Caenorhabditis cDNA encodes an eIF-4A-like protein

Deborah L. Roussell and Karen L. Bennett*

Department of Molecular Microbiology and Immunology, School of Medicine, University of Missouri, Columbia, MO 65212, USA

Submitted May 29, 1992

EMBL accession no. Z12116

The eukaryotic initiation factor 4A (eIF-4A) is one of the best characterized members of the 'DEAD box' family of RNA helicases. The function of eIF-4A protein within the initiation factor complex includes RNA binding and melting of RNA secondary structures that might interfere with the translation process (1). The eIF-4A protein also has an RNA-dependent ATPase activity (1). A modified ATPase B site (DEAD:asp, glu, ala, asp) distinguishes this family of RNA helicases (2). The DEAD helicase family members are a subgroup of a superfamily of (putative) helicases characterized by six conserved motifs (2, 3).

We designed primers to the conserved DEAD family motifs, Ia (PTRELA) and II (DEADXM), for use in polymerase chain reactions (PCR) to amplify genomic DNA of *Caenorhabditis elegans*, a free-living soil nematode. Sequence analysis of one PCR product revealed a predicted protein sequence that was similar to the mouse and yeast eIF-4As. With the PCR product probe, we isolated and sequenced a *C.elegans* cDNA that has 71% identity with the mouse eIF-4A gene product (4). The 1615 bp cDNA, called CeIF, corresponds to a 1.7 kb RNA. Like many *C.elegans* cDNAs, the 5' end shows evidence of RNA trans-splicing (5). The first 12 bp of the cDNA correspond to the last 12 of the 22 nucleotide spliced leader (SL1). There are 52 bp of additional 5' noncoding sequence and 337 bp of 3' noncoding sequences, including a 16 bp polyA tail. CeIF is probably a full length cDNA. Genomic Southern analysis does not distinguish between one or two very highly conserved genes. The corresponding genomic clone maps to linkage group III (left of *col8*) on the physical map of *C.elegans*, near another related, but distinct, putative RNA helicase identified by expressed sequence tag (EST) sequencing (6).

A comparison of the predicted protein of this nematode cDNA with that of the eIF-4A genes from mouse (4), yeast (7), and tobacco (8) shows sequence variability in the first 30 amino acids, minor differences in predicted protein lengths, conserved spacing of the helicase motifs and 61% to 72% overall identity (Figure 1). A comparison with other 'DEAD' helicases shows identity in the helicase domains, but only 25–35% overall identity. These data suggest that CeIF is the nematode homologue of eIF-4A. Confirmation, however, will require functional analyses of CeIF.

ACKNOWLEDGEMENTS

We thank Dr S.Gharib and M.Patterson for technical assistance, C.Christoffersen for helpful discussions, Dr S.Kim for the cDNA

library, and Dr A.Coulson for physical map data. This work was supported by a UMC Weldon Springs Grant, a March of Dimes Starter Award, a Council for Tobacco Research grant and a NSF grant (DCB9110298) to K.L.B.; D.L.R is an NIH Training Grant Postdoctoral Fellow (DHHSST32AL-07276-05).

REFERENCES

1. Rozen, F., et al. (1990) *Mol. Cell. Biol.* **10**, 1134–1144.
2. Linder, P., et al. (1989) *Nature* **337**, 121–122.
3. Hodgman, T.C. (1988) *Nature* **333**, 22–23; 578.
4. Nielsen, P.J. and Trachsel, H. (1988) *EMBO J.* **7**, 2097–2105.
5. Krause, M. and Hirsh, D. (1987) *Cell* **49**, 753–761.
6. McCombie, W.R., et al. (1992) *Nature Genet.* **1**, 124–131.
7. Linder, P. and Slonimski, P.P. (1988) *Nucleic Acids Res.* **16**, 10359.
8. Owttrim, G.W., Hofmann, S. and Kuhlemeier, C. (1991) *Nucleic Acids Res.* **19**, 5491–5496.

```

CeIF      NTDVKNQVNVSSVDADGLIEGHDVDFVDFDMLKEELLRGVGFGEKPSAIOGRAIPVCTTGQDVIAQA
eIF-4A11  MSGGSADYNREHGPGQDPGVIESMNEIVDFDMLKESLLRGVYATGFEKPSAIOGRAIPVCTTGQDVIAQA
eIF-4A1   NSASQDSRSRDGPGQDPGVIESMNEIVDFDMLKESLLRGVYATGFEKPSAIOGRAIPVCTTGQDVIAQA
TIF 182   MSEGITIEESQIOTYDQVYVYKDFDMLKESLLRGVYATGFEKPSAIOGRAIPVCTTGQDVIAQA
NeIF-4A2  MAGSAPGSGDFARDAKNTLLGTDEQEEFTSYDEVYDFDMLKESLLRGVYATGFEKPSAIOGRAIPVCTTGQDVIAQA
NeIF-4A3  NEEDRLVFETSKGVEPIASFAEAGICDGLKQVYVYKDFEKPASIOGRAIPVCTTGQDVIAQA
consensus      * F N * *LLRG** *GFEKPSAIOGR** * G DV* QA

CeIF      QSGTGKTATFSVSLQRIEDHEDPHVQALVNAPTRELAQIQKVMALGELYLNWILPCTGGTSVRDD.ORK.LEAGIHVVVGT
eIF-4A11  QSGTGKTATFATSILOOLEIEFKETQALVAPTRELAQQIQKVLALQDYMATGACIGTGHVNRHMKLQAEAP.HIVVGT
eIF-4A1   QSGTGKTATFATSILOOLEIDLKATQALVAPTRELAQQIQKVMALQDYMATGACIGTGHVNRHMKLQAEAP.HIVVGT
TIF 182   QSGTGKTATFATSILOOLEIDSVKAPALVAPTRELAQQIQKVMALQDYMATGACIGTGHVNRHMKLQAEAP.HIVVGT
NeIF-4A2  QSGTGKTATFCGVLQDLQSLVCEQALVAPTRELAQQIQKVMALQDYLGVYKVMATGACIGTSVRDD.ORKLQ.GVHVVGTP
NeIF-4A3  QSGTGKTSIALTVCOIVDTKSSEVQALILSPTRELAQQEKVILAIQDYLGVYKVMATGACIGTSVRDD.ORKLQ.GVHVVGTP
consensus  QSGTGKT* * Q ** QAL***PTRELA Q *KV A** * *CG* * * *V GTP

CeIF      GRVDFMIRNRLDTSR.IKQFVLDEADENSRGKQDIYEFVSRPQDVVLLSATNPSEVLQVTHFRNRPVILVCKDELT
eIF-4A11  GRVDFMIRNRLSPKY.IKQFVLDEADENSRGKQDIYEFVSRPQDVVLLSATNPSEVLQVTHFRNRPVILVCKDELT
eIF-4A1   GRVDFMIRNRLSPKY.IKQFVLDEADENSRGKQDIYEFVSRPQDVVLLSATNPSEVLQVTHFRNRPVILVCKDELT
TIF 182   GRVDFMIRNRLDTSR.IKQFVLDEADENSRGKQDIYEFVSRPQDVVLLSATNPSEVLQVTHFRNRPVILVCKDELT
NeIF-4A2  GRVDFMIRNRLDTSR.IKQFVLDEADENSRGKQDIYEFVSRPQDVVLLSATNPSEVLQVTHFRNRPVILVCKDELT
NeIF-4A3  GRVDFMIRNRLDTSR.IKQFVLDEADENSRGKQDIYEFVSRPQDVVLLSATNPSEVLQVTHFRNRPVILVCKDELT
consensus  GRV D * R IK* *LDE*DENLNRGK*QIY*** * QV * SAT* * L** *FV*HILVCK*ELT

CeIF      LEGIKQFYINVKDEKMFDCDLQYVHVHTQAVIFCHTRKRVDTLTKENTHOFVTSALHGDMQKQERDVINFRSGSSRVL
eIF-4A11  LEGIKQFYINVEREEKMLDLCQYETLTITQAVIFCHTRKRVDTLTKENTHOFVTSALHGDMQKQERDVINFRSGSSRVL
eIF-4A1   LEGIKQFYINVEREEKMLDLCQYETLTITQAVIFCHTRKRVDTLTKENTHOFVTSALHGDMQKQERDVINFRSGSSRVL
TIF 182   LEGIKQFYINVEREEKMLDLCQYETLTITQAVIFCHTRKRVDTLTKENTHOFVTSALHGDMQKQERDVINFRSGSSRVL
NeIF-4A2  LEGIKQFYINVKDEKMFDCDLQYETLTITQAVIFCHTRKRVDTLTKENTHOFVTSALHGDMQKQERDVINFRSGSSRVL
NeIF-4A3  LEGIKQFYINVKDEKMFDCDLQYETLTITQAVIFCHTRKRVDTLTKENTHOFVTSALHGDMQKQERDVINFRSGSSRVL
consensus  LEGIKQFYINVKDEKMFDCDLQYETLTITQAVIFCHTRKRVDTLTKENTHOFVTSALHGDMQKQERDVINFRSGSSRVL

CeIF      ITTDLARGIDVOOVSIVNYDLPNRENIHRIGRSGFRGKGVAINFVTEADARQLKIESYTTTQIEEMSEIADLI
eIF-4A11  ITTDLARGIDVOOVSIVNYDLPNRENIHRIGRSGFRGKGVAINFVTEADARQLKIESYTTTQIEEMSEIADLI
eIF-4A1   ITTDLARGIDVOOVSIVNYDLPNRENIHRIGRSGFRGKGVAINFVTEADARQLKIESYTTTQIEEMSEIADLI
TIF 182   ITTDLARGIDVOOVSIVNYDLPNRENIHRIGRSGFRGKGVAINFVTEADARQLKIESYTTTQIEEMSEIADLI
NeIF-4A2  ITTDLARGIDVOOVSIVNYDLPNRENIHRIGRSGFRGKGVAINFVTEADARQLKIESYTTTQIEEMSEIADLI
NeIF-4A3  ITTDLARGIDVOOVSIVNYDLPNRENIHRIGRSGFRGKGVAINFVTEADARQLKIESYTTTQIEEMSEIADLI
consensus  ITTDLARGIDVOOVSIVNYDLPNRENIHRIGRSGFRGKGVAINFVTEADARQLKIESYTTTQIEEMSEIADLI

```

Figure 1. Alignment of the CeIF predicted protein sequence with those of eIF-4A (-like) genes. (Pairwise alignments, University of Wisconsin Genetic Computer Group Package, were adjusted manually.) The percents identity/similarity with CeIF are: mouse eIF-4A II (72%/85%) and I (71%/84%) (4); yeast Tif 1 and 2 (64%/79%) (5); and tobacco NeIF-4A2 (63%/76%) and 3 (61%/79%) (7). A consensus of identical and conserved amino acids (*) is also shown.

* To whom correspondence should be addressed