 © 1990 Oxford University Press

Nucleotide sequence of cDNA covering the complete coding part of the human vimentin gene

B.Honoré, P.Madsen, B.Basse, A.Andersen, E.Walbum, J.E.Celis and H.Leffers
Institute of Medical Biochemistry and Bioregulation Research Centre, Aarhus University, Ole Worms Allé, Build. 170, DK-8000 Aarhus C, Denmark

Submitted October 16, 1990 EMBL accession no. X56134

Seven clones containing all or parts of the coding sequence of the human vimentin gene (1398 bp) were isolated from a λgt11 human cDNA library constructed by reverse transcription of mRNA purified from SV40 transformed MRC-5 fibroblasts (MRC-5 V2) (1). Together, the clones span a cDNA sequence of 1765 bp. Previous attempts to isolate cDNA for human vimentin have resulted in partial lengths below 1100 bp (2, 3).

A monoclonal antibody, 98C1, recognizing vimentin among other proteins, was used for preliminary screening of 2×10^5 plaques. Two positive clones were obtained. One clone contained a 1.4 kb insert. Both strands were sequenced with the dideoxy chain termination technique (4). This clone was used to re-screen 2×10^5 plaques giving approximately 80 positive clones. Five clones were isolated and end-sequenced. Two of these latter clones extended the preliminary sequence in both directions. Both strands of the extensions were sequenced. The overlapping sequences were identical except for 44 bp on one of the clones considered to be a cloning artefact. The complete cDNA sequence (1765 bp) is shown in Fig. 1. The sequence contained an open reading frame coding for a 466 amino acid protein found to be almost identical with the published sequence for human vimentin based on a combination of partial cDNA sequences and genomic DNA sequences (2, 3). Our sequence, however, shows some mismatches in the coding sequence with the previously published partial sequences. We have thus found Ser-42 and Leu-442 instead of Asp and Phe respectively (2), and Asn-201, Leu-265, Ser-278, Ser-339 and Asn-350 instead of Ser, Ser, Ile, Cys and Lys respectively (3). The 3′ untranslated region contains 322 bp from the stop codon to the poly A stretch. In agreement with previous findings (3) we found two putative polyadenylation signals, the first is located 40 bp downstream from the stop codon, the second is located 301 bp from the stop codon.

ACKNOWLEDGEMENTS

We would like to thank O. Sønderskov for reproductions. P. Madsen was supported by a research position from the Medical Faculty, Aarhus University and H. Leffers by a fellowship from the Danish Cancer Foundation. The work was supported by grants from the Danish Biotechnology Programme, the Danish Cancer Society, the EEC Science Programme, the Danish Medical Research Council, the Danish Rheumatoid Society, NOVO and the Fund for Lægevidenskabens Fremme.

REFERENCES

1. Huschtscha,L.I. and Holiday,R. (1983) *J. Cell Sci.* **63**, 77–99.
2. Ferrari,S., Battini,R., Kaczmarek,L., Rittling,S., Calabretta,B., de Riel,J.K., Philiponis,V., Wei,J.-F. and Baserga,R. (1986) *Mol. Cell. Biol.* **6**, 3614–3620.
3. Perreau,J., Lilienbaum,A., Vasseur,M. and Paulin,D. (1988) *Gene* **62**, 7–16.
4. Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.

```
   1 CGCGCCACCGCCGCCGCCCAGGCCATCGCCACCCTCCGCAGCCATGTCCACCAGGTCCGT
   1                                           M  S  T  R  S  V
  61 GTCCTCGTCCTCCTACCGCAGGATGTTCGGCGGCCCGGGCACCGCGAGCCGGCCGAGCTC
   7  S  S  S  S  Y  R  R  M  F  G  G  P  G  T  A  S  R  P  S  S
 121 CAGCCGGAGCTACGTGACTACGTCCACCCGCACCTACAGCCTGGGCAGCGCGCTGCGCCC
  27  S  R  S  Y  V  T  T  S  T  R  T  Y  S  L  G  S  A  L  R  P
 181 CAGCACCAGCCGCAGCCTCTACGCCTCGTCCCCGGGCGGCGTGTATGCCACGCGCTCCTC
  47  S  T  S  R  S  L  Y  A  S  S  P  G  G  V  Y  A  T  R  S  S
 241 TGCCGTGCGCCTGCGGAGCAGCGTGCCCGGGGTGCGGCTCCTGCAGGACTCGGTGGACTT
  67  A  V  R  L  R  S  S  V  P  G  V  R  L  L  Q  D  S  V  D  F
 301 CTCGCTGGCCGACGCCATCAACACCGAGTTCAAGAACACCCGCACCAACGAGAAGGTGGA
  87  S  L  A  D  A  I  N  T  E  F  K  N  T  R  T  N  E  K  V  E
 361 GCTGCAGGAGCTGAATGACCGCTTCGCCAACTACATCGACAAGGTGCGCTTCCTGGAGCA
 107  L  Q  E  L  N  D  R  F  A  N  Y  I  D  K  V  R  F  L  E  Q
 421 GCAGAATAAGATCCTGCTGGCCGAGCTCGAGCAGCTCAAGGGCCAAGGCAAGTCGCGCCT
 127  Q  N  K  I  L  L  A  E  L  E  Q  L  K  G  Q  G  K  S  R  L
 481 GGGGGACCTCTACGAGGAGGAGATGCGGGAGCTGCGCCGGCAGGTGGACCAGCTAACCAA
 147  G  D  L  Y  E  E  E  M  R  E  L  R  R  Q  V  D  Q  L  T  N
 541 CGACAAAGCCCGCGTCGAGGTGGAGCGCGACAACCTGGCCGAGGACATCATGCGCCTCCG
 167  D  K  A  R  V  E  V  E  R  D  N  L  A  E  D  I  M  R  L  R
 601 GGAGAAATTGCAGGAGGAGATGCTTCAGAGAGAGGAAGCCGAAAACACCCTGCAATCTTT
 187  E  K  L  Q  E  E  M  L  Q  R  E  E  A  E  N  T  L  Q  S  F
 661 CAGACAGGATGTTGACAATGCGTCTCTGGCACGTCTTGACCTTGAACGCAAAGTGGAATC
 207  R  Q  D  V  D  N  A  S  L  A  R  L  D  L  E  R  K  V  E  S
 721 TTTGCAAGAAGAGATTGCCTTTTTGAAGAAACTCCACGAAGAGGAAATCCAGGAGCTGCA
 227  L  Q  E  E  I  A  F  L  K  K  L  H  E  E  E  I  Q  E  L  Q
 781 GGCTCAGATTCAGGAACAGCATGTCCAAATCGATGTGGATGTTTCCAAGCCTGACCTCAC
 247  A  Q  I  Q  E  Q  H  V  Q  I  D  V  D  V  S  K  P  D  L  T
 841 GGCTGCCCTGCGTGACGTACGTCAGCAATATGAAAGTGTGGCTGCCAAGAACCTGCAGGA
 267  A  A  L  R  D  V  R  Q  Q  Y  E  S  V  A  A  K  N  L  Q  E
 901 GGCAGAAGAATGGTACAAATCCAAGTTTGCTGACCTCTCTGAGGCTGCCAACCGGAACAA
 287  A  E  E  W  Y  K  S  K  F  A  D  L  S  E  A  A  N  R  N  N
 961 TGACGCCCTGCGCCAGGCAAAGCAGGAGTCCACTGAGTACCGGAGACAGGTGCAGTCCCT
 307  D  A  L  R  Q  A  K  Q  E  S  T  E  Y  R  R  Q  V  Q  S  L
1021 CACCTGTGAAGTGGATGCCCTTAAAGGAACCAATGAGTCCCTGGAACGCCAGATGCGTGA
 327  T  C  E  V  D  A  L  K  G  T  N  E  S  L  E  R  Q  M  R  E
1081 AATGGAAGAGAACTTTGCCGTTGAAGCTGCTAACTACCAAGACACTATTGGCCGCCTGCA
 347  M  E  E  N  F  A  V  E  A  A  N  Y  Q  D  T  I  G  R  L  Q
1141 GGATGAGATTCAGAATATGAAGGAGGAAATGGCTCGTCACCTTCGTGAATACCAAGACCT
 367  D  E  I  Q  N  M  K  E  E  M  A  R  H  L  R  E  Y  Q  D  L
1201 GCTCAATGTTAAGATGGCCCTTGACATTGAGATTGCCACCTACAGGAAGCTGCTGGAAGG
 387  L  N  V  K  M  A  L  D  I  E  I  A  T  Y  R  K  L  L  E  G
1261 CGAGGAGAGCAGGATTTCTCTGCCTCTTCCAAACTTTTCCTCCCTGAACCTGAGGGAAAC
 407  E  E  S  R  I  S  L  P  L  P  N  F  S  S  L  N  L  R  E  T
1321 TAATCTGGATTCACTCCCTCTGGTTGATACCCACTCAAAAAGGACACTTCTGATTAAGAC
 427  N  L  D  S  L  P  L  V  D  T  H  S  K  R  T  L  L  I  K  T
1381 GGTTGAAACTAGAGATGGACAGGTTATCAACGAAACTTCTCAGCATCACGATGACCTTGA
 447  V  E  T  R  D  G  Q  V  I  N  E  T  S  Q  H  H  D  D  L  E
1441 ATAAAAATTGCACACACTCAGTGCAGCAATATATTACCAGCAAGAATAAAAAAGAAATCC
1501 ATATCTTAAAGAAACAGCTTTCAAGTGCCTTTCTGCAGTTTTTCAGGAGCGCAAGATAGA
1561 TTTGGAATAGGAATAAGCTCTAGTTCTTAACAACCGACACTCCTACAAGATTTAGAAAAA
1621 AGTTTACAACATAATCTAGTTTACAGAAAAATCTTGTGCTAGAATACTTTTTAAAAGGTA
1681 TTTTGAATACCATTAAAACTGCTTTTTTTTTTCCAGCAAGTATCCAACCAACTTGGTTCT
1741 GCTTCAATAAATCTTTGGAAAAACTA(n)
```