Human release factor eRF1: structural organisation of the unique functional gene on chromosome 5 and of the three processed pseudogenes

Lucienne Guenet[a,*], Bertrand Toutain[b], Isabelle Guilleret[a], Bruno Chauvel[a], Larry L. Deaven[c], Jonathan L. Longmire[c], Jean-Yves Le Gall[a], Véronique David[a], André Le Treut[b]

[a] *Département de Biochimie et Biologie Moléculaire et UPR41 CNRS, Faculté de Médecine, 2, Avenue du Pr. Léon Bernard, CS34317, 35043 Rennes Cedex, France*

[b] *Département de Biochimie et Biologie Moléculaire, Laboratoire de Biochimie Médicale A, Faculté de Médecine, 2, Avenue du Pr. Léon Bernard, CS34317, 35043 Rennes Cedex, France*

[c] *Center for Human Genome Studies, Los Alamos National Laboratory, Los Alamos, NM 87545, USA*

Received 19 April 1999; received in revised form 20 May 1999

Abstract In lower and higher eukaryotes, a family of tightly related proteins designated eRF1 (for eukaryotic release factor 1) catalyses termination of protein synthesis at all three stop codons. The human genome contains four eRF1 homologous sequences localised on chromosomes 5, 6, 7 and X. We report here the cloning and the structural analysis of the human eRF1 gene family. It appears that the gene located on chromosome 5 alone is potentially functional, whereas the other three sequences resemble processed pseudogenes. This is the first description of the structural organisation of the human eRF1 gene, which has been remarkably conserved during evolution and which is essential in the translation termination process.

Key words: Human eRF1; Genomic structure; Gene family

1. Introduction

Termination of protein synthesis occurs when a stop codon enters the A site of the ribosome. Hydrolysis of the last peptidyl tRNA, liberating the nascent polypeptide, requires specific release factors (RFs) and GTP [1,2]. In higher eukaryotes, it was reported over 20 years ago that a single release factor can recognise all three nonsense codons in an in vitro GTP-dependent termination assay [3]. This factor has been characterised as a member of a highly conserved protein family and designated eRF1 for eukaryotic release factor 1 [4]. No member of the eRF1 protein family contained GTP-binding motifs known to be present in proteins involved in GTP-dependent reactions, however, so it was suggested that an additional eRF should exist that confers GTP requirement for the termination process [4]. It has been demonstrated [5] that a second polypeptide factor exists, named eRF3, which is inactive per se as a release factor, but which can interact with eRF1 and greatly stimulate its activity in the presence of GTP and low stop codon concentration. These authors suggested that a quaternary complex composed of eRF1, eRF3, GTP and the ribosome was the catalytically active termination complex. It was then shown that eRF3, which displays GTP-binding motifs in its structure [5], is a GTPase with negligible intrinsic activity, but that it is greatly stimulated by the concomitant presence of eRF1 and the ribosome [6]. The N-terminal fragment of eRF3 proteins in higher eukaryotes proved to be non-essential for GTP binding [6] and furthermore, GTPase and eRF1-stimulating activities of the truncated eRF3 proteins are preserved [5,7]. The molecular anatomy of interaction between the two human eRFs was characterised in vivo [8]; it was shown that in each factor, two regions in the C-terminal parts are critical for reciprocal binding, and that GTP and eRF1 binding regions of eRF3 are non-overlapping.

The mechanisms by which eRF1 and eRF3 act in the termination process are not yet clearly understood. In the case of eRF1, stop-signal recognition involved protein-RNA recognition instead of mRNA codon-tRNA anticodon pairing (reviewed in [9,10]). A model of molecular mimicry between release factors and tRNA has been proposed [11], assuming that a domain in the RF protein structure could recognise the stop codon. It has also been suggested that an efficient stop signal might be more than just the codon (reviewed in [12,13]). The authors concluded that the +4 base nature influenced the efficiency with which the stop codon was decoded, and that the two bases immediately downstream could also have significant effects. The resulting hierarchy of efficiencies for stop signals could play a role in competition at the signal, and thus be important in the regulation of certain cellular events [12]. Hypotheses about the role(s) of eRF3 have also been formulated. In a 'GTP state', eRF3 should control the positioning of eRF1 towards the ribosome, while in a 'GDP state', it should facilitate the dissociation of the eRFs from the ribosome [6]. The RF-tRNA mimicry model also predicted that eRF3 should mimic an EF-Tu function and bring eRF1 to the A site of the ribosome for termination of the protein synthesis (reviewed in [10]). Additional roles have been postulated for eRF3 (reviewed in [14]): recycling of eRF1 and the ribosome at the end of the translation termination process [9,10] and a proof-reading function [10]. Although some hypotheses have not yet been experimentally confirmed, important progress in understanding the roles of eRF1 and eRF3 proteins has been made.

Numerous cDNAs encoding eRF1 protein have been described in higher eukaryotes: one from *Xenopus laevis*, Cl1 [15], two from Syrian hamster, MACl1-4 and MACl1-1 [16] and four from human. TB3-1 clone was isolated from a human carcinoma cell line [17], clone 58 from a normal human duodenal mucosa cDNA library [18], and HSCl1DHG7 and

*Corresponding author. Fax: (33) 2 99 33 68 98.
E-mail: lucienne.guenet@univ-rennes1.fr

PII: S0014-5793(99)00795-4

HSCl1DHG11 from a human Daudi cell line [16]. By using human-mouse somatic cell panel analysis, TB3-1-related sequences were localised to human chromosomes 5, 6, 7 and X [19]. Hybridisation of a human genomic Southern blot with clone 58 displayed four *Hin*dIII intense bands of about 14, 13, 12 and 8.6 kb [20]. The genomic organisation of these four human eRF1 homologous sequences remains unknown, however.

In this paper, we describe the molecular cloning and the structural organisation of the four human eRF1 homologous sequences. We report evidence that only the gene localised on chromosome 5 exhibits a typical exon-intron organisation, and suggest that this is probably the unique functional gene.

2. Materials and methods

2.1. Cosmids

- Cosmid ICRFc104C2049: ICRF chromosome X-specific cosmid library, Resource Centre Primary Database, RZPD.
- Cosmid ICRFc109F1245: ICRF chromosome 6-specific cosmid library, Resource Centre Primary Database, RZPD.
- Cosmid ICRFc113B1812Q4: ICRF chromosome 7-specific cosmid library, Resource Centre Primary Database, RZPD; cosmid Y33B20: LLNL chromosome 7-specific cosmid library, UK HGMP Resource Centre.
- Cosmid 48G4 and cosmid 239D3: chromosome 5-specific cosmid library constructed at the Center for Human Genome Studies, Los Alamos National Laboratory, Los Alamos, NM, USA, under the auspices of the US Department of Energy.

2.2. Probes

The following probes were used after labelling with [α-^{32}P]dCTP by the random priming method [21].

- Clone 58: a 1548-bp clone isolated from a cDNA library of human duodenal mucosa [12] and described as HSHCGVII (X80916) [20]. This sequence is identical to HSCl1DHG7 (X81625) from nucleotide 765 to nucleotide 2313.
- Three other probes were obtained by polymerase chain reaction (PCR) using Promega Taq DNA polymerase and the human cDNA clone Cl1DHG7 (kindly given by M. Kress) as target DNA: DHG7-A, 334 bp (position 120–453: forward, 5′-AGGAGGAGGCGAGAAGAT-3′, reverse, 5′-CTTTCCTTCTTCTGTTAC-3′), DHG7-B, 160 bp (position 534–693: forward, 5′-AGAGGCTCTTACAGCACTAC-3′, reverse, 5′-TGACTGACCTCCTCTACCGTG-3′), DHG7-C, 809 bp (position 2376–3184: forward, 5′-TGTGTTTCCTTGTGTTGG-3′, reverse, 5′-AAGCAAAAAAAGCCTGTTC-3′).
- M-5: a cDNA fragment of 257 bp, obtained by PCR using a murine cDNA clone (I.M.A.G.E. Consortium clone ID 1050740, LNLN, HGMP Resource Center) as target DNA and the following primers: forward, 5′-GGATCCTTGGGAGGAAGTGAG-3′ (position 3) reverse, 5′-GGCCCGCCTCCAAGCTCTTAAT-3′ (position 249). This probe is 97% identical to the first 217 nucleotides of the human HSCl1DHG7 cDNA and contains 43 additional nucleotides upstream.

2.3. Library screening

The chromosome-specific cosmid libraries were screened with 50 ng of clone 58 probe labelled with [α-^{32}P]dCTP, in the presence of 50 ng of the appropriate cosmid vector labelled with [^{35}S]dATPαS. Positive clones were purified by a secondary screening.

2.4. Restriction analysis of cosmid clones

Cosmid DNA, single (*Bam*HI, *Hin*dIII, *Eco*RI, *Eco*RV, *Xba*I, *Xho*I) or double (*Bam*HI+*Hin*dIII, *Bam*HI+*Eco*RI, *Bam*HI+*Eco*RV, *Hin*dIII+*Eco*RI, *Hin*dIII+*Eco*RV, *Eco*RI+*Eco*RV) digested, was size-fractionated by electrophoresis in 0.7% agarose gel, and the DNA fragments were transferred to Hybond N$^+$ (Amersham) according to the standard procedure for Southern blotting [22]. Blots were successively hybridised with several ^{32}P-labelled probes: clone 58, DHG7-A, DHG7-B and the appropriate vector. Restriction maps for each cosmid were elaborated and DNA fragments of interest were subcloned in pBluescript II KS$^+$ for sequencing.

2.5. DNA sequencing and computer analysis

DNA fragments were sequenced twice on both strands [23] using dRhodamine terminator cycle sequencing kit (Perkin Elmer) and an ABI 377 DNA sequencer. Sequence data were analysed using Sequencing Analysis, version 3.2 (ABI), and sequence assembly was performed using SeqMan (DNASTAR). Various search tools were also used: BLAST (NCBI, Bethesda, MD, USA), CENSOR [24].

3. Results

3.1. The human chromosome 6 eRF1 gene

Screening of the ICRF chromosome 6-specific cosmid library with clone 58 probe yielded three positive clones. Hybridisation with the same probe of a Southern blot of each cosmid DNA cut with *Hin*dIII revealed a 12-kb fragment. DNA fragments of interest were identified by restriction analysis using clone 58 probe and three other probes (DHG7-A, DHG7-B and DHG7-C), then subcloned and sequenced. A 8170-bp genomic sequence was obtained and compared with HSCl1DHG7 cDNA sequence (B. Chauvel, personal communication). The human chromosome 6 eRF1 gene sequence lacked the first 457 nucleotides of the cDNA, as well as the last 1360 nucleotides of the 3′ UTR, and showed a 302-bp insertion (Fig. 1A). There was only 95% identity between the cDNA sequence and the corresponding genomic sequence, and in view of the absence of GT/AG splice sites, it is probable that the inserted region does not correspond to an intron. A 356-kb sequence in the MHC class I region of chromosome 6, including the eRF1 gene, has now been achieved (GenBank AF055066). Neither the first 457 nucleotides nor the last 1360 nucleotides of the cDNA were found in this sequence, however.

3.2. The human chromosome X eRF1 gene

Four positive clones were isolated by screening the ICRF chromosome X-specific cosmid library with clone 58 as a probe. Digesting these clones by *Hin*dIII followed by hybridisation with clone 58 revealed a 13-kb fragment for two of them; one of the two cosmids (104C2049) was partly sequenced. Comparison of the resulting 8635-bp sequence with that of HSCl1DHG7 cDNA revealed that the first 275 nucleotides of the cDNA were absent, whilst a 331-nucleotide sequence had been inserted (Fig. 1B). In order to verify the absence of the 5′ part of the cDNA sequence, the 5′ extended probe (M-5 probe) was used to screen the chromosome X-specific cosmid library, but it revealed no positive clone. There was only 92% nucleotide identity between the cDNA and the human chromosome X eRF1 gene. The sequence has been deposited in the GenBank database under accession number AF110131.

3.3. The human chromosome 7 eRF1 gene

Two positive clones were obtained from screening of the ICRF chromosome 7-specific cosmid library with clone 58 probe and three positive clones from screening of the LLNL chromosome 7-specific cosmid library. Southern blot of *Hin*dIII-digested cosmid DNA hybridised with clone 58 showed a unique 8.6-kb fragment. Cosmid 113B1812Q4 (ICRF) and cosmid Y33B20 (LLNL) were partly sequenced using primers

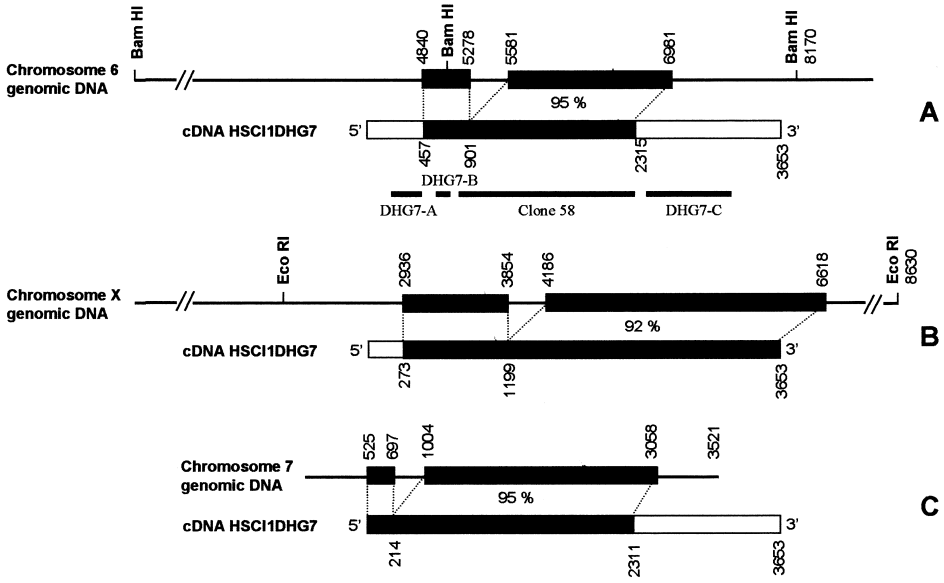

Fig. 1. Alignment of the human cDNA HSCl1DHG7 sequence and the eRF1 gene sequences on chromosome 6 (A), chromosome X (B) and chromosome 7 (C). Homologous regions between the cDNA and the genomic sequences are presented as solid boxes. cDNA sequences absent from the genomic sequences are shown as open boxes. The nucleotide identity between cDNA and genomic sequences is indicated for each chromosome in %. The position and the name of the different probes are indicated in A as black bars.

designed from the human HSCl1DHG7 cDNA. The resulting 3521-bp sequence was compared with the cDNA sequence. A 306-nucleotide region was inserted in the genomic sequence, whilst the last 1360 nucleotides of the 3′ UTR were missing (Fig. 1C). The DHG7-C probe, corresponding to a 809-bp DNA fragment of the 3′ UTR, gave no positive signal after hybridisation with the library filters, indicating that the human chromosome 7 eRF1 gene sequence stopped a few nucleotides downstream of the polyadenylation site AATAAA, as did that on chromosome 6. There was nearly 95% nucleotide identity between the cDNA and the corresponding genomic sequence, but numerous deletions of up to 40 nucleotides in length occurred on chromosome 7. The sequence has been deposited in the GenBank database under accession number AF112357.

3.4. The human chromosome 5 eRF1 gene

A chromosome 5-specific cosmid library constructed at the Center for Human Genome Studies (LANL) was screened with clone 58 probe and gave 13 positive clones, one of them (48G4) being positive also for M-5 probe. A Southern blot of the 13 cosmid DNA cut by *Hin*dIII and hybridised with clone 58 showed a unique band of 6.7 kb; although the *Hin*dIII fragment size was not as anticipated (14 kb), the analysis was continued. One cosmid, 239D3, positive for clone 58 alone, and the cosmid 48G4 positive for the two probes, were entirely sequenced. Comparison of the resulting 50 769-bp sequence with the cDNA sequence showed that the entire sequence of the human HSCl1DHG7 cDNA was retrieved.

By comparing the nucleotide sequence of the human chromosome 5 eRF1 gene and HSCl1DHG7 cDNA, we deduced

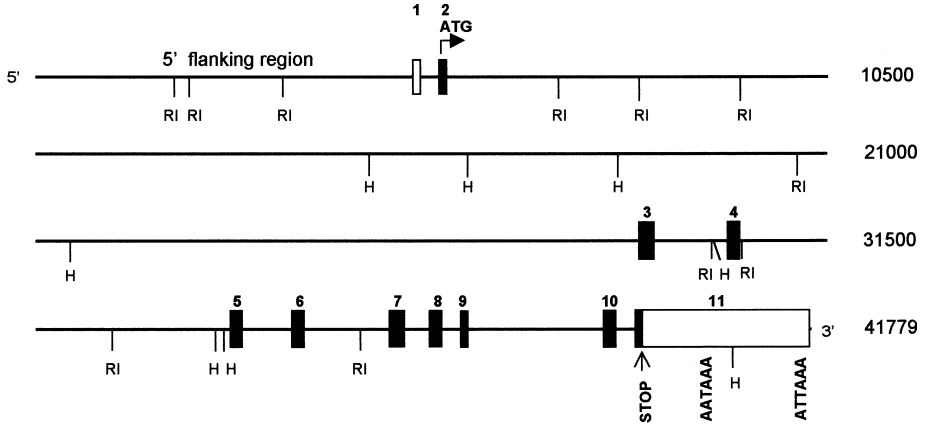

Fig. 2. Genomic organisation of the human chromosome 5 eRF1 gene. Untranslated exons are depicted as open and translated ones as closed boxes, lines represent introns. The translation start codon, the stop codon and the two polyadenylation sites are indicated. Restriction enzyme sites: RI, *Eco*RI; H, *Hin*dIII.

Table 1
Exon-intron organisation of the human chromosome 5 eRF1 gene

Exon number	Exon size (bp)	5′ Splice donor	Intron number	Intron size (bp)	3′ Splice acceptor
1	118	GGC CCT TAG gtaaggc	1	160	gctctag GGA GGA GGA
2	104	GCG GCC CGC GG gtgagcg	2	23582	tccacag C AAT GGC ACC
3	176	TAT AAC AAA G gtaatct	3	992	tttgcag TA CCT CCA AAT
4	140	CAT ACA GAG gtaagaa	4	3854	ttttcag GCT CTT ACA
5	139	AAG AAA CAC G gtaatac	5	613	gtcaaag GT AGA GGA GGT
6	191	TTT GAT CAG gtaagta	6	1160	gtgttag AGG TTA CAA
7	130	AAA TTA ATA G gtatcgg	7	274	atcgtag GA CGA TAC TTT
8	156	GGC ACA GAA G gtatgag	8	415	actacag AG GAG AAA ATT
9	65	GAC AAA GAG gtatgtg	9	1748	cctgtag ACC GGA CAG
10	148	GGA ATT GGA G gtgagta	10	281	tccccag GT ATC TTG CGG
11	2289				

the exon/intron organisation of the gene, encompassing 11 exons and 10 introns (Fig. 2). This genomic sequence was deposited in the GenBank database under accession number AF095901. As shown in Table 1, all exon-intron boundaries conform to consensus donor-acceptor sequences (gt/ag). The intron splice phasing is type 0 (the intron is located between codons) for intron 1, 4, 6 and 9, type 1 (the intron interrupts the first and second bases of the codon) for introns 3, 5, 7, 8 and 10, and type 2 (the intron interrupts the second and third bases of the codon) for intron 2. The intron size ranged from 160 to 23 582 bp, the largest being located between exons 2 and 3. The exons range from 65 bp for exon 9 to 2289 bp for exon 11 which contains a long 3′ UTR region. The first exon only contains untranslated sequence, exon 2 comprises the ATG initiation codon, and the TAG stop codon is located in exon 11. The two polyadenylation sites AATAAA and ATTAAA found in positions 2289 and 3637 respectively of the cDNA sequence, are also present in the 3′ UTR of exon 11. There are differences between the exonic and the cDNA sequences: an additional G is found in the untranslated exon 1 (between nucleotides 67 and 68 of the cDNA), two additional nucleotides T and C also occur in the 3′ UTR (between nucleotides 2382 and 2383 and nucleotides 2719 and 2720 respectively), and a G→T change (position 2655 of the cDNA) in the 3′ UTR gives a *Hin*dIII polymorphism site. That polymorphism generates two allelic fragments of 14

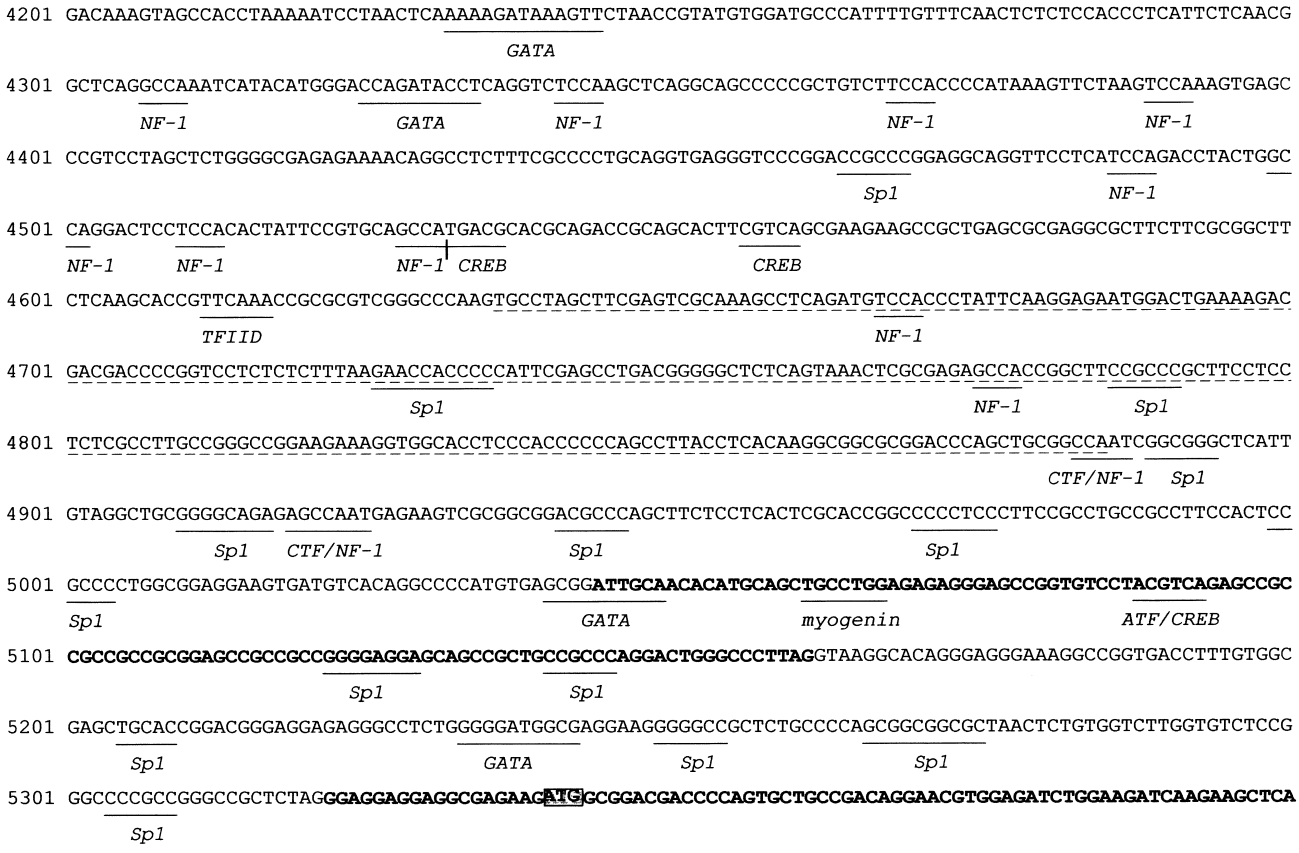

Fig. 3. Nucleotide sequence of portions of 5′ flanking chromosome 5 genomic DNA, the first exon, the first intron and proximal part of exon 2. Exon sequences are in boldface, the translation start site ATG is framed, the eukaryote Pol II putative promoter sequence is dot underlined and the predicted transcription factors binding sites are underlined with their names written below.

and 6.7 kb, the smaller being present on the cosmid DNA studied. The average allelic frequency in a normal population for the minor allele (6.7 kb) is about 4%.

The 5′-end flanking region (5040 bp long) was searched for putative promoters [25] and analysed for consensus factor binding motifs [26]. The sequence between positions 4356 and 5384 is GC-rich (65%), with an observed/expected ratio of CpG dinucleotides estimated to be 0.75 [27]. This region contains several predicted sequences that may function as Sp1 and Sp1-like protein binding sites and a putative eukaryote Pol II promoter region (4636–4885), but does not contain a canonical TATA or CCAAT box (Fig. 3). In addition, other potential regulatory elements are also present, including GATA, NF1, CTF and CREB.

4. Discussion

In this study, we report the cloning, sequencing and characterisation of the human eRF1 gene family, which contains four homologous sequences located on chromosomes 5, 6, 7 and X. The main point that emerges from this analysis is that only the gene present on chromosome 5, which corresponds to the SUP45L1 gene (as defined by the HUGO/GDB nomenclature committee), seems to be functional, whereas the three other sequences (SUP45L2 on chromosome 6, SUP45L3 on chromosome 7, SUP45L4 on chromosome X) look like non-functional copies of exonic sequences of the eRF1 gene.

These homologous sequences found on chromosomes 6, 7 and X accumulate base substitutions, small deletions and insertions, and nucleotide identity with HSCl1DHG7 cDNA sequence is only 95%, 95% and 92% respectively. Although each contains an inserted sequence and lacks a 3′-end A-stretch, these sequences may correspond to truncated processed pseudogenes. Such interspersed multigene families containing one or more functional genes and many processed pseudogenes are known, but there is no apparent evolutionary advantage in the addition of transcriptionally inactive gene copies. The significance of these three processed pseudogenes remains unknown.

As a consequence, only the gene located on chromosome 5 displays the organisation of a functional gene. Nucleotide identity between exonic sequences and cDNA HSCl1DHG7 is 99.9%. The gene spans approximately 40 kb and consists of 11 exons, the first untranslated, and 10 introns. A very long intron (23 582 bp) is located between exons 2 and 3. Upstream of the ATG, a 1100-bp sequence lacks repetitive elements and contains potential regulatory elements. Another region, 2900 bp long, downstream of exon 2 and corresponding to the proximal part of the large intron, also lacks repetitive elements and may play an important part in transcriptional regulation. A putative promoter sequence is predicted, without a TATA box, which is a major characteristic of housekeeping genes. This is consistent with the fact that eRF1 transcripts were detected by Northern blot analysis in a wide range of mammalian tissues [16,20], which suggests that the expression of eRF1 gene is essential for the function of most (or all) types of cells. Somewhat different results are found in the literature concerning the expression of eRF1 gene. Two transcripts of 3.6 and 2.2 kb were found in all of the tissues tested, the shorter of these in much lower abundance [16], whereas others [20] mentioned two major transcripts of about 4 and 2.8 kb and two minor transcripts of about 2 and 1.4 kb, with varying relative expression of the two major transcripts, depending on the tissue tested. Whatever the exact size of these two major transcripts, it is consistent with the two identified human cDNAs and the position of the two polyadenylation sites [16]. Alternative splicing regulation or alternative promoters have to be considered. Knowledge of the complete gene structure and of the 5′ flanking genomic sequence will allow the functional characterisation of the promoter region and of the transcription factors, and will therefore provide understanding of the mechanisms involved in expression at the transcriptional level.

There is a remarkably high degree of amino acid identity between members of the eRF1 protein family from yeast *Saccharomyces cerevisiae* (Sup45p) to human [4,16]. That sequence conservation corresponds to a functional conservation [16]: in fact, it has been demonstrated that either mammalian or amphibian eRF1 cDNA can replace the endogenous SUP45 gene in yeast with respect to its activity in translation termination and cell viability. Thus eRF1gene has successfully resisted a high selection pressure which is consistent with its indispensable function to the survival of the members of eukaryotic species during evolution.

Acknowledgements: We would like to thank Dr Phillip Jordan for his careful reading of the manuscript and Stephane Dreano for excellent technical assistance in sequencing.

References

[1] Caskey, C.T. (1980) Trends Biochem. Sci. 5, 234–237.
[2] Craigen, W.J., Lee, C.C. and Caskey, C.T. (1990) Mol. Microbiol. 4, 861–865.
[3] Konecki, D.S., Aune, K.C., Tate, W. and Caskey, C.T. (1977) J. Biol. Chem. 252, 4514–4520.
[4] Frolova, L., Le Goff, X., Rasmussen, H.H., Cheperegin, S., Drugeon, G., Kress, M., Arman, I., Haenni, A.L., Celis, J.E., Philippe, M., Justesen, J. and Kisselev, L. (1994) Nature 372, 701–703.
[5] Zhouravleva, G., Frolova, L., Le Goff, X., Le Guellec, R., Inge-Vechtomov, S., Kisselev, L. and Philippe, M. (1995) EMBO J. 14, 4065–4072.
[6] Frolova, L., Le Goff, X., Zhouravleva, G., Davydova, E., Philippe, M. and Kisselev, L. (1996) RNA 2, 334–341.
[7] Frolova, L.Y., Simonsen, J.L., Merkulova, T.I., Litvinov, D.Y., Martensen, P.M., Rechinsky, V.O., Camonis, J.H., Kisselev, L.L. and Justesen, J. (1998) Eur. J. Biochem. 256, 36–44.
[8] Merkulova, T.I., Frolova, L.Y., Lazar, M., Camonis, J. and Kisselev, L.L. (1999) FEBS Lett. 443, 41–47.
[9] Buckingham, R.H., Grentzmann, G. and Kisselev, L. (1997) Mol. Microbiol. 24, 449–456.
[10] Nakamura, Y. and Ito, K. (1998) Genes Cells 3, 265–278.
[11] Ito, K., Ebihara, K., Uno, M. and Nakamura, Y. (1996) Proc. Natl. Acad. Sci. USA 93, 5443–5448.
[12] Tate, W.P. and Brown, C.M. (1992) Biochemistry 31, 2443–2450.
[13] Tate, W.P., Poole, E.S., Dalphin, M.E., Major, L.L., Crawford, D.J. and Mannering, S.A. (1996) Biochimie 78, 945–952.
[14] Kisselev, L.L. and Frolova, L.Y. (1999) Biochemistry (Moscow) 64, 8–16.
[15] Tassan, J.P., Le Guellec, K., Kress, M., Faure, M., Camonis, J., Jacquet, M. and Philippe, M. (1993) Mol. Cell. Biol. 13, 2815–2821.
[16] Urbero, B., Eurwilaichitr, L., Stanfield, I., Tassan, J.P., Le Goff, X., Kress, M. and Tuite, M.F. (1997) Biochimie 79, 27–36.
[17] Grenett, H.E., Bounelis, P. and Fuller, G.M. (1992) Gene 110, 239–243.
[18] El Kahloun, A., Chauvel, B., Mauvieux, V., Dorval, I., Jouanolle, A.M., Gicquel, I., Le Gall, J.Y. and David, V. (1993) Hum. Mol. Genet. 2, 55–60.
[19] Grenett, H.E., Eipers, P.G., Kidd, V.J., Bounelis, P. and Fuller, G.M. (1992) Somat. Cell. Mol. Genet. 18, 97–102.

[20] Chauvel, B., Dorval, I., Fergelot, P., Pichon, L., Giffon, T., Gicquel, I., Le Gall, J.Y. and David, V. (1995) Immunogenetics 41, 380–382.
[21] Feinberg, A.P. and Vogelstein, B. (1983) Anal. Biochem. 132, 6–13.
[22] Southern, E.M. (1975) J. Mol. Biol. 98, 503–517.
[23] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[24] Jurka, J., Klonowski, P., Dagman, V. and Pelton, P. (1996) Comput. Chem. 20, 119–122.
[25] Prestridge, D.S. (1991) Cabios 7, 203–206.
[26] Prestridge, D.S. (1995) J. Mol. Biol. 249, 923–993.
[27] Milanesi, L. and Rogozin I.B. (1998) In: Guide to Human Genome Computing, 2nd edn. (Bishop, M.J., Ed.), pp 215–259, Academic Press, Cambridge.