Characterization of Multiple mRNAs That Encode Mammalian Translation Initiation Factor 5 (eIF-5)*

Kausik Si, Kallool Dast, and Umadas Maitra

From the Department of Developmental and Molecular Biology, Albert Einstein College of Medicine of Yeshiva University, Jack and Pearl Resnick Campus, Bronx, New York 10461

Eukaryotic translation initiation factor 5 (eIF-5) interacts with the 40 S initiation complex (40S-mRNA-Met-tRNAf, eF2-GTP) to promote the hydrolysis of bound GTP with the concomitant joining of the 60 S ribosomal subunit to the 40 S initiation complex to form a functional 80 S initiation complex. In this paper, the multiple mRNAs that encode mammalian eIF-5 have been characterized. In rat tissues, three major eIF-5 mRNAs of 3.5, 2.8, and 2.2 kilobases in length are detected. All major eIF-5 mRNAs are initiated from a single transcription initiation site, contain identical 5'-untranslated and coding regions, but differ from one another only in the length of their 3'-untranslated regions. The different lengths of the 3'-untranslated region of eIF-5 mRNAs are generated by the use of alternative polyadenylation signals. Additionally, we demonstrate tissue-specific variations in eIF-5 mRNA expression as well as preference for polyadenylation sites. These results should lead to increased understanding of the regulation of eIF-5 gene expression.

Eukaryotic translation initiation factor 5 (eIF-5) plays a key role in initiation of protein synthesis in eukaryotic cells (for reviews, see Refs. 1–4). During the initiation process, a 40 S preinitiation complex, consisting of a 40 S ribosomal subunit to which the initiator Met-tRNAf is bound as a Met-tRNAf\(^{32P}\)-GTP ternary complex, scans along the mRNA until it recognizes the initiation AUG codon to form the 40 S initiation complex (40S-mRNA-Met-tRNAf, eF2-GTP). The initiation factor eIF-5, a monomeric phosphoprotein of about 50 kDa (5–7), then interacts with the 40 S initiation complex to promote the hydrolysis of ribosome-bound GTP. Hydrolysis of GTP causes the release of eIF-2-GDP (and P\(_i\)) from the 40 S ribosomal subunit which is essential for the subsequent joining of the 60 S ribosomal subunit to the 40 S complex to form a functional 80 S initiation complex (80S-mRNA-Met-tRNAf\(^{32P}\)) that is active in peptidyl transfer (8–11).

To increase our understanding of the structure and function of eIF-5 protein and regulation of its activity, we have recently cloned, sequenced, and expressed both a rat cDNA and the Saccharomyces cerevisiae gene encoding functional eIF-5 of calculated Mr = 48,926, and 45,346, respectively (12–14). Although the derived amino acid sequences of the yeast and rat eIF-5 protein show considerable sequence homology and identity, analysis of the cDNA-deduced structure of rat eIF-5 mRNA shows several interesting features that are absent in yeast eIF-5 mRNA. First, the 5'-UTR of rat eIF-5 mRNA contains two small ORFs upstream of the eIF-5 coding region (12). Such a feature of the 5'-UTR is characteristic of many mRNAs of regulatory genes, e.g. genes for growth factors, transcription factors, oncogenes, and signal transduction components (4, 15). Second, while Northern analysis of yeast poly(A)\(^{+}\) mRNA showed only a single size class of eIF-5 mRNA of 1.75 kb (14), multiple mRNAs were found to encode rat eIF-5 (12). The mechanism of formation of these different forms of rat eIF-5 mRNAs was unclear.

In this paper, we have cloned and sequenced a human eIF-5 cDNA. Sequence analysis indicates that the rat eIF-5 mRNA, the deduced structure of human eIF-5 mRNA also contains two small ORFs preceding the coding region. Additionally, we have studied the pattern of eIF-5 expression in different mammalian tissues. Finally, we show that multiple mRNAs encoding mammalian (rat) eIF-5 are generated by alternative use of polyadenylation signals in the 3'-noncoding region.

EXPERIMENTAL PROCEDURES

cDNA Cloning and Sequencing—Specific DNA restriction fragments isolated from previously cloned partial 1.1-kb rabbit and complete 3.55-kb rat cDNA clones (12) were labeled with \(^{32P}\) by random-priming using \(\text{[c-}^{32P}\text{]}\text{dCTP (Boehringer Mannheim DNA labeling kit) and used as hybridization probes to screen a HeLa cell cDNA library in phage \text{λZAPII (Stratagene) as described by Sambrook et al. (16). Positive clones were plaque-purified to homogeneity, and the cDNA inserts present in recombinant plasmids were isolated by in vivo excision as a subclone in the plasmid vector pBlueScript SK(+) (Stratagene). The inserts present in the recombinant plasmids were sequenced from both ends by the dideoxy chain termination method (17) using U. S. Biochemical sequencing kit, and a series of appropriate 17-mer deoxyligounucleotide primers.}

Primer Extension Mapping of the Transcription Initiation Site(s) of eIF-5 mRNAs—The transcription initiation sites were mapped using the dinucleotide-directed primer extension method of Sambrook et al. (16) as follows. An 18-mer deoxyligounucleotide 5'-GGATCTCTTTT-GTCGAG-3' which is complementary to nucleotides 57 to 74 of the eIF-5 cDNA (see Fig. 1) was labeled at the 5' end with \(^{32P}\) using \(\text{[y-}^{32P}\text{]}\text{ATP (3000 Ci/mmol) and phage T4 polynucleotide kinase. The }^{32P}\)-labeled primer (about 100 ng) was mixed with about 1 \(\mu\)g of HeLa cell poly(A)\(^{+}\) RNA and the mixture was precipitated by adding 70% ethanol. The washed pellet was dissolved in 40 \(\mu\)l Pipes-\(HCl\), pH 6.4, containing 1 \(\mu\)l EDTA, 0.4 \(\mu\)l NaCl, and 80% formamide, and allowed to hybridize at 30 °C for about 16 h. The DNA-RNA hybrids formed were then precipitated by adding 70% ethanol. The washed pellet was dissolved in the reverse-transcription buffer (Promega) containing 40 \(\mu\)l NaPP, and 40 units of RNAse (Promega) and reverse-transcribed at 42 °C using 20 units of avian myeloblastosis virus reverse transcriptase. The \(^{32P}\)-labeled DNA-RNA hybrids formed were isolated,
treated with DNase-free RNase, and the labeled single-stranded DNA was isolated by phenol-chloroform extraction followed by analysis on a 6% polyacrylamide, 8M urea-sequencing gel.

Northern Analysis—The multiple tissue-Northern blot used in this study was purchased from Clontech. Each lane contained approximately 2 μg of poly(A)⁺ RNA isolated from each rat tissue as indicated in the text. Following prehybridization as described (16), the blot was hybridized in the same solution containing 10% dextran sulfate at 42 °C for 20 h with an appropriate 32P-labeled DNA fragment derived from the complete 3.55-kb rat cDNA (12), labeled with [α-32P]dCTP (6000 Ci/mmol) by random-priming using the Boehringer Mannheim kit. The blot was then washed three times in 0.1 3 SSC, 0.1% SDS (1 3 SSC is 0.15 M NaCl, 15 mM sodium citrate, pH 7.0) at 60°C for 1 h, air-dried, and then autoradiographed. Following completion of one Northern analysis, the hybridized probe was removed by immersing the blot for 2 min in 0.5% SDS. The temperature of the mixture was then brought down to about 20–22 °C by shaking the mixture at room temperature. The blot was then reused for another Northern analysis.

RESULTS

Molecular Cloning and Characterization of Human eIF-5 cDNA—To clone human eIF-5 cDNA, a λZAPII HeLa cell cDNA library (Stratagene) was screened using a 32P-labeled 515-bp PvuII-HpaI DNA restriction fragment isolated from the 3'-end of the previously cloned partial rabbit liver cDNA (12) as a hybridization probe. Five positive clones were obtained in this screening procedure. An additional three independent clones were isolated when a 544-bp HindIII-SphI fragment isolated from the 5'-end of the complete rat cDNA clone (12) was used as the probe. Analysis of the DNA insert size in each homogenous positive clone followed by partial sequencing at their ends indicated that these cDNAs encompass the complete coding sequence as well as the 5'-UTR of the human cDNA. These two cDNA inserts were

fig. 1. Comparison of the nucleotide sequence of human and rat eIF-5 cDNAs. The human cDNA sequence was obtained as described in the text while the rat sequence is from Ref. 12. The ATG assigned as the translation start codon (numbered +1) is boxed. The human cDNA contains the complete 5'-UTR and the unique transcription start site, marked by primer extension analysis, is denoted by a downward arrow (✓) at nucleotide position 318. There are two other ATG codons in the 5'-untranslated regions, both of which generate short ORFs represented by broken lines. The shaded nucleotide residues in the human sequence show that these residues are different from those in the rat sequence.
sequenced in their entirety and were found to contain the nucleotide sequence 335 to 1457 and the sequence 1439 to 11507, respectively, of the human eIF-5 cDNA. Fig. 1 shows the nucleotide sequence of human eIF-5 cDNA and is compared with the previously determined (12) rat eIF-5 cDNA. There is a high degree of sequence conservation in the 5'-UTR as well as in the N-terminal coding region between the human and rat eIF-5 cDNA. Furthermore, a characteristic feature of both the rat and human cDNAs is the presence of two short open reading frames preceding the ATG codon from which the translation of eIF-5 mRNA is known to be initiated (Fig. 1). Although we have not cloned the complete 3'-UTR of human eIF-5 cDNA, it appears that the rat and the human sequences diverge considerably in the 3'-UTR.

Fig. 2 shows the alignment of the amino acid sequence of human, rat (12), and the yeast (14) eIF-5 proteins. eIF-5 shows a high degree of sequence conservation at the amino-terminal half. The yeast protein is, however, shorter than the mammalian protein at the carboxyl-terminal end. Of particular interest is the observation that the derived amino acid sequence of eIF-5 from all three sources contains sequence motifs that are similar to the consensus GTP-binding domains that are characteristic of proteins of the GTPase superfamily (18) (Table I). The role of these sequence motifs in the hydrolysis of GTP mediated by eIF-5 is unclear at present.

Characterization of eIF-5 mRNAs—We have previously used 32P-labeled DNA restriction fragments derived from the coding region of the rat eIF-5 cDNA to carry out Northern analysis of total poly(A)+ RNA isolated from rat liver and HeLa cells (12). Three distinct transcripts of 3.5, 2.8, and 2.2 kb were detected in rat liver, whereas HeLa cells contained a 3.5- and a 2.2-kb transcript (12). The mode of formation of multiple eIF-5 transcripts in mammalian cells was unknown. Since the rat cDNA corresponding to the longest 3.5-kb eIF-5 mRNA contained multiple potential polyadenylation signals (12), the possibility exists that some or all of the eIF-5 transcripts may arise by alternative processing at the 3'-noncoding region of the eIF-5 mRNAs. Alternatively, since the 5'-noncoding region of the 3.55-kb rat cDNA contained an excellent 3'-splice site, TCCCT-TCTTCTCCAG, preceding the initiation ATG codon (12) the possibility also exists that some of the shorter eIF-5 transcripts are derived by efficient splicing of eIF-5 mRNAs at this site. It is to be noted that many vertebrate cDNA sequences that have upstream ATG codons contain an unspliced 5'-intron (19). Finally, the multiple eIF-5 transcripts in mammalian cells may arise from multiple transcription initiation sites.

To distinguish between the above possibilities, we first carried out a primer extension analysis to map the transcription initiation site(s) of eIF-5 mRNAs. Fig. 3 shows that when a 18-nucleotide deoxyoligonucleotide primer complementary to nucleotides 257 to +47 of the eIF-5 cDNA was labeled with 32P at the 5'-end and used for primer extension analysis using HeLa cell poly(A)+ RNA as a template, a single 262-nucleotide long

1. Fig. 2. Sequence comparison of the predicted human, rat, and the yeast S. cerevisiae eIF-5 proteins. The amino acid sequences were aligned for maximum homology. The conserved residues are highlighted. Gaps are represented by broken lines.

Table I

Comparison of the putative G1-G4 GTP-binding domains in the eIF-5 amino acid sequence with the conserved sequence motifs in the GTPase superfamily. The single letter amino acid code is used. The numbers preceding the sequence motifs represent the position of the first amino acid in each binding motif.

| Motifs | G1 | G2 | G3 | G4 |
|--------|----|----|----|----|
| Consensus | GXXGKGG(S/T) | D-(X)-T107 | DXXG | (N/T)(K/Q)XD |
| Human eIF-5 | 27GKNGIKT34 | 29D-D-(X)-T107 | 202 DSG592 | 422 NKKD524 |
| Rat eIF-5 | 27GKNGIKT34 | 29D-D-(X)-T107 | 209 DWDG202 | 422 NKKD522 |
| Yeast eIF-5 | 27GKNGIKT34 | 29D-D-(X)-T107 | 209 DGTG188 | 249 TQDL499 |
Multiple mRNAs for Mammalian eIF-5

DISCUSSION

Our interest to clone the human eIF-5 cDNA stems from our earlier observation (12) that in rat eIF-5 cDNA, the ATG codon from which translation of eIF-5 protein is initiated is preceded by two short open reading frames. The presence of short ORFs in the 5'-UTR is characteristic of mRNAs of many regulatory genes (4, 15, 19). Results presented in this paper show that the 5'-UTR of mammalian eIF-5 cDNAs is remarkably conserved (~98% identical), including the presence of two short ORFs. Furthermore, primer extension analysis clearly shows that this 5'-UTR is part of the mature eIF-5 mRNA and does not contain an unspliced 5'-intron. It is interesting to note that in contrast to mRNAs of housekeeping genes and many other
vertebrate genes which have a GC-rich 5'-leader sequence (19), the 3'-UTR of eIF-5 mRNA is relatively GC-poor (only 40% G + C). Furthermore, the putative initiation ATG codons in these short ORFs are in relatively poor sequence context with respect to translation initiation (pyrimidines at both -3 and +4 positions). Thus a significant fraction of the scanning 40 S preinitiation complex is expected to bypass these upstream ATG codons and initiate translation from the ATG codon at +1 of the known coding sequence of eIF-5. This ATG codon is in good sequence context for initiation having an A at the -3 position. Further work is clearly necessary to understand the role of these upstream ORFs, if any, in the regulation of translation of eIF-5 mRNAs in mammalian cells.

In earlier studies reported from several laboratories, eIF-5 isolated from rabbit reticulocyte lysates, was described to be a protein of about 150 kDa (for a review, see Refs. 2 and 3). In contrast, later work published from this laboratory on purification and characterization of the protein from mammalian cells (5-7) as well as cloning and expression of its cDNA (12, 13) clearly showed that mammalian eIF-5 is a protein of about 50 kDa. However, Northern analysis of total poly(A)+ RNA isolated from rat liver and HeLa cells (12) showed that multiple mRNAs encode mammalian eIF-5. The presence of a 3'-splice site in the 5'-UTR immediately preceding the ATG start codon in the eIF-5-cDNA raises the possibility that one of the eIF-5 mRNAs could be a splice variant that might encode the high molecular weight isoform of eIF-5. Distinct molecular forms of the protein (isozymes) have been observed for wheat germ eIF-4F (20, 21), and mammalian and yeast eIF-4A (22, 23).

Experiments presented in this paper are consistent with the conclusion that all eIF-5 transcripts originate from a single initiation site, contain an identical 5'-UTR and coding region but differ from one another in the length of their 3'-UTR. The different lengths of the 3'-UTR of eIF-5 mRNAs are most likely generated by the use of alternative polyadenylation signals. It is also clear that the 2.2-kb eIF-5 mRNA, which is a relatively minor species in most rat tissues, is highly expressed in rat testis indicating a tissue-specific polyadenylation preference. The biological significance of alternative polyadenylation in the generation of multiple mRNAs encoding the same protein is not clear. However, a large number of mammalian genes including that for translation initiation factor eIF-4E have been shown to use differential polyadenylation signals to generate multiple mRNAs encoding the same protein (24, 25). It is particularly noteworthy that recent experiments in a variety of eukaryotic systems have shown that the 3'-UTR and the length of the poly(A) tail are important determinants in the translational regulation of gene expression particularly during embryonic development and differentiation (for a review, see Refs. 26 and 27). Further work will undoubtedly focus on whether the activity of eIF-5 and expression of this initiation factor gene are regulated during development and differentiation.

Acknowledgment—We thank Dr. Todd Evans of this institution for critically reading the manuscript.

REFERENCES
1. Maitra, U., Stringer, E. A., and Chaudhuri, A. (1982) Annu. Rev. Biochem. 51, 869-900.
2. Hershey, J. W. B. (1991) Annu. Rev. Biochem. 60, 717-755.
3. Merrick, W. C. (1992) Microbiol. Rev. 55, 291-315.
4. Kozak, M. (1992) Annu. Rev. Cell Biol. 8, 197-225.
5. Raychaudhuri, P., Chaudhuri, A., and Maitra, U. (1985) J. Biol. Chem. 260, 2132-2139.
6. Ghosh, S., Chevesich, J., and Maitra, U. (1989) J. Biol. Chem. 264, 5134-5140.
7. Chevesich, J., Chaudhuri, A., and Maitra, U. (1993) J. Biol. Chem. 268, 20659-20667.
8. Trachsel, H., and Stachelin, T. (1978) Proc. Natl. Acad. Sci. U. S. A. 75, 204-208.
9. Peterson, D. T., Safer, B., and Merrick, W. C. (1979) J. Biol. Chem. 254, 7730-7735.
10. Raychaudhuri, P., Chaudhuri, A., and Maitra, U. (1985) J. Biol. Chem. 260, 2140-2145.
11. Chakrabarti, A., and Maitra, U. (1991) J. Biol. Chem. 266, 14039-14045.
12. Das, K., Chevesich, J., and Maitra, U. (1993) Proc. Natl. Acad. Sci. U. S. A. 90, 3058-3062.
13. Chaudhuri, J., Das, K., and Maitra, U. (1994) Biochemistry 33, 4794-4799.
14. Chakrabarti, D., and Maitra, U. (1993) J. Biol. Chem. 268, 10524-10533.
15. Sambrook, J., Fritsch, E. F., and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
16. Sanger, F., Nicklen, S., and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U. S. A. 74, 5463-5467.
17. Bourne, H. R., Sanders, D. A., and McCormick, F. (1991) Nature 349, 117-126.
18. Kozak, M. (1991) J. Cell Biol. 115, 887-903.
19. Brownlee, K. S., Lax, S. R., and Ravel, J. M. (1987) J. Biol. Chem. 262, 11228-11232.
20. Allen, M. L., Metz, A. M., Timmer, R. T., Rhoads, R. E., and Browning, K. S. (1993) Trends Biochem. Sci. 18, 159-164.
21. Allen, M. L., Metz, A. M., Timmer, R. T., Rhoads, R. E., and Browning, K. S. (1992) J. Biol. Chem. 267, 23232-23236.
22. Nielsen, P. J., and Trachsel, H. (1988) EMBO J. 7, 2097-2105.
23. Linder, P., and Slonimski, P. P. (1989) Proc. Natl. Acad. Sci. U. S. A. 86, 2286-2290.
24. Leff, S. E., Rosenfeld, M. G., and Evans, R. M. (1986) Annu. Rev. Biochem. 55, 1091-1117.
25. Jaramillo, M., Pelletier, J., Edery, I., Nielsen, P. J., and Sonenberg, N. (1991) J. Biol. Chem. 266, 10446-10451.
26. Jackson, R. J., and Standart, N. (1990) Cell 62, 15-24.
27. Curtis, D., Lehan, R., and Zamore, P. D. (1995) Cell 81, 171-179.