The EMBO Journal vol.7 no.7 pp.2097 - 2105, 1988

The mouse protein synthesis initiation factor 4A gene family includes two related functional genes which are differentially expressed

Peter J.Nielsen and Hans Trachsel[1]

Max-Planck-Institut für Immunbiologie, Stübeweg 51, D-7800 Freiburg, FRG and [1]Institut für Biochemie und Molekularbiologie der Universität Bern, CH-3012 Bern, Switzerland

Communicated by G.Köhler

We have cloned and characterized a family of mouse genomic sequences hybridizing to mouse cDNA probes coding for eIF-4A, one of the protein synthesis initiation factors involved in the binding of mRNA to the ribosome. We estimate that there is a total of ~9–13 eIF-4A pseudogenes. We also found an eIF-4A intronless retroposon which, when compared to the cDNA, contains a single nucleotide difference. This possibly functional gene contains a mouse repetitive B1 element integrated in the promoter region. Furthermore, we have cloned two intron-containing eIF-4A genes (termed eIF-4AI and eIF-4AII). The eIF-4AII gene codes for a previously unknown form of eIF-4A. Northern blot hybridization with RNA from several mouse organs shows a variation in eIF-4AI expression within a factor of 7. In contrast, relative to liver, eIF-4AII expression is 20- to 30-times higher in brain and kidney, 10- to 17-fold higher in lung and heart, and is about equally abundant in liver, spleen and thymus. These data suggest that the relative efficiency of protein synthesis initiation for different mRNAs, as reflected by discrimination in messenger 5′-terminal cap recognition and binding to ribosomes, varies in different tissues.

Key words: B1 repetitive element/gene family/initiation factor 4A/protein synthesis/pseudogene

Introduction

Eukaryotic initiation factor 4A (eIF-4A), a single polypeptide of 45 kd, is one of the factors required for the binding of mRNA to 43S preinitiation complexes during initiation of protein synthesis (for reviews see Moldave, 1985; Kozak, 1983; Maitra *et al.*, 1982). Several activities have been ascribed to eIF-4A including the ATP-dependent binding to mRNA (Grifo *et al.*, 1982; Abramson *et al.*, 1987), specific cross-linking to the oxidized 5′ cap of mRNAs in the presence of ATP (Sonenberg, 1981; Edery *et al.*, 1983), and the RNA-dependent hydrolysis of ATP (Seal *et al.*, 1983; Grifo *et al.*, 1984). Since the binding of inosine-substituted reovirus mRNA to ribosomes is less dependent on ATP and the cap structure than native reovirus mRNA (Kozak, 1980; Morgan and Shatkin, 1980), it has been suggested that cap-binding factors, including eIF-4A, may be involved in ATP-dependent melting of secondary structure near the 5′ end of mRNA (Sonenberg, 1981; Abramson *et al.*, 1987). eIF-4A exists both as a component of a high mol. wt protein complex (called eIF-4F) involved in cap recognition (Tahara *et al.*, 1981; Grifo *et al.*, 1983) and as a single polypeptide required for mRNA binding to ribosomes (Daniels-McQueen *et al.*, 1983). To what extent the free and complex-associated forms of eIF-4A differ, physically and functionally, and whether they exchange, is not known. We present evidence here which shows that in the mouse there are two functional eIF-4A genes coding for proteins whose amino acid sequences are 91% identical. These two genes are differentially expressed in different tissues. We also show that there exists a family of eIF-4A-related genes in the mouse, most of which are non-functional pseudogenes.

Results

Southern blot analysis of eIF-4A sequences

In order to characterize genomic sequences coding for eIF-4A, total mouse liver DNA was digested with several different restriction endonucleases, fractionated by agarose-gel electrophoresis, transferred to a nylon membrane and hybridized with radiolabeled eIF-4A cDNA. The resulting autoradiogram (Figure 1a) shows numerous hybridizing bands for all restriction enzymes tested. This result is consistent with either a single eIF-4A gene containing many introns, or with the presence of many eIF-4A genes, or both. In order to better characterize eIF-4A sequences, similarly digested genomic DNA was probed with radiolabeled oligonucleotides. Figure 1b and c shows the resulting autoradio-

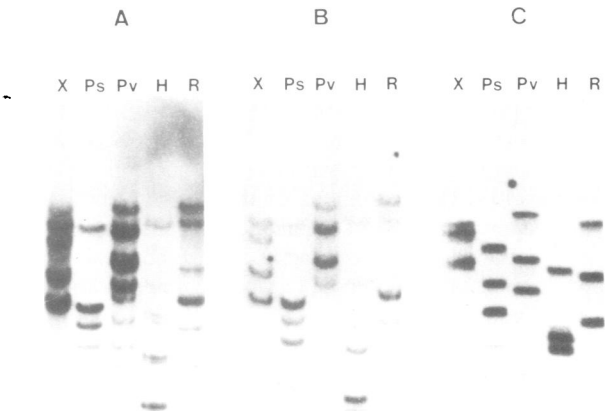

Fig. 1. Hybridization of genomic DNA with eIF-4A cDNA and oligonucleotide probes. Mouse genomic liver DNA (6 μg) digested separately with five different restriction endonucleases was electrophoresed, blotted onto a nylon membrane, hybridized and washed under high stringency conditions as described in Materials and methods. Radiolabeled probes were eIF-4A cDNA spanning nt −45 to +363 (**A**), oligonucleotide 4A5S spanning nt −42 to −23 (**B**), and oligonucleotide 4AC2 spanning nt 7 to 24 (**C**) [see Nielsen *et al.* (1985) for numbering]. Restriction endonuclease abbreviations are X: *Xba*I, Ps: *Pst*I, Pv: *Pvu*II, H: *Hin*dIII, and R: *Eco*RI. Autoradiography was for 9 days (A and B) or 3 days (C) with an intensifying screen.

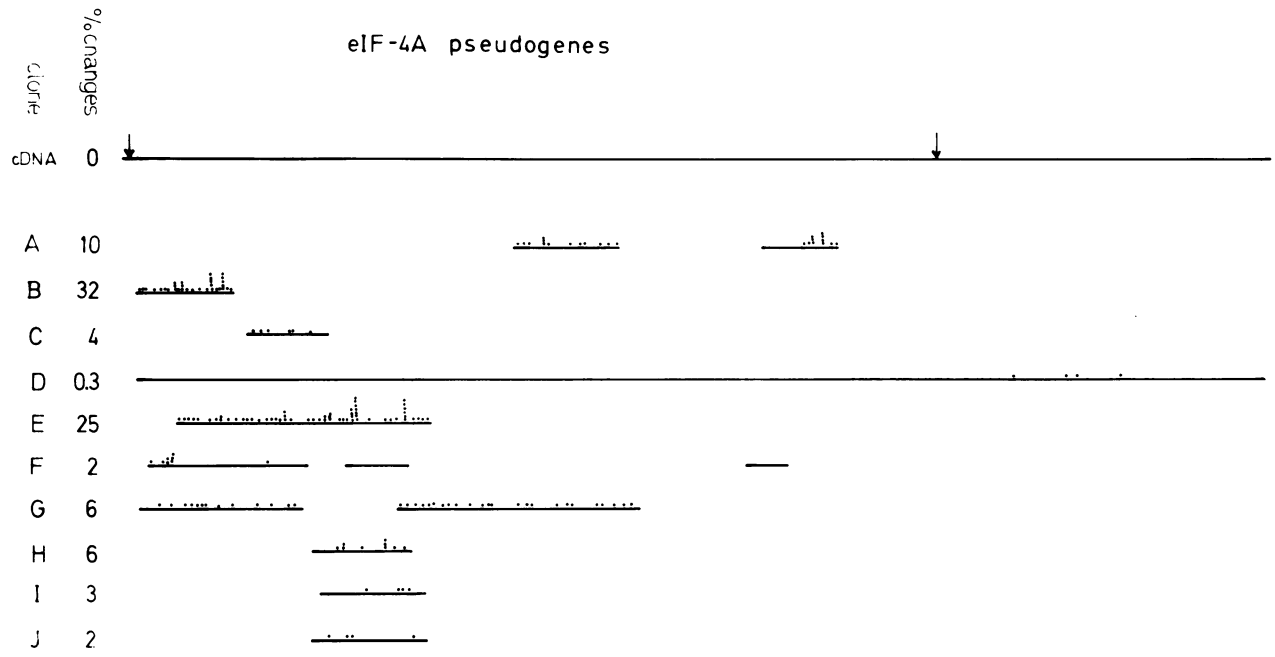

Fig. 2. eIF-4A pseudogenes. Sequenced regions of eIF-4A genes A–J are linearly depicted underneath eIF-4AI cDNA. The arrows on the cDNA line indicate the beginning and end of the coding region. Each dot above the processed gene lines corresponds either to a nucleotide exchange, deletion, or insertion. The percent to which each gene deviates (for that portion of the gene analyzed) from the cDNA sequence is listed next to the clone letter.

gram following stringent hybridization with oligonucleotides 4A5S (spanning nt −42 to −23, see Nielsen *et al.*, 1985 for numbering) and 4AC2 (spanning nt 7–24). The presence of several bands hybridizing with the oligonucleotide probes demonstrates unequivocally that there is more than one eIF-4A-related genomic sequence. The fact that this hybridization occurs under stringent conditions implies that some of the eIF-4A sequences are very similar to each other.

Cloning and analysis of eIF-4A genes

In order to isolate eIF-4A genes, ~1 × 10^6 recombinant phage from two mouse genomic libraries (generously provided by J.Epplen) were screened with mouse eIF-4A cDNA. Approximately 120 positive phages were isolated and most were further characterized in a variety of ways including hybridization to various oligonucleotides corresponding to different regions of the cDNA, restriction enzyme mapping and DNA sequencing. The analyzed clones could be grouped into three classes: non-functional retroposons or pseudogenes, a possibly functional retroposon and functional genes.

eIF-4A pseudogenes

A summary of data from the pseudogenes analyzed is shown in Figure 2. Based on restriction mapping and DNA sequence analysis, we have identified seven pseudogenes and estimate that there may be three to six more pseudogenes which we have cloned but not analyzed in detail. The fact that we have isolated at least two independent clones for each pseudogene family suggests that we have isolated all of the mouse eIF-4A genes. Portions of most of the pseudogenes were partially sequenced. The extent to which they deviated from the cDNA sequence ranged from 2 to 32% (see Figure 2). All pseudogenes contained numerous alterations which eliminate the ability to code for eIF-4A.

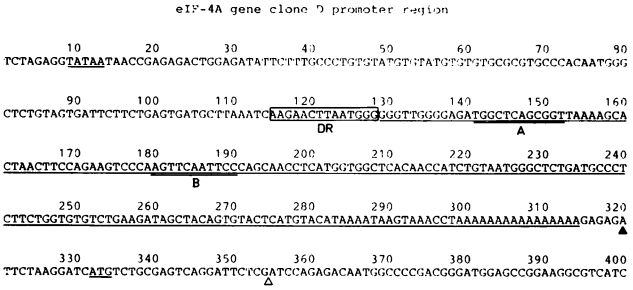

Fig. 3. eIF-4A gene D promoter region. The genomic DNA sequence immediately preceding the eIF-4A gene D is shown. The positions of the AUG (underlined, nt 332), the cap site (▲), and the first splice junction (△) found in the authentic eIF-4AI mRNA are indicated. The 14 bp direct repeat also found at the 3′ end of gene D is boxed. The mouse repetitive B1 element is underlined and the RNA polymerase III consensus motifs A and B are double underlined.

A possibly functional retroposon

One of the eIF-4A genes isolated (clone D) was sequenced in its entirety and found to contain a single nucleotide difference within the coding region when compared to eIF-4A cDNA. This G → A transition at position 486 would change the encoded amino acid from methionine to isoleucine. There were four changes in the 3′ non-coding region. This gene bears characteristics of a retroposon including the lack of introns, a poly(A) region located at the same position as the poly(A) in eIF-4A mRNA, and a 14 bp direct repeat 5′ and 3′ to the 4A sequence. In addition, this gene has a mouse B1 repetitive element inserted between the 5′ direct repeat and the beginning of the 4A sequence (Figure 3). The canonical box A and box B associated with DNA polymerase III promoters (Fowlkes and Shenk, 1980) are present. A TATA sequence is present 5′ to the direct

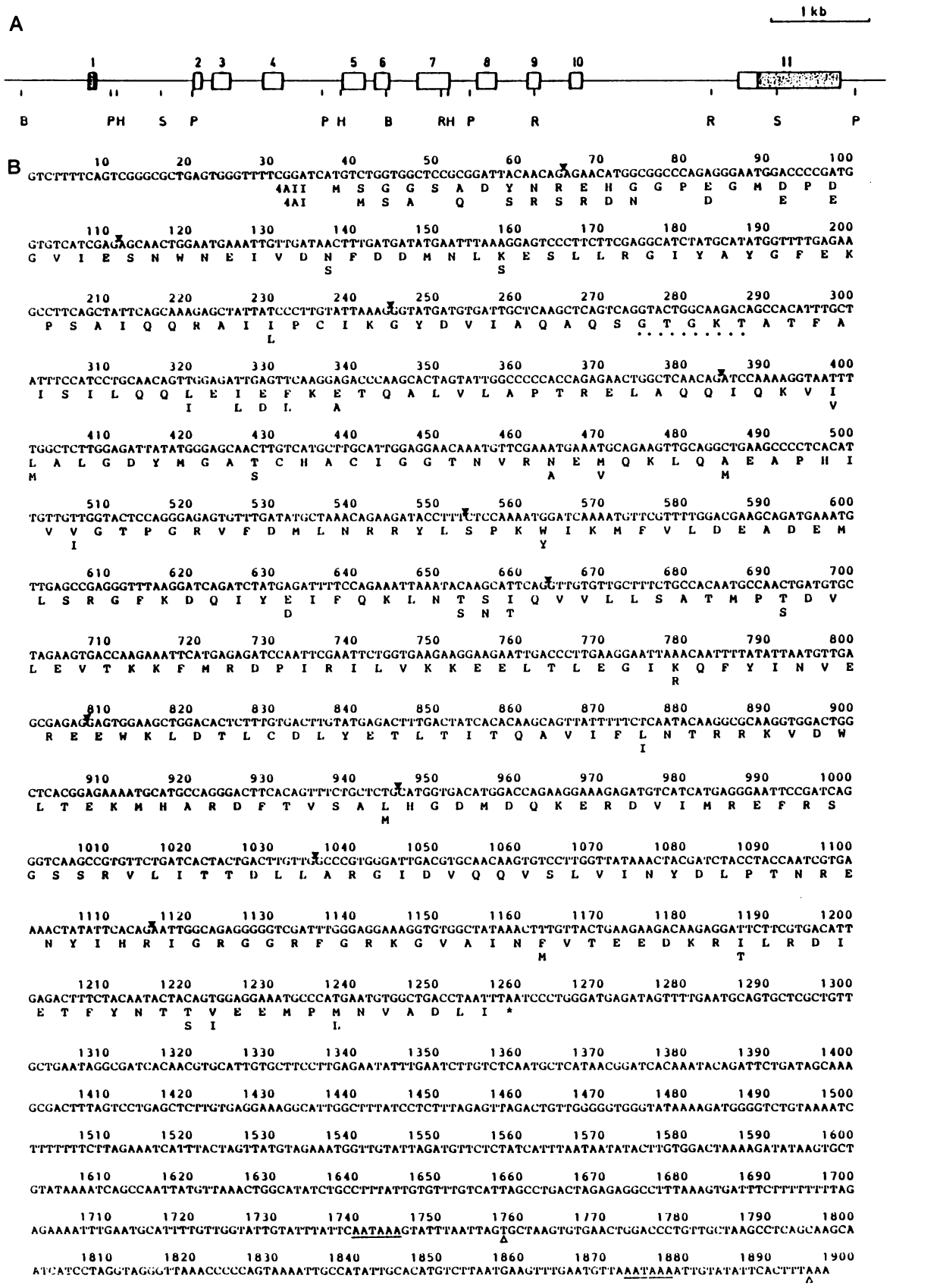

Fig. 4. Genomic organization and exon sequence of the eIF-4AII gene. (**A**) The exonic portions of the eIF-4AII gene are depicted as boxes. The stippled regions of exons I and II are the 5′ and 3′ untranslated sequences respectively. Restriction endonuclease recognition sites are indicated. B: *Bam*HI, H: *Hin*dIII, P: *Pst*I, R: *Eco*RI, S: *Sac*I. (**B**) The nucleotide sequence corresponding to eIF-4AII mRNA is shown. The cap site is nucleotide 1. The probable start AUG is at position 37. The predicted amino acid sequence of eIF-4AII as well as the difference between eIF-4AI and eIF-4AII are shown beneath the DNA sequence. The filled triangles above the sequence indicate the splice junctions, the open triangles beneath the sequence indicate the position of poly(A) addition for both forms of eIF-4AII messenger. The polyadenylation consensus signals for both messengers are underlined. The possible ATP binding domain of eIF-4A is underlined with dots.

Table I. Quantitative analysis of eIF-4A expression in various mouse organs[a]

Organ	eIF-4AI[b]	Long/short[c]	eIF-4AII[b]	eIF-4AII/eIF-4AI[d]
Liver	1	1.7	1	0.13
Spleen	2.9	1.7	1.6	0.06
Thymus	6.9	1.8	1.5	0.03
Heart	2.9	4.5	17.3	0.9
Lung	4.1	4.8	10.1	0.3
Kidney	2.9	8.3	32.2	1.4
Brain	2.0	11.6	20.3	1.3

[a]Autoradiograms of Northern blots containing total RNA isolated from various organs and hybridized with probes specific for eIF-4AI and eIF-4AII mRNA (see Materials and methods) were optically scanned and the peak areas integrated. The values presented are the average of two separate Northern gels.
[b]The abundance of eIF-4A mRNA (the sum of short and long forms) in each organ was corrected for small variations in RNA applied to each lane using the signal obtained after hybridization with 28S rRNA and then expressed relative to the abundance in liver.
[c]Refers to the relative abundance of the long and short forms of eIF-4AI mRNA. Although the eIF-4AI probe spans nt 1205–1402 of eIF-4AI cDNA and the short form mRNA ends at nt 1351, there is probably only a minor bias in hybridization signal strength in favor of the longer form since the portion of the probe specific for the long form is relatively short and the length of the radioactively labeled probe (labeled by random priming with hexamers) was short.
[d]The relative abundance of eIF-4AI mRNA compared with eIF-4AII mRNA in each organ was estimated by taking into account that the specific activities of both probes were approximately the same and that the length of eIF-4AI hybridizable sequences was 40% the length of the sequences hybridizing with the eIF-4AII probe. Roughly the same values (within a factor of 2) are obtained assuming that the amount of eIF-4AII mRNA in the brain is four times the abundance of eIF-4AI mRNA in liver (based on frequency of eIF-4A clones in the corresponding cDNA banks, P.Nielsen unpublished).

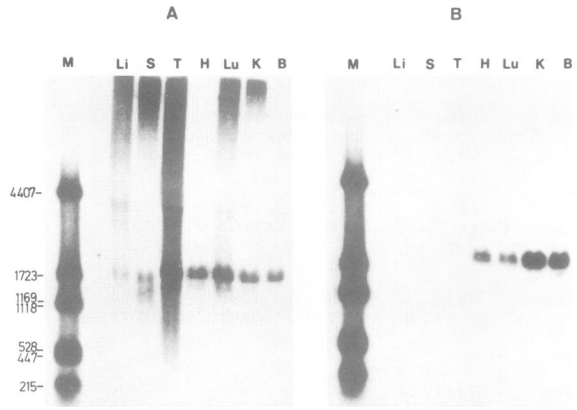

Fig. 5. Northern analysis of eIF-4AI and eIF-4AII mRNA expression in various mouse organs. Total cellular RNA was isolated from various mouse organs, and 18 μg from each organ was electrophoresed, transferred to a nylon membrane, hybridized and washed as described in Materials and methods. The marker lanes (M) were *Hin*dIII-digested and end-labeled PB40 (Nielsen *et al.*, 1985) and the sizes are indicated in basepairs. RNA samples were from liver (Li), spleen (S), thymus (T), heart (H), lung (Lu), kidney (K) and brain (B). (**A**) Hybridization was with eIF-4AI cDNA spanning the 3′ untranslated sequence from nt 1205 to nt 1402 (see Nielsen *et al.*, 1985 for numbering). Exposure was for 4 days with an intensifying screen. (**B**) Hybridization was with eIF-4AII genomic probe spanning the 3′ untranslated sequence from nt 1418 to the end of the mRNA (see Figure 4 for numbering). Exposure was for 27 h with an intensifying screen.

repeat. We do not know whether this gene is transcribed either by polymerase II or III. The fact that the coding sequence is virtually identical to eIF-4A cDNA suggests either that this gene is expressed, or that it represents a recent retroposition.

Two functional eIF-4A genes

Two eIF-4A genes were isolated which contain introns. Partial sequence analysis revealed that one gene codes for the eIF-4A cDNA previously isolated from a mouse liver cDNA bank (Nielsen *et al.*, 1985). This gene has also been isolated and completely sequenced by N.S.Reddy and A.J.Wahba (personal communication). The other intron-containing eIF-4A gene isolated is a second functional gene coding for a previously unknown form of eIF-4A which we will refer to as eIF-4AII. The entire gene was sequenced and found to consist of eleven exons (Figure 4A) distributed over ~8 kb. All splice junctions show the consensus intronic sequences of GT for the splice donor and AG for the splice acceptor. The nucleotide sequence within the coding region is 78% conserved between eIF-4AI and eIF-4AII. The predicted amino acid differences between eIF-4AI and eIF-4AII (9%) are distributed throughout the protein (see Figure 4B). Allowing for conservative changes, the identity between eIF-4AI and eIF-4AII increases to 98%. A hypothetical ATP-binding domain (Figure 4B, position 280–290), identified based on consensus sequences from nucleotide binding proteins (Walker *et al.*, 1982), remains unchanged. cDNA clones for eIF-4AII were also isolated and sequenced. They agree completely with the genomic sequence and also revealed that, as is the case for eIF-4AI, there are two mRNA forms for eIF-4AII which probably result from the use of alternative polyadenylation sites (see Figure 4B).

Tissue specificity of eIF-4AI and eIF-4AII expression

The expression of both eIF-4AI and eIF-4AII genes was examined in several mouse tissues using probes derived from the non-conserved 3′ untranslated regions of each cDNA. An example of the results obtained is shown in Figure 5. Autoradiograms from two separate Northern gels were optically scanned in order to estimate the relative abundance of eIF-4AI and eIF-4AII mRNA in each tissue. The results are summarized in Table I. When compared to liver, the amount of eIF-4AI mRNA in other tissues varies within a factor of 7 with the highest expression in thymus (Figure 5A). The expression of eIF-4AII, however, shows dramatic tissue variation (Figure 5B). Highest expression is seen in brain and kidney where it is 20- to 30-times the level in liver. Moderate expression is seen in lung and heart, and low expression in liver, thymus and spleen. (Recently we have observed high expression of eIF-4AII mRNA also in testis, H.Beckmann and P.Nielsen, not shown.) Since eIF-4A can be isolated together with eIF-4E (24 kd) and a 220 kd polypeptide in a complex called eIF-4F which is active in 5′-terminal cap recognition (see Discussion), we examined the tissue distribution of eIF-4E mRNA using a mouse cDNA probe (generously provided by J.Pelletier and N.Sonenberg). We find that the pattern of eIF-4E mRNA abundance is similar to that of eIF-4AII with a 40-fold higher expression in kidney relative to liver (data not shown).

Also striking is the fact that the relative amounts of short and long form eIF-4AI mRNA seem to correlate with the expression of eIF-4AII mRNA (see Table I). Thus, in liver,

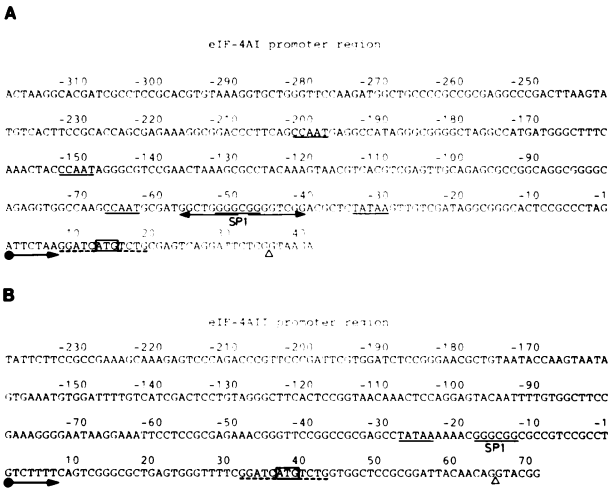

Fig. 6. eIF-4AI and eIF-4AII promoter regions. The genomic DNA sequence preceding the transcriptional start sites (filled circle with arrow) of eIF-4AI (**A**) and eIF-4AII (**B**) are shown. The respective initiating AUG for each gene is boxed, the first exon−intron junction is indicated (△) and the conserved sequence surrounding the start AUG is underlined with a dashed line. Possible promoter-related sequence motifs are underlined and include TATAA, CAATT and transcription factor SP1 consensus binding sequence GGGCGG (Gidoni *et al.*, 1984). An 8 nt mirror repeat centered on the SP1 consensus binding site in the eIF-4AI gene is indicated.

spleen and thymus, where eIF-4AII expression is low, the ratio of long to short form eIF-4AI mRNA is ~1−2. In heart and lung, which show moderate levels of eIF-4AII mRNA, the eIF-4AI mRNA long to short form ratio is 4−5. In kidney and brain, where eIF-4AII mRNA levels are high, the long form of eIF-4AI mRNA is 8−12 times more abundant than the short form. The resolution of the long and short forms of eIF-4AII mRNA in Figure 5B is not sufficient to allow an estimation of their abundance. The approximate estimation of the relative abundance of eIF-4AI and eIF-4AII in each tissue (Table I) suggests that with the possible exception of brain and kidney, eIF-4AI is in excess of eIF-4AII. Based on the frequency of positive cDNA clones in mouse brain cDNA banks, we estimate eIF-4AII mRNA to represent ~0.2 to 0.3% of total poly(A)$^+$ mRNA. In liver, we obtain an estimate of the abundance of eIF-4AI mRNA of ~0.06 to 0.1%.

The sequences of the promoter regions from eIF-4AI and eIF-4AII genes are shown in Figure 6. Although eIF-4AI is ubiquitously expressed and eIF-4AII shows strong differences in tissue expression, both promoters include TATA and SP1 consensus sequences. In addition, several CCAAT elements can be found upstream of the transcriptional start site of the eIF-4AI gene (Figure 6A).

The initiator AUG for eIF-4AII is most likely the first AUG following the cap site at nucleotide 37. This AUG is in frame with the subsequent open reading frame coding for eIF-4AII. Thus, the primary eIF-4AII translation product is predicted to contain 407 amino acids, one more than eIF-4AI. Strong amino acid similarity between eIF-4AI and eIF-4AII begins with the 13th amino acid in eIF-4AI and the 14th amino acid in eIF-4AII (Figure 4B). The 12 nucleotides centered on the putative start AUG are identical in both genes (underlined Figure 6), although the 5′-untranslated and first 30−40 coding nucleotides of both messengers show otherwise no obvious relatedness. The position of the start of transcription for eIF-4AI and eIF-4AII was determined by

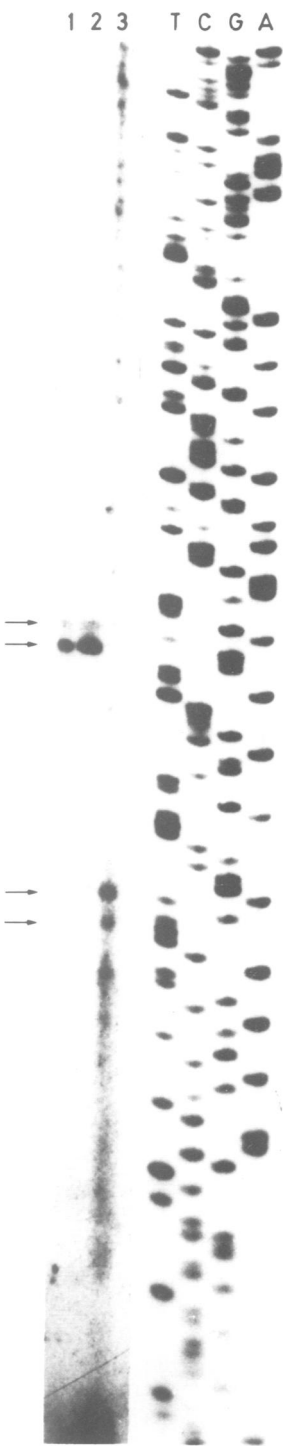

Fig. 7. Primer extension analysis of the 5′ end of eIF-4AI and eIF-4AII mRNA. The synthetic oligonucleotide complementary to eIF-4AII mRNA nucleotides 153 to 172 (see Figure 4) was end-labeled, gel isolated and annealed to 1.5 μg poly(A)-containing mRNA from the mouse preB cell line 70Z/3 (**lane 1**) or to 18 μg total RNA isolated from mouse brain (**lane 2**). After polymerization with reverse transcriptase, the products were electrophoresed in parallel with a dideoxy sequencing reaction of pseudogene clone D promoter region primed with an oligonucleotide spanning nt 310−328 as size markers (see Materials and methods for details). Alternatively, the synthetic oligonucleotide complementary to eIF-4AI mRNA nt 68−85 (see Nielsen *et al.*, 1985 for numbering) was end-labeled, gel isolated, and annealed to 5 μg of poly(A)-containing mRNA from the mouse T cell lymphoma cell line BW 5147 (**lane 3**). Exposure was for 5 days with an intensifying screen. The extension products are indicated with arrows.

primer extension experiments using radiolabeled oligonucleotides complementary to the corresponding mRNA. For eIF-4AI and eIF-4AII messengers, two products separated by 3 and 2 nt respectively were observed (Figure 7). The smaller of the two products is often observed in primer extension experiments and probably results from premature terminations following inhibition of the reverse transcriptase by the cap structure (Williams and Mason, 1985). The position of the start of transcription for both genes is indicated in Figure 6.

Discussion

We have shown here that the mouse genome contains numerous pseudogenes for eIF-4A which presumably are the result of retrotranspositions. Although most pseudogenes were only partially sequenced, they all contained sufficient mutations to render them unable to code for eIF-4A. One exception is the eIF-4A retroposon D which shows a single nucleotide difference when compared to the eIF-4AI coding region. This could be interpreted either as evidence for the functionality of this gene or for a relatively recent retrotransposition event. The promoter region of this gene is interrupted by a mouse B1 repetitive element which may or may not be involved in the expression of this gene. Further experiments are needed to resolve this.

Several examples exist in the literature of pseudogenes carrying integrated repetitive elements (Man *et al.*, 1987, and references therein). There are also several examples of retroposon-like genes whose sequences are identical or nearly identical to the cDNAs derived from their related functional genes, although it is not known whether they are expressed. Examples include rat and mouse ribosomal protein genes (Dudov and Perry, 1984; Kazumaki *et al.*, 1987), human dihydrofolate reductase (Chen *et al.*, 1982), human metallothionein II (Karin and Richards, 1982) and rat cytochrome *c* (Scarpulla, 1984). In contrast, for the chicken calmodulin (Stein *et al.*, 1983) and human phosphoglycerate kinase (McCarney and Thomas, 1987), there appear to be two functional genes, one with introns and one apparently processed. Interestingly, in both cases the intron-containing gene is ubiquitously expressed whereas the retroposed gene shows restricted tissue-specific expression.

Sequence comparison of all eIF-4A retroposons examined, including clone D, reveals that they are in all likelihood exclusively derived from the eIF-4AI gene. This suggests that the cells in which the retroposition occurred, presumably germ cell precursors or early embryonic cells, were expressing the eIF-4AI gene and not the eIF-4AII gene.

We show here that in mouse there are two functional intron-containing eIF-4A genes which we call eIF-4AI and eIF-4AII. The eIF-4AI gene codes for the previously isolated eIF-4A cDNA (Nielsen *et al.*, 1985) and the eIF-4AII gene codes for a heretofore unknown form of eIF-4A. The nucleotide sequence of the coding regions of these two genes shows an overall similarity of 78%. However, the changes which have occurred are primarily of the 'silent' type. A calculation of the nucleotide differences between eIF-4AI and eIF-4AII coding regions beginning from eIF-4AII amino acid 14 until the stop codon yields a total uncorrected silent (synonomous) site substitution frequency of 82% and a replacement (non-synonomous) site substitution frequency of 4% (see Materials and methods for substitution frequency calculation). Although a correction of the silent site frequency for 'multiple hits' and back mutations is very uncertain for such a high substitution frequency, extrapolation of the correction function described by Dayhoff (1978) gives a value of at least 400%. Assuming an average silent site substitution rate of 4.7×10^{-9} substitutions per site per year (Li *et al.*, 1985), the time of divergence of the two eIF-4A genes can be estimated to be at least 850 million years ago. This is considerably older than the invertebrate–vertebrate divergence 500–600 million years ago. The fact that the mouse eIF-4AI and eIF-4AII non-coding sequences show no recognizable similarity and that in yeast there are also two functional eIF-4A genes (P.Linder, personal communication) also speaks for the antiquity of the eIF-4A gene duplication. A comparison of the predicted amino acid sequences between the two mouse genes reveals 91% identity and allowing for conservative changes, the two forms of eIF-4A are 98% conserved. Clearly, the amino acid sequence of eIF-4A is highly conserved. Consistent with this, comparison of the mouse eIF-4A amino acid sequence with the predicted sequence of yeast eIF-4A recently cloned (P.Linder, personal communication) also shows a high degree of conservation (65% identity). The non-conserved amino acids between yeast and mouse eIF-4A are almost always the same amino acids which differ between mouse eIF-4AI and eIF-4AII.

It is worth noting that both eIF-4AI and eIF-4AII have very short 5′ untranslated sequences. Since eIF-4A itself may be involved in the melting of 5′ secondary structure, and thereby influence the accessibility of the initiator AUG for the 40S ribosomal subunit (Sonenberg, 1981), the lack of an extensive 5′ untranslated sequence in eIF-4A mRNA may render these messengers less dependent on eIF-4A. This hypothesis needs to be examined. Systematic substitutions of various 5′ untranslated sequences followed by characterization of translational efficiency and initiation factor dependence would be one way to do this.

Also conserved between eIF-4AI and eIF-4AII is the expression of mRNA with two alternative 3′ untranslated sequences. Although the 3′ untranslated regions of both genes show no obvious similarity, both have retained two polyadenylation signals separated by ~310 nt in eIF-4AI and 130 nt in eIF-4AII. Both polyadenylation sites are used in both genes since we have isolated and sequenced cDNAs corresponding to all four possible transcripts. Although we do not know the significance of the alternative 3′ ends, our data from Northern analysis of eIF-4AI expression in various mouse tissues (Figure 5A and unpublished) shows a reduced abundance of the short form eIF-4AI mRNA in those tissues which express eIF-4AII (see Table I). The significance of this correlation is unclear.

We previously published the isolation of eIF-4AI cDNAs containing different 5′ untranslated sequences (Nielsen *et al.*, 1985). We now believe the 5′ sequences reported to be unique to the eIF-4AI long form are an artifact resulting from cDNA synthesis. Although both eIF-4AI and eIF-4AII mRNAs can exist with two different 3′ untranslated sequences, we believe the 5′ non-translated sequences of both forms are the same. This conclusion is based on three arguments. First, primer extension analysis using labeled primers specific for eIF-4AI and eIF-4AII reproducibly give a single doublet for each primer suggesting a single transcriptional start site for each gene. The doublet could be interpreted as two start sites one or two nucleotides apart,

or due to premature termination during primer extension as a result of interference by the cap structure. Cap interference is often observed during primer extension (Williams and Mason, 1985). Second, two different synthetic oligonucleotides specific for the putative eIF-4AI long form-specific 5′ untranslated sequences failed to hybridize with mRNA in Northern blots and with genomic DNA in Southern blots although they did hybridize with the cDNA clone. This also supports the notion of the artifactual origin of these sequences. Third, within the putative long-form-specific 5′ sequence is a 20 nt inverted repeat of a nearby eIF-4AI 5′ sequence, an artifactual feature often observed at the 5′ ends of cDNAs.

The significance of two functional eIF-4A genes is not known. Our data suggest that eIF-4AI is moderately expressed with some variation in all mouse tissues and may represent a basal eIF-4A activity required for protein synthesis by all cells. It should be mentioned that in contrast to most of the other initiation factors whose abundance lies under one copy per ribosome, eIF-4A is rather unusual in that there are an estimated three copies of eIF-4A per ribosome (Duncan and Hershey, 1983). One explanation for the apparent excess eIF-4A could be that multiple copies of eIF-4A are required per initiation per messenger. These estimates from HeLa, rabbit reticulocyte and Erlich ascites cells are based on protein quantitation after two-dimensional electrophoresis and initiation factor purification yields. It is not clear whether this estimation includes both forms of eIF-4A since the level of expression of both genes is not known in these tissues. Two or three different forms of eIF-4A have been observed on immunoblots following two-dimensional electrophoresis of total HeLa lysates (Duncan and Hershey, 1983). Since the mature amino termini of neither eIF-4AI nor eIF-4AII is known (eIF-4AI appears to be N-terminally blocked; Nielsen *et al.*, 1985; W.Merrick, unpublished), it is difficult to predict their relative isoelectric points. However, the primary translation product of eIF-4AII mRNA should be more acidic than that of eIF-4AI. Whether these multiple forms of eIF-4A reflect post-translational modification or simultaneous eIF-4AI and eIF-4AII expression remains to be determined.

The tissue-associated variation in eIF-4AII expression shown here could mean that the cellular requirement for translational initiation varies between organs with the two forms of eIF-4A being functionally equivalent. Increased eIF-4A levels (either eIF-4AI or eIF-4AII) would reflect either a general increase in initiation rate or an increase in the translation of mRNAs particularly dependent on eIF-4A.

It is also conceivable, however, that eIF-4AI and eIF-4AII are functionally not interchangeable and that the presence of eIF-4AII in certain tissues provides a translational advantage for a specific class of messengers. It is known that a small proportion of the cellular eIF-4A is present as one of the three subunits of the cap recognition factor eIF-4F (Edery *et al.*, 1983; Grifo *et al.*, 1983). (The other two subunits of eIF-4F have mol. wts of 220 000 and 24 000, the latter polypeptide being the previously characterized cap binding protein or eIF-4E.) That eIF-4F plays a key role in translational regulation of mRNA expression is suggested by its low abundance (Duncan *et al.*, 1987; Hiremath *et al.*, 1985), and by its ability to discriminate between different mRNAs (Ray *et al.*, 1983; Sarkar *et al.*, 1984). Whether the complex-bound and free forms of eIF-4A exchange is not known.

Several groups have observed that the proteolytic fragments derived from the free eIF-4A and the 46 kd subunit of the cap binding factor eIF-4F are virtually identical with the exception of one or two fragments (Edery *et al.*, 1983; Merrick *et al.*, 1987). These differences could be explained by post-translational modifications of eIF-4A or by the presence of two different eIF-4A gene products. If these differences are due to differences between eIF-4AI and eIF-4AII, then we predict that eIF-4AII is the eIF-4F-associated form of eIF-4A since the amino acid sequence of the free form of eIF-4A purified from rabbit reticulocytes (W.Merrick, personal communication) corresponds to eIF-4AI. Such a model where, for the most part, eIF-4AII is associated with eIF-4F and eIF-4AI is free (schematized in Figure 8), implies that eIF-4AI and eIF-4AII are functionally not equivalent. If eIF-4AII is present primarily in the eIF-4F cap recognition complex, the pattern of eIF-4AII mRNA expression reflects regulation of eIF-4F abundance in different tissues. In support of this, the abundance of mRNA for eIF-4E, a second subunit of eIF-4F, shows tissue variation which is similar to that of eIF-4AII mRNA (data not shown).

Important experiments which should be done to help elucidate the functions of eIF-4AI and eIF-4AII include the *in vitro* comparison of the activities of eIF-4AI and eIF-4AII. These activities include ATP binding and hydrolysis, the ability to participate in the formation of the eIF-4F cap binding complex, possible differences in binding to different mRNAs, and differences in eIF-4A-dependent stimulation of protein synthesis associated with mRNAs containing varying degrees of secondary structure. The subcellular distribution of both forms of eIF-4A should also be studied, as well as the physiological signals involved in the regulation of eIF-4A gene expression.

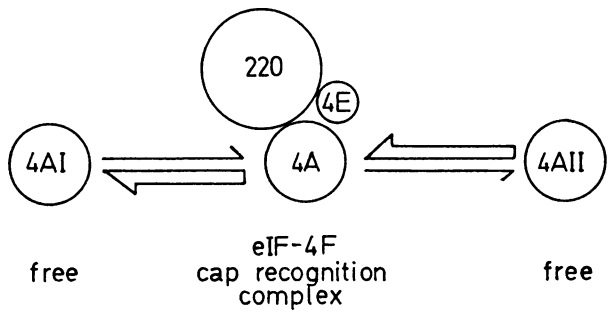

Fig. 8. Possible distribution of eIF-4AI and eIF-4AII between free and complex associated forms. The exchange of eIF-4AI and eIF-4AII between the free form and the cap recognition complex or eIF-4F-associated form is schematized. The large and small arrows depict the proposed preference of eIF-4AII for the cap recognition complex and eIF-4AI for the free form.

Materials and methods

Materials

Restriction endonucleases were obtained from New England Biolabs Inc., Anglican Biotechnology and IBN and were used as recommended by the suppliers. The Klenow fragment of *Escherichia coli* polymerase I was from Promega Biotech. Molony mouse mammary tumor virus-reverse transcriptase was from BRL, T4 ligase, polymerase and polynucleotide kinase were from New England Biolabs. Calf intestine phosphatase was from Boehringer Mannheim. [γ-^{32}P]ATP (>3000 Ci/mmol), [α-^{32}P]dATP (>3000 Ci/mmol) and [α-^{35}S]dATP (1200 Ci/mmol) were obtained from Amersham

Corp. Oligodeoxynucleotides were synthesized with an automated DNA synthesizer from Applied Biosystems Inc. and HPLC purified using an automated system described previously (Birsner *et al.*, 1987).

Bacterial strains, phage and plasmids

E.coli strains BMH 71-18 (Messing *et al.*, 1977), RR1 and XL-Blue (Stratagene) were used as hosts for plasmid cloning. All phage screening was done using the bacterial hosts Y1090 (Young and Davis, 1983) or Q358 (Karn *et al.*, 1980) using ^{32}P end-labeled oligonucleotides or randomly primed (Feinberg and Vogelstein, 1984) ^{32}P-labeled probes and standard methods. Two mouse genomic banks were screened (both generously provided by J.Epplen). The first was a *Sau*3A partial digestion cloned into the *Bam*HI site of the lambda Charon 4A. The second was an *Eco*RI partial digestion cloned into the *Eco*RI site of the lambda Charon 30. A mouse brain cDNA library cloned into the *Eco*RI site of λgt11 (generously provided by G.K.McMaster) was also screened.

Sequencing

All sequencing was performed by first subcloning into the vectors pTZ (Pharmacia) or Bluescript (Stratagene) followed by exonuclease III digestion (Henikoff, 1984). Single-stranded templates were produced by helper phage superinfection as recommended by the vector suppliers and sequenced by dideoxy-chain termination (Sanger *et al.*, 1977) as described (Biggin *et al.*, 1983).

Southern blot analysis

Genomic DNA (6 μg) was digested with restriction endonucleases as described in the figure legends and fractionated on 0.6% agarose gels. For hybridization with radiolabeled, gel-purified oligonucleotides (10^9 c.p.m./μg), gels were dried under vacuum and low heat, rehydrated in 6 × SSC (6 × SSC = 0.9 M sodium chloride, 0.9 M sodium citrate, pH 7.0) and hybridized for 4 h at T_m −5 [T_m = 4(G+C) + 2(A+T), (Suggs *et al.*, 1981)] in 6 × SSC containing 10 × Denhardt's solution [10 × Denhardt's solution = 0.2% Ficoll, 0.2% polyvinylpyrrolidone, 0.2% BSA (pentax fraction X)], 1 mM tetrasodium pyrophosphate, 15 mM disodium monobasic phosphate, 0.1% SDS, 50−100 μg/ml sheared, denatured salmon sperm DNA and 5 × 10^5 to 1 × 10^6 c.p.m./ml radiolabeled oligonucleotide. Gels were then washed in 6 × SSC for several hours at 37−42°C followed by a stringent wash for 1 min at the T_m and by a 30 min wash at 37−42°C, all in 6 × SSC. For hybridization with longer probes, DNA was first denatured, neutralized and transferred to nitrocellulose (Schleicher and Schuell or Millipore) or to Gene Screen (New England Nuclear), baked for 2 h at 80°C according to standard procedures. Blots were then prehybridized for 2 h at 64°C in the hybridization buffer described above, or at 42°C in the same buffer but containing 50% deionized formamide. Hybridization was at the same temperature as prehybridization for 8−14 h using randomly primed ^{32}P-labeled probes (Feinberg and Vogelstein, 1984) (10^8 c.p.m./μg and 5−10 × 10^5 c.p.m./ml). Blots were washed twice for 15 min with 2 × SSC at room temperature, then once for 30 min in 0.2 × SSC/0.1% SDS at 42°C or more stringently with 0.1 × SSC/0.1% SDS at 62°C for 30 min. The probes used for hybridization and the exposure of blots or gels was as indicated in the figure legends.

Northern blot analysis

Total RNA was extracted as described (Davis *et al.*, 1986) from quick frozen mouse organs by first hammering to a powder and then immediately adding ice-cold guanidinium isothiocyanate buffer. After homogenization, the sample was loaded on a guanidinium−CsCl cushion and centrifuged 21 h at 32 000 r.p.m. (174 000 *g*) in a Sorvall TH-641 rotor (DuPont). The RNA pellet was redissolved, phenol extracted, ethanol precipitated and resuspended in water. Total (18 μg) RNA from each organ (described in the figure legend) was denatured with glyoxal and separated on a 0.9% agarose gel, transferred to a membrane support and hybridized as previously described (Nielsen *et al.*, 1985). Hybridization probes are described in the figure legends.

Primer extension

Synthetic oligonucleotides to be used as primers were first end-labeled with [^{32}P]ATP using T4 polynucleotide kinase and then purified by excision from 12% urea−polyacrylamide gels. For the extension reaction, RNA (see figure legend for amount) and ~15 pmol (2 × 10^6 c.p.m.) oligonucleotide primer were heated at 90°C for 2 min and then annealed at 55°C for 15 min in 10 μl containing 10 mM Pipes (pH 6.4) and 0.4 M NaCl. The annealing reaction was cooled to room temperature and the reaction volume brought to 50 μl containing 50 mM Tris (pH 7.5), 10 mM DTT, 75 mM KCl, 3 mM MgCl, 50 μg/ml actinomycin D, 0.5 mM in all four dNTPs, 10 U RNasin (Promega), and 250 U cloned Moloney murine leukaemia virus reverse transcriptase. Polymerization was for 1 h at 42°C. The resulting products were fractionated on a 6% urea−polyacrylamide gel in parallel with radiolabeled DNA size markers. The resulting gel was dried and autoradiographed.

Calculation of substitution frequency

The calculation of nucleotide differences between the coding regions of eIF-4AI and eIF-4AII was done as described (Lomedico *et al.*, 1979). The total number of possible single step mutations was 3 × 1183 nucleotides = 3549. For both sequences, these changes were classified according to whether they would (replacement) or would not (silent) change the amino acid. The average of the total silent changes between both sequences was calculated and then divided by three to give the number of 'available' silent sites. The acutal number of observed silent changes was then divided by the 'available' silent site which resulted in a silent site nucleotide difference of 82%. A similar calculation for replacement site differences between eIF-4AI and eIF-4AII gave the value of 4%.

Acknowledgements

We would like to thank B.Groetschel and T.Mandel for valuable technical assistance, J.Epplen and G.McMaster for recombinant DNA libraries, J. Pelletier and N.Sonenberg for eIF-4E cDNA, and G.Köhler for discussions, support and critical reading of the manuscript. This work was supported in part by grant 3.233-0.85 from the Swiss National Science Foundation to H.T.

References

Abramson,R.D., Dever,T.E., Lawson,T.G., Ray,B.K., Thach,R.E. and Merrick,W.C. (1987) *J. Biol. Chem.*, **262**, 3826−3832.

Biggin,M.D., Gibson,T.J. and Hong,G.F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 3963−3965.

Birsner,U., Gilles,U., Nielsen,P.J. and McMaster,G.K. (1987) *J. Chromat.*, **402**, 381−386.

Chen,M.-J., Shimada,T., Moulton,A.D., Harrison,M. and Nienhuis,A.W. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 7435−7439.

Daniels-McQueen,S., Detjen,B.M., Grifo,J.A., Merrick,W.C. and Thach, R.E. (1983) *J. Biol. Chem.*, **258**, 7195−7199.

Davis,L.G., Dibner,M.D. and Battey,J.F. (1986) *Basic Methods in Molecular Biology*. Elsevier, Amsterdam.

Dayhoff,M.O. (1978) *Atlas of Protein Sequence and Structure 5, supplement 3*. National Biomedical Research Foundation, Washington, DC.

Dudov,K.P. and Perry,R.P. (1984) *Cell*, **37**, 457−468.

Duncan,R. and Hershey,J.W.B. (1983) *J. Biol. Chem.*, **258**, 7228−7235.

Duncan,R., Milburn,S.C. and Hershey,J.W.B. (1987) *J. Biol. Chem.*, **262**, 380−388.

Edery,I., Hümbelin,M., Darveau,A., Lee,K.A.W., Milburn,S., Hershey, J.W.B., Trachsel,H. and Sonenberg,N. (1983) *J. Biol. Chem.*, **258**, 11398−11403.

Feinberg,A.P. and Vogelstein,B. (1984) *Anal. Biochem.*, **137**, 266−267.

Fowlkes,D.M. and Shenk,T. (1980) *Cell*, **22**, 405−413.

Gidoni,D., Dynan,W.S. and Tjian,R. (1984) *Nature*, **312**, 409−413.

Grifo,J.A., Tahara,S.M., Leis,J.P., Morgan,M.A., Shatkin,A.J. and Merrick,W.C. (1982) *J. Biol. Chem.*, **257**, 5246−5252.

Grifo,J.A., Tahara,S.M., Morgan,M.A., Shatkin,A.J. and Merrick,W.C. (1983) *J. Biol. Chem.*, **258**, 5804−5810.

Grifo,J.A., Abramson,R.D., Satler,C.A. and Merrick,W.C. (1984) *J. Biol. Chem.*, **259**, 8648−8654.

Henikoff,S. (1984) *Gene*, **28**, 351−359.

Hiremath,L.S., Webb,N.R. and Rhoads,R.E. (1985) *J. Biol. Chem.*, **260**, 7843−7849.

Karin,M. and Richards,R.I. (1982) *Nature*, **299**, 797−802.

Karn,J., Brenner,S., Barnett,L. and Cesareni,G. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 5172−5176.

Kazumaki,T., Tanaka,T., Ishikawa,K. and Ogata,K. (1987) *Biochim. Biophys. Acta*, **909**, 99−106.

Kozak,M. (1980) *Cell*, **19**, 79−90.

Kozak,M. (1983) *Microbiol. Rev.*, **47**, 1−45.

Li,W.-H., Luo,C.-C. and Wu,C.-I. (1985) In MacIntyre,R.J. (ed.), *Molecular Evolutionary Genetics*. Plenum, New York, pp. 1−94.

Lomedico,P., Rosenthal,N., Efstratiadis,A., Gilbert,W., Kolodner,R. and Tizard,R. (1979) *Cell*, **18**, 545−558.

Maitra,U., Stringer,E.A. and Chaudhuri,A. (1982) *Annu. Rev. Biochem.*, **51**, 869−900.

Man,Y.M., Delius,H. and Leader,D.P. (1987) *Nucleic Acids Res.*, **15**, 3291–3304.
McCarney,J.R. and Thomas,K. (1987) *Nature*, **326**, 501–505.
Merrick,W.C., Abramson,R.D., Anthony,D.D.,Jr, Dever,T.E. and Caliendo,A.M. (1987) In Ilan,J. (ed.), *Translational Regulation of Gene Expression*. Plenum, New York, pp. 265–286.
Messing,J., Gronenborn,B., Müller-Hill,B. and Hofschneider,H.P. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 3642–3646.
Moldave,K. (1985) *Annu. Rev. Biochem.*, **54**, 1109–1149.
Morgan,M.A. and Shatkin,A.J. (1980) *Biochemistry*, **19**, 5960–5966.
Nielsen,P.J., McMaster,G.K. and Trachsel,H. (1985) *Nucleic Acids Res.*, **13**, 6867–6880.
Ray,B.K., Brendler,T.G., Adya,S., Daniels-McQueen,S., Miller,J.K., Hershey,J.W.B., Grifo,J.A., Merrick,W.C. and Thach,R.E. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 663–667.
Sanger,F., Nicklen,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Sarkar,G., Edery,I., Gallo,R. and Sonenberg,N. (1984) *Biochim. Biophys. Acta*, **783**, 122–129.
Scarpulla,R.C. (1984) *Mol. Cell. Biol.*, **4**, 2279–2288.
Seal,S.N., Schmidt,A. and Marcus,A. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 6562–6565.
Sonenberg,N. (1981) *Nucleic Acids Res.*, **9**, 1643–1656.
Stein,J.P., Munjaal,R.P., Lagace,L., Lai,E.C., O'Malley,B.W. and Means, A.R. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 6485–6489.
Suggs,S.V., Hirose,T., Miuake,T., Kawashime,E.H., Johnson,M.J., Itakura,K. and Wallace,R.B. (1981) In Brown,D.D. (ed.), *ICN–UCLA Symp. Dev. Biol.* Academic Press, New York, Vol. 23, pp. 683–693.
Tahara,S.M., Morgan,M.A. and Shatkin,A.J. (1981) *J. Biol. Chem.*, **256**, 7691–7694.
Walker,J.E., Saraste,M., Runswick,M.J. and Gay,N.J. (1982) *EMBO J.*, **1**, 945–951.
Williams,J.G. and Mason,P.J. (1985) In Hames,B.D. and Higgins,S.J. (eds), *Nucleic Acid Hybridisation: A Practical Approach*. IRL Press, Oxford, pp. 139–160.
Young,R.A. and Davis,R.W. (1983) *Science*, **222**, 778–782.

Received on March 25, 1988; revised on April 25, 1988