The EMBO Journal vol.7 no.1 pp.29-35, 1988

A gene encoding a protein with zinc fingers is activated during G_0/G_1 transition in cultured cells

Philippe Chavrier, Marino Zerial, Patrick Lemaire, José Almendral[1], Rodrigo Bravo and Patrick Charnay

Differentiation Programme, European Molecular Biology Laboratory, Postfach 10.2209, 6900 Heidelberg, FRG

[1]Present address: CIEMAT, 28040, Madrid, Spain

Communicated by R.Cortese

Zinc fingers are DNA-binding domains present in several eukaryotic regulatory proteins. We have identified a mouse gene, *Krox*-20, encoding a protein with three zinc fingers and whose expression is activated during G_0/G_1 transition in cultured cells. Serum stimulation of quiescent cells leads to rapid and transient accumulation of *Krox*-20 mRNA, with kinetics similar to those of the c-*fos* proto-oncogene. The induction does not require *de-novo* protein synthesis. In the mouse, *Krox*-20 is expressed at low levels in tissues which contain rapidly dividing cells. These properties suggest that *Krox*-20 encodes a transcription control factor, possibly involved in the modulation of cell proliferation.

Key words: zinc fingers/*Krox*-20 gene/G_0-G_1 transition/transcription factor/gene regulation

Introduction

The study of transcriptional control proteins has led, so far, to the identification of two types of DNA-binding domains involved in the recognition of target sequences. The so-called 'helix-turn-helix' domain has been noticed in a number of prokaryotic regulatory proteins (reviewed in Pabo and Sauer, 1984), in the product of the yeast mating-type gene loci and in the 'homeodomain' of the products of a number of *Drosophila* segmentation genes (Laughon and Scott, 1984; Shepherd *et al.*, 1984). It was recently also observed in cyclin (PCNA), an auxiliary protein of DNA polymerase delta (Almendral *et al.*, 1987; Bravo *et al.*, 1987a; Prelich *et al.*, 1987). The structure of a second type of DNA-binding domain has emerged from the study of *Xenopus laevis* transcription factor IIIA (TFIIIA). TFIIIA interacts with a region ~50 nucleotides long, internal to the *X.laevis* 5S RNA gene and is required for initiation of its transcription (Bogenhagen *et al.*, 1980; Engelke *et al.*, 1980; Sakonju *et al.*, 1980; Sakonju and Brown, 1981, 1982). The DNA-binding domain consists of tandemly repeated units of 28 to 30 amino acids, containing two cysteines and two histidines at invariant positions (Brown *et al.*, 1985; Ginsberg *et al.*, 1984; Miller *et al.*, 1985). It was proposed that each of these units folds as an independent domain, centered on a zinc ion coordinated by the cysteines and histidines, and interacts with five nucleotides on the DNA molecule (Brown *et al.*, 1985; Miller *et al.*, 1985; Rhodes and Klug, 1986). These units have been referred to as 'zinc fingers'. Recently, genetic evidence supporting this model has been provided (Blumberg *et al.*, 1987; Frankel *et al.*, 1987; Johnston, 1987). Zinc fingers have now been observed in a number of yeast and *Drosophila* regulatory proteins, including the product of the *Drosophila* segmentation 'gap' gene *Krüppel* (*Kr*) (Preiss *et al.*, 1985; Rosenberg *et al.*, 1985).

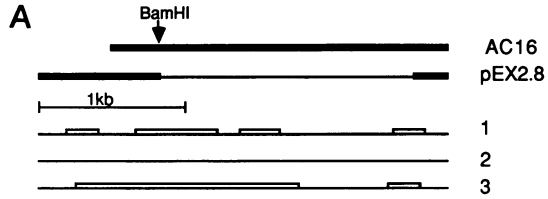
The combination of major advances in *Drosophila* molecular genetics and of the strong evolutionary conservation of essential protein domains permits the isolation of mammalian genes potentially involved in the regulation of development and pattern formation, as exemplified in the case of the genes containing the homeobox (Levine *et al.*, 1984; Colberg-Poley *et al.*, 1985; Hart *et al.*, 1985; Jackson *et al.*, 1985; Joyner *et al.*, 1985; Duboule *et al.*, 1986; Rubin *et al.*, 1986; Wolgemuth *et al.*, 1986). Recently, it was shown that the *Kr* DNA sequence encoding the finger region hybridized under low stringency conditions to several sequences in the genomes of vertebrates (Schuh *et al.*, 1986). This suggested that the use of this probe might allow the identification and isolation of genes encoding proteins with zinc fingers and potential regulatory functions. Indeed, the screening of mouse genomic libraries has led to the cloning of a number of such genes (Chowdhury *et al.*, 1987; Chavrier *et al.*, submitted).

We have now applied this approach to investigate whether genes encoding zinc fingers might be activated during G_0/G_1 transition in cultured cells. Mouse NIH3T3 fibroblasts can be brought to a quiescent state (G_0) by serum deprivation. They can be subsequently stimulated to re-enter G_1 and proliferate by exposure to serum or purified growth factors (Pardee *et al.*, 1978). It is well established that the activation of some genes is required for quiescent cells to respond to mitogens and proliferate (reviewed in Baserga, 1985). Among these, the immediate-early genes do not require *de novo* protein synthesis for activation. This category includes the proto-oncogenes c-*fos* and c-*myc* (Kelly *et al.*, 1983; Cochran *et al.*, 1984; Greenberg and Ziff, 1984; Kruijer *et al.*, 1984; Müller *et al.*, 1984; Bravo *et al.*, 1987b) and is thought to include many other genes as well, whose identities and functions are not yet known (Cochran *et al.*, 1983; Lau and Nathans, 1985, 1987; Almendral *et al.*, submitted). In this paper, we describe one such primary response gene which is activated very early during G_0/G_1 transition and encodes a putative protein containing zinc fingers. This protein is therefore likely to be a transcription factor possibly involved in the control of cell proliferation.

Results

Cloning of Krox-20 specific cDNAs

A cDNA library was prepared in phage λ 1149 with RNA extracted from NIH3T3 cells serum-stimulated in the presence of cycloheximide for 4 h. A collection of more than



B

```
     5'.....AGCGATTGATTAATAGCTGGGCGAGGGGACACACTGACTGTTATAATAACACTACACCAGCAACTCCTGGCTCCCCAACA
  81 GCCGGATCACAGGCAGGAGAGAGTCAGTGACGGATAGACTTTTTTTTTCTTTAAGAAGCCAACAACTTGGTTGCTAGTTTTATTTCTGTT
 171 AATTTTTTTCTTTTTTTTGGTGTGTGTGGATGTGTTGTGGTGGTCTTTTCTAAGTGTGGAGGGCAAAAGGAGATACCATCCTAGGCTCAG
 261 TTCAACCCCTCTCCAAAAAAAACGGCTTCTCTGGCGCTCCAGCTGCCTGACAGCCTCTACCCGGTGGAAGACCTCGCCGCCTCGTCGGTG
 351 ACCATCTTCCCCAATGGTGAACTGGGAGGCCCCTTTGACCAGATGAACGGAGTGGCGGGAGATGGCATGATCAACATTGACATGACTGGA
                                                  MetAsnGlyValAlaGlyAspGlyMetIleAsnIleAspMetThrGly
                                                                                              10
 441 GAGAAGAGACCCCTGGATCTCCCGTATCCGAGTAGCTTCGCTCCCATCTCTGCACCTAGAAACCAGACCTTCACCTACATGGGCAAATTC
     GluLysArgProLeuAspLeuProTyrProSerSerPheAlaProIleSerAlaProArgAsnGlnThrPheThrTyrMetGlyLysPhe
              20                                  30                                  40
 531 TCCATTGACCCACAGTACCCTGGTGCCAGCTGCTATCCAGAAGGTATCATCAATATTGTGAGTGCGGGCATCTTGCAAGGGGTCACCCCT
     SerIleAspProGlnTyrProGlyAlaSerCysTyrProGluGlyIleIleAsnIleValSerAlaGlyIleLeuGlnGlyValThrPro
              50                                  60                                  70
 621 CCAGCTTCAACCACAGCCTCCTCCAGCGTCACCTCCGCCTCCCCCAACCCACTGGCCACGGGACCCCTGGGTGTGTGTACCATGTCCCAG
     ProAlaSerThrThrAlaSerSerSerValThrSerAlaSerProAsnProLeuAlaThrGlyProLeuGlyValCysThrMetSerGln
              80                                  90                                  100
 711 ACTCAGCCTGAACTGGACCACCTCTACTCTCCGCCACCACCTCCTCCTCCTTATTCGGGCTGTACAGGAGATCTCTACCAGGATCCTTCA
     ThrGlnProGluLeuAspHisLeuTyrSerProProProProProProProTyrSerGlyCysThrGlyAspLeuTyrGlnAspProSer
              110                                 120                                 130
 801 GCATTCTTATCGCCGCCATCCACCACCTCCTTCCTACCCATCCCCCAAGCCAGCTATGGACCCAGGTCTCATTCCTATGATCCCAGACTA
     AlaPheLeuSerProProSerThrThrSerPheLeuProIleProGlnAlaSerTyrGlyProArgSerHisSerTyrAspProArgLeu
              140                                 150                                 160
 891 TCCTGGATTTTTTCCATCTCCGTGCCAGAGGAGATCCACACGGTGCCTGCTGGCCCAGATCGAAAGCCGTTTCCCTGTCCTCTGGACTCC
     SerTrpIlePheSerIleSerValProGluGluIleHisThrValProAlaGlyProAspArgLysProPheProCysProLeuAspSer
              170                                 180                                 190
 981 CTGCGAGTGCCCCCTCCACTCACGCCACTCTCTACCATCCGTAATTTTACTCTGGGGGGTCCCGGTGCTGGAGTCACGGGACCAGGAGCA
     LeuArgValProProProLeuThrProLeuSerThrIleArgAsnPheThrLeuGlyGlyProGlyAlaGlyValThrGlyProGlyAla
              200                                 210                                 220
1071 AGTGGAGGTGGTGAGGGACCTCGGCTGCCTGGCAGTGGGTCTGCAGCAGTGACTGCCACCCCTTATAATCCGCACCACCTGCCATTGCGG
     SerGlyGlyGlyGluGlyProArgLeuProGlySerGlySerAlaAlaValThrAlaThrProTyrAsnProHisHisLeuProLeuArg
              230                                 240                                 250
1161 CCCATCCTGCGACCTCGAAAGTACCCTAACAGGCCCAGCAAAACGCCAGTGCACGAAAGGCCCTATCCCTGCCCAGCAGAAGGTTGTGAT
     ProIleLeuArgProArgLysTyrProAsnArgProSerLysThrPro[ValHisGluArgProTyrProCysProAlaGluGlyCysAsp
              260                                 270                                 280
1251 AGGAGGTTCTCACGCTCTGATGAGCTGACCAGGCACATCCGAATCCACACGGGCCACAAGCCCTTCCAGTGTCGGATCTGCATGCGAAAC
     ArgArgPheSerArgSerAspGluLeuThrArgHisIleArgIleHis[ThrGlyHisLysProPheGlnCysArgIleCysMetArgAsn
              290                                 300                                 310
1341 TTCAGCCGAAGTGACCACCTTACTACTCACATCCGAACCCACACCGGGGAGAAGCCCTTTGCCTGTGACTATTGTGGCCGCAAGTTTGCC
     PheSerArgSerAspHisLeuThrThrHisIleArgThrHis[ThrGlyGluLysProPheAlaCysAspTyrCysGlyArgLysPheAla
              320                                 330                                 340
1431 AGGAGTGACGAAAGGAAGCGCCACACCAAGATCCACTTCGGCAGAAGGAACGGAAGAGCAGTGCTCCCTCTGCACCTCCATCTGCCCAGT
     ArgSerAspGluArgLysArgHisThrLysIleHis]PheGlyArgArgAsnGlyArgAlaValLeuProLeuHisLeuHisLeuProSer
              350                                 360                                 370
1521 CTTCAGCCTCTGGTCCTGGGGGCTCGCAGGCCGGGGGCAGCCTGTGCGGTAACAGCGCCATTGGAGGACCACTGGCCTCCTGCACCTCTC
     LeuGlnProLeuValLeuGlyAlaArgArgProGlyAlaAlaCysAlaValThrAlaProLeuGluAspHisTrpProProAlaProLeu
              380                                 390                                 400
1611 GAACCAGGACACCGTGAGATGAAGCTCTGGCTGACACACCAGTTTCTCCAGGCCCCAGAGGCCCTCTGTCCGAGCTGCCAACACTACCAC
     GluProGlyHisArgGluMetLysLeuTrpLeuThrHisGlnPheLeuGlnAlaProGluAlaLeuCysProSerCysGlnHisTyrHis
              410                                 420                                 430
1701 CCTTCCCTGTTCCTCCCCATGATCCCGTGATCTGGGCAAAGGACCTTGATGGAGCCCAGCTCTGTCCCACCTTCTCACGGACGGCCTTCC
     ProSerLeuPheLeuProMetIleProEnd
              440         445
1791 GAAAACTTAGGCCATTTGAAGGGAGTTGACTGTCACTCCAAGAAATGGGGGAGCAAAAAGAGGGCTGGGTGAGGGCCCCTGGCCTACAGG
1881 GCTGCGCTCTGACCCTGACTGAGAGATGTCTGACTATGGTCTGCTAGCCCTTTCCGTTGACCCTGGATGCCAGTTGTTCTGAGACTTTTT
1971 CTACAATAGGTTGGGAGTTGCTGATTCCTTTGATCGAGGACAGCGGAAAAGACTAAATTAAAGCAAAACTGATGTGGCACTTTAATGGCT
2061 TGGGACTGACTTGGGGGTGGGGGTGGGGAGTTGTACAGTGAGCAGAGCCTCATCCTGGCTGCTGCACTCTGGCCCTAGAGCAGTGAATGG
2151 AGGTTTCTCTGGCCATCTCAACCCTTAAGCAATATGTCCTAGAAACTCAAGAGAATGGAAGTGCAATGTCGGGCAGGACAAAGCAATATT
2241 GGCTCCTTTTTTTGTTGTAGTTGAGGAACAAAGATTATTTTTTCAGTGTATATCCATTTAGATTTTGTGTATTTTTGATGTGCACTGCTC
2331 TCCGAGTTCTGAACCTTCGGGAAAAAAAATGTAAAGCATTTATGATCTCTTGAAATGAGTCAAAGGTTAACTAACTTATTTAAAGGGGGAC
2421 GTACATAGGATGCATGCAGTGAGTGGTGTTGCAAGTGTCCTCTGTGCCTTGTGTGATGTGGGCAGTGTTAACAGGGTCTGCATGTGTACA
2511 GGATGCCTTACTATGGGAACAGAAAAATCACTCCTTGGGTTTAAGTATGGCTGTATATTTCTGCCTATTAATATTTGGAATTTTTTTAGA
2601 AAGTATATTTTTGTATGCTCTGTTTTGTGACTTAAAAGTGTTACCTTTGTAGCCAAATTTCAGGTGAGAATGTGCTTCAATGTCACTGCC
2691 GCTGATTTGTTTGGTTATTAGCTCTTAATAGTTGTGGAGAGAGAAACAATCTATTCTAACATAAAAAAACCACTAACTGGAGTTCAGATA
2781 ATGGATGGCTTATTGACTATGGTGTAAATAAATACTTTTCAACAATAAAAAAAA
```

Fig. 1. Nucleotide sequence of *Krox*-20 cDNA and predicted amino acid sequence of the protein. **A**. The complete cDNA sequence is a composite of the overlapping sequences derived from the cDNA clones AC16 and pEX2.8. The regions sequenced in each clone are shown in bold. The positions of open reading frames larger than 201 nucleotides are presented underneath (open boxes). **B**. In the nucleotide sequence, the consensus signal for poly(A) addition is shown in bold characters and the sequences ATTTA potentially involved in mRNA selective degradation are underlined. In the amino acid sequence, the three potential initiation methionines are shown in bold characters and the zinc finger repeats are bracketed. The numbers on the left refer to nucleotide position and those under the amino acid sequence to amino acid position.

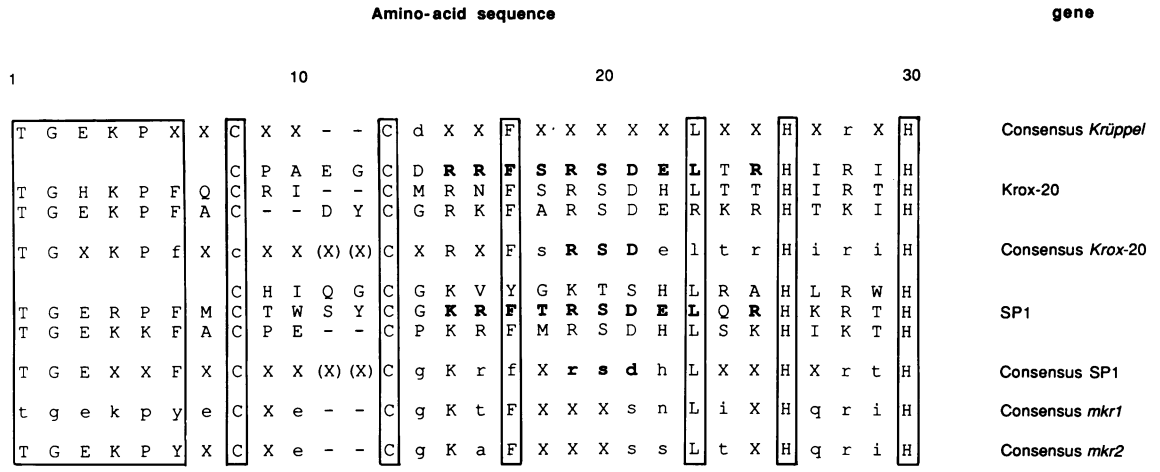


Fig. 2. Amino acid sequence comparison of *Kr*, *Krox*-20, Sp1, *mkr*1 and *mkr*2 finger motifs. The sequences are in the one letter code (Dayhoff, 1978). They are aligned to show the repeated units. The position of the amino acids which are strictly conserved in the *Kr* finger motif are boxed, as well as the interfinger H/C link (Schuh *et al.*, 1986). For each gene, a consensus for the repeated motif is displayed: capital letters correspond to strictly conserved amino acids, lower case letters to amino acids conserved over 50% and X to no conservation over 50%. The amino acids conserved or corresponding to conservative changes between finger 1 of *Krox*-20 and finger 2 of Sp1 are shown in bold characters. The sequence R-S-D conserved in *Krox*-20 and Sp1 consensus is also shown in bold characters. The consensus sequence of *Kr* and of *mkr*1 and *mkr*2 are derived from the works of Rosenberg and collaborators (1986) and Chowdhury and collaborators (1987) respectively. The Sp1 finger sequence is a personal communication of R.Tjian.

70 non-overlapping clones corresponding to genes activated during G_0/G_1 transition was isolated from this library using differential screening (Almendral *et al.*, submitted). We screened this collection by hybridization with the *Kr* finger probe under low stringency conditions. One single clone, with a 2.3 kb insert referred to as AC16, gave a positive signal and was studied in greater detail. We had previously isolated from a mouse genomic library a collection of 23 non-overlapping recombinant phages containing sequences homologous to the *Kr* finger probe (Chavrier *et al.*, submitted). A combination of Southern blot and restriction pattern analyses indicated that the AC16 cDNA corresponded to one of the genes previously cloned, *Krox*-20 (data not shown).

The AC16 cDNA was then used to isolate several other cDNAs corresponding to the *Krox*-20 gene from the original λ 1149 library as well as from a library prepared in plasmid pEX-1 (Stanley and Luzio, 1984) using the same RNA preparation. In the screening of the pEX-1 library, 20 positive clones were isolated from 2×10^5 colonies screened. The clone containing the longest insert (2.8 kb) was selected for further analysis and named pEX2.8. The AC16 sequence was completely included within the pEX2.8 cDNA, as judged by the analysis of the pattern of restriction fragments (data not shown). In addition, a comparison of the restriction patterns of AC16, pEX2.8 and two other cDNA clones was carried out. No difference was detected in the common regions, suggesting that no major rearrangement occurred during the cloning of AC16 and pEX2.8 sequences.

cDNA and protein sequences

The complete nucleotide sequences of AC16 and of parts of pEX2.8 were established as described in Figure 1A. Identical sequences were observed within the analysed overlapping regions of the two clones. The composite complete nucleotide sequence is presented in Figure 1B. The sequence is 2834 nucleotides long and terminates with a stretch of A residues located 14 nucleotides downstream from a AATAAA polyadenylation signal consensus sequence (Birnstiel *et al.*, 1985). The position of the open reading frames in the sense orientation is also shown in Figure 1A. The largest open reading frame extends from nucleotide positions 255 to 1727 (Figure 1B). If the first encountered methionine codon (position 393) is used for initiation of translation, this reading frame will encode a polypeptide of 445 amino acids with an unmodified mol. wt of 48 434 daltons. This coding sequence is preceded within the cDNA by a 5′ flanking region of 392 nucleotides containing stop codons in all reading frames. From the estimated size of *Krox*-20 mRNA [~3200 nucleotides including the poly(A) tail, see below], it can be deduced that the entire 5′ untranslated region covers ~700 nucleotides. The coding region is followed by a long 3′ untranslated region [1099 nucleotides plus the poly(A)] containing stop codons in all reading frames and rich in A and T (57%).

Krox-20 encodes zinc fingers

As expected from the cross-hybridization of *Krox*-20 with the *Kr* finger probe, the predicted polypeptide contains sequences repeated in tandem with similarity to the *Kr* finger consensus: Cys-X_2-Cys-X_3-Phe-X_5-Leu-X_2-His-X_3-His. *Krox*-20 encodes three zinc fingers, located between amino acid positions 280 and 358 (Figure 1B). Figure 2 shows an alignment of these fingers with the *Kr* consensus. The position of the highly conserved amino acids is strictly maintained, except that the first repeat contains four amino acids in between the two cysteines and the leucine is not present in the third finger. In addition, and as expected from the work of Schuh and collaborators (1986), a stretch of six amino acids located between successive fingers, Thr-Gly-Glu-Lys-Pro-Tyr, the so-called H/C link, is highly conserved (Figure 2). Part of this sequence is also observed in front of the first finger and close to the amino terminus of the

protein (Figure 1: positions 15–18 of the protein sequence).

According to the model of Miller and collaborators (1985), the portion of the repeated unit defining the specificity of recognition of the target DNA sequence is the finger, a loop located between Cys 13 and His 26 (positions measured within the repeated unit sequence, see Figure 2). No strong similarity is observed when *Krox*-20 fingers are compared to the other mouse zinc fingers so far described (Chowdhury *et al.*, 1987; Chavrier *et al.*, submitted) apart from the conservation of Phe 17, Leu 23 and a basic amino acid at position 15 (Figure 2 and data not shown). However, a significant similarity is observed with the fingers of the human transcription factor Sp1 (R.Tjian, personal communication; Figure 2). Sp1 binds to the SV40 GC boxes and is required for efficient transcription from the early promoter (Dynan and Tjian, 1983a,1983b; Gidoni *et al.*, 1984). Sp1, like the predicted product of *Krox*-20, contains three zinc fingers. The strongest similarity is observed between the first finger of *Krox*-20 and the second finger of Sp1: at 10 out of 12 positions within the DNA-binding loop, either the amino acids are identical or conservative changes occur (Figure 2). The sequence Arg-Ser-Asp-Glu is observed in fingers 1 and 3 of *Krox*-20 and in finger 2 of Sp1. The presence of acidic amino acids at positions 21 and 22 is common in the fingers of the two proteins, while it is exceptional in the other mouse fingers described so far (Figure 2; Chowdhury *et al.*, 1987; Brown and Argos, 1986; Chavrier *et al.*, submitted).

Although *Krox*-20 might be similar to Sp1 and possibly to other genes within the finger region, Southern blot analysis carried out under stringent hybridization conditions, with a probe consisting in the entire AC16 cDNA, only detected DNA fragments corresponding to the *Krox*-20 cDNA gene in the mouse genome (data not shown).

Structural features of Krox-20 protein

The analysis of the predicted protein sequence revealed several additional features. The protein is basic (estimated pI: 8.98). The basic amino acids are clustered in one region, located between positions 256 and 365 (Figure 1B). This region covers the finger domains and in addition extends toward the N- and C-termini. The distribution of proline residues is also quite atypical within the *Krox*-20 protein: the overall percentage is high (15.5%) and this amino acid is concentrated within two regions located respectively between positions 19 and 281, and 369 and 445. In particular, a stretch of seven prolines is observed between positions 117 and 123. In contrast, the finger domains are almost devoid of proline residues. These features allow the division of the protein into three main regions. Region A (positions 19 to 254) is highly enriched in proline, serine and threonine residues (overall percentage: 38%) and poor in charged amino acids. Region B (positions 280 to 365) is highly basic (25% of arginine and lysine) and very poor in proline; it includes the finger repeats and is likely to constitute a DNA-binding domain. The C-terminal region C (positions 369 to 445) is rich in proline residues and nevertheless is the only one with a high potential for forming α-helices (as determined according to the Chou-Fasman method). It is possible that the stretch located between regions A and B, which is both basic and rich in proline residues might constitute a hinge between an amino-terminal domain and the DNA-binding domain.

A search of the NBRF protein data base for protein sequences similar to that of *Krox*-20 protein revealed, as expected, a strong similarity with TFIIIA. This similarity was strictly limited to the finger domains. In addition we detected

```
180 TVPAGPDRKPFPC...PLDSLRVPPPLTPLSTIRNFTLGGPGAGVTGPGA Krox-20
    | |  ||||  ||   | |||   || ||||  |    || |||  |  |
160 TPAHTAGRRRNPCVAEPDDSISPDPPRTPVSRKRPRPAGATGGGGGGVHA Py T

227 SGGGEGPRLPGSGSAAVTATPYNPH Krox-20
    |||   |  || | |  | ||| |
210 NGGSVFGHPTGGTSTPAHPPPYHSQ Py T
```

Fig. 3. Amino acid sequence similarity between parts of the predicted *Krox*-20 protein and the polyoma large T antigen. Large T antigen sequence is from the work of Soeda and collaborators (1980). The sequences were aligned using the Bestfit program of the UWGCG package.

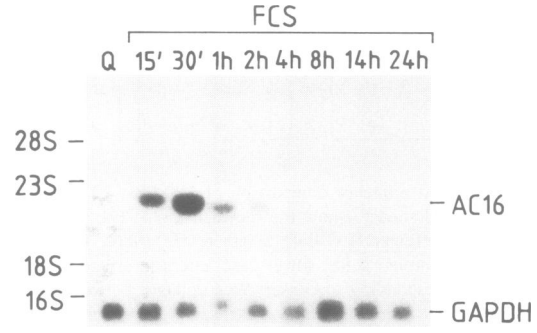


Fig. 4. Northern blot analysis of *Krox*-20 gene expression in NIH3T3 cells. Quiescent NIH3T3 cells were serum-stimulated for the indicated length of time. Q corresponds to no serum stimulation. Poly(A)$^+$ RNA (0.2 μg) was fractionated by electrophoresis on a 1% agarose-formaldehyde gel, transferred and hybridized with a probe corresponding to the complete AC16 cDNA. The blot was subsequently rehybridized with the GAPDH probe. The positions of RNA markers are shown.

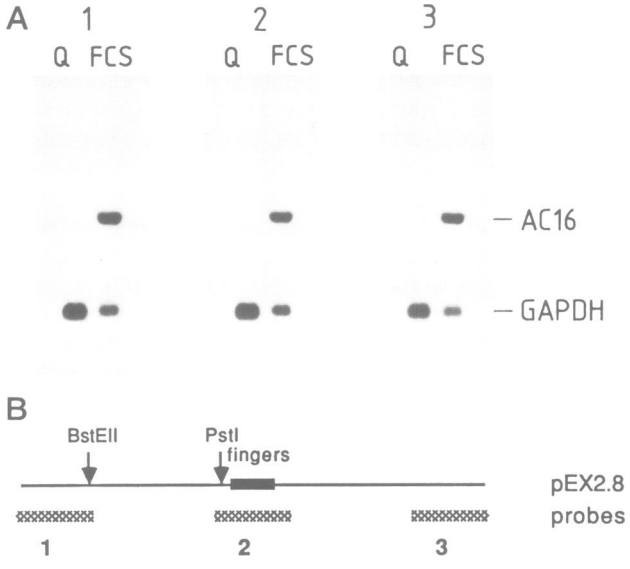


Fig. 5. Detection of *Krox*-20 mRNA with different cDNA probes. **A.** Poly(A)$^+$ RNA samples (0.2 μg) from quiescent cells (Q) or cells serum-stimulated in the presence of cycloheximide for 4 h (FCS) were fractionated by electrophoresis on 1% agarose-formaldehyde gels, transferred and hybridized with the three pEX2.8 non-overlapping probes described at the bottom. The number at the top refers to the probe used. The blots were subsequently rehybridized with the GAPDH probe. **B.** Localization of the different DNA fragments derived from pEX2.8 and used as probes.

a weaker similarity with a segment (amino acids 160−234) of the polyoma large T antigen (Figure 3). This observation may be of interest since this region of the large T is within the amino-terminal 40% part of the molecule described to be sufficient for cell immortalization and reduction in serum level requirement (Rassoulzadegan *et al.*, 1983). Furthermore, the region of the polyoma large T antigen corresponding to amino acids 154 to 220 has been shown to be required for cell immortalization (Asselin and Bastin, 1985). No significant similarity was detected with other proteins present in the data base.

The *Krox*-20 open reading frame contains two other ATG codons downstream and close to the first one (Figure 1). The environments of the two downstream ATG codons show a better fit to the consensus sequence for translational initiation in higher eukaryotes: (GCC)GCCA/GCCATGG (Kozak, 1986, 1987). This raises the possibility that initiation may also occur at the level of the downstream ATGs. Consistent with this idea, we have observed that *in vitro* translation of *Krox*-20 RNA prepared by transcription with T7 RNA polymerase gives rise to at least two polypeptides, slightly differing in size (data not shown).

In addition to the large open reading frame encoding the putative finger protein, a second open reading frame is seen in the cDNA over a length of 527 nucleotides (Figure 1A). This open reading frame overlaps with the large one between nucleotide positions 592 and 1119. It is possible to speculate that alternative splicing of the RNA might incorporate this reading frame in the coding sequence to generate another gene product.

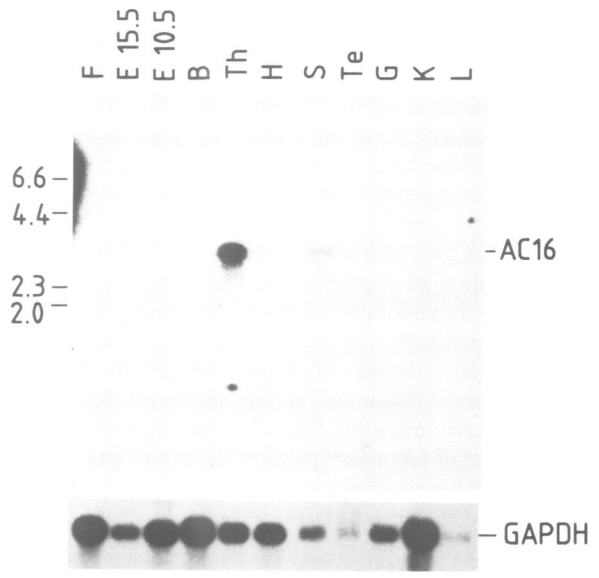


Fig. 6. Northern blot analysis of *Krox*-20 gene expression in adult mouse tissues and whole embryos. Poly(A)$^+$ RNA (2 μg) was fractionated by electrophoresis on a 1% agarose gel and hybridized with a 5′ AC16 probe not containing the finger region (nucleotide positions 525 to 1140). The blot was subsequently rehybridized with the GAPDH probe. The origins of the RNAs are as follows: F: F9 cells, E15: 15.5 day old embryo, E10: 10.5 day old embryo, B: brain, Th: thymus, H: heart, S: spleen, Te: testis, G: gut, K: kidney, L: liver. The positions of λ phage DNA *Hin*dIII restriction fragments are indicated.

Regulation of Krox-20 expression

We have examined the regulation of *Krox*-20 expression during G_0/G_1 transition by Northern blot analysis of poly(A)$^+$ RNA prepared from quiescent NIH3T3 cells exposed to fetal calf serum for different lengths of time. The blot was hybridized with a probe corresponding to the entire AC16 cDNA. An mRNA molecule with a size of ~3.2 kb was detected in cells within 15 min of serum treatment (Figure 4). The amount of RNA increased up to 30 min and decreased to undetectable levels after 2 h. Addition of serum plus cycloheximide to the cells instead of serum alone prevented the decrease in RNA level and led to super-induction (compare Figures 4 and 5, and data not shown). At 30 min the level of *Krox*-20 mRNA was at least 90-fold higher than that of quiescent cells, as estimated by densitometer scanning of the autoradiogram. Subsequent rehybridization of the filter with a probe corresponding to the glyceraldehyde 3-phosphate dehydrogenase (GAPDH), a gene whose expression does not vary during G_0/G_1 transition, confirmed that similar amounts of poly(A)$^+$ RNA were present in the different lanes (Figure 4). The size of the *Krox*-20 mRNA decreased slightly after 1 h of serum induction, suggesting that the poly(A) tail was trimmed off, as observed in the case of other serum-inducible genes (Treisman, 1985; Fort *et al.*, 1987).

Krox-20 belongs to a multigene family encoding homologous finger domains (Chavrier *et al.*, submitted). To exclude the possibility that the band observed in the Northern blot could correspond to cross-hybridization of the finger region of *Krox*-20 with a mRNA corresponding to another member of the gene family, we performed additional Northern blots, with three probes derived respectively from the 5′ region of pEX2.8, from its finger region and from its 3′ non-coding region (Figure 5B). Bands of similar intensities and at identical positions were observed with the three probes on blots containing poly(A)$^+$ RNA from NIH3T3 cells stimulated with serum and cycloheximide for 4 h (Figure 5A). No hybridization was observed with RNA from quiescent cells. Rehybridization of the same blots with the GAPDH probe indicated that similar amounts of poly(A)$^+$ RNA were present in the different lanes (Figure 5A). It is very unlikely that other members of the gene family should be highly similar to *Krox*-20 along the entire cDNA sequence, especially within the 3′ non-coding region. Therefore, our results strongly suggest that the detected mRNA molecule corresponds to the *Krox*-20 gene.

The expression of *Krox*-20 was then examined in different adult mouse organs and in whole embryos using the Northern blotting technique with a 5′ AC16 probe not containing the finger region (nucleotide positions 525 to 1140). An mRNA molecule of the same size as the one detected in NIH3T3 cells was detected only in RNA samples from thymus, spleen and testis (Figure 6). The lowest level of expression was observed in testis and the highest in thymus. The latter, however, was 20- to 50-fold lower than the level observed in serum-stimulated NIH3T3 cells. An identical pattern of expression was observed with a 3′ non-coding AC16 probe (positions 1930 to 2843), suggesting again that cross-hybridization with a mRNA corresponding to another gene was not involved (data not shown). Rehybridization of the blots with the GAPDH probe indicated that tissue-specific detection of *Krox*-20 mRNA was not due to larger amounts of poly(A)$^+$ RNA applied in the respective lanes (Figure 6).

In conclusion, the *Krox*-20 gene is activated transiently during G_0/G_1 transition in fibroblasts. The induction occurs very early after serum addition and does not require *de novo* protein synthesis, since it is not prevented by cycloheximide treatment of the cells. In the mouse, the gene appears to be expressed at low level in a tissue-specific manner.

Discussion

G_0/G_1 transition has been the focus of numerous studies in recent years. It is thought to be a critical regulatory period, involving activation or repression of several master regulatory genes whose expression is essential in the control of cell proliferation and differentiation. Some of these genes are likely to encode DNA-binding factors regulating the expression of secondary genes at the transcriptional level. Among the genes analysed so far and known to be activated during G_0/G_1 transition, none has been shown to encode a transcription factor, although indirect evidence suggests that this might be the case for the proto-oncogenes c-*fos* and c-*myc* (Kingston *et al.*, 1984; Setoyama *et al.*, 1986a, 1986b; Distel *et al.*, 1987; Kaddurah-Daouk, 1987). In this paper, we describe a gene, *Krox*-20 which is activated very early during G_0/G_1 transition. Its similarity to known transcription factors suggests that it could constitute one of these master regulatory genes.

The rapid accumulation of *Krox*-20 mRNA following serum stimulation and the fact that this effect does not require *de novo* protein synthesis suggest that the expression of *Krox*-20 is directly regulated by one or several growth factors present in the serum. The kinetics of accumulation and subsequent disappearance of *Krox*-20 mRNA are similar to those of c-*fos* (Greenberg and Ziff, 1984; Kruijer *et al.*, 1984; Müller *et al.*, 1984). The rapid disappearance of the mRNA after 30 min indicates that it is highly unstable. The stabilization of its level in the presence of cycloheximide suggests that cycloheximide treatment reduces mRNA turnover. These properties are consistent with the high percentage of A and T observed within the 3′ non-translated region and the presence in this region of three copies of the sequence ATTTA (Figure 1B). These characteristics have been found to be common to a number of transiently expressed genes and are thought to mediate selective mRNA degradation (Shaw and Kamen, 1986).

In the mouse, low levels of *Krox*-20 mRNA were detected in thymus, spleen and testis. It is striking that these are organs containing rapidly dividing cells. It is therefore possible that the low level of *Krox*-20 mRNA observed in the entire organ is due to high expression of the gene in a subpopulation of proliferating cells. Alternatively, the pattern of expression of *Krox*-20 might be due to cell-type specificity and may not be related to cell proliferation. Further experiments will be required to decide between these two different possibilities.

The fact that Sp1 and *Krox*-20 contain three zinc fingers and that some of these fingers are highly similar strongly supports the idea that *Krox*-20 encodes a DNA-binding protein with transcription control activity. It is not known, however, to which degree finger similarity implies similarity of the target sites on the DNA. Their particular finger similarity sets Sp1 and *Krox*-20 apart from the other known members of the mouse multigene family encoding *Kr*-type finger structures and suggests that they are part of a subfamily. In this respect, it is interesting to note that a factor prepared from HeLa cells and different from Sp1 has recently been shown to bind to the SV40 21 bp repeats, which contain the Sp1 binding site (Kim *et al.*, 1987). In addition, although we detected only the gene *Krox*-20 in a mouse genomic Southern blot with the AC16 probe, hybridization was observed under non-stringent conditions between a probe corresponding to the 5′ part of AC16 (and not containing the finger region) and genomic clones containing other mouse *Kr*-related sequences (Chavrier *et al.*, submitted). This observation suggested that there are other mouse genes with weak similarity to *Krox*-20 outside of the finger region.

Krox-20 sequences appear to be conserved during higher vertebrate evolution, since we could detect unique sequences hybridizing to the AC16 probe in stringent conditions in human DNA (data not shown). This strong conservation suggests a possible important function of the gene. In contrast, no hybridization was observed with *Drosophila* DNA (data not shown).

In conclusion, activation of *Krox*-20 expression by growth factors, although suggestive, does not necessarily imply that the gene is involved in the control of cell proliferation. However, two other indirect pieces of evidence are consistent with this idea: (i) *Krox*-20 is expressed, although at low level, in tissues in which intense cell proliferation occurs; (ii) the only sequence similarity detected between *Krox*-20 and the other known proteins, apart from the finger proteins, is within a region of the polyoma large T antigen known to be required for the modulation of cellular serum requirements. The availability of *Krox*-20 cDNA now enables us to test directly the possible involvement of the gene in the control of cell proliferation.

Materials and methods

Cell culture, DNA and RNA extraction

NIH3T3 and F9 cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum (FCS). Quiescent state was obtained by maintaining the cells for 72 h in DMEM containing 1% FCS. Serum stimulation was carried out with DMEM containing 20% FCS. When used, cycloheximide was at 10 μg/ml. Total RNA was isolated from cell lines and mouse tissues according to the guanidinium thiocyanate procedure of Chirgwin and collaborators (1979), as modified by Maniatis and collaborators (1982). Poly(A)$^+$ RNA was selected by oligo(dT)-cellulose (Collaborative Research) column chromatography (Maniatis *et al.*, 1982).

cDNA library screening

A λ 1149 and a pEX-1 mouse cDNA library (Almendral *et al.*, submitted) were screened as described (Maniatis *et al.*, 1982), using 10^6 c.p.m./ml of [^{32}P]oligo-labeled AC16 probe (sp. act. $\sim 10^9$ c.p.m./μg). A total of 25 clones were isolated from the two libraries. The inserts were characterized by restriction analysis and Southern blotting. The average size of the inserts was 1.5 kb. Clone pEX2.8, containing the longest insert (2.8 kb) was used for further studies. In addition, the restriction patterns of two other clones, also containing long inserts, were determined and compared to those of AC16 and pEX2.8.

Northern and Southern blots

Poly(A)$^+$ RNA was fractionated by electrophoresis on 1% agarose-formaldehyde gel according to Maniatis and collaborators (1982) and transferred to GeneScreen membrane (Dupont). The membrane was baked for 2 h under vacuum at 80°C and the RNA was cross-linked to the membrane by exposure to UV light (1.6 KJ/m^2) for 2 min. Pre-hybridization (2 h) and hybridization (20 h) were carried out at 65°C in 0.5 M $NaPO_4$ (pH 7.2), 1 mM EDTA, 7% SDS, 50 μg/ml denatured sonicated mouse genomic DNA, using $2-4 \times 10^6$ c.p.m./ml of [^{32}P]oligo-labeled DNA probe (sp. act. $\sim 10^9$ c.p.m./μg). The filters were washed twice at 65°C in 0.1 × SSC, 1% SDS for 30 min. Kodak XAR5 films were exposed for

1–4 days at −70°C, in the presence of an intensifying screen. For rehybridization, the GeneScreen membrane was stripped of the probe by washing for 1 h at 80°C in 1 mM Tris (pH 8), 0.1 mM EDTA, 0.05% SDS.

Southern blots of restriction digests of plasmid, phage and genomic DNA were carried out as described (Southern, 1975), using GeneScreen membrane. Hybridization under stringent conditions was performed as in the case of Northern blots, except that the mouse genomic DNA was omitted from the hybridization buffer. Hybridization under low stringency conditions was carried out at 60°C in 5 × SSPE, 5 × Denhardt's solution containing 0.2% SDS and 100 μg/ml denatured salmon sperm DNA. In this case, washing was for 1 h at 50°C in 2 × SSPE, 0.2% SDS.

DNA sequencing and sequence analysis

DNA fragments to be sequenced were cloned in both orientations in the *Eco*RI or *Bam*HI sites of pUC19. Serial external deletions were created in the inserts according to the following procedure: the plasmids were linearized by opening at the unique *Xba*I site located within the polylinker. The 5′ protruding ends were filled in using T4 DNA polymerase in the presence of 2′-deoxyadenosine 5′-O-[1-thiotriphosphate] (Pharmacia) and dCTP, dGTP and dTTP. This treatment protects the ends against a subsequent digestion with *Exo*III nuclease. The DNAs were then digested with *Sma*I and treated with *Exo*III nuclease (Stratagene) for variable lengths of time. After digestion with mung bean nuclease, the plasmids were circularized by ligation and used to transform HB101 cells. The deleted inserts were subsequently transferred into M13 derived vectors. Single stranded DNA was prepared (Messing, 1983) and the nucleotide sequence was established using the dideoxynucleotide procedure (Sanger *et al.*, 1977) and the Sequenase™ procedure (USB). The entire sequence was read on both strands. Nucleotide and amino acid sequences analyses were carried out using the UWGCG Sequence Analysis Software Package (version 5).

Acknowledgements

We thank H.Jäckle and J.M.Blanchard for the gifts of the *Krüppel* and GAPDH probes, respectively, R.Tjian and J.Kadomaga for the communication of unpublished information, R.Brown and S.Courtneidge for helpful discussions, and P.Argos, R.Cortese, K.Simons, C.Steward and E.Wagner for critical review of the manuscript. P.Chavrier, M.Z., P.L. and A.L. were supported by fellowships from the Association pour la Recherche sur le Cancer, the European Molecular Biology Laboratory (M.Z. and P.L.) and the European Molecular Biology Organization respectively.

References

Almendral,J.M., Huebsch,D., Blundell,P.A., Macdonald-Bravo,H. and Bravo,R. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 1575–1579.
Asselin,C. and Bastin,M. (1985) *J. Virol.*, **56**, 958–968.
Baserga,R. (1985) *The Biology of Cell Reproduction*. Harvard University Press, Cambridge, MA, USA.
Birnstiel,M.L., Busslinger,M. and Strub,K. (1985) *Cell*, **41**, 349–359.
Blumberg,H., Eisen,A., Sledziewski,A., Bader,D. and Young,E.T. (1987) *Nature*, **328**, 443–445.
Bogenhagen,D.F., Sakonju,S. and Brown,D.D. (1980) *Cell*, **19**, 27–35.
Bravo,R., Frank,R., Blundell,P.A. and Macdonald-Bravo,H. (1987a) *Nature*, **326**, 515–517.
Bravo,R., Macdonald-Bravo,H., Müller,R., Hübsch,D. and Almendral,J.M. (1987b) *Exp. Cell Res.*, **170**, 103–115.
Brown,R.S. and Argos,P. (1986) *Nature*, **324**, 215.
Brown,R.S, Sander,C. and Argos,P. (1985) *FEBS Lett.*, **186**, 271–274.
Chirgwin,J.M., Przybyla,A.E., McDonald,R.J. and Rutter,W.J. (1979) *Biochemistry*, **18**, 5294–5299.
Chowdhury,K., Deutsch,U. and Gruss,P. (1987) *Cell*, **48**, 771–778.
Cochran,B.H., Reffel,A.C. and Stiles,C.D. (1983) *Cell*, **33**, 939–947.
Cochran,B.H., Zullo,J., Verma,I.M. and Stiles,C.D. (1984) *Science*, **226** 1080–1082.
Colberg-Poley,A.M., Voss,S.D., Chowdhury,K. and Gruss,P. (1985) *Nature*, **314**, 713–719.
Dayhoff,M.O. (1978) *Atlas of Protein Sequence and Structure*. National Biomedical Research Foundation, Washington, DC, Vol. 5, Suppl. 3.
Distel,R.J., Ro,H.-S., Rosen,B.S., Groves,D.L. and Spiegelman,B.M. (1987) *Cell*, **49**, 835–844.
Duboule,D., Baron,A., Mähl,P. and Galliot,B. (1986) *EMBO J.*, **5**, 1973–1980.
Dynan,W.S. and Tjian,R. (1983a) *Cell*, **35**, 79–87.
Dynan,W.S. and Tjian,R. (1983b) *Cell*, **32**, 669–680.
Engelke,D.R., Ng,S.-Y., Shastry,B.S. and Roeder,G. (1980) *Cell*, **19**, 717–728.
Fort,P., Rech,J., Vie,A., Piechaczyk,M., Bonnieu,A., Jeanteur,P. and Blanchard,J.-M. (1987) *Nucleic Acids Res.*, **15**, 5657–5667.
Frankel,A.D., Berg,J.M. and Pabo,C.O. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 4841–4845.
Gidoni,D., Dynan,W.S. and Tjian,R. (1984) *Nature*, **312**, 409–413.
Ginsberg,A.M., King,B.A. and Roeder,R.G. (1984) *Cell*, **39**, 479–489.
Greenberg,M.E. and Ziff,E.B. (1984) *Nature*, **311**, 433–438.
Hart,C.P., Awgulewitsch,A., Fainsod,A., McGinnis,W. and Ruddle,F.H. (1985) *Cell*, **43**, 9–18.
Jackson,I.J., Schofield,P. and Hogan,B. (1985) *Nature*, **317**, 745–748.
Johnston,M. (1987) *Nature*, **328**, 353–355.
Joyner,A.L., Kornberg,T., Coleman,K.G., Cox,D.R. and Martin,G.R. (1985) *Cell*, **43**, 29–37.
Kaddurah-Daouk,R., Greene,J.M., Baldwin,A.S. and Kingston,R.E. (1987) *Genes and Development*, **1**, 347–357.
Kelly,K., Cochran,B.H., Stiles,C.D. and Leder,P. (1983) *Cell*, **35**, 603–610.
Kim,C.H., Heath,C., Bertuch,A. and Hansen,U. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 6025–6029.
Kingston,R.E., Baldwin,A.S.,Jr and Sharp,P.A. (1984) *Nature*, **312**, 280–282.
Kozak,M. (1986) *Cell*, **44**, 283–292.
Kozak,M. (1987) *J. Mol. Biol.*, **196**, 947–950.
Kruijer,W., Cooper,J.A., Hunter,T. and Verma,I.M. (1984) *Nature*, **312**, 711–716.
Lau,L.F. and Nathans,D. (1985) *EMBO J.*, **4**, 3145–3151.
Lau,L.F. and Nathans,D. (1987) *Proc. Natl. Acad. Sci. USA*, **84**, 1182–1186.
Laughon,A. and Scott,M.P. (1984) *Nature*, **310**, 25–31.
Levine,M., Rubin,G.M. and Tjian,R. (1984) *Cell*, **38**, 667–674.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) In *Molecular Cloning*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Messing,J. (1983) *Methods Enzymol.*, **101**, 20–78.
Miller,J., McLachlan,A.D. and Klug,A. (1985) *EMBO J.*, **4**, 1609–1614.
Müller,R., Bravo,R., Burckhardt,J. and Curran,T. (1984) *Nature*, **312**, 716–720.
Pabo,C.O. and Sauer,R.T. (1984) *Annu. Rev. Biochem.*, **53**, 293–321.
Pardee,A.B., Dubrow,R., Hamlin,J.L. and Kletzien,R.F. (1978) *Annu. Rev. Biochem.*, **47**, 715–750.
Preiss,A., Rosenberg,U.B., Kienlin,A., Seifert,E. and Jäckle,H. (1985) *Nature*, **313**, 27–32.
Prelich,G., Tan,C.K., Kostura,M., Mathews,M.B., So,A.G., Downey, K.M. and Stillman,B. (1987) *Nature*, **326**, 517–520.
Rassoulzadegan,M., Naghashfar,Z., Cowie,A., Carr,A., Grisoni,M., Kamen,R. and Cuzin,F. (1983) *Proc. Natl. Acad. Sci. USA*, **80**, 4354–4358.
Rhodes,D. and Klug,A. (1986) *Cell*, **46**, 123–132.
Rosenberg,U.B., Preiss,A., Seifert,E., Jäckle,H. and Knipple,D.C. (1985) *Nature*, **313**, 703–706.
Rubin,M.R., Toth,L.E., Patel,M.D., D'Estachio,P. and Nguyen-Huu,M.C. (1986) *Science*, **233**, 663–667.
Sakonju,S. and Brown,D.D. (1981) *Cell*, **23**, 665–669.
Sakonju,S. and Brown,D.D. (1982) *Cell*, **31**, 395–405.
Sakonju,S., Bogenhagen,D.F. and Brown,D.D. (1980) *Cell*, **19**, 13–25.
Sanger,F., Nicklen,A.R. and Coulson,A. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
Schuh,R., Aicher,W., Gaul,U., Coté,S., Preiss,A., Maier,D., Seifert,E., Nauber,U., Schröder,C., Kemler,R. and Jäckle,H. (1986) *Cell*, **47**, 1025–1032.
Setoyama,C., Frunzio,R., Liau,G., Mudryj,M. and De Crombrugghe,B. (1986a) *Proc. Natl. Acad. Sci. USA*, **83**, 3213–3217.
Setoyama,C., Hatamochi,A., Peterkofsky,B., Prather,W. and de Crombrugghe,B. (1986b) *Biochem. Biophys. Res. Commun.*, **3**, 1042–1048.
Shaw,G. and Kamen,R. (1986) *Cell*, **46**, 659–667.
Shepherd,J.C.W., McGinnis,W., Carrasco,A.E., De Robertis,E.M. and Gehring,W.J. (1984) *Nature*, **310**, 70–71.
Soeda,E., Arrand,J.R., Smolar,N., Walsh,J.E. and Griffin,B.E. (1980) *Nature*, **283**, 445–453.
Southern,E.M. (1975) *J. Mol. Biol.*, **98**, 503–517.
Stanley,K.K. and Luzio,J.P. (1984) *EMBO J.*, **3**, 1429–1434.
Treisman,R. (1985) *Cell*, **42**, 889–902.
Wolgemuth,D.J., Engelmyer,E., Duggal,R.N., Gizang-Ginsberg,E., Mutter,G.L., Ponzetto,C., Viviano,C. and Zakeri,Z.F. (1986) *EMBO J.*, **5**, 1229–1235.

Received on October 21, 1987