Nucleotide sequence of a cDNA encoding murine CSF-1 (Macrophage-CSF)

John F.DeLamarter, Cathy Hession[1], Dominique Semon, Nicholas M.Gough[2], Rene Rothenbuhler and Jean-Jacques Mermod

BIOGEN, SA, Case Postale, 1211 Geneva 24, Switzerland, [1]BIOGEN Research Corporation, 14 Cambridge Center, Cambridge, MA 02142, USA and [2]Ludwig Institute for Cancer Research, Royal Melbourne Hospital, Melbourne, Victoria 3050, Australia

Submitted January 29, 1987

A human Colony Stimulating Factor-1 (CSF-1 or Macrophage-CSF) encoding cDNA was previously isolated from the pancreatic tumor cell line MIA-PACA-2 (1). This cDNA encodes a protein of 26 Kd molecular weight. To isolate a murine cDNA homologous to human M-CSF, a probe from the coding region of the human cDNA was prepared by nick-translation. A recombinant lambda phage library made from the mRNA of the murine fibroblast cell line, L929, was screened with the human M-CSF cDNA probe. The cDNA isolated from the murine library has an open reading frame of 1656 nucleotides which encodes a protein of 552 amino acids. The first 32 amino acids encode a putative signal peptide which when cleaved yields a mature protein beginning with the amino acid sequence: Lys-Glu-Val-Ser. This deduced sequence is identical to the amino terminal residues reported for M-CSF as purified from L-cell supernatants (2). The mature protein deduced from the cDNA sequence has a molecular weight of 57,261 d. All ten of the cysteine residues are found in approximately the amino-terminal half of the molecule. There are three potential N-linked glycosylation sites (Asn-X-Thr/Ser) located at residues 154, 172, and 378. From the first amino acid encoded by the cDNA to residue 180 and from residue 476 to the carboxy-terminal residue, approximately 80% homology is found between the murine and human cDNA's at both the nucleotide and amino acid levels. For comparison the deduced human M-CSF protein sequence is aligned below that of the murine sequence. The open reading frame in the murine cDNA encodes an additional 295 amino acids which are not found in the deduced amino-acid sequence of human M-CSF (1). However, higher molecular weight forms of human M-CSF mRNA have also been identified by Northern analysis of MIA-PACA-2 mRNA using the 26 Kd encoding cDNA as a probe (3). Furthermore, the murine M-CSF encoding cDNA we isolated was incorporated into an animal cell expression vector under SV40 early promoter transcriptional control. When the vector was transfected into the simian cell line, COS-7, murine M-CSF biological activity, as measured on mouse bone marrow cells, was observed in the transfected-culture supernatants.

```
  GAATTCGAGCTCGAGCGCGGCCGGAAAGTGAAAGTTTGCCTCGGTGCTCTCGGTGTCGCTGCGGCTCTCTGCATCCCAGGACAGCGGCGTGGCCCTCGACCGGGGCGCGGGC

TCTTCAGCCACTAGCGAGCAAGGGAGCGAGCGAACCAGGGCGGCCAACACGCCGTGCCGGGACCCAGCTGCCCGTATGACCGCGCGGGGCCGCGCGGGGCGCTGCCCTTCTTCG
                                                                    murine   MetThrAlaArgGlyArgAlaGlyArgCysProSerSer
                                                                    human    MetThrAlaProGlyAlaAlaGlyArgCysProProThr
                                                                             1                                  10

ACATGGCTGGGCTCCCGGCTGCTGCTGGTCTGTCTCCTCATGAGCAGGAGTATTGCCAAGGAGGTGTCAGAACACTGTAGCCACATGATTGGGAATGGACACCTGAAGGTCCTG
ThrTrpLeuGlySerArgLeuLeuLeuValCysLeuLeuMetSerArgSerIleAlaLysGluValSerGluHisCysSerHisMetIleGlyAsnGlyHisLeuLysValLeu
ThrTrpLeuGlySerLeuLeuLeuLeuValCysLeuLeuAlaSerArgSerIleThrGluGluValSerGluTyrCysSerHisMetIleGlySerGlyHisLeuGlnSerLeu
                  20                  30                  40                  50

CAGCAGTTGATCGACAGTCAAATGGAGACTTCATGCCAGATTGCCTTTGAATTTGTAGACCAGGAACAGCTGGATGATCCTGTTTGCTACCTAAAGAAGGCCTTTTTTCTGGTA
GlnGlnLeuIleAspSerGlnMetGluThrSerCysGlnIleAlaPheGluPheValAspGlnGluGlnLeuAspAspProValCysTyrLeuLysLysAlaPheLeuVal
GlnArgLeuIleAspSerGlnMetGluThrSerCysGlnIleThrPheGluPheValAspGlnGluGlnLeuLysAspProValCysTyrLeuLysLysAlaPheLeuLeuVal
                        60                  70                  80

CAAGACATAATAGATGAGACCATGCGCTTTAAAGACAACACCCCCAATGCTAACGCCACCGAGAGGCTCCAGGAACTCTCCAATAACCTGAACAGCTGCTTCACCAAGGACTAT
GlnAspIleIleAspGluThrMetArgPheLysAspAsnThrProAsnAlaAsnAlaThrGluArgLeuGlnGluLeuSerAsnAsnLeuAsnSerCysPheThrLysAspTyr
GlnTyrIleMetGluAspThrMetArgPheArgAspAsnThrProAsnAlaIleAlaIleValGlnLeuGlnGluLeuSerLeuArgLeuLysSerCysPheThrLysAspTyr
90                  100                 110                 120

GAGGAGCAGAACAAGGCCTGTGTCCGAACTTTCCATGAGACTCCTCTCCAGCTGCTGGAGAAGATCAAGAACTTCTTTAATGAAACAAAGAATCTCCTTGAAAAGGACTGGAAC
GluGluGlnAsnLysAlaCysValArgThrPheHisGluThrProLeuGlnLeuLeuGluLysIleLysAsnPheAsnGluThrLysAsnLeuLeuGluLysAspTrpAsn
GluGluHisAspLysAlaCysValArgThrPheTyrGluThrProLeuGlnLeuLeuGluLysValLysAsnValPheAsnGluThrLysAsnLeuLeuAspLysAspTrpAsn
      130                 140                 150                 160
```

```
ATTTTTACCAAGAACTGCAACAACAGCTTTGCTAAGTGCTCTAGCCGAGATGTGGTGACCAAGCCTGATTGCAACTGCCTGTACCCTAAAGCCACCCCTAGCAGTGACCCGGCC
IlePheThrLysAsnCysAsnAsnSerPheAlaLysCysSerSerArgAspValValThrLysProAspCysAsnCysLeuTyrProLysAlaThrProSerSerAspProAla
IlePheSerLysAsnCysAsnAsnSerPheAlaGluCysSerSer
          170                 180                 190                 200

TCTGCCTCCCCTCACCAGCCCCCGCCCCCTCCATGGCCCCTCTGGCTGGCTTGGCTTGGGATGATTCTCAGAGGACAGAGGGCAGCTCCCTCTTGCCCAGTGAGCTTCCCCTT
SerAlaSerProHisGlnProProAlaProSerMetAlaProLeuAlaGlyLeuAlaTrpAspAspSerGlnArgThrGluGlySerSerLeuLeuProSerGluLeuProLeu
                210                 220                 230                 240

CGCATAGAGGACGCAGGCAGTGCCAAGCAGCGACCACCCAGGAGTACCTGCCAGACCCTCGAGTCAACAGAGCAACCAAACCATGGGGACAGACTCACTGAGGACTCACAACCT
ArgIleGluAspAlaGlySerAlaLysGlnArgProProArgSerThrCysGlnThrLeuGluSerThrGluGlnProAsnHisGlyAspArgLeuThrGluAspSerGlnPro
                250                 260                 270

CATCCTTCTGCGGGGGGGCCCGTCCCTGGGGTGGAAGACATTCTTGAATCTTCACTGGGCACTAACTGGGTCCTAGAAGAAGCTTCTGGAGAGGCTAGTGAGGGATTTTTGACC
HisProSerAlaGlyGlyProValProGlyValGluAspIleLeuGluSerSerLeuGlyThrAsnTrpValLeuGluGluAlaSerGlyGluAlaSerGluGlyPheLeuThr
280                 290                 300                 310

CAGGAAGCAAAGTTTTCCCCCTCCACGCCTGTAGGGGGCAGCATCCAGGCAGAGACTGACAGACCCAGGGCCCTCTCAGCATCTCCATTCCCTAAATCAACAGAGGACCAAAAG
GlnGluAlaLysPheSerProSerThrProValGlyGlySerIleGlnAlaGluThrAspArgProArgAlaLeuSerAlaSerProPheProLysSerThrGluAspGlnLys
      320                 330                 340                 350

CCAGTGGATATAACAGACAGGCCGTTGACAGAGGTGAACCCTATGAGACCCATTGGCCAGACACAGAATAATACTCCTGAGAAGACTGATGGTACATCCACGCTGCGTGAAGAC
ProValAspIleThrAspArgProLeuThrGluValAsnProMetArgProIleGlyGlnThrGlnAsnAsnThrProGluLysThrAspGlyThrSerThrLeuArgGluAsp
            360                 370                 380                 390

CACCAGGAGCCAGGCTCTCCCCATATTGCGACACCGAATCCCCAACGAGTCAGCAACTCAGCCACCCCCGTTGCTCAGTTACTGCTTCCCAAAAGCCACTCTTGGGGCATTGTG
HisGlnGluProGlySerProHisIleAlaThrProAsnProGlnArgValSerAsnSerAlaThrProValAlaGlnLeuLeuLeuProLysSerHisSerTrpGlyIleVal
                400                 410                 420                 430

CTGCCCCTTGGGGAGCTTGAGGGCAAGAGAAGTACCAGGGATCGAAGGAGCCCCGCAGAGCTGGAAGGAGGATCAGCAAGTGAGGGGGCAGCCAGGCCTGTGGCCCGTTTTAAT
LeuProLeuGlyGluLeuGluGlyLysArgSerThrArgAspArgArgSerProAlaGluLeuGluGlySerAlaSerGluGlyAlaAlaArgProValAlaArgPheAsn
                440                 450                 460

TCCATTCCTTTGACTGACACAGGCCATGTGGAGCAGCATGAGGGATCCTCTGACCCCCAGATCCCTGAGTCTGTCTTCCACCTGCTGGTGCCGGGCATCATCCTAGTCTTGCTG
SerIleProLeuThrAspThrGlyHisValGluGlnHisGluGlySerSerAspProGlnIleProGluSerValPheHisLeuLeuValProGlyIleIleLeuValLeuLeu
                GlnGlyHisGluArgGlnSerGluGlySerSerSerProGlnLeuGlnGluSerValPheHisLeuLeuValProSerValIleLeuValLeuLeu
470                 480                 490                 500

ACTGTTGGGGGCCTCCTGTTCTACAAGTGGAAGTGGAGGAGCCATCGAGACCCTCAGACATTGGATTCTTCTGTGGGGCGACCAGAGGACAGCTCCCTGACCCAGGATGAGGAC
ThrValGlyGlyLeuLeuPheTyrLysTrpLysTrpArgSerHisArgAspProGlnThrLeuAspSerSerValGlyArgProGluAspSerSerLeuThrGlnAspGluAsp
AlaValGlyGlyLeuLeuPheTyrArgTrpArgArgArgSerHisGlnGluProGlnArgAlaAspSerProLeuGluGlnProGluGlySerProLeuThrGlnAsp   Asp
      510                 520                 530                 540

AGACAGGTGGAACTGCCAGTATAGAAAGGATTCTATGACCCCTCACCATCCTGGACACACTCGTTTGTCAATGTCCCTCTGAAAATGTGGCGCCCAGCCCTGGACACAGTACTC
ArgGlnValGluLeuProValEnd
ArgGlnValGluLeuProValEnd
            550

CAGATGTTGTCTGACCAGCTCAGAGTACAGTGGGACGGTTGTCTTCCTTGATCTGGACAGTACTCTTCTACTCGTGCAGATTAAGATCACATTAGTTTTAACAGCTGCATCATA

TATTGTCATATGTTGAGCTTGTAGTCTATTAAAAACCCCAGTTCTATTTCCTGTGAAAAAAAAAAAAAAAAA
```

References

(1) Kawasaki, E.S. et al. (1985) Science 230, 291-296. (2) Ben-Avram, C.M. et al. (1985) Proc. Natl. Acad. Sci. USA 82, 4486-4489. (3) Ralph, P. et al. (1986) Blood 68, 633-639.