human (*B*) cDNAs encoding UOG-1 and GDF-1. The specific clones sequenced to assemble the complete sequences shown are described in text. Numbers indicate nucleotide position relative to the 5′ end. Predicted amino acid sequences of UOG-1 and GDF-1 are shown below.

transcript, immunohistochemical analysis of day-14.5 mouse embryo sections was done by using antibodies prepared against bacterially produced GDF-1 protein. To avoid possible cross-reactivity with other members of the TGF-β superfamily, these antibodies were directed against the N-terminal pro- region of GDF-1. The antibodies detected GDF-1 protein exclusively in the developing brain, spinal cord, and peripheral nerves (Fig. 6). Within these tissues, the staining was restricted to the regions containing postmitotic neurons. Whether the staining reflects sites of GDF-1 synthesis or target sites for GDF-1 (or both) is not known. In either case, because Northern analysis of RNA prepared from whole day-14.5 embryos as well as from day-14.5 fetal brain detected only a single 3.0-kb transcript (Fig. 1 *A* and *C*), these data indicate that GDF-1 protein is translated from this larger transcript.

Sequence of Human GDF-1. To look for conserved regions in the GDF-1 mRNA and protein sequences, cDNAs encoding human GDF-1 were isolated by using the murine GDF-1 probe. Eight hybridizing clones were isolated from screening human adult cerebellum [oligo(dT)-primed] and fetal brain [oligo(dT)/random hexanucleotide-primed] cDNA libraries (Fig. 2*B*). Fig. 3*B* shows the 2510-bp human cDNA sequence obtained by determining the entire nucleotide sequence of clone hBr-5 and the 5′-most 400 nt of clones hBr-6, hBr-7, and hBr-8. The 3′ half of the sequence contains a long ORF beginning with an ATG codon at nt 1347 and potentially encoding a 372-amino acid protein, the sequence of which is homologous to murine GDF-1 (Fig. 7*A*). Like the murine GDF-1 sequence, the human sequence contains a pair of basic residues (Arg-Arg) at amino acids 252 and 253, which presumably represents a site for proteolytic processing. After the predicted cleavage site, the sequence contains all of the invariant and most of the highly conserved amino acids characteristic of all members of the TGF-β superfamily, including the seven cysteine residues. The murine GDF-1

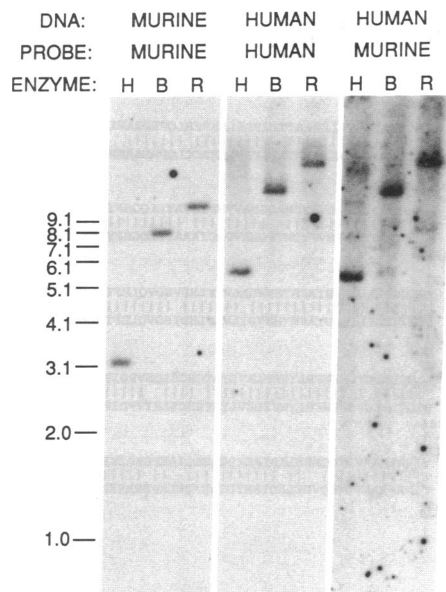

FIG. 4. Genomic Southern analysis of GDF-1. Ten micrograms of genomic DNA isolated from BNL cells (murine) or BeWo cells (human) was digested with *Hin*dIII (H), *Bam*HI (B), or *Eco*RI (R), electrophoresed, blotted, and probed with the entire murine or human GDF-1-coding sequences. Numbers at left indicate sizes of standards in kb.

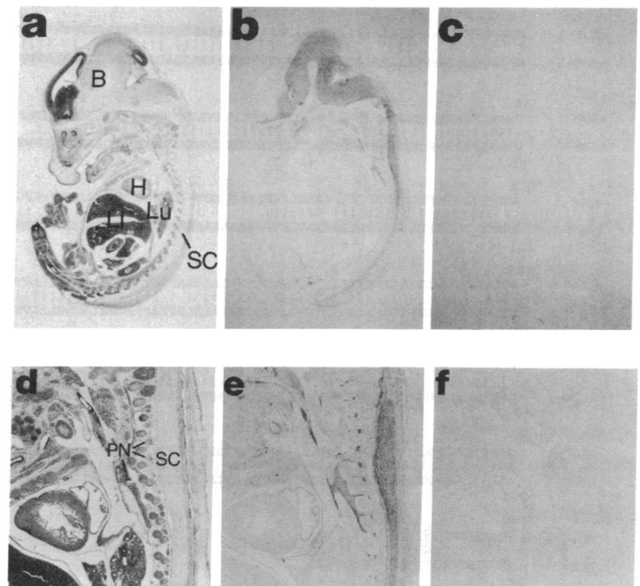

FIG. 6. Immunolocalization of GDF-1 protein in mouse embryos. Sections of day-14.5 mouse embryos were stained with hematoxylin and eosin (*a* and *d*) or incubated with anti-GDF-1 (*b* and *e*) or preimmune (*c* and *f*) serum and analyzed using the Vectastain method. (*d*, *e*, and *f*) Higher magnifications of specific regions of the embryos shown in *a*, *b*, and *c*, respectively. B, brain; SC, spinal cord; H, heart; Li, liver; and Lu, lung; and PN, peripheral nerves.

sequence and the human sequence are 87% identical in the region beginning with the first conserved cysteine and extending to the C terminus and 69% identical throughout the entire length of the protein. Because other members of the TGF-β superfamily show a much higher degree of sequence conservation across species (for example, see ref. 18), genomic Southern analysis was done to determine whether the murine and human sequences represent the same gene. As shown in Fig. 4, both murine and human probes derived from the GDF-1 ORF hybridized to the same pattern of bands in human DNA, suggesting that the human gene is, indeed, the homolog of murine GDF-1.

Conservation of a Bicistronic Structure. Like the murine sequence, the human sequence also contains a second long ORF, potentially encoding 350 amino acids in the region upstream of the GDF-1-coding sequence. An alignment of this upstream ORF (hUOG-1) with that present in the murine sequence showed that the upstream ORF is even more highly conserved than that for GDF-1 (Fig. 7*B*), with the overall amino acid sequence identity between mUOG-1 and hUOG-1 being 81% [although the ORFs for both mUOG-1 and hUOG-1 extend upstream of the putative

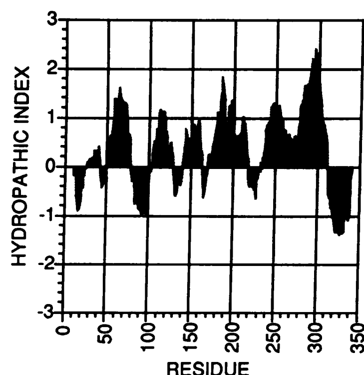

FIG. 5. Hydropathicity profile of mUOG-1. Average hydrophobicity values were calculated using the method of Kyte and Doolittle (16) with a window of 21. Positive numbers indicate increasing hydrophobicity.

initiating methionine to the very 5′ ends of the sequences, two lines of reasoning suggest that these may be the true initiation codons: (*i*) multiple cDNAs primed by random hexanucleotides at various distances from the 3′ end terminated very close to the 5′ ends of both the murine and human sequences (Fig. 2); (*ii*) as discussed below, the murine and human sequences show much less conservation upstream of the putative initiation codon for UOG-1 than in the coding sequence itself].

In contrast to the high degree of conservation seen between mUOG-1 and hUOG-1 and between murine GDF-1 and human GDF-1, the intervening spacer region and the putative 5′ and 3′ untranslated regions show much less similarity between murine and human sequences. This selective conservation of the two ORFs is most clearly evident in a DIAGON plot comparing the murine and human nucleotide sequences (Fig. 8). The two sequences begin to diverge in the intervening spacer region precisely after the stop codons for UOG-1 and in the 3′-untranslated region 2 nt after the stop codons for GDF-1. Moreover, the intervening spacer region in the murine sequence is 401 nt in length, whereas the corresponding region in the human sequence is only 269 nt in length. The conservation of the amino acid sequence of UOG-1 is also evident in the nonrandom pattern of nucleotide differences between the murine and human sequences spanning the UOG-1 ORFs. Of the 209-nt differences in this region, 57 represent first-position differences, 29 represent second-position differences, and 123 represent third-position differences; of the 123 third-position differences, 89 do not result in differences in the predicted amino acid sequence.

The data presented here suggest that two proteins, UOG-1 and GDF-1, are encoded by a single mRNA molecule. The selective conservation of the two ORFs in the murine and human sequences suggests that both proteins are synthesized *in vivo*. Although the immunohistochemical data indicate that the downstream ORF (GDF-1) is translated from this mRNA, further experiments will be required to show that the upstream ORF (UOG-1) is translated. To my knowledge, this type of bicistronic organization is unprecedented among mammalian cellular mRNAs. However, polycistronic

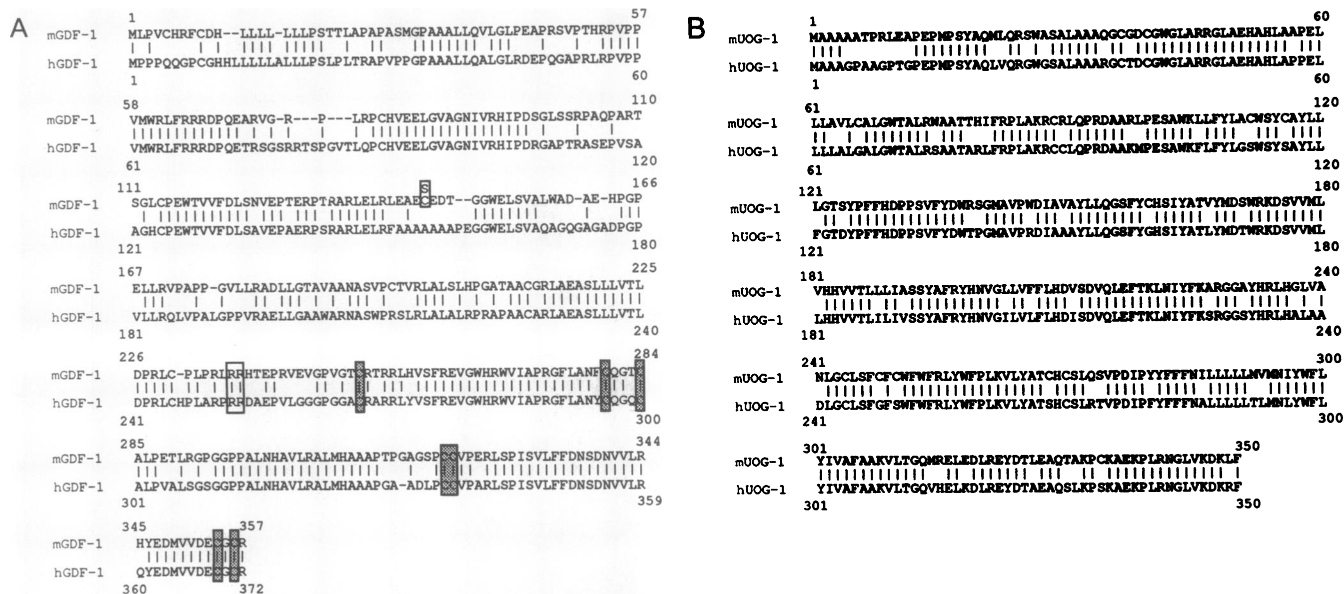

FIG. 7. Alignment of murine and human sequences. Alignment of murine GDF-1 with human GDF-1 (*A*) or mUOG-1 with hUOG-1 (*B*) were done with the SEQHP local homology program. Numbers indicate amino acid number relative to the N terminus of each protein. Dashes denote gaps introduced to maximize alignment. The seven invariant cysteines in the GDF-1 sequences are shaded. The predicted dibasic cleavage sites are boxed. The box at position 145 in the mGDF-1 sequence shows the alternative amino acids at this position for the two classes of cDNA clones isolated (see text).

mRNAs in prokaryotes often encode proteins carrying out related biological functions. Therefore, it seems reasonable to hypothesize that UOG-1 and GDF-1 may also functionally interact. From the known properties of the other members of the TGF-β superfamily, GDF-1 is likely to be an extracellular factor mediating cell-differentiation events. The hydropathicity analysis suggests that UOG-1 may be a transmembrane protein. It is tempting to speculate that UOG-1 may be a receptor—perhaps a receptor for GDF-1 itself.

GDF-1 is unusual among TGF-β family members in its specificity for the nervous system. This expression pattern of GDF-1 taken together with the demonstration that another TGF-β family member—namely, activin A—can promote neuronal survival (19) suggests that this family of extracellular signaling factors may play a more general role in the maintenance and function of the nervous system than has been previously appreciated. Finally, the sequence homology between GDF-1 and the other inducing molecules of the TGF-β superfamily raises the possibility that GDF-1 may play an inductive role in neural development.

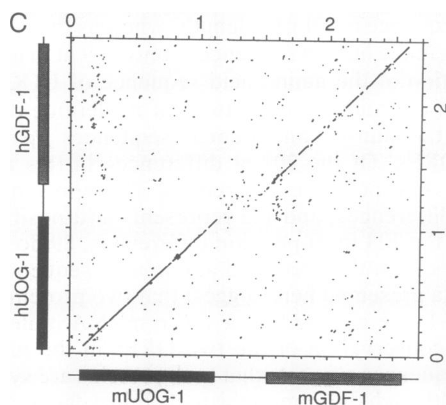

FIG. 8. DIAGON plot of murine (m) and human (h) nucleotide sequences was done with a window of 20 and a stringency of 14. Locations of UOG-1 and GDF-1 ORFs are shown by the solid and stippled boxes,respectively. Numbers indicate nucleotide positions in thousands.

I thank Steve McKnight, Jeremy Nathans, Daniel Nathans, Donald Brown, Anthony Lanahan, Tom Moench, Andy McMahon, and Joshua Farber for helpful discussions. This work was supported by the Carnegie Institution of Washington, by a Basil O'Conner Starter Scholar Research Award from the March of Dimes Birth Defects Foundation, and by Junior Faculty Research Award JFRA-254 and Research Grant CD-457 from the American Cancer Society.

1. Jost, A. (1947) *Arch. Anat. Microsc. Morphol. Exp.* **36**, 271–315.
2. Cate, R. L., Mattaliano, R. J., Hession, C., Tizard, R., Farber, N. M., Cheung, A., Ninfa, E. G., Frey, A. Z., Gash, D. J., Chow, E. P., Fisher, R. A., Bertonis, J. M., Torres, G., Wallner, B. P., Ramachandran, K. L., Ragin, R. C., Manganaro, T. F., MacLaughlin, D. T. & Donahoe, P. K. (1986) *Cell* **45**, 685–698.
3. Behringer, R. R., Cate, R. L., Froelick, G. J., Palmiter, R. D. & Brinster, R. L. (1990) *Nature (London)* **345**, 167–170.
4. Padgett, R. W., St. Johnston, R. D. & Gelbart, W. M. (1987) *Nature (London)* **325**, 81–84.
5. Mason, A. J., Hayflick, J. S., Ling, N., Esch, F., Ueno, N., Ying, S.-Y., Guillemin, R., Niall, H. & Seeburg, P. H. (1985) *Nature (London)* **318**, 659–663.
6. Forage, R. G., Ring, J. M., Brown, R. W., McInerney, B. V., Cobon, G. S., Gregson, R. P., Robertson, D. M., Morgan, F. J., Hearn, M. T. W., Findlay, J. K., Wettenhall, R. E. H., Burger, H. G. & de Kretser, D. M. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 3091–3095.
7. Smith, J. C., Price, B. M. J., van Nimmen, K. & Huylebroeck, D. (1990) *Nature (London)* **345**, 729–731.
8. van den Eijnden-van Raaij, A. J. M., van Zoelent, E. J. J., van Nimmen, K., Koster, C. H., Snoek, G. T., Durston, A. J. & Huylebroeck, D. (1990) *Nature (London)* **345**, 732–734.
9. Thomsen, G., Woolf, T., Whitman, M., Sokol, S., Vaughan, J., Vale, W. & Melton, D. A. (1990) *Cell* **63**, 485–493.
10. Wozney, J. M., Rosen, V., Celeste, A. J., Mitsock, L. M., Whitters, M. J., Kriz, R. W., Hewick, R. M. & Wang, E. A. (1988) *Science* **242**, 1528–1534.
11. Ozkaynak, E., Rueger, D. C., Drier, E. A., Corbett, C., Ridge, R. J., Sampath, T. K. & Oppermann, H. (1990) *EMBO J.* **7**, 2085–2093.
12. Sampath, T. K., Coughlin, J. E., Whetstone, R. M., Banach, D., Corbett, C., Ridge, R. J., Ozkaynak, E., Oppermann, H. & Rueger, D. C. (1990) *J. Biol. Chem.* **265**, 13198–13205.
13. Lee, S.-J. (1990) *Mol. Endocrinol.* **4**, 1034–1040.
14. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
15. Henikoff, S. (1984) *Gene* **28**, 351–359.
16. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157**, 105–132.
17. Nathans, J., Thomas, D. & Hogness, D. S. (1986) *Science* **232**, 193–202.
18. Derynck, R., Jarrett, J. A., Chen, E. Y. & Goeddel, D. V. (1986) *J. Biol. Chem.* **261**, 4377–4379.
19. Schubert, D., Kimura, H., LaCorbiere, M., Vaughan, J., Karr, D. & Fischer, W. H. (1990) *Nature (London)* **344**, 868–870.