Highly conserved sequences in the 3' untranslated region of mRNAs coding for homologous proteins in distantly related species

David Yaffe, Uri Nudel, Yitzhak Mayer and Sara Neuman

Department of Cell Biology, The Weizmann Institute of Science, Rehovot 76100, Israel

Received 28 January 1985; Revised and Accepted 29 April 1985

ABSTRACT

Comparison of the nucleotide sequence of mRNAs coding for several vertebrate actins revealed a high degree of sequence homology in the 3' untranslated region (3' UTR) between those mRNAs coding for homologous (isotypic) actins in different organisms but not between mRNAs coding for very similar isoforms differing in their function or tissue specificity. A similar pattern of sequence conservation in the 3' UTR is also found in several other genes. Furthermore, while there is a great variation in the size of the 3' UTR of mRNAs coding for different proteins, mRNA coding for isotypic proteins in distantly related organisms often have 3' UTR of similar size.

The data suggest that the 3' UTR may play an important role in the regulation of expression of at least some genes at the transcriptional or posttranscriptional level.

INTRODUCTION

The rate of accumulation of neutral point mutations in the nucleotide sequence of DNA during evolution has been estimated to be in the order of 1% per one million years (my; 1, 2). A significant deviation from this value should indicate the involvement of a selective pressure to preserve or to change the sequence. The changes during evolution in the nucleotide sequences in the coding regions of genes, resulting in amino acid replacements (replacement site mutations), seem to be determined predominantly by the selective pressure operating on the protein level. Therefore, the rate of accumulation of these mutations varies greatly and may be different for each gene. In contrast, the accumulation of mutations which do not result in a change in the amino acids sequence (synonymous or silent site mutations) was assumed to be free of this constraint and was found to be similar for a great variety of genes and organisms. The accumulation of these changes has been used as a biological clock to study the evolutionary history of genes (2-5).

The function of the noncoding regions of genes, i.e., the 5' and 3' untranslated regions (5' UTR, 3' UTR), and of the introns is unknown. Sequence comparisons have indicated that in general the rate of nucleotide

replacements in these regions is more rapid than in the coding regions, with the exception of short sequences involved in RNA splicing, polyadenylation, initiation of transcription and translation which are highly conserved (reviewed in Ref. 6). However, Miyata et al. (7) have pointed out that while the rate of synonymous substitutions in the coding regions of a great variety of genes was similar (0.5% per 1 my) and the rate of the substitutions in the introns and the proximal half of the 3' UTR was similar to that of the synonymous site substitutions, the rate of substitutions in the 5' UTR and the distal half of the 3' UTR of the mRNA was approximately half of that rate, suggesting the existence of a selective pressure to conserve these regions.

Here we compare the nucleotide sequence of mRNAs coding for several actins and show a high degree of sequence homology in a great portion of the 3' UTR of mRNA coding for homologous actins in distantly related vertebrates, but not among mRNA coding for very similar isoforms differing in their tissue specificity. Reviewing published data on other genes reveals a high degree of sequence homology between the 3' UTRs of several other mRNAs coding for isotypic proteins in distantly related species. These observations suggest that the nucleotide sequence of the 3' UTR of many mRNAs fulfills an important function related to the specificity of expression of the genes or their transcripts.

MATERIALS AND METHODS

The homology in the 3' UTRs was calculated by aligning the sequences to obtain maximal homology and minimal single base deletions. The homology was calculated as 100x the ratio between matching nucleotides and total number of nucleotides in the sequence. Each gap was scored as one mismatch and therefore was counted as one nucleotide in the overall calculation of the length of DNA. In the calculation of the expected homology (Fig. 2), correction for multiple changes was done according to Miyata et al. (7). The homology in the silent sites was calculated as 100x the ratio between matching nucleotides in the silent sites and the total number of silent sites.

RESULTS AND DISCUSSION

Sequence conservation in homologous actin genes

The actins are a group of highly conserved proteins encoded by a multigene family, and they seem to be prevalent in all eukaryotes. Six different mammalian actins were identified by amino acid sequencing: skeletal muscle, heart muscle, 2 smooth muscle and the cytoplasmic beta and gamma actins (9, 10). The muscle actins are tissue specific and participate in the construction and

function of the contractile apparatus. Their synthesis is developmentally regulated (11-15). The cytoplasmic actins are involved in cytoskeleton structure, mitosis, cytoplasmic streaming and other cell movements, and appear to be present in all cells. The complete DNA sequence is available for the genes coding for the rat and chicken skeletal muscle actins (16, 17), and the rat and chicken beta actins (18, 19). Partial DNA sequences have been reported for the human β -actin mRNA (20, 21), the human cardiac actin gene (22), and the rat cardiac actin mRNA, including the entire 3' untranslated region (14). There are 4 amino acid replacements between skeletal muscle and cardiac muscle actins and 25 between these muscle actins and the nonmuscle actins. No differences in amino acid sequence have been found between the skeletal muscle actins of different mammals and chicken, between the mammalian and chicken cardiac actins, nor between the mammalian and chicken cytoplasmic β -actins. This high sequence conservation simplifies the identification and comparison of homologous genes in distantly related organisms.

β -actin genes. Since the estimated rate of accumulation of neutral point mutations during evolution is about 1% per my, the expected homology of sequences in 75 my (the divergence time between rat and human) in the absence of selective pressure is about 52% (after correction for multiple changes). The sequence homology between chicken and rat, which diverged ca. 300 my ago, in DNA regions which are not subjected to a selective pressure should be approximately the same as in unrelated sequences (<30%).

The 3' UTR of human, rat and chicken β -actin mRNAs are exceptionally large (592 bp, 664 bp, 591 bp, respectively). The over all homology in the 3' UTR between rat and human β -actin genes is 85%. The homology in the silent sites in the sequenced part of the coding region is 78%. Alignment of the sequence to obtain maximal homology (Fig. 1) reveals the existence in the 3' UTR of many long stretches with homology of 90-100%.

Even more striking is the existence of long stretches of highly conserved sequences between the mammalian and the avian 3' UTR, indicating very strong selective pressure to conserve these sequences (Fig. 1). The overall homology between rat and chicken β -actin mRNA in the 3' UTRs is 60%. The conserved sequences are spread over the entire 3' UTRs and the sequences in the 5' section of the 3' UTR are somewhat more conserved than in the distal half. These data are not in agreement with the observation made by Miyata et al. (2), that in general the 3' halves of the 3' UTR are more conserved than the silent sites in the coding region, while the 5' halves of the 3' UTR are not. This differs also from the homologies found in the skeletal muscle and cardiac

```

C: ACCGGACTGTTACCAACACCC-----ACACCCCTGTGATGAACAAAAACCATAAAT--GCCA-TAAACAGACGAGA--TTGCC-ATGGC-----
R: GC-GGACTGTTACTGAGCTGCGTTTACACCCCT--TCTTAGACAAAACCT--AACTTGGCGAGAAAAAAATGAGACATTGG-ATGGCTTTATT-
H: GC-GGACTATGACTTAGTTGCG-TTACACCCCT-TCTT-GACAAAACCT--AACTTGGCGAGAAAACAA--TGAGA--TTGGCATGGCTTTATT

C: -----TTTATTGTGTTTTCTTTTGGCGC--TTGACTCA-GGATTA AAAA AACTG
R: GTTTTTGTGTTTTGTGTTTTTAAATTTTTTTTAAAGGTTTTTTTTTGTGTTGTTTGGCCCTTTTGACTCAAGGATTTAAAAACTG
H: GTTTTTTGTGTTGTTT-----GGTTTTTTTTTT-----TTTGGC--TTGACTC-AGGATTTAAAA-CTG

C: GAATGG--TGAAGGTGTCAGCAGCAGTCTTTTTGAAACATGTTGGAGCGAACGCCCCAAAGTCTACAATGCATCTGAGGACTTTGATTGTACATT
R: GAACGGTGTGAAGGCGACCGCAGTTGGTTGGAGC--AAACAT-----CCCCAAAGTCTACAATGTGGCTGAGGACTTTGATTGTACATTG
H: GAACGGTG--AAGGTGACGACAGTCGGTTGGAGC--GAGCAT-----CCCCAAAGTTC-ACAATGTGGCCGAGGACTTTGATTGCACATTG

C: GTTC-----TTTTTAATAGTCATTCCAAATATTGTTATAATGCATTGTTACAGGAAGTTACT-CGCTCTGTG-----A
R: TTTTTGTGTTTGGTTTTTTAATAGTCAGTCCCAAGTATCCACGCGCATAGTGGTTACAGGAAGTCCCTCACCTCCCAAAGCCACCCC-CAACTC-CT
H: TT-----GTTTTTTAATAGTCATTCCAAATATGATGCAT--TGTTACAGGAAGTCCCTTGCCATCTAAAAGCCACCCCCTCTCTCT

C: AGGCAACAGCCCACTGGGAGGAGCGGTACCAATTAAGTGTGATGATAATGCTGTGTGTAATATTATGAACCAACAGGTGCTTTTGTATC
R: AAGGGAGGATGGCTGCATCCATGCCCTGAGTCCACACCGGAGGTGACAGCATTGCTTCTGTGTAATATTATGA--CTTGCAACATTTTTTAAATC
H: AAGGAGAATGCCCA--GTCCTCTCCC--AAGTCCACAGGGGAGGTGATAGCATTGCTTTCGTGTAATATTATGA--TGCAAA-ATTTTTTAA-TC

C: TTCCGCCCTTAAAAACAAACACACTTGATCCTTTTTGGTTTGTCAAGCAAGCGGGCTGTGTTCCCACTGA--TAGATGTG-----AA
R: TTCCGCCCTTAATAC-----TTCATTTTTGTTTTTAATTTCTGAATGGTCAGCCATTGCTGGCTGCCCC--TTTTTTTGTCCCCCAACTTGATGA
H: TTCCGCCCTTAATAC-----TTTTTTATTTGTTTATTT-TGAATGATGAGCC-TTCGTGCCCCCTTCCCCCTTTTGT-CCCCCAACTTGATGATG

C: TG--AAGGCTTT-ACAGTCCCCACAGTCTAGGAGTAAAGTCCAGTAT-GTGGGGAGGAGGGCTACCTGTACACTGACTTAAGACCAAGTTGAATAA
R: TG--AAGGCTTT-GGTCTCCCTGGGAGT--GGTTGAGGTGTTGAGGC-GCCAG--GGCTGGCCTGTACACTGACGTGAGACC-GTTTAATAA
H: TATGAAGCTTTTGGTCTCCCTGGGAGT-----GG-CTGGAGGCAGCCAG--GGCTTACCTGTACACTGACTTGAGACC-AGTTGAATAA

C: AAGTGCACACAATAGAGGCTGACTGGTGTGTTGTTTTATTCTGTGCTGCGCTGCTTGGCGTGGTAGTGTCTCATCTAGCCTGCCAGCCTGTGT
R: AAGTGCACACCTTACAAACAAGTTGTGGCTGTGCTGCTTCTACTGGGTGTGGGAGCAGGCTGGGTGGGTGTGNACTCCACGTGGGGAAGGGCAAT
H: AAGTGCACACCTTA

C: GGGTCAGCTATCTGCATGGGCTGCGTGTGCTGTGCTGTGCTGCAGAGTTGGATAAACCGGGATGATATTCAGCA
R: TTAGTGGGCTGGCCTTGCTGATAGCTAGTGGGAGGCTAAAGGATCATGATCTTTATCAGGTCTCATAAATACC

```

Figure 1. Comparison of the 3' UTR of the mRNAs coding for rat, chicken and human β -actins. The alignment was made to obtain maximal homology. Identical nucleotides are marked by vertical lines. The polyadenylation signal AATAAA is marked by a horizontal line above the sequence. The sequences of the chicken (C) and rat (R) 3UTR were obtained from genomic clones (18, 19) and the human (h) from cDNA clones (20, 21). The comparison between the rat and human sequence has also been described recently by Ponte et al. (21).

actin mRNA described later. The region of high sequence homology ends 12 nucleotides after the polyadenylation signal, i.e., it is confined to the sequence which is preserved in the mature mRNA. The sequence homology between

rat and chicken β -actin genes in the silent positions of the coding region is 71%.

Cardiac actin genes. The comparison of the sequence of the rat and human cardiac actin mRNA has recently been reported elsewhere (14, 23) and is shown schematically in Fig. 2. The overall homology between the 3' UTR of rat and human cardiac actin mRNA is 84%. A portion greater than a half of the 3' UTR distal to the coding region shares more than 92% homology and another region proximal to the coding region shows 85% homology. Interestingly, a stretch of 17 nucleotides in the latter region shares 85% homology also with the rat skeletal muscle actin mRNA. The sequence homology in the silent sites in the sequenced part of the coding region is 77% (14).

Considerable sequence homology in the terminal region of the 3' UTR exists also between human and xenopus cardiac actin mRNAs which have diverged 300-400 my ago (8).

Skeletal muscle actin genes. We have previously reported that a cDNA probe containing the 3' UTR of the rat skeletal muscle actin mRNA hybridizes specifically with skeletal muscle mRNA from rat, rabbit and dog, suggesting a considerable sequence conservation in this region of the mRNA (24). As reported by Ordahl and Cooper (25) and as shown in Fig. 2, the 3' half of the 3' UTR of chicken and rat skeletal muscle actin mRNA share over 75% homology, including several short stretches (9-16 nucleotides long) of 100% homology.

Figure 2.

B

Figure 2. A schematic presentation of sequence homologies in the 3' UTR between evolutionary related genes. A. mRNAs coding for actins; B. mRNA coding for other proteins. Thin dashed lines connect regions that show sequence homology and the percentage of homology between the regions is indicated (only when the homology is >30%). The arrows indicate the polyadenylation signals; the polyA addition sites are 20-30 nucleotides downstream from the polyadenylation signal. Bars describing data derived from cDNA clones are closed in the

3' end, while bars describing data from genomic clones are open in the 3' end.

The numbers to the left indicate the expected homology between neutral sequences which had been diverged at the time of divergence between each pair of genes. The rate of divergence of neutral sequences is about 1% per million years (1, 2, 43). The calculation of expected homology for neutral sequences includes a correction for multiple changes. The estimated divergence times between the compared genes are as follows, man/rodent 75 my (59); man/rabbit 46 my (2); mouse/rabbit 75 my (2); mammals/birds 300 my (59, 60). The divergence times between the various human β -type globins were taken from Efstratiadis et al. (5) and are as follows, β/γ , ϵ , 200 my; γ/ϵ , 100 my; β/δ , 40 my.

The numbers in parentheses indicate the sources of the data.

The sequence homology in the silent sites in the coding region is 72%, which is also remarkably high. A high degree of sequence homology in the same region between the 3' UTR of rat and human skeletal muscle actin mRNA has been described recently by Gunning et al. (8) and Gillespie et al. (64).

Very limited sequence homology between 3' UTR of mRNA coding for nonhomologous actins

The skeletal muscle and cardiac actins differ by only 4 amino acids (1.1% divergence). In newborn rats and mice the 2 genes are co-expressed in the heart and skeletal muscle. However, during development, the amount of mRNA coding for the skeletal muscle actin decreases rapidly in the heart (14), while in the skeletal muscle the amount of the cardiac actin mRNA decreases rapidly (13, 14). As has been shown elsewhere (14) and in Fig. 2, very limited sequence homology is found between the 3' UTR of the skeletal muscle and cardiac actin mRNAs. This consists of 2 short sequences. One is 25 nucleotides long and is located 17 nucleotides downstream from the translation termination codon (85% homology). This sequence is also partially conserved in chicken skeletal muscle actin mRNA. The second sequence is 10 nucleotides long and includes the polyadenylation signal (75% homology). The overall sequence homology between the 3' UTR of rat cardiac and skeletal muscle actin mRNAs is lower than between the silent sites in the coding region (codons 282-374) of these genes (38% and 58%, respectively).

No obvious sequence homology is observed between the 3' UTR of the β -actin mRNA and the muscle actin mRNAs (not shown).

Strong conservation of sequences in 3' untranslated regions of mRNAs coding for isotypic proteins is not confined to the actin gene family

Surveying and analyzing the recently published data reveals that extensive sequence homology in the 3' UTR between mRNA coding for isotypic proteins in different organisms is quite common. The extent of sequence conservation and

3'

M : AAATTTTC-AAAGCATCC-GTATACCGAA
|||||
H : AAATTTTGGAAAGCATCCTGTGTGCTGA

the regions which are most conserved vary greatly. Some of these data are presented schematically in Fig. 2b. A striking example for strong sequence conservation is the mRNAs coding for mouse and human β -nerve growth factor (NGF). Alignment of the published nucleotide sequences of the human and mouse NGF mRNA (26) to obtain maximal homology shows sequence conservation in the 3' UTR which is even greater than the conservation between the rat/human β -actin genes. The overall sequence homology is 86.5%. In the central part (nucleotides 22-143) the homology is 93.5%, with long sequences sharing 100% homology (Fig. 3).

the mouse and human c-fos mRNA is discussed below.

Table 1: Length of 3' untranslated region of mRNAs coding for isotypic proteins (nucleotides)

<u>β-actin</u>		<u>Atrial natriuretic polypeptide</u>		<u>β-globin</u>	
Chicken	591			Human	132
Human	592	Rat	264	Rabbit	92
Rat	664	Human	234	Mouse	131
<u>Skeletal actin</u>		<u>α-tubulin</u>		<u>α-globin</u>	
Rat	238	Human (bal)	216	Human	110
Human	251	Rat (pilaTl)	195	Rabbit	87
Chicken	271			Mouse	100
				Chicken	96
<u>Cardiac actin</u>		<u>Nerve growth factor</u>		<u>Growth hormone</u>	
Rat	163	Mouse	154	Rat	102
Human	181	Human	154	Human	104
				Bovine	102
<u>c-fos</u>		<u>Prolactin</u>			
Human	813	Rat	148		
Mouse	823	Human	145		
<u>Fibronectin</u>		<u>Whey acidic protein</u>			
Human	689	Mouse	130		
Bovine	686	Rat	132		
Rat	688?				

Sources of data: Human β -actin, Ref. 21; rat and human bovine fibronectin, Ref. 58 and R. Hynes, personal communication; atrial natriuretic polypeptide, Ref. 57; whey protein, Ref. 61; bovine growth hormone, Refs. 30 and 63. The sources for the data on the other mRNAs are indicated in Fig. 2.

homology between the 3' UTR of avian and mammalian isotypic globin genes. When genes coding for isotypic globins in the different mammals are compared, the overall homology and in particular the homology in certain parts of the 3' UTR of these genes is usually greater than the homology in silent sites of the same genes, and certainly greater than expected of neutral sequences. On the other hand, it seems that the rate of divergence of the 3' UTR between the nonhomologous human β -type globin genes, β -, γ -, δ - and ϵ -globin genes (which differ in their expression during development) is somewhat higher than the rate of divergence of silent sites in the same genes.

Conservation of the size of the 3' untranslated region

Early studies on the actin isoproteins indicated that the mammalian and

chicken cytoplasmic actins are encoded by mRNAs which are considerably larger than the mRNA coding for the skeletal muscle actin (11, 28). Sequence analysis showed that this was mainly due to the length of the 3' UTRs; the β -actin mRNA has a remarkably large 3' untranslated region (16-19). Table 1 presents the available data on the sizes of the 3' UTR of mRNA coding for actins and several randomly selected mRNAs coding for other proteins. It can be seen that (a) the mRNA coding for the cardiac, skeletal muscle and β -actins vary greatly in size, the 3' UTR of the β -actin being >2.5 times larger than the 3' UTR of muscle actin mRNA; (b) mRNA coding for homologous actins in rat, human and chicken have 3' UTR of similar size; and (c) conservation of the size of the 3' UTR of homologous mRNAs seems to exist also in other genes.

This apparent conservation of the length is probably not only a consequence of sequence conservation since it exists also between homologous genes showing a considerable degree of sequence divergence, including deletions or insertions, in the 3' UTR. For example, the length of the 3' UTR is similar for mammalian and chicken skeletal muscle actin mRNA, though there is little homology in sequence in the first 140 nucleotides of the 3' UTR between the mammalian and chicken skeletal muscle actin mRNA. The alignment of the sequence to give maximal homology suggests a number of deletions or insertions in the 3' UTR of the rat and chicken β -actin mRNAs. The conservation of size of the 3' UTR may reflect constraints to maintain structures which are involved in some way in the specificity of expression of the different mRNA.

Possible role of the conserved sequence

The great homology in the nucleotide sequence of the 3' UTR of β -actin mRNAs from distantly related vertebrates like rat and chicken, and the sequence conservation between chicken and rat skeletal muscle actin and rat and human cardiac actin genes indicates a very strong evolutionary constraint to preserve these sequences. It is clear that this situation is not confined to the actin gene family since a comparable gene-specific sequence conservation is found also in the 3' UTRs of several other genes. The extent of sequence conservation seems to vary, suggesting that the constraint to preserve the nucleotide sequence in this region of the gene is stronger in some genes than in others. Moreover, a great variation is observed in the distribution of the conserved sequences along the 3' UTR. In the skeletal and cardiac muscle actin genes the 3' half of the 3' UTR region is much more conserved than the 5' half. In the β -actin genes the homology is distributed along the entire 3' UTR, the 5' half being slightly more conserved. Moreover, the rate of sequence divergence in the 3' UTR may change during evolution. As pointed out by Martin et al. (29), the rate of divergence of the nucleotide

sequence in the 3' UTR of globin mRNA is significantly lower in primates than in other vertebrates. Another example is the sequence homology in the 3' UTR of mRNAs coding for the pre- α subunit of the pituitary glycoprotein hormone (30). There is 82% sequence homology between the 3' UTR of mRNA coding for rat and mouse pre- α subunit of the pituitary glycoprotein hormone and 72% sequence homology between the 3' UTR of the mRNA coding for this hormone in the human and bovine. However, there is very little, if any, homology between the 3' UTR of the mRNA coding for this protein in the rodents and in the human or bovine.

The fact that in many cases the sequence conservation in the 3' UTR is higher than in the introns and in the silent sites in the coding region makes it very unlikely that this conservation is a result of a strong selective pressure to preserve the coding region. Furthermore, comparison of the nucleotide sequence of the mRNA coding for the mouse and rat whey protein (a milk constituent) revealed that the conservation of the nucleotide sequence in the 3' UTR (91% homology) is even greater than the conservation of the nucleotide sequence in the replacement sites of the region coding for the c-terminal half of the mature protein (81-83% homology; 31). This suggests that the constraint to conserve the nucleotide sequence of the 3' UTR was stronger than the constraint to conserve the amino acids sequence.

The great differences in the nucleotide sequence and size of the 3' untranslated region between non-homologous genes belonging to the same or to different gene families suggest that the conservation of the sequence and perhaps the size of the 3' UTR of mRNA coding for isotypic proteins is not related to the ubiquitous mechanisms of transcription, processing or translation, but rather to some as yet unidentified function associated with the specificity of expression of the genes at the transcriptional or posttranscriptional levels.

An increasing number of observations suggest that primary gene transcripts may contain information for posttranscriptional regulation of expression of the gene. Perhaps the most striking example is the production by alternative splicing of the primary transcripts of different mRNAs with different coding information. E.g., in *Drosophila*, a single myosin heavy chain gene produces by alternative splicing mRNA coding for 3 different developmentally regulated and probably tissue-specific myosin heavy chain isoforms (32). Alternative RNA processing of calcitonin gene transcripts in the brain and in the parathyroid generates mRNA coding for different polypeptides (33). These and many other data accumulated recently suggest that primary gene transcripts are

endowed with information for different splicing in different cellular environments.

Since the evolution of nucleotide sequences of coding regions of the genes is dominated primarily by the selective pressure operating on the protein level, it is conceivable that sequences or structures necessary for the regulation of expression of the gene or mRNA will be conserved mainly in non-coding regions of the gene. Recently it was found that introns in genes coding for the constant region of immunoglobulins contain enhancer-like sequences associated with the tissue-specific activation of the rearranged immunoglobulin genes (34-37). Unlike introns, which are removed during the processing of the primary transcript, the 5' and 3' untranslated regions are conserved in the mature mRNA. Therefore, they can serve as sites for encoding information associated with posttranscriptional regulation of expression.

The function of these regions is still unknown. Experiments in which most of the 3' UTR of β -globin mRNA was eliminated failed to abolish the translational activity (38). However, there are other observations suggesting the involvement of the 3' UTR in regulation of translation. For example, in the human, there are two α -globin genes, coding for identical proteins (α_1 and α_2). The entire nucleotide sequences of the two mRNAs are identical except for the 3' UTR in which the divergence is 17.2%. However, there is a 3-fold difference in the translation efficiency of the two mRNAs (39). Since the nucleotide sequences of the 2 mRNAs differ only in the 3' UTR, these results suggest that the difference in translation efficiency of these mRNAs is determined by the sequence of the 3' UTR.

Another example, suggesting the possible involvement of the 3' UTR in the regulation of gene expression at the translational level is provided by the cellular gene coding for the c-fos protein, homologous to the FBJ murine osteosarcoma oncogene (v-fos). Transfection of 208F rat cells with recombinant v-fos/mouse c-fos chimeric genes indicated that the presence of the viral enhancer was not sufficient to transform cells transfected by chimeric genes. Rather, replacement of a DNA sequence 3' to the coding region with v-fos 3' noncoding sequence or with SV40 early polyadenylation signal conferred transformability on the chimeric genes. Similar results were obtained with v-fos/human c-fos chimeric genes. Analysis of the RNA and proteins produced in cells transfected with the various chimeric genes brought to the conclusion that some interaction between the 3' UTR and other parts of the c-fos mRNA inhibited the synthesis of the fos protein in the transfected cells, indicating the involvement of the 3' UTR in translational control of synthesis of the c-fos protein in normal cells (40). Comparison of the nucleotide sequences of

the mouse (41) and human (42) c-fos genes shows a very strong sequence homology in the 3' UTR (Fig. 2b). The overall sequence homology in the very long 3' UTR is 92%. This includes an uninterrupted sequence of 428 nucleotides with 96% homology, which terminates 27 nucleotides 3' to the AATAAA polyadenylation signal (i.e., in the region of the polyA addition site). The fact that the conserved sequence does not extend beyond the polyadenylation site suggests that the selective pressure to preserve the sequence is operating at the mRNA level. The possibility that the exceptionally high sequence homology in the 3' UTR between the mouse and human c-fos mRNA is due to recent exchange of DNA sequences between mouse and human c-fos genes mediated by recombination with the viral gene seems very unlikely, since the 3' UTR of the v-fos gene does not contain sequences homologous to the last 350 nucleotides of the c-fos mRNA 3' UTR.

As described above, the comparison between the chicken and rat β -actin genes (the only pair of homologous actin genes for which data were available for the comparison of the 3' flanking DNA) showed also that the high homology was confined to the sequences which are preserved in the mature mRNA. However, in some other genes, such as human and mouse β -globin (Fig. 2b), the available data indicated that the conservation extends into the region 3' to the polyadenylation site. More data is needed to determine the significance of these observations.

The great variability in the location of the conserved sequence and in the extent of conservation may hint at the involvement of 3' UTR sequences in several different functions which may vary from gene to gene. In some genes there is also a very strong conservation of the sequences in the 5' UTR (e.g., in the human and mouse c-fos genes).

ACKNOWLEDGMENTS

We wish to thank Mrs. Malvine Baer and Dr. Rebecca Aft-Koenigsberg for editorial assistance. This work was supported by the the National Institutes of Health (Grant No. GM-22767), the Muscular Dystrophy Association, Inc., U.S.A., the U.S.-Israel Binational Science Foundation, Jerusalem, and the Leo and Julia Forchheimer Center for Molecular Biology at the Weizmann Institute of Science. U. Nudel is the incumbent of the A. and E. Bloom Career Development Chair for Cancer Research at the Institute.

REFERENCES

1. Miyata, T. and Yasunaga, T. (1981) Proc. Natl. Acad. Sci. USA 78, 450-453.
2. Miyata, T., Hayashida, H., Hasegawa, M., Kobayashi, M. and Koike, K. (1982) J. Mol. Evol. 19, 28-35.
3. Hayashida, H. and Miyata, T. (1983) Proc. Natl. Acad. Sci. USA 80, 2671-2675.
4. Shen, S.I., Slightom, J.L. and Smithies, O. (1981) Cell 26, 191-203.

5. Efstratiadis, A., Posakony, J.W., Maniatis, T., Lawn, R.M., O'Connell, C., Spritz, R.A., DeRiel, J.K., Forget, B.G., Weissmann, S.M., Slightom, J.L., Blechl, A.E., Smithies, D., Baralle, F.E., Shoulders, C.C. and Proudfoot, N.J. (1980) *Cell* 21, 653-668.
6. Breathnach, R. and Chambon, P. (1981) *Ann. Rev. Biochem.* 50, 349-383.
7. Miyata, T., Yasunaga, T. and Nishida, T. (1980) *Proc. Natl. Acad. Sci. USA* 77, 7328-7332.
8. Gunning, P., Mohum, T., Ng, S.-Y., Ponte, P. and Kedes, L. (1984) *J. Mol. Evol.* 20, 202-214. *Cell* 20, 555-566.
9. Vanderkerckhove, J. and Weber, K. (1979) *Differentiation* 14, 124-133.
10. Elzinga, M. and Lu, R. (1976) In "Contractile Systems in Nonmuscle Tissues", eds. Perry, S.V., Margreth, A. and Adelstein, R., p. 29 (North-Holland, Amsterdam)
11. Garrels, J. and Gibson, W. (1976) *Cell* 9, 793-805.
12. Shani, M., Zevin-Sonkin, D., Saxel, O., Carmon, Y., Katcoff, D., Nudel, U. and Yaffe, D. (1981) *Dev. Biol.* 86, 483-492.
13. Minty, J.A., Alonso, S., Caravatti, M. and Buckingham, M.E. (1982) *Cell* 30, 185-192.
14. Mayer, Y., Czosnek, H., Zeelon, P.E., Yaffe, D. and Nudel, U. (1984) *Nucl. Acids Res.* 12, 1087-1100.
15. Melloul, D., Aloni, B., Calvo, J., Yaffe, D. and Nudel, U. (1984) *EMBO J.* 3, 983-990.
16. Zakut, R., Shani, M., Givol, D., Neuman, S., Yaffe, D. and Nudel, U. (1982) *Nature* 298, 857-859.
17. Fornwald, J.A., Kuncio, G., Peng, I. and Ordahl, C.P. (1982) *Nucl. Acids Res.* 10, 3861-3876.
18. Nudel, U., Zakut, R., Shani, M., Neuman, S., Levy, Z. and Yaffe, D. (1983) *Nucl. Acids Res.* 11, 1759-1771.
19. Kost, T.A., Theodorakis, N. and Hughes, S.H. (1983) *Nucl. Acids Res.* 11, 8287-8301.
20. Hanukoglu, I., Tanese, N. and Fuchs, E. (1983) *J. Mol. Biol.* 163, 673-678.
21. Ponte, P., Ng, S., Engel, J., Gunning, P. and Kedes, L. (1984) *Nucl. Acids Res.* 12, 1687-1690.
22. Hamada, H., Petrino, M.G. and Kakunaga, T. (1982) *Proc. Natl. Acad. Sci. USA* 79, 5901-5905.
23. Nudel, U., Mayer, Y., Zakut, R., Shani, M., Czosnek, H., Aloni, B. and Yaffe, D. (1984) In "Experimental Biology and Medicine, Vol. 9, eds. Eppenberger, H.M. and Perriard, J.C. Karger Press, Basel, pp. 219-227.
24. Shani, M., Nudel, U., Zevin-Sonkin, D., Zakut, R., Givol, D., Katcoff, D., Carmon, Y., Reiter, J., Frischauf, A.M. and Yaffe, D. (1981) *Nucl. Acids Res.* 9, 579-589.
25. Ordahl, C.P. and Cooper, T.A. (1983) *Nature* 303, 348-349.
26. Ullrich, A., Gray, A., Berman, C. and Dull, T.J. (1981) *Nature* 303, 821-825.
27. Cowan, P.J., Dorner, P.R., Fuchs, E.V. and Cleveland, D.W. (1983) *Mol. Cell. Biol.* 3, 1738-1745.
28. Katcoff, D., Nudel, U., Zevin-Sonkin, D., Carmon, Y., Shani, M., Lehrach, H., Frischauf, A.M. and Yaffe, D. (1980) *Proc. Natl. Acad. Sci. USA* 77, 960.
29. Martin, S.L., Zinner, E.A., Davidson, W.S. and Wilson, A.C. (1981) *Cell* 25, 7637-741.
30. Nilson, J.H., Thomason, A.R., Cserbak, M.T., Moncman, C.I. and Woychik, R.P. (1983) *J. Biol. Chem.* 258, 4679-4682.
31. Hennighausen, L.G., Sippel, A.E., Hobbs, A.A. and Rosen, J.M. (1982) *Nucl. Acids Res.* 10, 3733-3744.
32. Rozek, C.E. and Davidson, N. (1983) *Cell* 32, 23-34.
33. Amara, S.G., Jonas, V., Rosenfeld, M.G., Ong, E.S. and Evans, R.M. (1982)

- Nature 298, 240-244.
34. Emorine, L., Kuehl, M., Weir, L., Leder, P. and Max, E.E. (1983) Nature 304, 447-449.
35. Banerji, J., Olson, L. and Schaffner, W. (1983) Cell 33, 729-740.
36. Gillies, S.O., Morrison, S.L., Ol, V.T. and Tonegawa, S. (1983) Cell 35, 137-148.
37. Queen, C. and Baltimore, D. (1983) Cell 33, 741-748.
38. Kronenberg, H.M., Roberts, B.E. and Efstratiadis, A. (1979) Nucl. Acids Res. 6, 153-165.
39. Liehaber, S.A. and Kan, Y. (1982) J. Biol. Chem. 257, 11852.
40. Miller, A.D., Curran, T. and Verma M. (1984) Cell 36, 51-60.
41. Van Beveren, C., van Straaten, F., Curran, T., Muller, R. and Verma, I.M. (1983) Cell 32, 1241-1255.
42. Van Straaten, F., Muller, R., Curran, T., Van Beveren, C. and Verma, I.M. (1983) Proc. Natl. Acad. Sci. USA 80, 3183-3185.
43. Li, W.H., Cojocori, T. and Nei, M. (1981) Nature 292, 237-239.
44. Lemischke, I.R., Farmer, S., Racaniello, V.R. and Sharp, P.A. (1981) J. Mol. Biol. 151, 101-120.
45. Ginsburg, I., Behar, L., Givol, D. and Littauer, U.Z. (1981) Nucl. Acids Res. 9, 2691-2697.
46. Cooke, N.E. and Baxter, J.D. (1982) Nature 297, 603-606.
47. Cooke, N.E., Coit, D., Shine, J., Baxter, J.D. and Martial, J.A. (1981) J. Biol. Chem. 256, 4007-4016.
48. Lawn, R.M., Efstratiadis, A., O'Connell, C. and Maniatis, T. (1980) Cell 21, 647-651.
49. Dolan, M., Dodgson, I.B. and Engel, J.D. (1983) J. Biol. Chem. 258, 3983-3990.
50. Page, S.G., Smith, S. and Goodman, H.M. (1981) Proc. Natl. Acad. Sci. USA 78, 2087-2103.
51. De Noto, F.M., Moore, D.D. and Goodman, H. (1981) Nucl. Acids Res. 9, 3719-3730.
52. Heindell, H.C., Liu, A., Paddock, G.V., Studnicka, G.M. and Salsen, W.A. (1978) Cell 15, 43-54.
53. Nishioka, Y. and Leder, P. (1979) Cell 18, 875-882.
54. Spritz, R.A., De Riel, J.K., Forget, B.G. and Weissmann, S.M. (1980) Cell 21, 639-646.
55. Slightom, J.L., Blechl, A.E. and Smithies, O. (1980) Cell 21, 627-636.
56. Baralle, F.E., Shoulders, C.C. and Proudfoot, N.J. (1980) Cell 21, 621-626.
57. Schwarzbauer, J.E., Tankun, J.W., Lemischke, I.R. and Hynes, R.O. (1983) Cell 35, 421-431.
58. Kornblitt, A.R., Vibe-Paderson, K. and Baralle, F.E. (1983) Proc. Natl. Acad. Sci. USA 80, 3218-3222.
59. Dayhoff, M.O. (1978) in "Atlas of Protein Sequence and Structure," Vol. 5, Suppl. 3 (ed.: Dayhoff, M.O.), National Biomedical Research Foundation, USA, pp. 1-8.
60. Dickerson, R.E. (1971) J. Mol. Evol. 1, 26-45.
61. Maki, M., Takayanagi, R., Misono, K.S., Tibbetts, C. and Inagani, T. (1984) Nature 309, 722-724.
62. Oikawa, S., Imi, M., Veno, A., Tanake, S., Noguchi, T., Nakuzato, H., Kangawa, K., Fukuda, A. and Matsuo, H. (1984) Nature 309, 724-726.
63. Miller, W.L., Martial, J.A. and Baxter, J.D. (1980) J. Biol. Chem. 255, 7521-7524.
64. Gillespie, G., Lloyd, J., Hopkinson, D., and Edwards, Y. (1984) J. Muscle Res. Cell Mot. 5, 457-464.