BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS **243,** 273–276 (1998)
ARTICLE NO. RC978091

Characterisation of a New Human and Murine Member of the DnaJ Family of Proteins

Ginters Silins,[1,2] Sean Grimmond,[2] and Nicholas Hayward
Queensland Cancer Fund Research Unit, Joint Experimental Oncology Program, Queensland Institute of Medical Research, Herston, QLD 4029, Australia

Received December 24, 1997

We report the characterisation of a human gene, designated *MCG18* (multiple endocrine neoplasia type 1 candidate gene 18), that encodes a new member of the DnaJ family of proteins. Database searches indicate that *MCG18* also has the locus name *HSPF2*. *MCG18* lies ~250bp centromeric of the *VRF/VEGFB* gene on chromosome 11q13. The *MCG18* cDNA is predicted to encode a 241 amino acid product that has partial homology to *Escherichia coli dnaJ* in that it contains the J domain. However, MCG18 has greatest similarity to a functionally undefined protein from *Caenorhabditis elegans*, both of which are predicted to have a membrane-spanning region adjacent to their J domains. The cDNA encoding the murine homolog (*Mcg18*) was also cloned and sequenced, and the encoded protein shares ~81% similarity to MCG18. The coding region of *MCG18* is interrupted by 4 introns and the mRNA is expressed as a 1.4kb message in all tissues examined, including those derived from the breast, ovary, bladder, lung and keratinocytes.

The *E. coli* heat shock protein DnaJ is the founding member of a family of proteins that are associated with protein folding, complex assembly and export [reviewed in 1-3]. Prokaryotic and eukaryotic DnaJ homologues have a modular organisation, consisting of the J domain, a Gly/Phe-rich domain, $[Cys(X)_2Cys(X)Gly(X)Gly]_4$ repeats and a loosely conserved C-terminal region, as well as perhaps auxillary domains such as for protein targeting. The DnaJ family also includes a number of genes with a broad range of functions that encode only the ~70 amino acid long J domain, which is thought to mediate interaction with heat shock 70 proteins (HSP70). In addition to the DnaJ homologue from humans (HDJ2/HSDJ) at least 4 other genes that encode the J domain (HSJ1, HDJ1/HSP40, P58 [4] and CSP [5]) have been characterised.

We had previously identified an expressed sequence tag (EST) clone corresponding to the 3′ end of a novel gene within 560bp of the translation start codon of *VRF/VEGFB* [6], which we had designated multiple endocrine neoplasia type 1 candidate gene 18 (*MCG18*). An incomplete and incorrect cDNA sequence for the gene, called heat shock 40kD protein 2 (*HSPF2*), has subsequently been deposited in the GenBank database under accession number AF012106. We report here independent cloning and the complete coding sequence of the human J domain-containing gene *MCG18* and its murine homolog (*Mcg18*).

MATERIALS AND METHODS

Nucleotide sequencing and analysis. Human cosmid clone CLGW4 containing the *MCG18* gene has been described previously [6]. Expressed sequence tag (EST) clones 108172 (human), 350966 (murine) and 385535 (murine) were obtained from Genome Systems Inc. Genomic and cDNA templates for *MCG18* and *Mcg18* were sequenced with both vector-derived and gene-specific oligonucleotides using an Applied Biosystems Incorporated (ABI) dye terminator sequencing kit, as described previously [7]. The sequence data were complied using MacVector 4.2.1 software (IBI-Kodak). Sequence similarity searches were performed using the program GAP-BLAST [8] at the National Centre for Biotechnology Information (http://www.ncbi.nih.gov.nlm). Alom analyses for transmembrane domains [9] and peptide homology alignments using the program BESTFIT (GCG, Wisconsin) were conducted using the Australian National Genome Information Service computer faculty (http://www.angis.su.oz.au).

Northern analysis. Total cellular RNA from human cell lines derived from the breast, ovary, bladder, lung and keratinocytes were

[1] To whom correspondence should be sent at Queensland Institute of Medical Research, P.O. Royal Brisbane Hospital, Herston 4029, Australia. Fax: (61) 7 3362 0107; E-mail: gintS@qimr.edu.au.
[2] These authors contributed equally.

Sequences presented in this article have been submitted to the GenBank database and appear under accession numbers AF036874 and AF036875.

Abbreviations: *dnaJ* - gene encoding the *E. coli* heat-shock protein DnaJ; HSP70 - heat shock 70 protein; *HSPF2* - gene encoding the human heat shock 40kD protein 2; HSPF2 - protein product of *HSPF2; MCG18* - human multiple endocrine neoplasia type 1 candidate gene 18; MCG18 - protein product of *MCG18; Mcg18* - murine *MCG18* gene homolog; Mcg18 - protein product of *Mcg18;* UTR - untranslated region.

0006-291X/98 $25.00

prepared using RNeasy Mini Kits (Qiagen) following the manufacturer's protocol. 15μg of total RNA were electrophoresed through a 1.2% agarose gel containing formaldehyde, blotted onto nylon membrane (Amersham) and hybridised to a radiolabeled (^{32}P-dCTP) cDNA probe (derived from the insert of EST clone number 108172) as described previously [7]. The membranes were washed at 65°C in 0.1 × SSC (20 × SSC is 3M NaCl/0.3M trisodium citrate) and 0.1% SDS, prior to exposing to X-ray film.

RESULTS AND DISCUSSION

Characterisation of the Human MCG18 cDNA

During sequence characterisation of the *VRF/VEGFB* promoter region on human cosmid CLGW4 [6] which maps to chromosome 11q13, we identified a region that matched the 3′ untranslated region (UTR) of numerous human cDNA entries of the EST database—a gene which we designated *MCG18* and has subsequently been given the locus name *HSPF2*. EST clone 108172, which contains the region from the 5′ UTR to the poly(A) tail, was obtained and sequenced. However, comparison of the sequence data to other entries of the EST database (GenBank accession numbers AA524179 (clone 936891) and AA447986 (clone 782725)) suggested that clone 108172 contained an unprocessed intron. Sequencing of cosmid CLGW4 [10] that contained the entire gene led to the identification of 4 introns within the coding sequence and confirmed that clone 108172 contained an unspliced intron 4. It is also possible that clone 108172 represents an alternate splice form, in which case retention of the intron would create a new C-terminus for MCG18.

The cDNA sequence of *MCG18* shown in Fig. 1, beginning within the 5′ UTR and ending at the poly(A) tail, was translated in all possible reading frames and compared to the GenBank non-redundant protein database using the program GAP-BLASTX. The coding region of 726bp was identified on the basis of showing homology to the DnaJ family of proteins (Fig. 2). The putative initiation codon occurs within a strong Kozak context [11] and is preceded by an in-frame stop codon. The predicted 3′ UTR of human *MCG18* contains the polyadenylation signal AATAAA that precedes a poly(A) tract by 12 nucleotides. Northern analysis showed that *MCG18* is expressed as an ~1.4kb message in all cell lines examined, including those derived from the breast, ovary, bladder, lung and keratinocytes (Fig. 3).

```
tgaagtctagccccatcctggtccaatgcgctcttggtagcctcctttcccagctgcccg 60
 *  s  l  a  p  s  w  s  n  a  l  l  v  a  s  f  p  s  c  p
cccgccgccATGCCGCCCTTACTGCCCCTGCGCCTGTGCCGGCTGTGGCCCCGCAACCCT 120
 p  a  a  M  P  P  L  L  P  L  R  L  C  R  L  W  P  R  N  P  17
                          exon 1 ↓ exon 2
CCCTCCCGGCTCCTCGGAGCGGCCGCCGGGCAGCGGTCCAGACCCAGTACTTATTATGAA 180
 P  S  R  L  L  G  A  A  A  G  Q  R  S  R  P  S  T  Y  Y  E  37
CTGTTGGGGGTGCATCCTGGTGCCAGCACTGAGGAAGTTAAACGAGCTTTCTTCTCCAAG 240
 L  L  G  V  H  P  G  A  S  T  E  E  V  K  R  A  F  F  S  K  57
       ↓     exon 3
TCCAAAGAGCTGCACCCAGACCGGGACCCTGGGAACCCAAGCCTGCACAGCCGCTTTGTG 300
 S  K  E  L  H  P  D  R  D  P  G  N  P  S  L  H  S  R  F  V  77
GAGCTGAGCGAGGCATACCGTGTGCTCAGCCGTGAGCAGAGCCGCCGCAGCTATGATGAC 360
 E  L  S  E  A  Y  R  V  L  S  R  E  Q  S  R  R  S  Y  D  D  97
CAGCTCCGCTCAGGTAGTCCCCCAAAGTCTCCACGAACCACAGTCCATGACAAGTCTGCC 420
 Q  L  R  S  G  S  P  P  K  S  P  R  T  T  V  H  D  K  S  A  117
CACCAAACACACAGCTCCTGGACACCCCCCAACGCACAGTACTGGTCCCAGTTTCACAGC 480
 H  Q  T  H  S  S  W  T  P  P  N  A  Q  Y  W  S  Q  F  H  S  137
GTGAGGCCACAGGGGCCCCAGTTGAGGCAGCAGCAACACAAACAAAACAAACAAGTGCTG 540
 V  R  P  Q  G  P  Q  L  R  Q  Q  Q  H  K  Q  N  K  Q  V  L  157
                                                          ↓
GGGTACTGCCTCCTCCTCATGCTGGCGGGCATGGGCCTGCACTACATTGCCTTCAGGAAG 600
 G  Y  C  L  L  L  M  L  A  G  M  G  L  H  Y  I  A  F  R  K  177
GTGAAGCAGATGCACCTTAACTTCATGGATGAAAAGGATCGGATCATCACAGCCTTCTAC 660
 V  K  Q  M  H  L  N  F  M  D  E  K  D  R  I  I  T  A  F  Y  197
             exon 4  ↓  exon 5
AACGAAGCCCGGGCACGGGCCAGGGCCAACAGAGGCATCCTTCAGCAGGAGCGACAACGG 720
 N  E  A  R  A  R  A  R  A  N  R  G  I  L  Q  Q  E  R  Q  R  217
CTAGGGCAGCGGCAGCCGCCACCATCCGAGCCAACCCAAGGCCCCGAGATCGTGCCCCGG 780
 L  G  Q  R  Q  P  P  P  S  E  P  T  Q  G  P  E  I  V  P  R  237
GGCGCCGGCCCCTGAGGGGCTCACCTGGATGGGGCCTGCAGTGCGTTCCCGCTTTGCTTC 840
 G  A  G  P  *                                               241
CTTCCCTGGACGGCCCGCTCCCCGAAACGCGCGCAATAAAGTGATTCGCAG(A)n       892
```

FIG. 1. Nucleotide sequence and conceptual translation of the human *MCG18* cDNA, derived from EST clones 108172 and 936891, as well as from sequencing of the genomic cosmid clone CLGW4. The 5′ UTR and in-frame amino acids prior to the putative initiation codon are shown in lowercase lettering. The polyadenylation signal is underlined and arrows indicate the positions of introns within the coding sequence.

```
E. coli DnaJ (residues 6-70)                ---YYEILGVSKTAEE
                                                .|||:||| ..|..
Human     MPPL...LPLRLCRLWPRNPPSRLLGAAAGQRSRPSTYYELLGVHPGAST 47
          ||.|   ||||||||||:. . |||.||.|||| |..||||||||||||.
Murine    MPSLLLQLPLRLCRLWPHSLSIRLLTAATGQRSVPTNYYELLGVHPGASA 50

E. coli   REIRKAYKRLAMKYHPDRNQGDKEAEAKFKEIKEAYEVLTDSQKRAAYDQ
           |:::|: . ....||||:.|:.. ..:| |:.||| ||. .|.| .||:
Human     EEVKRAFFSKSKELHPDRDPGNPSLHSRFVELSEAYRVLSREQSRRSYDD 97
          ||:|||||.|||||||||||||.||||||||.|||||||||:|||.||.
Murine    EEIKRAFFTKSKELHPDRDPGNPALHSRFVELNEAYRVLSREESRRNYDH 100

E. coli   YG---

Human     QLRSGSPPKSPRTTVHDKSAHQTHSSWTPPNAQYWSQFHSVRPQGPQLRQ 147
          ||:|:|||||. .|...| .:||||||.||||||.||||||||||: |.
Murine    QLHSASPPKSSGSTAEPKYTQQTHSSWEPPNAQYWAQFHSVRPQGPESRK 150
                    .         .         .         .
Human     QQHKQNKQVLGYCLLLMLAGMGLHYIAFRKVKQMHLNFMDEKDRIITAFY 197
          ||:|:|..|||||||||:|||||||:||||:.|:| .||||||||||:|
Murine    QQRKHNQRVLGYCLLLMVAGMGLHYVAFRKLEQVHRSFMDEKDRIITAIY 200
                  .         .         .         .
Human     NEARARARANRGILQQERQRLGQRQPPPSEPTQGPEIVPRGAGP 241
          |:.|||||||: :|||||. .|..:.|| |.::. |:|.:.:|
Murine    NDTRARARANRARIQQERQQRQQPRAEPSLPPESSRIMPQDTSP 244
```

FIG. 2. BESTFIT alignment of human MCG18 and murine Mcg18, and alignment of MCG18 with the J domain of heat shock protein DnaJ from *E. coli* (SWISS-PROT accession number P08622). Perfect matches (|), conservative pairings (:) and semi-conservative pairings (.) are shown.

The mRNA size indicates that approximately 400 nucleotides of 5′ UTR remain uncharacterised.

Characterisation of the Murine Mcg18 cDNA

A GAP-BLASTN search of the EST database with the *Mcg18* cDNA sequence identified numerous entries encoding the murine homolog, of which clones 350966 and 385535 were obtained and sequenced with murine gene-specific oligonucleotides. The full-length coding sequence can be found under GenBank accession number AF036875 and the predicted murine Mcg18 protein sequence is shown in Fig. 2. EST clone 350966 begins within the 5′ UTR, contains a fully spliced 735bp coding region of *Mcg18* and ends at the poly(A) tail. Clone 385535 begins within exon 3 and ends at the same point, and differs to clone 350966 in that it encodes an additional Ser after the His residue at position 124 (Fig. 2). MCG18, however, does not contain the additional Ser and the sequence of clone 350966 is therefore represented in Fig. 2.

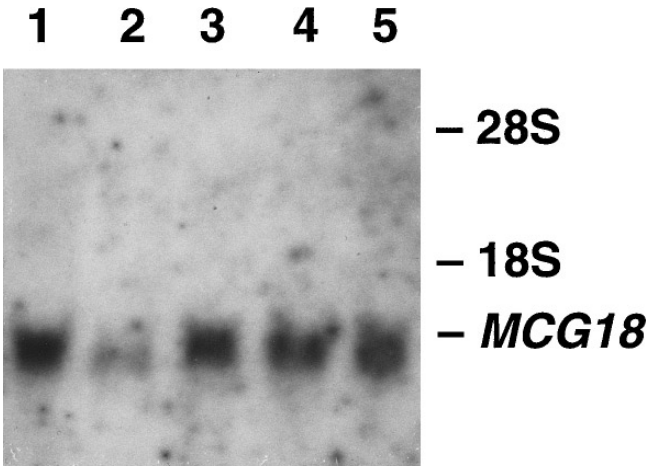

FIG. 3. Autoradiogram of a Northern blot with 15μg total RNA from the human cell lines H69-lung carcinoma (lane 1); JAM-ovary carcinoma (lane 2); BT20-breast carcinoma (lane 3); HaCaT-keratinocytes (lane 4); and, T24-bladder (lane 5) - probed with an *MCG18* cDNA clone. The positions of 28S and 18S ribosomal RNA are indicated, as is the transcript size of *MCG18*.

Structural Analyses of MCG18

MCG18 and Mcg18 aligned by the program BESTFIT in Fig. 2 show ~71% identity and ~81% similarity, and the human protein of 241 amino acids is shorter in length by 3 residues. The highly conserved region between residues 35 to 99 corresponds to the J domain of *E. coli* DnaJ as shown in Fig. 2, but all other hallmarks of the DnaJ homolog are absent [reviewed in 1-3]. GAP-BLASTP alignment of MCG18 residues 100 to 241 (lacking the J domain) to the non-redundant protein database fails to detect members of the J domain-containing family, including HSJ1 [1-3], HDJ1/HSP40 [1-3], the P58 PKR inhibitor (which has tetratricopeptide repeats and weak homology to eIF-2α) [4], the cys-

```
MCG18:    28  QRSRPSTYYELLGVHPGASTEEVKRAFFSKSKELHPDRDPGNPSLHSRFVELSEAYRVLS 87
              ++ R  T+YE+LGV   A+  E+K AF+++SK++HPD +    S  + F+EL  AY VL
C01G8.4: 22  KKIRQRTHYEVLGVESTATLSEIKSAFYAQSKKVHPD-NSSEESATASFLELKNAYDVLR 80

MCG18:    88  REQSRRSYDDQLRSGSP--PKSPRTTVHDKSAHQTHSSWTPPNAQYWSQFHSVRPQGPQ- 144
              R   RR YD QLR G    P   +   + +A Q   S      + WS + S  P   +
C01G8.4: 81  RPADRRLYDYQLRGGGGRYPNGGQRYQYPNTAPQYDFS------RDWSTYWSQNPDNSRS 134

MCG18:    145 LRQQQHKQNKQVLGYCL------LLMLAGMGLHYIAFRKVKQMHLNFMDEKDRIITAFYNEARAR 203
               R+++ K +++ +   +      L+++AG    Y+      Q  L+ + ++D I   F  +   R
C01G8.4: 135 SREERDKSSREFMKSIVKWTAIGLVLVAGYNGGYLYLLAYNQKQLDKLIDEDEIAKCFLRQKEFR 199
```

FIG. 4. GAP-BLASTP alignment [8] of MCG18 (residues 28-203) with a DnaJ-like protein from *C. elegans* (gene product C01G8.4) (residues 22-199). The symbols '+' and '–' represent a conservative substitution and a gap in the alignment, respectively. The J domain is double underlined and the transmembrane domains predicted by Alom [9] are underlined.

teine string protein (which features a cysteine-rich region) [5] and the MIDA1 protein (which has homology to Zuotin and tryptophan-mediated repeats) [12]. Furthermore, targeting signals and a CAAX box [1] appear to be absent from MCG18.

GAP-BLASTN alignment of *MCG18* to the non-redundant database shows a perfect match over a combined region of ~445bp to a putative cDNA encoding a DnaJ protein from humans called HSPF2 (GenBank accession number AF012106). This unpublished cDNA sequence however is erroneous in that it actually represents a genomic DNA fragment of *MCG18* spanning introns 1 to 3. The assigned open reading frame begins within intron 2, spans exon 3 which encodes part of the J domain, and terminates within intron 3. With the exception of HSPF2, GAP-BLASTP analysis of the non-redundant protein database (performed to find previously characterised proteins) and GAP-TBLASTN analysis of the EST database (performed to detect undefined homologs in other organisms) indicates that MCG18 has greatest homology to a functionally undefined gene product from *C. elegans* (GenBank accession number U80439, gene product C01G8.4) since homology extends beyond the J domain. The alignment of MCG18 and C01G8.4 (Fig. 4) shows 54/185 (29%) identity and 91/185 (49%) similarity overall. These proteins have no obvious motifs apart from the J domain, with one possible exception; the program Alom [9] identified a potential membrane spanning domain between residues 159 to 175 of MCG18 and 146 to 162 of C01G8.4, as shown in Fig. 4.

In summary, we have identified a novel DnaJ-like protein from humans and mice. The elucidation of the gene structure of *MCG18* and *Mcg18* provides the initial characterisation of MCG18/HSPF2.

ACKNOWLEDGMENTS

This work was supported by the National Health and Medical Research Council of Australia and the Queensland Cancer Fund. The authors thank Drs G. Weber and C. Larsson for providing cosmid CLGW4.

REFERENCES

1. Caplan, A. J., Cyr, D. M., and Douglas, M. G. (1993) *Mol. Biol. Cell.* **4,** 555–563.
2. Cyr, D. M., Langer, T., and Douglas, M. G. (1994) *Trends Biochem. Sci.* **19,** 176–181.
3. Silver, P. A., and Way, J. C. (1993) *Cell* **74,** 5–6.
4. Lee, T. G., Tang, N., Thompson, S., Miller, J., and Katze, M. G. (1994) *Mol. Cell. Biol.* **14,** 2331–2342.
5. Coppola, T., and Gundersen, C. (1996) *FEBS Lett.* **391,** 269–272.
6. Silins, G., Grimmond, S., Egerton, M., and Hayward, N. (1997) *Biochem. Biophys. Res. Commun.* **230,** 413–418.
7. Townson, S., Lagercrantz, J., Grimmond, S., Silins, G., Nordenskjöld, M., Weber, G., and Hayward, N. (1996) *Biochem. Biophys. Res. Commun.* **220,** 922–928.
8. Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) *Nucleic Acids Res.* **25,** 3389–3402.
9. Klein, P., Kanehisa, M., and DeLisi, C. (1985) *Biochim. Biophys. Acta* **815,** 468–476.
10. Grimmond, S., Lagercrantz, J., Drinkwater, C., Silins, G., Townson, S., Pollock, P., Gotley, D., Carson, E., Rakar, S., Nordenskjöld, M., Ward, L., Hayward, N., and Weber, G. (1996) *Genome Res.* **6,** 124–131.
11. Kozak, M. (1987) *Nucleic Acids Res.* **15,** 8125–8148.
12. Shoji, W., Inoue, T., Yamamoto, T., and Obinata, M. (1995) *J. Biol. Chem.* **270,** 24818–24825.