Volume 251, number 1,2, 201–206 FEB 07388 July 1989

Analysis of 3′-untranslated regions of seven c-myc genes reveals conserved elements prevalent in post-transcriptionally regulated genes

Sophie Vriz and Marcel Méchali

Institut Jacques Monod, Tour 43, 2 place Jussieu, 75251 Paris Cédex 05, France

Received 19 April 1989

We have characterized the complete sequence of two c-myc cDNAs from the amphibian *Xenopus laevis*, and could thus compare the 3′-non-coding sequences of 7 myc cDNAs from 6 species spread over 350 million years of evolution. Although the size of these sequences is heterogeneous, we identified three completely conserved sequences of 10, 11 and 12 contiguous nucleotides. We observed that two of these elements may be contained in conserved stem-loop structures previously implicated in mRNA turnover. The length of these motifs, their existence in conserved predicted structures, and their presence in regulated eukaryote mRNA with a frequency greater than predicted by chance, suggest that they are functionally important.

Oncogene, myc; mRNA stability

1. INTRODUCTION

Little is known about the determinants or sequence elements involved in mRNA stability. Most data support the conclusion that the 3′-untranslated region of eukaryotic mRNA plays an important role in this process (reviewed in [1,2]). The poly(A) tail and the processing linked to the generation of the 3′-terminus might be one important determinant (reviewed in [3]). Another element is the A-U rich segment found in the 3′-non-coding region of mRNAs encoding for lymphokines, cytokines and proto-oncogenes and the conserved AUUUA motif present in this A-U rich sequence [4–7].

Post-transcriptional regulation of the cellular proto-oncogene c-myc has been well documented [8]. The half-life of c-myc mRNA is 10 to 20 min in various cell types [9] and can be modulated in several ways. Destabilization of c-myc mRNA was observed when cell growth is inhibited by interferon [10,11], during differentiation of teratocarcinoma cells [12,13] or Friend erythroleukemia cells [14], without change in the transcription rate. Stabilization of c-myc RNA is observed in fibroblasts stimulated by growth factors [15], and during oogenesis in *Xenopus laevis* where it was followed by destabilization of the stored RNA after fertilization [16]. Recent experiments involving various deletions and hybrid c-myc genes have shown that the sequences directly responsible for the short c-myc mRNA half-life are localized in the 3′-untranslated region [17]. We postulated that these 3′-end sequences should be conserved throughout the evolution. To increase the significance of a comparative analysis we have sequenced the complete 3′-untranslated region of two c-myc genes of the amphibian *Xenopus laevis*, a species distant from humans by 350 million years. We show that although the overall homology between the different c-myc 3′-untranslated regions was low, three blocks of significantly conserved sequences are present. The different properties of such blocks suggest that they could be involved in post-transcriptional regulation of the c-myc RNA.

Correspondence address: S. Vriz, Institut Jacques Monod, Tour 43, 2 place Jussieu, 75251 Paris Cédex 05, France

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/89/$3.50 © 1989 Federation of European Biochemical Societies

2. MATERIALS AND METHODS

2.1. *Cloning and sequencing of the* Xenopus *c-myc gene*

A cDNA library prepared from *Xenopus* oocyte poly(A)$^+$ RNA [18] was used to identify the *Xenopus* c-myc cDNA [16]. Two different c-myc coding sequences were found and completely sequenced on both strands (Vriz et al., to be published) by the method of Sanger [19], using both single-stranded DNA M13 mp18 and mp19 recombinants and Bluescribe (Stratagene) double-stranded DNA recombinants.

2.2. *Sequence alignments and comparisons*

As the length of the 3′-untranslated regions analysed was highly heterogeneous, the optimal alignment between the seven c-myc sequences was done both by the method of Kanehisha [20] and visually. The Genbank data base was searched for exact matches to the conserved nucleotide blocks found. At the time of the search, the total data base contained 16752872 base pairs. Potential secondary structures were searched by using the program of Zucker and Stiegler [21].

3. RESULTS AND DISCUSSION

3.1. *Three blocks of nucleotides are perfectly conserved in the 3′-untranslated sequence of c-myc RNA*

The proto-oncogene c-myc has been found in the genome of amphibians, where it is highly expressed during oogenesis [16,22]. Recently we have characterized a second c-myc gene in *Xenopus laevis* and the complete sequence of the 3′-untranslated region of these two *Xenopus* c-myc genes was obtained. This permitted comparison of the 3′-non-coding regions of 7 cDNAs from 6 different species spread over 350 million years of evolution (fig.1). Several points deserve emphasis.

First, the size of these 3′-untranslated regions is highly heterogeneous, varying from 290 bases for rat to 950 bases for *Xenopus* myc 2. Second, the polyadenylation signal of *Xenopus* myc 1 and 2 is 549 b and 568 b, respectively, downstream of the polyadenylation signal of the five other species. These *Xenopus* additional sequences are not related to the other species (unpublished data on request). Therefore the alignment was done from the stop codons until the relatively conserved polyadenylation signal region (fig.1A). Third, the first bases downstream of the coding sequence are poorly homologous and there may be a deletion downstream of the human stop codon, as compared to the other species. Fourth, the overall homology between these sequences is low, ranging from 45 to 51% when the *Xenopus* myc 3′-end was compared to chicken or human 3′-ends, although in a number of cases the conversion of one base to another during evolution was clear in the alignment (fig.1A). In contrast, the overall divergency observed in these 3′-untranslated regions highlights blocks of sequences that are totally conserved. Fig.1B is a histogram of the nucleotides conserved at the same position for all seven myc sequences aligned in fig.1A. Three uninterrupted runs of 100% homology for 12 nucleotides (block A), 10 nucleotides (block B), and 11 nucleotides (block C) are apparent. Considering the length of the sequences analyzed this conservation appeared significant.

3.2. *The 3′ c-myc motifs are present in other genes*

The Genbank data base was searched for matches to the 12 nucleotide block A, allowing one single mismatch to compensate for possible sequencing errors in these non-coding regions. A total of 16752872 base pairs (38% G+C) representing both prokaryote and eukaryote sequences were examined. A 12 nucleotide sequence (block A) would be expected to occur 1.8 times [23] but was found in 22 genes, all eukaryotic (table 1). For two of these genes (human growth hormone gene and human plasminogen activator) the block A is localized in Alu type sequences where transcription occurs on the antisense strand of the human growth factor gene by RNA polymerase III [24]. Transcription of the antisense strand of the c-myc gene in this region has been found for the human and mouse myc genes [25–27].

Fig.1. Comparison of 3′ c-myc untranslated sequences. (A) Optimal alignment of 3′ c-myc untranslated sequences. The 3′-untranslated sequence from different myc species were aligned according to evolution and from the termination codon to the first polyadenylation signal for human (H), cat (C), rat (R), mouse (M) and chicken (Ch) and from the termination codon to 547 and 568 bases before the polyadenylation signal for *Xenopus* (X1 and X2). H, human [38]; C, cat [39]; R, rat [40]; M, mouse [38,41]; Ch, chicken [42]; X1 and X2, *Xenopus laevis* contain two c-myc genes (to be published elsewhere); T, trout [43]. Shaded areas are for an identical nucleotide at the same position for at least 4 out of 7 sequences. (B) Graphic representation of the 3′ c-myc untranslated sequences. Histogram of nucleotides conserved at the same position for all seven myc species.

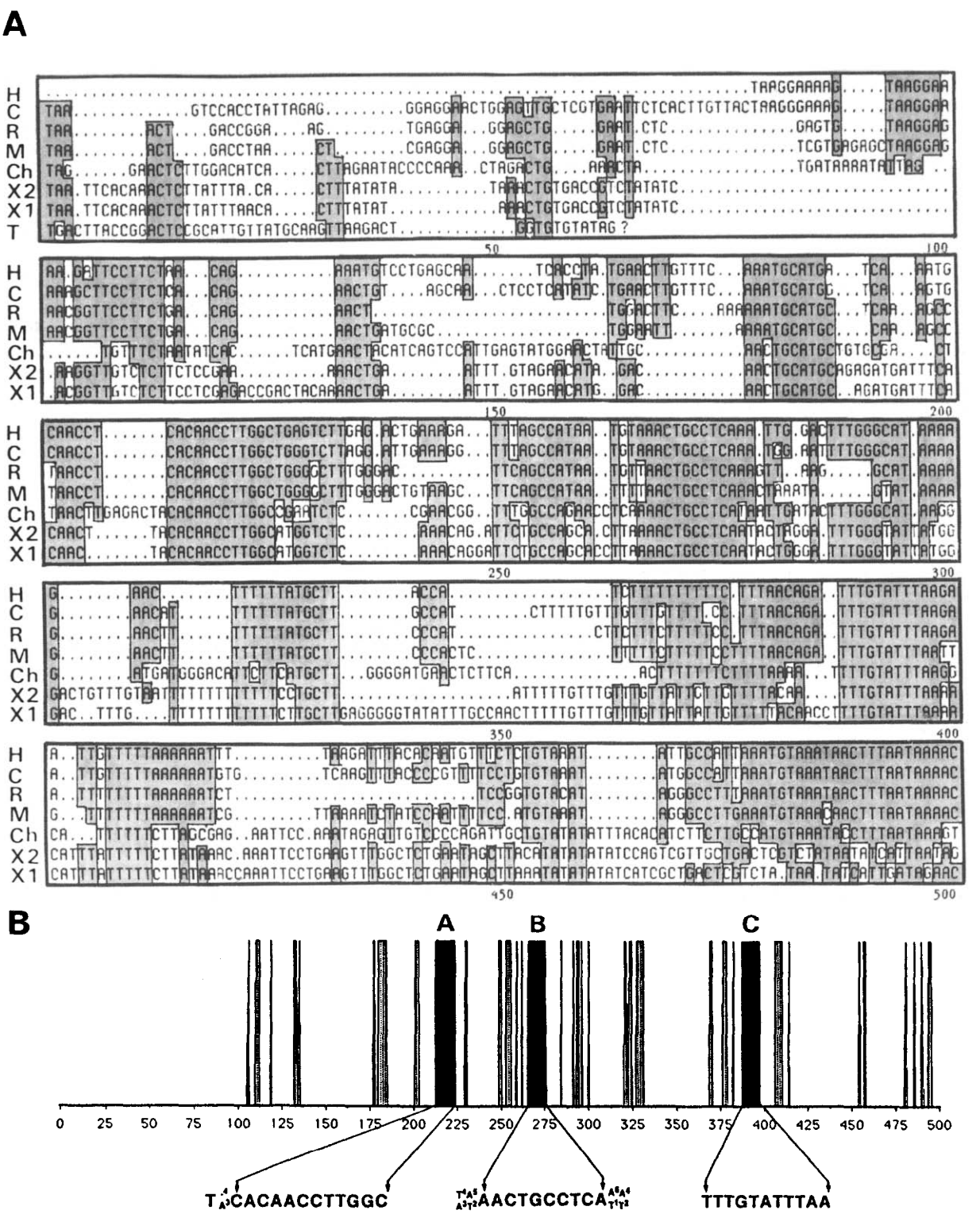
A
H
C
R
M
Ch
X2
X1
T
50
100
150
200
250
300
350
400
450
500
B
A
B
C
0 25 50 75 100 125 150 175 200 225 250 275 300 325 350 375 400 425 450 475 500
TACACAACCTTGGC
AACTGCCTCA
TTTGTATTTAA

Table 1 : Genes containing at least 11 out of the 12 nucleotides of the block A.
The Genbank data base was searched for at least 11 out 12 nucleotides of block A.

GENE	SEQUENCE	GENBANK	POSITION
c-myc conserved sequence	**C A C A A C C T T G G C**		
Rat parathyroïd hormone gene	C A C A A C C T g G G C	RATPTH3	mRNA coding sequence
Mouse ß globin Major gene	C A C c A C C T T G G C	MUSHBBMAJ	mRNA coding sequence
Spinach chloroplaste photosystem II	t A C A A C C T T G G C	SPICPD2CB	mRNA coding sequence
Mouse SLP gene	t A C A A C C T T G G C	M12384	mRNA coding sequence
Porcin inhibin ßa sub. = TGF ß	C A C A A C t T T G G C	PIGINHBAR	mRNA 5' untranslated region
Rat phosphoenol pyruvate carboxylase	C A C A c C C T T G G C	RATPECG1	mRNA 5' untranslated region
Schzophyllum 1G2 gene	C A g A A C C T T G G C	SCO1G2	mRNA 3' untranslated region
Human growth hormone gene HGH-N	C A C A A t C T T G G C	HUMGHN	genomic 3' end
Wheat a/ß gliadingene	C A C A t C C T T G G C	WHTGLIABD	genomic 3' end
Human tissue plasminogene activator	C A C A A t C T T G G C	HUMTPA	intron
Rat metallothionein 2 and 1	C A C A A C C c T G G C	RATMT12C	genomic 5' end
Human insulin gene	C A C c A C C T T G G C	HUMINS1	genomic 5' end
Yeast (S. cerevisiae) pho5 gene	t A C A A C C T T G G C	YSCPHO5A	genomic 5' end
Mouse His-tRNA gene	C A C A A C C T T G G C	MUSTRHMT1	genomic 5' end
Human interferon ß	C A a A A C C T T G G C	HUMIFNB3	unpublished from Genbank

The 10 nucleotide block B was not found in the prokaryotic data base. Allowing one mismatch it was expected to occur 25 times in the total eukaryotic data base only by chance. It was found 44 times, and only in transcribed regions of eukaryotic genes encoding for regulatory functions as the interferon family (27 different genes), acetylcholine receptors (2 genes), the epidermal growth factor, the proto-oncogene raf, yeast GCN4, procollagen, elongation factor 2, *Drosophila* Notch, mouse adipocyte lipid-binding protein and chicken embryonic myosine (data from Genbank, available on request).

The 11 nucleotide block C has a high (A + T) content (91%) which contrasts with the A + T content of the two other blocks (41.6 and 50%). It is an extension of the UAUUUA motif previously described to confer instability to mRNAs [4–6]. Although the seven c-myc 3'-untranslated regions contain one (mouse) to six AUUUA motifs (*Xenopus* myc 1) according to the species, only one of these motifs is conserved during evolution, namely the one found in the 11 nucleotide block C, defining a AUUUA c-myc element enlarged to 11 nucleotides totally conserved.

3.3. *Two types of AUUUA motifs*

Our further analysis of the AUUUA motifs in both c-myc genes and other genes revealed that they are of two types, either with the motifs organized in clusters or with a single AUUUA motif isolated in the sequence. The first category includes cytokines, lymphokines, interferons and it is the repeat of AUUUA motifs which is responsible for the A-U richness of the sequence [4,5]. The second category includes the proto-oncogenes c-myc, L-myc, N-myc, fos, myb, p53, the cyclin, PDGFA, and bovine transducin. For these sequences, it is not the AUUUA motif which confers A-U richness, but rather the even repeats of U close to this motif. Interestingly, all sequences containing the block B, including non-myc sequences, also possess the AUUUA motif.

3.4. *Secondary structure involving B and C blocks: possible relationship with the pathway of mRNA degradation*

The involvement of two conserved sequences in a common regulatory signal has been previously demonstrated for 3'-mRNA processing of the 3'-region of sea urchin histone H3 mRNA [28] as a consequence of a common secondary structure [3,29]. The formation of stem-loop secondary structures has been implicated in 3'-end processing or RNA decay ([2,3] for reviews). Indeed a search for a secondary structure around the block C of

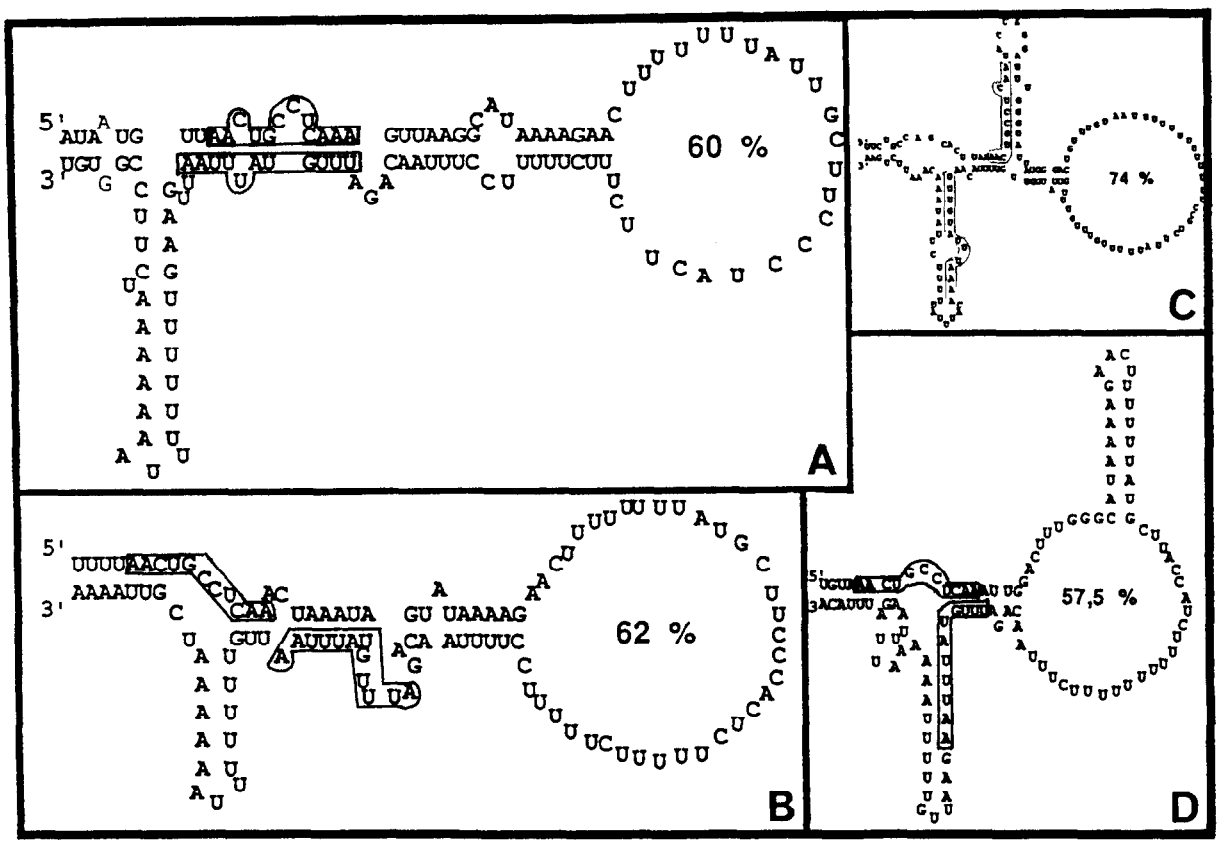

Fig.2. Potential secondary structures. Potential secondary structures for the c-myc sequences around the blocks B and C were according to Zucker algorithm [21]. (A) Rat; (B) mouse; (C) *Xenopus* 2; (D) human. The proportion of uridine in the loop is indicated and blocks B and C are underlined.

the myc RNA [21] revealed the potential formation of a stem-loop structure engaging both the B and C elements in all myc sequences analyzed from different species (fig.2). A triple helix formation in this A-U region is also formally possible. These potential structures are highly similar with a single-stranded U rich loop adjacent to a double-stranded structure holding the B and C conserved elements.

Endonucleolytic cleavages at uridine-rich single-stranded regions have been reported both for mRNA degradation in *E. coli* [30] and in eukaryotes [2,31,32] and double-stranded structures close to the A-U rich element have been postulated to be involved in the (2′-5′)oligoadenylate synthetase-RNase L pathway of mRNA degradation [33,34]. Thus, oligoadenylate synthetase is activated by double-stranded RNA (reviewed in [33]), and the oligoadenylate products activate RNase L, a latent endonuclease which cleaves RNA only in single-stranded U rich regions [31,35]. Among the homopolymers of the four common nucleotides, RNase L cleaves only poly(U) [31]. Moreover, the covalent association of a double-stranded RNA segment to a single-stranded RNA segment was shown to induce a *cis*-activation of RNase L by the (2′-5′)oligo(A) synthetase in vitro [36,37]. Jones and Cole [17] have recently constructed various deletions and hybrid molecules within the 3′-end of myc mRNA and concluded that the 3′-untranslated region was directly involved in the turnover of myc mRNA. When the half-life of the mRNA encoded by their deletion mutants is corrected for that of the control β2-microglobulin stable transcript, their results are in agreement with a role of the B and C elements in the control of mRNA stability.

The availability of c-myc sequences in the 3′-untranslated region from distant species in

evolution enabled one to reveal three totally conserved blocks of 12, 10 and 11 nucleotides. The relatively low expected frequency of such sequences contrasts with their finding in a number of mRNAs encoded by regulatory genes, suggesting that their presence is not accidental. All transcripts containing the block B also have the AUUUA motif. For the myc transcripts these two elements can be involved in a potential double-stranded structure close to a single-stranded loop uridine-rich. The presence of these two elements in a number of genes that undergo post-transcriptional regulation is significant and suggests that they might be involved in the control of mRNA stability. We hope that these observations will direct experimental analyses both at the level of the conserved sequences and at the structure level.

Acknowledgements: We thank P. Brooks, A.L. Haenni, F. Schweisguth and M. Nadal for critical reading of the manuscript. This work was supported by grants from the Association pour la Recherche sur le Cancer, INSERM and Fondation pour la Recherche Médicale.

REFERENCES

[1] Shapiro, D.J., Blume, J.E. and Nilsen, D.A. (1987) BioEssays 6, 221–226.
[2] Brawerman, G. (1987) Cell 48, 5–6.
[3] Birnstiel, M.L., Busslinger, M. and Strub, K. (1985) Cell 41, 349–359.
[4] Caput, D., Beutler, B., Hartog, K., Thayer, R., Brown-Shimer, S. and Cerami, A. (1986) Proc. Natl. Acad. Sci. USA 83, 1670–1674.
[5] Shaw, G. and Kamen, R. (1986) Cell 46, 659–667.
[6] Reeves, R., Spies, A.G., Nissen, M.S., Back, C.D., Weinberg, A.D., Barr, P.J., Magnunson, N.S. and Magnunson, J.A. (1986) Proc. Natl. Acad. Sci. USA 83, 3228–3232.
[7] Clemens, M.J. (1987) Cell 49, 157–158.
[8] Cole, M.D. (1986) Annu. Rev. Genet. 20, 361–384.
[9] Dani, C., Blanchard, J.M., Piechaczyk, M., El Sabouty, S., Marty, L. and Jeanteur, P. (1984) Proc. Natl. Acad. Sci. USA 81, 7046–7050.
[10] Dani, C., Mechti, N., Piechaczyk, M., Lebleu, B., Jeanteur, P. and Blanchard, J.M. (1985) Proc. Natl. Acad. Sci. USA 82, 4896–4899.
[11] Knight, E., jr, Anton, E.D., Fahey, D., Friedland, B.K. and Jonack, G.J. (1985) Proc. Natl. Acad. Sci. USA 82, 1151–1154.
[12] Dony, C., Kessel, M. and Gruss, P. (1985) Nature 317, 636–639.
[13] Dean, M., Levine, R.A. and Campisi, J. (1986) Mol. Cell. Biol. 6, 518–524.
[14] Mechti, N., Piechaczyk, M., Blanchard, J.M., Marty, L., Bonnieu, A., Jeanteur, P. and Lebleu, B. (1986) Nucleic Acids Res. 14, 9653–9666.
[15] Blanchard, J.M., Piechaczyk, M., Dani, C., Chambard, J.C., Franchi, A., Pouyssegur, J. and Jeanteur, P. (1985) Nature 317, 443–445.
[16] Taylor, M.V., Gusse, M., Evan, G., Dathan, N. and Méchali, M. (1986) EMBO J. 5, 3563–3570.
[17] Jones, T.R. and Cole, M.D. (1987) Mol. Cell. Biol. 7, 4513–4520.
[18] Rebagliati, M.R., Weeks, D.L., Harvey, R.P. and Melton, D.A. (1985) Cell 42, 769–777.
[19] Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. USA 74, 5463–5467.
[20] Kanehisa, M. (1984) Nucleic Acids Res. 12, 203–213.
[21] Zucker, M. and Stiegler, P. (1981) Nucleic Acids Res. 9, 133–148.
[22] King, M.W., Roberts, J.M. and Eisenman, R.L. (1986) Mol. Cell. Biol. 6, 4499–4508.
[23] Nei, M. and Li, W.H. (1979) Proc. Natl. Acad. Sci. USA 76, 5269–5273.
[24] Seeburg, P.H. (1982) DNA 1, 239–249.
[25] Bentley, D. and Groudine, M. (1986) Nature 321, 702–706.
[26] Nepveu, A., Marcu, K.B., Skoultchi, A.I. and Lachman, H.M. (1987) Genes Dev. 1, 938–945.
[27] Kindy, M.S., McCormack, J.E., Buckler, A.I.J., Levine, R.A. and Sonenshein, G.E. (1987) Mol. Cell. Biol. 7, 2857–2862.
[28] Georgiev, O. and Birnstiel, M.L. (1985) EMBO J. 4, 481–489.
[29] Strub, K., Galli, G., Busslinger, M. and Birnstiel, M.L. (1984) EMBO J. 3, 2801–2807.
[30] Cannistraro, V.J., Subbarao, M.N. and Kennel, D. (1986) J. Mol. Biol. 192, 257–274.
[31] Floyd-Smith, G., Slattery, E. and Lengyel, P. (1981) Science 212, 1030–1032.
[32] Albrecht, G., Krowczynska, A. and Brawerman, G. (1984) J. Mol. Biol. 178, 881–896.
[33] Lengyel, P. (1982) Annu. Rev. Biochem. 51, 251–282.
[34] Mechti, N., Bisbal, C., Leonetti, J.P., Salehzada, T., Affabris, B., Bayard, M., Piechaczyk, M., Blanchard, J.M., Jeanteur, P. and Lebleu, B. (1988) Biochimie 70, 869–875.
[35] Wreschner, D.H., McCauley, J.W., Skehel, J.J. and Kerr, I.M. (1981) Nature 289, 414–417.
[36] Nielsen, T.W. and Baglioni, C. (1979) Proc. Natl. Acad. Sci. USA 76, 2600–2604.
[37] Nielsen, T.W., Weissman, S.G. and Baglioni, C. (1980) Biochemistry 19, 5574–5579.
[38] Bernard, O., Cory, S., Gerondakis, S., Webb, E. and Adams, J.A. (1983) EMBO J. 2, 2375–2383.
[39] Stewart, M.A., Forrest, D., McFarlane, B., Onions, D., Wilkie, N. and Neil, J.C. (1986) Virology 154, 121–134.
[40] Hayashi, K., Makino, R., Kawamura, H., Arisawa, A. and Yoneda, K. (1987) Nucleic Acids Res. 15, 6419–6436.
[41] Stanton, L.W., Fahrlander, P.D., Tesser, P.M. and Marcu, B. (1984) Nature 310, 423–425.
[42] Watson, D.K., Reddy, E.P., Duesberg, P.H. and Papas, T.S. (1983) Proc. Natl. Acad. Sci. USA 80, 2146–2150.
[43] Van Beneden, R.J., Watson, D.K., Chen, T.T., Laulenberger, J.A. and Papas, T.S. (1986) Proc. Natl. Acad. Sci. USA 83, 3698–3702.