

Retropseudogenes constitute the major part of the human elongation factor 1α gene family

Hans O.Madsen, Knud Poulsen[1], Otto Dahl[2], Brian F.C.Clark[3] and J.Peter Hjorth
Department of Molecular Biology and Plant Physiology, Aarhus University, DK-8000 Aarhus C, [1]Department of Oral Biology, Royal Dental College, Vennelyst Boulevard, DK-8000 Aarhus C, [2]Department of General and Organic Chemistry, Copenhagen University, DK-2100 Copenhagen Ø and [3]Division of Biostructural Chemistry, Department of Chemistry, Aarhus University, DK-8000 Aarhus C, Denmark

Received November 29, 1989; Revised and Accepted February 19, 1990 EMBL accession nos X16869–X16873 (incl.)

ABSTRACT

The elongation factor 1a (EF-1α) is a protein which promotes the GTP-dependent binding of aminoacyl-tRNA to ribosomes in the protein synthesis process. A human gene coding for EF-1α has previously been cloned and sequenced along with a pseudo-gene [1]. Here, we have further analyzed the family of human EF-1α genes. Using an EF-1α cDNA as probe twelve genomic EF-1α-like clones were isolated and analyzed. Four of these were sequenced and found to contain EF-1α retropseudogenes. A Southern blot analysis indicated that the remaining eight clones also contained retropseudogenes. Genomic Southern blot analysis revealed at least twenty loci in the human genome with sequence homology to the EF-1α cDNA. Besides the already described active gene only one potentially active locus was found. The others appeared to be retropseudogenes. EF-1α retropseudogenes were also found to be abundant in the mammalian species mouse and pig, while the chicken contained only one presumably active EF-1α gene.

INTRODUCTION

Processed genes [2] or processed retropseudogenes [3] are terms for pseudogenes that resemble a cDNA copy of a fully processed mRNA species. They include the 3′ terminal poly(A) tract of the mRNA and lack any introns present in the parental gene. In addition, they often extend to the ultimate 5′ end of the mRNA and are surrounded by short direct repeat sequences. These pseudogenes are believed to be DNA copies of reverse transcribed mRNA's, inserted into staggered chromosomal breaks [4]. The events must have taken place in germ cells or their precursors in order to establish the inheritance of these pseudogenes. Therefore, retropseudogenes usually are remnants of genes expressed in such cells, e.g. genes coding for housekeeping proteins [3].

Retropseudogenes seem to be most frequent in mammals. For example, there are 20–30 tubulin retropseudogenes in mammals but none in the fruit fly *Drosophila* and the chicken [3]. Similarly, retropseudogenes for glyceraldehyde-3-phosphate dehydrogenase are very abundant in several mammalian species but are absent in the chicken [3].

The elongation factor 1a (EF-1α) is a housekeeping protein which has been intensively studied because of its central role in the protein synthesis process. It is part of the elongation factor 1 (EF-1) complex, which in addition to EF-1α contains the EF-1β and EF-1γ proteins. EF-1α promotes the binding of aminoacyl-tRNA to the acceptor site of the ribosome under hydrolysis of GTP while EF-1β/γ catalyzes the exchange of GDP for GTP on EF-1α [5].

EF-1α genes have been described for a variety of eukaryotes. The yeast *Saccharomyces cerevisiae* contains two EF-1α genes [6], the fungus *Mucor racemosus* has three [7] and the brine shrimp *Artemia salina* may contain as many as four copies of the gene [8]. The fruit fly *Drosophila melanogaster* contains two genes, one of which is found to be expressed only in certain stages of development [9]. In mammals, EF-1α genes have been described only for mouse and man. A cDNA [10] and a partial genomic sequence [11] have been identified in mouse. In humans two cDNAs [1,12] and one active gene together with five retropseudogenes [1] have been described.

Here, we describe EF-1α-like sequences in the human genome with homology to an EF-1α cDNA clone using Southern blot analysis and sequencing of genomic clones combined with genomic Southern blot analysis. This revealed more than twenty EF-1α-like loci, of which only two can be intron-containing EF-1α genes. The rest are most likely retropseudogenes.

MATERIALS AND METHODS

Bacterial strains. E. coli strain Y1088 or Y1090 [13] and *E. coli* strain K802 [14] were used for propagation of λgt11 [15] and λL47.1 [15] recombinant phages, respectively. The M13mp18/19 phages and recombinant derivatives hereof were propagated in *E. coli* JM105 [16].

Preparation of genomic DNA. High-molecular weight genomic DNA was prepared essentially as described [17]. Human DNA was isolated from blood cells of one individual (JPH), while

murine, porcine and chicken DNA were isolated from liver tissue of single animals.

Libraries. The library of human liver cDNA's cloned into the Eco RI site of λ gt11 was kindly provided by George A. Ricca, Meloy Laboratories, Virginia, USA. The human genomic library was constructed from 12−20 kb fragments of partially Sau 3A digested human genomic DNA, using λL47.1 [15] as Bam HI substitution vector [18].

Hybridizations. Southern blot [19] hybridizations and *in situ* hybridizations were performed essentially as described [20], using stringent conditions. Double stranded DNA probes were labelled by nick translation [21] using [α-^{32}P]-dATP and DNA Polymerase I. Synthetic oligonucleotide probes prepared by the *in situ* phosphoramidite method [22] and purified by PAGE gel electrophoresis were end-labelled [23] using [γ-^{32}P]-ATP and T4 Polynucleotide Kinase.

DNA sequencing. cDNA and genomic DNA fragments to be sequenced were subcloned in the M13mp18/19 phages and sequenced as described [24,25]. Ten oligonucleotides homologous to a published human EF-1α cDNA sequence [12] were used together with the universal sequencing primer in the sequencing strategy.

RESULTS AND DISCUSSION

Analysis of cDNA and genomic clones

In order to analyse the homogeneity of the EF-1α mRNA in human liver, we compared a number of individually derived

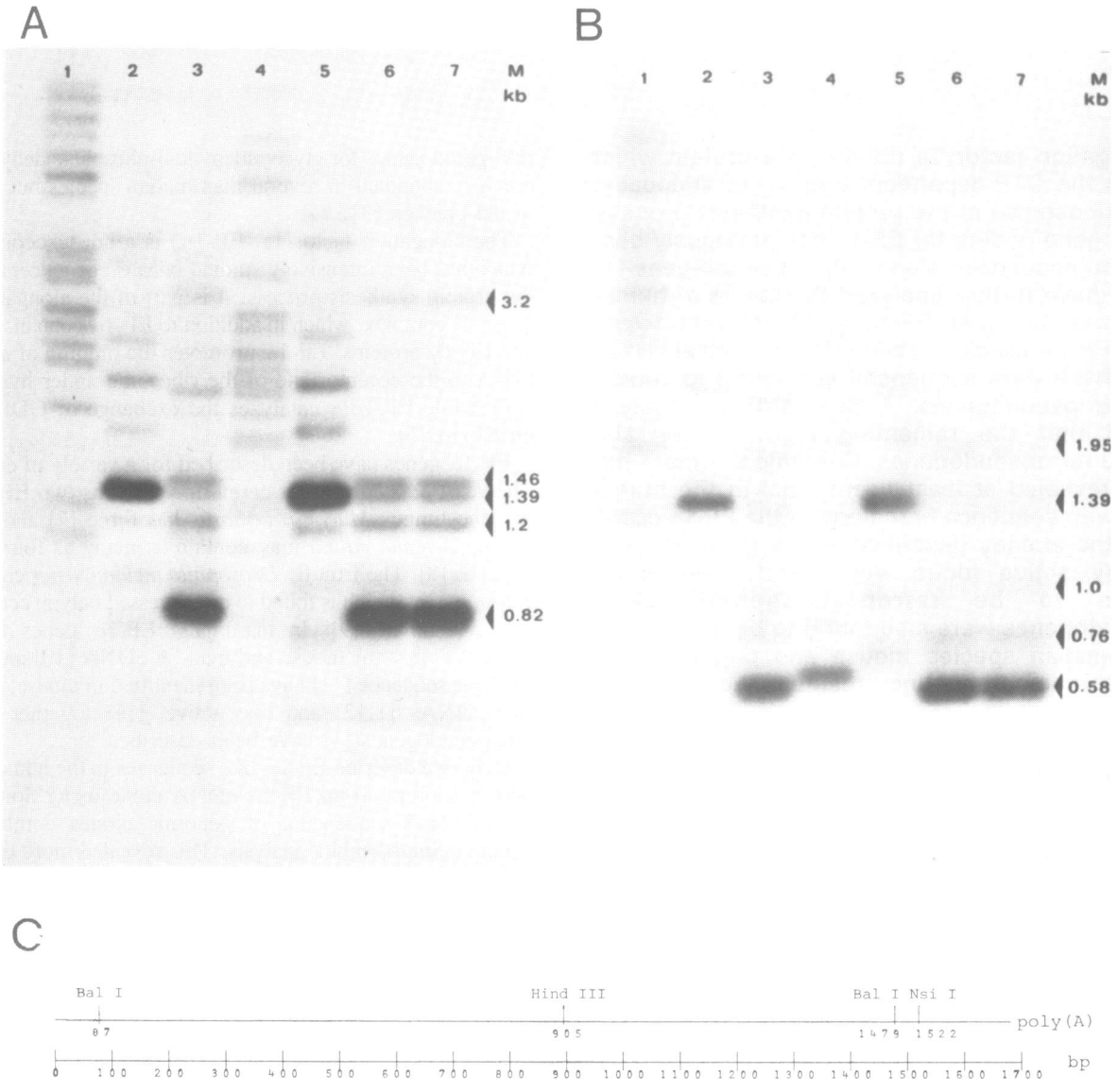

Figure 1. Southern blot analysis of human genomic DNA, using the 5′-end Bal I/Hind III fragment of CEF4 (A) and the 3′-end Bal I/Hind III fragment of CEF4 (B) as hybridization probes. 5 micrograms of restriction enzyme digested human genomic DNA was loaded in each lane and electrophoresed in a 1% agarose gel. The digestions are as follows: Hind III (lane 1), Bal I (lane 2), Hind III/Bal I (lane 3), Hind III/Nsi I (lane 4), Bal I/Nsi I (lane 5), Hind III/Bal I/Nsi I (lane 6) and Hind III/Bal I/Nsi I with double the amount of restriction enzyme and double the digestion time that of lane 6 (lane 7). Size markers (M) are indicated to the right of the figure. The hybridizing fragments marked with arrows are discussed in the text.The restriction enzyme map of the human EF-1α cDNA clone CEF4 with the 5′-end to the left (C) shows the restriction sites which were used in the gene analysis.

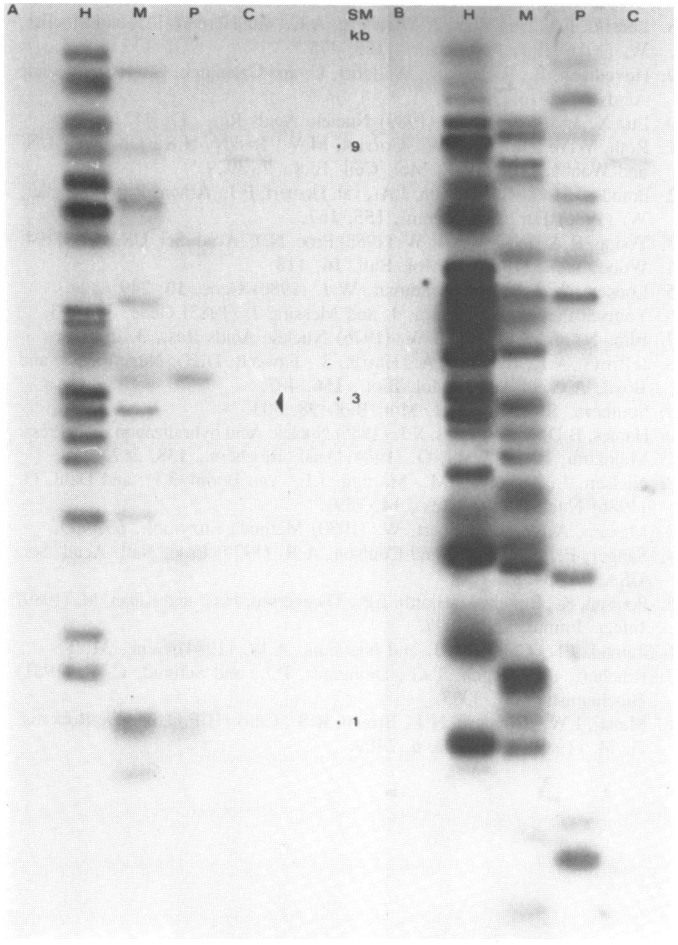

Figure 2. Southern blot analysis of genomic DNA from different species, using the 5'-end Bal I/Hind III fragment of CEF4 (A) and the 3'-end Bal I/Hind III fragment of CEF4 (B) as hybridization probes. 5 micrograms of human (H), murine (M), porcine (P) and chicken (C) genomic DNA was digested with Hind III and electrophoresed in a 1% agarose gel. The size markers (SM) are placed between Figs. A and B. The arrow in Fig. A shows the smallest of the two hybridizing fragments in chicken.

cDNA clones. The published sequence of a human EF-1α cDNA [12], enabled the preparation of a synthetic 55-oligomer, which was used to screen a human liver cDNA library. Nine positive clones were purified and sequenced.

The largest clone, CEF4, shown in Fig. 1C, is 1696 bp and yet not full-length. The sequence of the cDNAs are identical in overlapping regions. Comparison with the published sequences [1,12] reveals only a single base difference, which is found in the 3' untranslated end , and therefore this is most likely an allelic difference.

The sequence homogeneity among the 9 EF-1α cDNA clones derived from liver, the one from fibroblast cells [1], and the one from a lymphoid cell line [12], indicates that there is only one EF-1α gene active in human.

The cDNA insert of the CEF4 clone was used as probe in the screening of a human genomic library. Approximately 1/10,000 of the phages contained sequences homologous to the probe, and this high frequency agrees with that found by Uetsuki *et al.* [1]. Twelve randomly selected positive plaques (λ1 to λ12) were purified. Restriction enzyme mapping of the recombinant phage DNA combined with Southern blot hybridization using fragments of the CEF4 cDNA insert as probe indicated that the twelve clones represent ten different gene copies.

Four of the clones were sequenced. λ1 (EMBL accession no. X16870) represented a group of three clones which contained seemingly identical gene copies with the unique Hind III site found in the cDNA (Fig. 1). λ9 (EMBL accession no. X16872) was randomly chosen from the rest of the clones that contained the unique Hind III site. The other two were the aberrant clones λ4 (EMBL accession no. X16871), which hybridized very weakly to the 5' end of CEF4 and λ6 (EMBL accession no. X16873), which hybridized only to the 3' end of CEF4.

The sequences contained in λ1 and λ9 have a common structure characterized by the lack of introns, a poly(A) tail at the 3' end and direct repeats surrounding the gene. Moreover, both genes are incapable of producing normal EF-1α due to nucleotide substitutions, deletions, insertions, and duplications. These sequences therefore qualify as retropseudogenes.

λ4 and λ6 contain truncated EF-1α retropseudogenes. Both genes are interrupted by sequences which are identified as Alu elements. These elements are also recognized as retropseudogenes which seem to be derived from human 7SL RNA [3].

To reveal the gene type in the remaining eight genomic clones, the restriction enzymes Bal I, Hind III and Nsi I were used together in a Southern blot analysis of DNA from each clone, using the two Bal I/Hind III fragments from CEF4 as probes (Fig. 1). One of the clones, λ12, hybridized only weakly to the probes. The remaining seven clones showed hybridizing fragments of sizes indicating that they contain intron-free EF-1α genes, presumably retropseudogenes.

Genomic Southern blot analysis

To estimate the proportion of retropseudogenes among human EF-1α genes human genomic DNA was also digested with Bal I, Hind III and Nsi I in different combinations. The digestions were analyzed by Southern blot hybridizations as for the genomic clones (Fig. 1).

When the genomic DNA is digested with all three restriction enzymes together and hybridized to the 5'-Bal I/Hind III fragment probe of 820 bp (Fig. 1A, lanes 6 and 7) most EF-1α genomic sequences reveal the 820 bp fragment which indicates that they are retropseudogenes defined by the lack of introns in this 5' region. A similar result is seen when the 3' Hind III/Bal I cDNA fragment of 580 bp is used as probe (Fig. 1). The small very weakly hybridizing fragments presumably represent highly mutated EF-1α retroposons. Three of the remaining four fragments hybridizing to each of the two probes (1.39 kb, 3.2 kb, and 1.46 kb in Fig. 1A and 1.39 kb, 1.95 kb, and 0.76 kb in Fig. 1B) can be accounted for by a retropseudogene which has lost the Hind III site, the intron-free pseudogene contained in the genomic clone λ10, and the EF-1α gene sequenced by Uetsuki *et al.* [1], respectively. The last Hind III/Bal I fragments of 1.2 kb and 1.0 kb represent a retropseudogene or a second intron-containing EF-1α gene.

EF-1α genes of other species

Since retroposition seems to take place primarily in the germ cell lineage of mammals [3], we compared EF-1α genes of an avian species with 3 different mammalian species (Fig. 2). Genomic DNA from chicken, man, mouse and pig was digested with Hind III and equal quantities were analyzed by Southern blot hybridization, using both Bal I/Hind III fragments of CEFR4 (Fig. 1) as probes.

Based on the number of hybridizing fragments, each of the 3 mammalian species contains a family of EF-1α-like genes, while there apparently exists only one EF-1α gene in the chicken (Fig. 2). In addition, a similar genomic Southern blot hybridization shows the existence of a low number of *Xenopus laevis* EF-1α gene copies (data not shown). These observations indicate, that besides man other mammals also contain a large number of EF-1α retropseudogenes.

A copy number of EF-1α-like sequences in the different species can be estimated when the hybridizations are compared. In man, mouse, pig and chicken the copy numbers are visually estimated to be 20, 15, 10 and 1, respectively. In man the EF-1α-like copy number could be still higher than the given estimate, since as many as ten different genomic sequences were found in twelve randomly selected genomic clones with homology to the EF-1α cDNA probe.

CONCLUDING REMARKS

Based on the analysis of genomic clones and genomic Southern blotting we find that the human genome contains a family of at least twenty distinct EF-1α-like sequences, of which the large majority are retropseudogenes. Also mouse and pig contain a high number of EF-1α-like sequences. The large majority of these are presumably also retropseudogenes, since our results indicate the presence of only one EF-1α gene copy in chicken and other non-mammalian eukaryotes also contain a low number of EF-1α genes.

The fact that two of the analyzed human genomic clones contain inserted Alu elements is interesting, because it may indicate the existence of 'hot spot' regions for retropseudogene insertion. An Alu element has previously been found inserted into a human dihydrofolate reductase retropseudogene [26]. On the other hand, Alu elements constitute about 5% of the humane genome [27], so these instances could just be a matter of chance.

The analysis of human liver cDNA clones indicates the existence of only one active EF-1α gene copy per genome, but the genomic Southern blot analysis opens up the possibility of the existence of an extra active EF-1α gene in addition to the one identified by Uetsuki *et al.* [1]. An EF-1α-like gene which is expressed in early stages of development has actually been found in *D. melanogaster* and in *X. laevis* [28].

ACKNOWLEDGEMENT

We thank S.I.S. Rattan and J. Cavallius for providing the HeLa cell RNA and H.C. Thøgersen for purifying synthetic primers. We also thank M.E. Clark for correcting the manuscript and L. Heilesen for typing the manuscript. This work was supported by the Carlsberg Foundation and Senetek A/S.

REFERENCES

1. Uetsuki, T., Naito, A., Nagata, S. and Kaziro, Y. (1989) J. Biol. Chem., **264**, 5791.
2. Hollis, G.F., Hieter, P.A., McBride, O.W., Swan, D. and Leder, P. (1982) Nature, **296**, 321.
3. Weiner, A.M., Deininger, P.L. and Efstratiatis, A. (1986) Ann. Rev. Biochem., **55**, 631.
4. Lemischka, I. and Sharp, P.A. (1982) Nature, **300**, 330.
5 Moldave, K. (1985) Ann. Rev. Biochem., **54**, 1109.
6. Nagata, S., Nagashima, K., Tsunetsugu-Yokota, Y., Fujimura, K., Miyazaki, M. and Kaziro, Y. (1984) EMBO J., **3**, 1825.
7. Linz., J.E., Katayama, C. and Sypherd, P.S. (1986) Mol. Cell. Biol., **6**, 593.
8. Lenstra, J.A., van Vliet, A., Arnberg, A.C., van Hemert, F.J. and Moeller, W. (1986) Eur. J. Biochem., **155**, 475.
9. Hovemann, B., Richter, S., Walldorf, U. and Cziepluch, C. (1988) Nucleic Acids Res., **16**, 3175.
10. Lu, X. and Werner, D. (1989) Nucleic Acids Res., **17**, 442.
11. Roth, W.W., Bragg, P.W., Corrias, M.V., Reddy, N.S., Dholakia, J.N. and Wahba, A.J. (1987) Mol. Cell. Biol., **7**, 3929.
12. Brands, J.H.G.M., Maassen, J.A., van Hemert, F.J., Amons R. and Moeller, W. (1986) Eur. J. Biochem., **155**, 167.
13. Young, R.A. and Davis, R.W. (1983) Proc. Natl. Acad. Sci. USA, **80**, 1194.
14. Wood, W.B. (1966) J. Mol. Biol. **16**, 118.
15. Loenen, W.A.M. and Brammar, W.J. (1980) Gene, **10**, 249.
16. Yanisch-Perron, C., Vieira, J. and Messing J. (1985) Gene, **33**, 103.
17. Blin, N. and Stafford, D.W. (1976) Nucleic Acids Res., **3**, 2303.
18. Jeffreys, A.J., Barrie, P.A., Harris, S., Fawcett, D.H., Nugent, Z.J. and Boyd, A.C. (1982) J. Mol. Biol., **156**, 487.
19. Southern, E.M. (1975) J. Mol. Biol. **98**, 503.
20. Hames, B.D. and Higgens, S.J. (1985) Nucleic Acid hybridization. IRL Press.
21. Meinkoth, J. and Wahl, G. (1984) Anal. Biochem., **138**, 267.
22. Nielsen, J., Taagaard, M., Marugg, J.E., van Boom, J.H. and Dahl, O. (1986) Nucleic Acids Res., **14**, 7391.
23. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol., **65**, 499.
24. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA, **74**, 5463.
25. Poulsen, K., Brandt, J., Hjorth, J.P., Thoegersen, H.C. and Kilian, M. (1989) Infect. Immun., **57**, 3097.
26 Shimada, T., Chen, M.-J. and Nienhuis, A.W. (1984) Gene, **31**, 1.
27. Rinehart, F.P., Ritch, T.G., Deininger, P.L. and Schmid, C.W. (1981) Biochemistry, **20**, 3003.
28. Mattaj, I.W., Coppard, N.J., Brown, R.S., Clark, B.F.C. and De Robertis, E. M. (1987) EMBO J., **6**, 2409.