MOLECULAR AND CELLULAR BIOLOGY, Mar. 1988, p. 1319-1326
0270-7306/88/031319-08$02.00/0

Vol. 8, No. 3

Characterization of a Mouse Multigene Family That Encodes Zinc Finger Structures

PHILIPPE CHAVRIER, PATRICK LEMAIRE, OLIVIER REVELANT, RODRIGO BRAVO, AND PATRICK CHARNAY*

Differentiation Programme, European Molecular Biology Laboratory, Postfach 10.2209, 6900 Heidelberg, Federal Republic of Germany

Received 17 August 1987/Accepted 15 December 1987

The *Drosophila* segmentation gene Krüppel encodes multiple tandemly repeated units predicted to form DNA-binding zinc fingers. We have isolated 23 bacteriophages, containing nonoverlapping inserts from a mouse genomic DNA library, on the basis of cross-hybridization under nonstringent conditions to a probe corresponding to the Krüppel finger region. Nucleotide sequence analysis of six phage DNAs indicated that they all contained regions with similarity to Krüppel and potentially encoded zinc finger domains. Within these regions, the level of similarity to Krüppel was particularly high between successive fingers. Northern (RNA) blotting analysis suggested that the mouse sequences belonged to different genes, the expression of some of which was modulated during cell differentiation and development. Hybridization experiments suggested that the similarity between some of the genes extended outside of the finger regions. In conclusion, our data suggest that the mouse genome contains a large family of evolutionarily related genes encoding possible *trans*-acting factors. These genes are likely to play a regulatory role at the transcriptional level.

Identification and cloning of genes involved in the control of differentiation and development in mammals were almost impossible until recently because of the complexity of the genomes and the insufficient power of the available genetic techniques. The discovery of the presence of the homeo box sequence within several *Drosophila melanogaster* homeotic genes and of the conservation of this sequence within mammalian genomes has radically changed the situation (16, 30). The use of *Drosophila* homeo boxes as probes has allowed the cloning of a number of mouse and human genes potentially involved in the regulation of development (1, 8, 9, 12, 18, 20, 22, 24, 28, 31, 38, 51). One characteristic feature of the homeo box is that it encodes a domain with similarity to the helix-turn-helix domain of procaryotic gene-regulatory proteins (27, 30, 34, 45). This suggests that homeo-box-containing genes encode DNA-binding proteins which regulate the expression of other genes at the transcriptional level. The discovery of the homeo box thus demonstrates the potential of the application of *Drosophila* molecular genetics to the study of mammalian development.

Recently, a second class of eucaryotic DNA-binding proteins has emerged from the analysis of transcription factor IIIA (TFIIIA), a protein which interacts with a region approximately 50 nucleotides long, internal to the *Xenopus laevis* 5S RNA gene, and is required for initiation of its transcription (3, 14, 40–42). Analysis of the amino acid sequence revealed the existence of 9 similar tandemly repeated units, consisting of approximately 30 residues and containing 2 cysteines and 2 histidines at invariant positions (5, 17, 33). It was proposed that each of these units folds as an independent domain, centered on a zinc ion, and interacts with about five nucleotides (5, 33, 36). These domains have been referred to as zinc fingers. Finger-type domains were later observed in the amino acid sequences deduced from the nucleotide sequences of several *Saccharomyces cerevisiae* and *D. melanogaster* regulatory genes, suggesting that it might be a common structure among a number of eucaryotic transcription control proteins (4, 19, 37, 48, 49). Recently, in the case of yeast and *X. laevis* regulatory finger proteins, it was shown by genetic analysis that the invariant cysteine and histidine are essential for the regulatory function of the protein (2) and that zinc is an essential cofactor (15, 23).

Krüppel (*Kr*), a gene involved in segmentation control in *Drosophila* embryos, encodes a finger protein (35, 37). *Kr* is a gap gene, and strong *Kr* mutations lead to embryos lacking all the thoracic and five abdominal segments (35, 50). Under low-stringency hybridization conditions, a probe derived from the *Kr* finger-coding sequence hybridized to multiple DNA sequences in the genomes of *D. melanogaster* and other eucaryotes (44). This property was used to isolate and characterize two murine genes containing finger-encoding regions (7). We extended this analysis by isolating from a murine genomic DNA library 23 nonoverlapping lambda recombinant bacteriophages which hybridized to the *Drosophila Kr* finger probe under low-stringency conditions. Characterization of some of the cloned mouse sequences suggested that they belonged to a large family of evolutionarily related genes, which encoded zinc fingers. This property and the similarity with *Kr* suggested that the products of these genes were DNA-binding proteins involved in genetic control at the transcriptional level.

MATERIALS AND METHODS

Cell culture and DNA and RNA extraction. The different cell lines were maintained in Dulbecco modified Eagle medium supplemented with 10% fetal calf serum and 0.2% glucose. F9 cells (25), P19 cells (13), and PC13 cells were obtained from E. Wagner. Drug-induced differentiation of embryonal carcinoma cells was obtained according to Rudniki and McBurney (39). Total RNA was isolated from cell lines and mouse tissues according to the guanidinium thiocyanate procedure of Chirgwin et al. (6), as modified by Maniatis et al. (29). Poly(A)$^+$ RNA was selected by oligo(dT)-cellulose (Collaborative Research, Inc., Waltham, Mass.) column chromatography (29).

* Corresponding author.

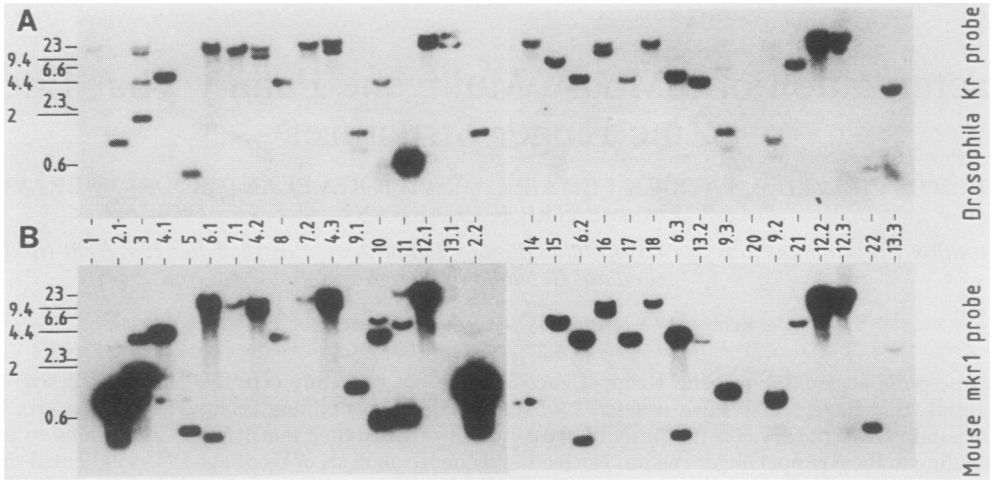

FIG. 1. Southern blotting analysis of the cloned *Kr*-related mouse sequences with *Drosophila* and mouse finger probes. *Bam*HI digests of the different phage DNAs (0.5 to 2 μg) were analyzed by hybridization under low-stringency conditions with a finger probe derived from the *Drosophila Kr* gene (A). The filters were then stripped of the probe and subsequently rehybridized under low-stringency conditions with a finger probe from the mouse gene *mkr1* (B). Phage numbers are indicated. The low amount of phage 20 DNA present on the filter (<0.1 μg) prevented detection of this DNA. Exposure times were 2 days for the *Drosophila* probe and 1 h for the mouse probe.

Genomic library screening and Southern blotting. A lambda EMBL3 mouse genomic DNA library (generous gift of T. Grünfeld) was screened as previously described (29), except for the hybridization conditions. A total of 1.2×10^6 phage plaques were analyzed. Hybridization was performed for 16 h at 60°C in a 5× SSPE –5× Denhardt solution containing 0.2% sodium dodecyl sulfate [SDS] and 100 μg of denatured salmon sperm DNA per ml with 10^6 cpm of ^{32}P-oligo-labeled probe (specific activity, approximately 10^9 cpm/μg) per ml. The nitrocellulose filters were then washed for 1 h at 50°C in 2× SSPE–0.2% SDS. For Southern blotting experiments, restriction fragments from phage DNAs were separated by electrophoresis on 1.2% agarose gel and transferred to a GeneScreen membrane (Du Pont Co., Wilmington, Del.) according to Southern (46). Hybridization and washing were carried out as described for library screening. In some cases, the probe was hybridized to an excess of mouse genomic DNA before filter hybridization to prevent recognition of repetitive sequences.

Northern blotting. Poly(A)$^+$ RNA was fractionated by electrophoresis on a 1% agarose-formaldehyde gel according to Maniatis and co-workers (29) and transferred to a GeneScreen membrane (Du Pont). 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. Prehybridization and hybridization were carried out for 4 and 20 h, respectively, at 65°C in a solution containing 0.5 M $NaPO_4$ (pH 7.2), 1 mM EDTA, 7% SDS, and 50 μg of mouse liver DNA per ml with 2×10^6 to 4×10^6 cpm of ^{32}P-oligo-labeled DNA probe (specific activity, approximately 10^9 cpm/μg). The filters were washed twice at 65°C in 0.1× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate)–1% SDS for 30 min. For rehybridization, the GeneScreen membrane was stripped of the probe by washing for 1 h at 80°C in a solution containing 1 mM Tris (pH 8), 0.1 mM EDTA, and 0.05% SDS. Uniformly labeled, single-stranded RNA probes were derived from a 3.5-kilobase (kb) *Hin*dIII DNA fragment subcloned from phage 6.1 into the pGEM-1 vector (Promega Biotec, Madison, Wis.). In vitro transcriptions by SP6 or T7 RNA polymerases were carried out in the presence of [α-^{32}P]CTP according to the instructions of the manufacturer (Boehringer GmbH, Mannheim, Federal Republic of Germany). Alkali breakage of the RNA probe was performed as previously described (21).

Fingerprint analysis of recombinant phage DNA. The procedure was carried out according to Coulson and co-workers (10). Phage DNA was digested with the restriction enzyme *Hin*dIII or *Bgl*II. The DNA fragments were end labeled with ^{32}P by using the Klenow fragment of DNA polymerase I and were subsequently digested with the restriction enzyme *Sau*3A. Labeled fragments were separated by electrophoresis on polyacrylamide-urea gels, and the sizes of the different fragments corresponding to each phage were estimated using a digitalizer and a program implemented on the EMBL VAX computer by H. Lehrach.

DNA sequencing and sequence analysis. DNA fragments to be sequenced were cloned in M13-derived vectors. Single-stranded DNA was prepared (32) and the nucleotide sequence was established by using the dideoxynucleotide procedure (43) and the Sequenase procedure (USB, Cleveland, Ohio). Nucleotide and amino acid sequences analyses were carried out using the University of Wisconsin Genetics Computer Group (Madison, Wis.) Sequence Analysis Software Package (version 5).

RESULTS

Cloning of mouse genomic sequences related to the finger region of the *Drosophila* gene *Kr*. Southern blotting analysis of mouse genomic DNA, carried out at low stringency with a probe consisting of a 561-base-pair (bp) *Bam*HI-*Sal*I fragment containing the *Kr* finger domain-coding sequence, revealed the presence of a series of hybridizing bands (44). We used the same probe and slightly different hybridization conditions (see Materials and Methods) to screen a mouse genomic DNA phage library. Thirty-five recombinant phages positive after second screening were isolated. DNA was purified from each clone, digested with different restriction enzymes, and analyzed by Southern blotting with the *Kr* finger probe. At least one fragment from each digest gave a positive signal (Fig. 1A). When the blots were subsequently stripped of the *Drosophila* probe and rehybridized with

another finger probe, consisting of a 562-bp *Eco*RI fragment derived from the mouse gene *mkr1* isolated by Chowdhury and co-workers (7), the same bands were detected, although with different relative intensities (Fig. 1B). This suggested that the two probes recognized the same regions. The recognition of at least one sequence in each recombinant phage by two different finger probes strongly supported the idea that the basis for the cross-hybridization was the presence of finger regions within these sequences. The higher intensities of the signals observed with the mouse probe suggested that this sequence was more similar to the detected genomic fragments than was the *Drosophila* probe (Fig. 1). The strong hybridization signal to the *mkr1* probe observed with phages 2.1 and 2.2, as well as the sizes of the hybridizing fragments obtained after digestion with other enzymes (data not shown), suggested that these phages contained sequences from the *mkr1* gene.

The potential overlaps between the different phages were detected by the following procedures. (i) The restriction patterns of the different phage DNAs were compared. (ii) One restriction fragment hybridizing to the *Kr* finger probe was subcloned from each recombinant phage genome. It was subsequently used as a probe in Southern blotting analyses of digests of the different phage DNAs under stringent hybridization conditions. (iii) Similar Southern blotting experiments were carried out with probes derived from the entire DNA of some of the recombinant phages. (iv) A fingerprint analysis was performed on the DNAs of the different phages by the procedure of Coulson et al. (10). These different analyses allowed us to detect the presence of overlapping regions within some of the recombinant phages and to restrict our subsequent work to 23 nonoverlapping phages. The phages have thus been numbered from 1 to 23, identical numbers with different decimals indicating overlapping phages.

Presence of finger motifs in *Kr*-related sequences. To determine whether the sequences which hybridized to the *Kr* finger probe did actually encode finger domains, we subcloned several of them as *Sau*3A restriction fragments into a M13 vector and established their nucleotide sequences. In addition, the *Kr*-related regions from two phage clones isolated from a mouse cDNA library with the *Drosophila Kr* finger hybridization probe (5a) were also sequenced. The first cDNA corresponded to a gene contained in one of the genomic clones, phage 20 (5a). The second cDNA corresponded to a gene contained in phages 4.1, 4.2, and 4.3, as suggested by Southern blotting analysis (data not shown). The different sequences are displayed in Fig. 2. In each case, it was possible to find an open reading frame which, when translated into amino acids, gave rise to imperfect tandem repeats of a 28-amino-acid sequence. These repeats contained an exact fit with the basic consensus for the *Kr* finger: Cys-X_2-Cys-X_3-Phe-X_5-Leu-X_2-His-X_3-His (Fig. 3). Since this motif was observed in seven out of seven fragments derived from six different phages, this strongly supported the idea that most, if not all, the phages isolated on the basis of cross-hybridization with the *Kr* probe contained sequences encoding finger motifs. We will now refer to these sequences and to the putative genes which they are part of, under the generic name of *Krox*, for *Kr* box, followed by a number corresponding to the original phage isolate.

Amino acid sequence comparison of the different finger domains. A comparison of the amino acid sequences of the finger motifs analyzed in the present study, as well as those from two other mouse genes isolated independently by Chowdhury et al. (7), was performed. As mentioned above, a perfect alignment was observed between each of the mouse fingers and the consensus for the *Kr* finger (Fig. 3). The two cysteines (positions 8 and 11) and the two histidines (positions 24 and 28) were conserved in each repeat. The phenylalanine (position 15) was conserved in all but one repeat, and the leucine (position 21) was conserved in all but two repeats. In addition and as expected from the work of Schuh et al. (44), a stretch of six amino acids located between successive fingers, Thr-Gly-Glu-Lys-Pro-Tyr, the so-called H/C link, was highly conserved. Indeed, the conservation of this sequence and of the flanking histidine and cysteine probably constituted the basis for nucleotide sequence conservation allowing cross-hybridization between the different mouse sequences and the *Kr* finger probe.

The similarity observed between the different *Krox* fingers and the *Drosophila Kr* fingers was essentially limited to the elements mentioned above. In contrast, more extensive similarity was noticed between the different mouse fingers. Thus, while the levels of similarity between the *Kr* fingers and the mouse fingers were on the order of 40%, these levels were in the range of 50 to 80% when fingers from different mouse genes were compared (Fig. 3 and data not shown). There was, however, one exception, *Krox-20*, whose fingers appeared as distant as *Kr* fingers from the other mouse fingers. In particular, several amino acids highly conserved among the other mouse genes, like Glu-10, Lys-13, and Ser-19, were not observed in the *Krox-20* sequence (Fig. 3).

Interestingly, an even higher similarity was observed between the different fingers encoded by each particular gene (Fig. 3). For instance, for gene *Krox-6*, several fingers had identical or almost identical sequences.

Nucleotide sequence similarity outside of the finger-encoding regions. To determine whether *Kr*-related genes contained similar sequences outside of the finger-encoding regions, we made use of the cDNA clone corresponding to the gene *Krox-20* (5a). A series of 3′ external deletions were created in the cDNA with exonuclease III and mung bean nuclease (5a). One clone retaining 615 bp of *Krox-20* sequences but with the finger region deleted was selected. It contained the 5′ part of the coding region of the gene with the 3′ end of the fragment being located 110 bp 5′ to the finger region (5a). Southern blotting experiments were carried out with digests of the different genomic clones of *Kr*-related sequences by using the purified 615-bp fragment as a probe under low-stringency hybridization conditions (Fig. 4). In addition to phage 20, which also hybridized to the probe under stringent conditions (data not shown), several phage DNAs gave rise to fragments which hybridized to the probe (Fig. 4). This suggested that they contained domains with similarity to the non-finger 5′ part of the coding region of the *Krox-20* gene and that within these domains the different genes were related to one another.

Expression of *Kr*-related sequences. To determine whether the different *Kr*-related sequences were expressed as RNA in mouse cells, we carried out Northern blotting experiments by using different subclones containing these sequences as hybridization probes. If mouse *Kr*-related sequences play a role in regulating development of embryos similar to that of the *Drosophila* gene, they should be expressed during early development. Therefore, probes containing the *Kr*-related sequences and derived from nine phages were hybridized with poly(A)$^+$ RNA extracted from different mouse embryonal carcinoma cell lines, as well as different mouse organs. With seven of these probes, described in the legend to Fig. 5, one or several discrete bands were observed, suggesting that most of the *Kr*-related sequences were transcribed and were

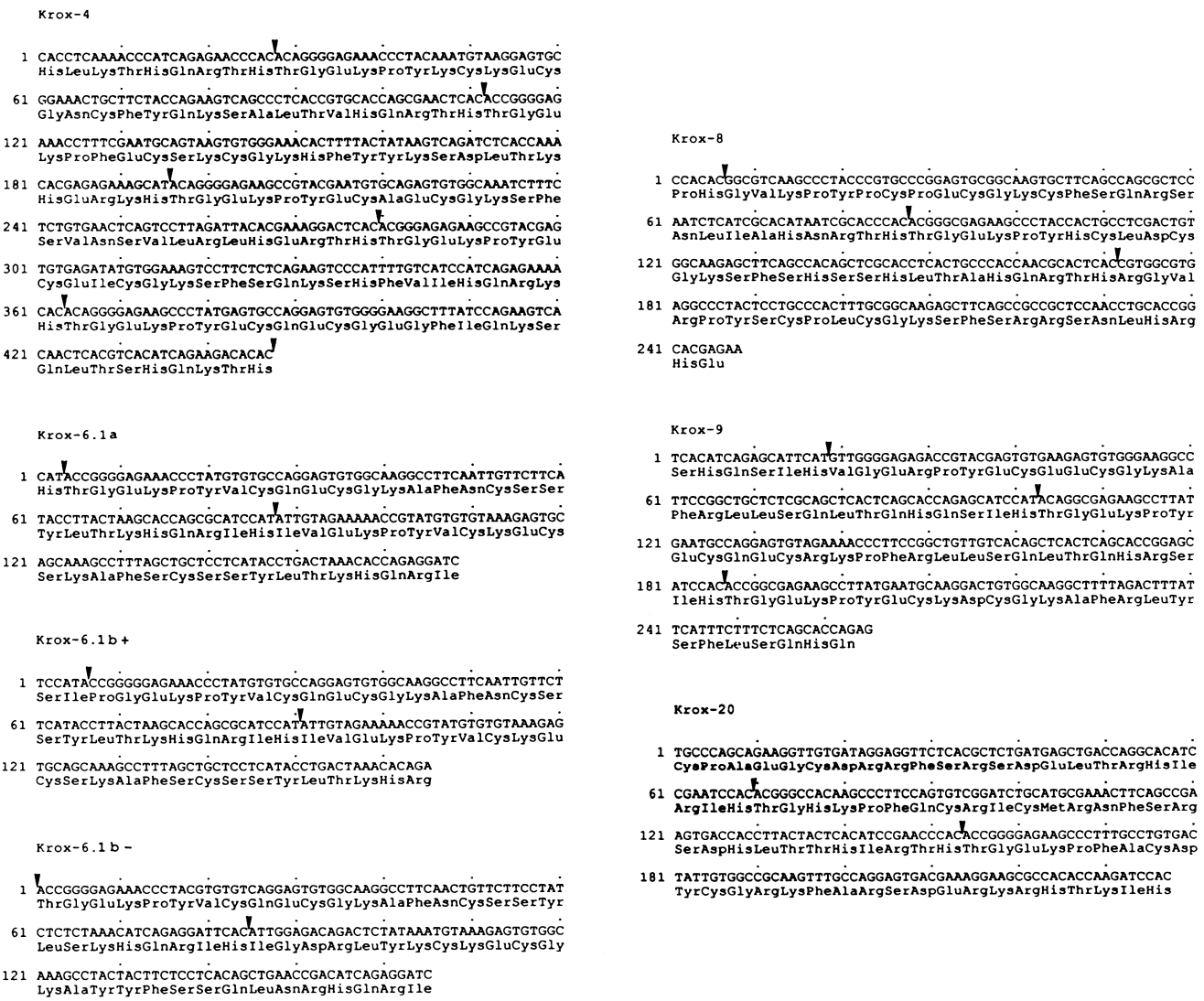

FIG. 2. Nucleotide sequence of *Kr*-related mouse DNA regions. Portions of fragments containing the region of cross-hybridization with the *Kr* finger probe were sequenced. The number after Krox refers to the phage from which the fragment was isolated, and the + and − arbitrarily refer to the extremity of the fragment. Two different fragments, 6.1a and 6.1b, were sequenced from phage 6.1. Sequences 4 and 20 were derived from cDNA clones corresponding to the genomic phages 4.1 and 20, respectively (5a). Each nucleotide was sequenced at least two times. The sequences of fragments, 4, 6.1a, and 20 were established on both strands. Sequence 9 was obtained independently from homologous fragments derived from phages 9.1 and 9.2. The translation of the reading frame giving rise to a finger motif is presented. The arrowheads mark the beginnings of the different finger motifs.

part of mouse genes (Fig. 5; 5a). In addition, with RNA derived from the same cell line or the same organ, the different probes gave rise to bands corresponding to RNA molecules of different sizes (Fig. 5). This suggested that the different probes did not recognize the same RNA species and therefore corresponded to distinct genes. In all cases, the signals observed were weak, indicating a low level of expression of the genes. The amount and the quality of the RNA present in the different lanes were analyzed by subsequent rehybridization with a probe corresponding to the ubiquitously expressed gene for the glyceraldehyde 3-phosphate dehydrogenase (Fig. 5B and data not shown).

For probe 15, a unique band corresponding to an RNA molecule of 1.2 kb was observed in F9 cells (Fig. 5A). A stronger band, corresponding to an RNA molecule of the same size, was detected in F9 cells that had differentiated into parietal endoderm cells, suggesting that the expression of the gene was regulated during cell differentiation. Among the four genes whose expression was analyzed in mouse tissues, *Krox-5* and *Krox-8* appeared to be expressed in a ubiquitous manner; probe 5 detected a unique transcript of 2.6 kb, and probe 8 detected two RNA transcripts, 1.9 and 4.6 kb long (Fig. 5B). The other two genes showed tissue-specific patterns of expression. Gene *Krox-20* gave rise to a unique 3.2-kb mRNA detected specifically in thymus, and to lower levels in spleen and testis (5a). Probe 6 revealed a 1.8-kb transcript, mainly restricted to liver and kidney. In the latter case, the RNA analysis was repeated with single-stranded RNA probes. While no RNA was detected by the sense probe (data not shown), the same 1.8-kb transcript was detected by the antisense probe (Fig. 5B). This indicated that the expressed RNA corresponded to the sense strand. In conclusion, our data suggested that most *Krox* genes were expressed, although at low levels, and that at least some of

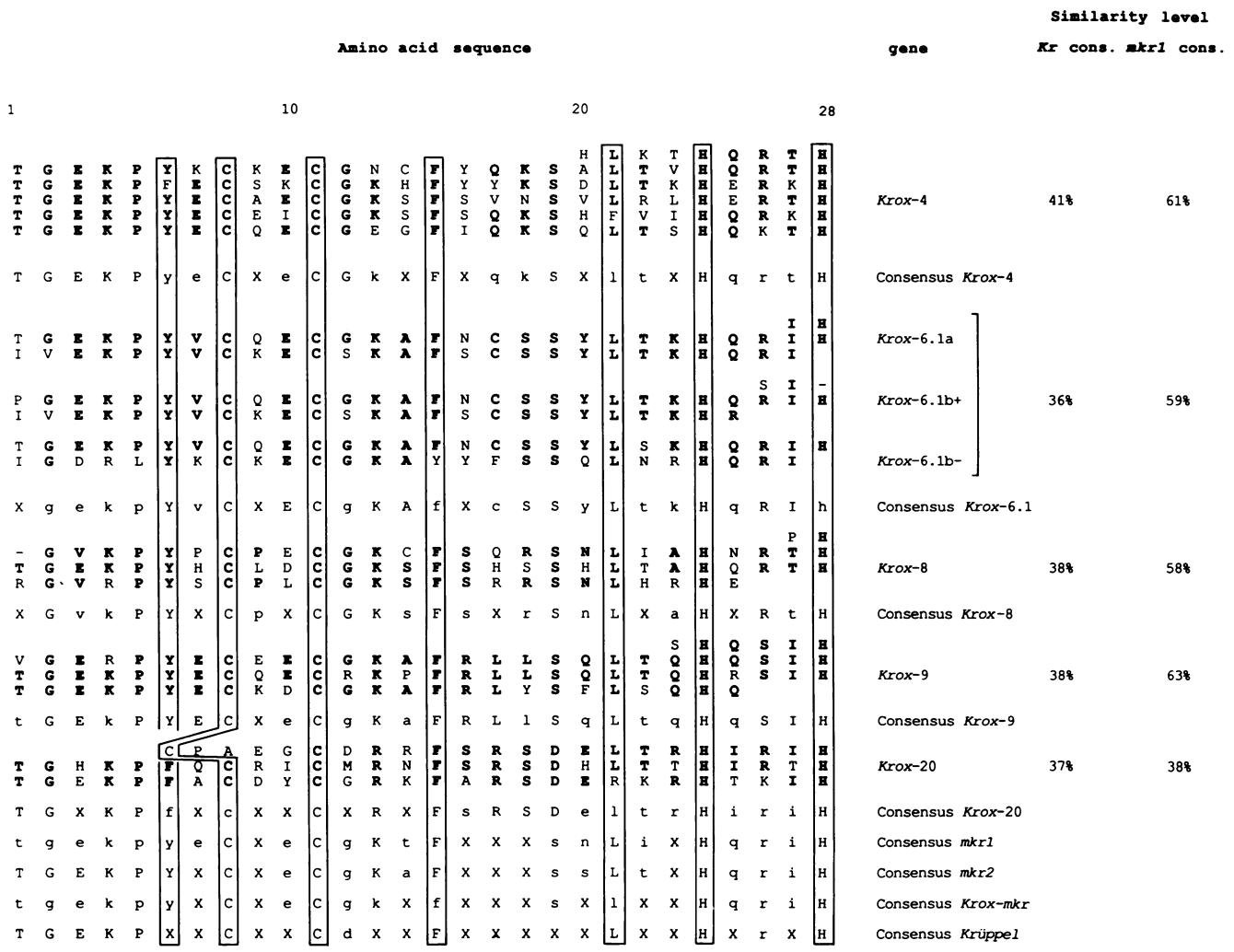

FIG. 3. Amino acid sequence comparison of mouse finger motifs. The sequences are in the one-letter code (11). They are aligned to show the repeated units. The positions of the amino acids which are highly conserved in the TFIIIA motif are boxed. For each gene, a consensus for the repeated motif is displayed; capital letters indicate strictly conserved amino acids, lowercase letters indicate amino acids conserved over 50%, and X indicates no conservation over 50%. Within each sequence, amino acids identical to the consensus are shown in bold characters. Consensus for *mkr1* and *mkr2* and for *Kr* are derived from the works of Chowdhury et al. (7) and Schuh et al. (44), respectively. A general consensus for the known mouse genes, including *mkr1*, *mkr2*, and those presented in this work, is also displayed. The similarity levels (percent positional identity) between the different *Krox* sequences and either the *Kr* consensus (*Kr* cons.) or the *mkr1* consensus (*mkr1* cons.) are presented.

them were regulated during cell differentiation and development.

DISCUSSION

In their publications proposing the zinc finger model for the structure of TFIIIA and describing the presence of a pattern within the DNA sequence recognized by TFIIIA, Klug and co-workers (33, 36) predicted that the finger structure might be largely used by eucaryotic organisms in DNA-binding domains of transcription control factors. Indeed, there is now a growing list of regulatory genes from yeast and *D. melanogaster* which encode finger motifs of the TFIIIA type, including in particular the two *Drosophila* segmentation gap genes sequenced so far, *Kr* and Hunchback (37, 47). In the present work, we provide evidence for the existence of a large family of evolutionarily related genes containing sequences with similarity to the finger region of *Kr* and which we have named *Krox* genes. The capacity of encoding zinc fingers similar to those of TFIIIA and the close similarity with *Kr* suggest that these mouse genes encode proteins with DNA-binding activity and are involved in transcriptional control of gene expression.

The conservation of the *Kr* finger motif among mouse *Krox* genes is striking. Unlike the case of TFIIIA, the spacing between the key amino acids (Cys, Phe, Leu, and His) is rigorously maintained. In addition, the sequence of the H/C link located between successive fingers is also highly conserved. This suggests that the mouse genes detected with the *Drosophila Kr* probe constitute only a small subset of the genes encoding TFIIIA-type fingers, probably those containing multiple fingers with the conserved H/C link in between. The possible function of the H/C link, which might justify its strong conservation, is unknown. Does it merely constitute a structural component, necessary for the correct positioning of the fingers, or does it play a more

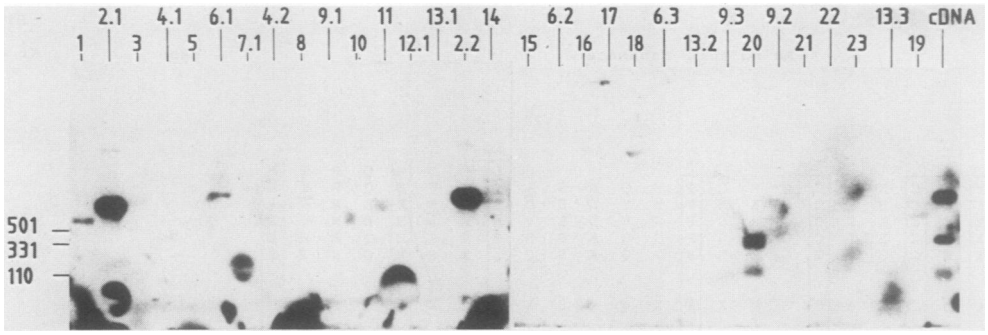

FIG. 4. Southern blotting analysis of cloned *Kr*-related sequences with the non-finger probe derived from gene *Krox-20*. *Sau*3A digests of different phage DNAs (approximately 1 μg) were analyzed by hybridization under low-stringency conditions with a non-finger probe derived from a cDNA clone corresponding to gene *Krox-20*. Phage numbers are indicated. Lane cDNA corresponds to *Sau*3A digests of the *Krox-20* cDNA clone (10 ng). The amount of phage DNA is only approximately 0.1 μg for phage 20. The numbers to the left of the gel indicate the lengths of size markers in base pairs.

active role, interacting with the DNA backbone or constituting a target site for specific recognition factors? Also unclear is the reason for the conservation of the precise *Kr* spacing between critical amino acids, in addition to the conservation of the H/C link.

Comparison of amino acid sequences of the finger motifs of the different mouse genes indicates an even higher similarity, involving conservation of several amino acids, in addition to the critical amino acids of the finger structure. This observation suggests that the different *Kr*-related mouse genes have appeared by successive duplications of a common ancestor gene late after the divergence between protostomes and deuterostomes, marking the separation of future insects and mammals (26), or that the mouse *Krox* genes have evolved in a concerted manner after the duplications of the ancestor gene.

One gene, *Krox-20*, has a finger amino acid sequence less similar to the sequences of the other mouse fingers. This might constitute an indication of the existence of subfamilies within the *Krox* gene family. This idea is supported by our observation of a high similarity between the fingers of *Krox-20* and those of the transcription factor Sp1 (5a). In addition, we have recently isolated a cDNA corresponding to another *Kr*-related gene, which we have called *Krox-24*. *Krox-24* encodes zinc fingers highly similar to those of *Krox-20*, although the genes are clearly distinct (Lemaire et al., manuscript in preparation). The existence of *Krox* gene subfamilies is also supported by our finding of possible similarity between *Krox-20* and some of the other *Krox* genes outside of the finger region. The latter observation suggests that the products corresponding to these genes share similar domain(s), in addition to the DNA-binding domain. Indeed, *Krox-24* presents several regions of strong similarity with *Krox-20* outside the finger domains as determined by direct comparison of amino acid sequences derived from the nucleotide sequences (Lemaire et al., submitted).

The very high similarity of the sequences of the different fingers corresponding to particular genes (*Krox-6*, for example) suggests that these fingers might recognize either identical or very similar DNA sequences, if the finger amino acid sequence dictates the DNA recognition sequence as expected. In such a case, the entire DNA target site would be constituted by more or less perfect repetitions of a unique motif.

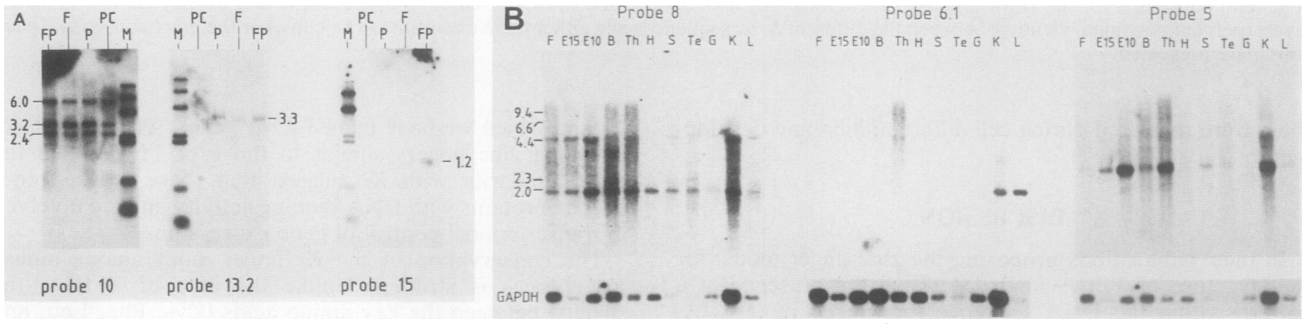

FIG. 5. Northern blotting analysis of the transcription of *Kr*-related sequences. (A) Poly(A)$^+$ RNA (5 μg) from mouse embryonal carcinoma cell lines was analyzed by hybridization under stringent conditions with different probes as indicated. The origin of the RNA is as follows: F, F9 cells; FP, F9 cells differentiated into parietal endoderm cells; P, P19 cells; PC, PC13 cells; M, lambda DNA digested with *Hin*dIII. The sizes of the RNA molecules are indicated. (B) Poly(A)$^+$ RNA (2 μg) from different mouse organs or whole embryos was analyzed with probes derived from phages 8, 6.1, and 5. The origin of the RNA is 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 blots were subsequently stripped of the *Krox* probe and rehybridized with a glyceraldehyde 3-phosphate dehydrogenase (GAPDH) probe. The positions of lambda DNA *Hin*dIII restriction fragments are indicated. The different *Krox* probes were fragments containing *Kr*-related sequences and derived from the genomic phage clones as follows: probe 5, an 8-kb *Eco*RI fragment; probe 8, a 6-kb *Bam*HI fragment; probe 10, a 3.2-kb *Eco*RI fragment; probe 13.2, a 3-kb *Eco*RI fragment; probe 15: a 2.7-kb *Bgl*II fragment; for gene *Krox-6*, the probe was a single-stranded RNA derived from a 3.5-kb *Hin*dIII fragment and corresponding to the antisense orientation.

A majority of the *Krox* genes tested (seven of nine) appear to be expressed in embryonal carcinoma cells or in mouse tissues. The probes hybridized to RNAs of different sizes, suggesting that the genes are distinct. Two of the probes (8 and 10) detected several transcripts. These RNA molecules might encode related gene products. The level of at least one of the RNAs (*Krox-15*) varies during differentiation of F9 cells, suggesting that the expression of the gene is regulated. In addition, analysis of mouse tissue RNA with four of the probes indicates tissue-specific expression of two of them (*Krox-6* and *Krox-20*). In the cases of probes 5 and 6, detection of the same transcripts in polysomal and whole-cell poly(A)$^+$ liver-derived RNAs indicated that the hybridizing sequences are not contained within introns (data not shown). Since the *Krox* genes belong to a multigene family, we cannot exclude that some of the bands detected in Northern blots are due to cross-hybridization of the probe with RNAs derived from other genes. Indeed, definitive attribution of an RNA molecule to a particular gene in the family will require cloning and sequencing of the corresponding cDNA, as we did for *Krox-20*. However, several observations argue in favor of specific detection by the different probes. (i) No cross-hybridization between the different probes and the nonrelated phages from our collection of genomic clones was detected, except within highly repetitive regions (data not shown). (ii) The probes derived from *Krox-5*, *Krox-6*, *Krox-8*, and *Krox-20* detected only unique restriction fragments corresponding to the homologous sequences in Southern blotting experiments carried out with genomic DNA (data not shown). (iii) For gene *Krox-20* (5a) and *Krox-24* (Lemaire et al., submitted), identical transcripts were detected in Northern blotting experiments with probes derived from different portions of the cDNAs, in particular with probes which did not contain any finger sequences (5a; Lemaire et al., submitted). Whether or not all the RNA transcripts detected are the exact homologs of the probes, a conservative interpretation of our results is that some of the *Krox* genes are expressed in a tissue-specific manner.

In conclusion, we have cloned a series of genes which belong to a large multigene family. Since several genes are apparently represented only once in our collection, the total number of genes is likely to exceed the number of our nonoverlapping clones (23 clones). In addition, use of mouse *Krox* finger hybridization probes might give access to other genes not detected with the *Drosophila* probe. As a result of their similarity to TFIIIA and *Kr*, the *Krox* genes are likely to encode transcription control factors. The determination of their precise functions and of the signification of their evolutionary relationship will constitute a challenging task.

ACKNOWLEDGMENTS

We thank H. Jäckle for the gift of the *Kr* finger probe, P. Gruss for the *mkr1* finger probe, T. Grünfeld and B. Herrmann for the mouse genomic library, members of A.-M. Frischauf' and H. Lehrach's group for technical help, L. Henry for excellent technical assistance, and S. Courtneidge, E. Wagner and P. Argos for critical reading of the manuscript.

P. Chavrier and P.L. were supported by fellowships from the Association pour la Recherche sur le Cancer and the European Molecular Biology Laboratory, respectively.

LITERATURE CITED

1. **Awgulewitsch, A., M. F. Utset, P. H. Charles, W. McGir, and F. H. Ruddle.** 1986. Spatial restriction in expression of a mouse homoeo box locus central nervous system. Nature (London) **320**:328–335.
2. **Blumberg, H., A. Eisen, A. Sledziewski, D. Bader, and E. T. Young.** 1987. Two zinc fingers of a yeast regulatory protein shown by genetic evidence to be essential for its function. Nature (London) **328**:443–445.
3. **Bogenhagen, D. F., S. Sakonju, and D. D. Brown.** 1980. A control region in the center of the 5S RNA gene directs specific initiation of transcription. II. The 3' border of the region. Cell **19**:27–35.
4. **Brown, R. S., and P. Argos.** 1986. Fingers and helices. Nature (London) **324**:215.
5. **Brown, R. S., C. Sander, and P. Argos.** 1985. The primary structure of transcription factor TFIIIA has 12 consecutive repeats. FEBS Lett. **186**:271–274.

5a. **Chavier, P., M. Zerial, P. Lemaine, J. Almendral, R. Bravo, and P. Charnay.** 1988. A gene encoding a protein with zinc fingers is activated during G0/G1 transition in cultured cells. EMBO J. **7**:29–35.

6. **Chirgwin, J. M., A. E. Przybyla, R. J. McDonald, and W. J. Rutter.** 1979. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry **18**:5294–5299.
7. **Chowdhury, K., U. Deutsch, and P. Gruss.** 1987. A multigene family encoding several "finger" structures is present and differentially active in mammalian genomes. Cell **48**:771–778.
8. **Colberg-Poley, A. M., S. D. Voss, K. Chowdhury, and P. Gruss.** 1985. Structural analysis of murine genes containing homoeo box sequences and their expression in embryonal carcinoma cells. Nature (London) **314**:713–719.
9. **Colberg-Poley, A. M., S. D. Voss, K. Chowdhury, C. L. Stewart, E. F. Wagner, and P. Gruss.** 1985. Clustered homeo boxes are differentially expressed during murine development. Cell **43**:39–45.
10. **Coulson, A., J. Sulston, S. Brenner, and J. Karn.** 1986. Toward a physical map of the genome of the nematode Caenorhabditis elegans. Proc. Natl. Acad. Sci. USA **83**:7821–7825.
11. **Dayhoff, M. O.** 1978. Atlas of protein sequence and structure, vol. 5, suppl. 3. National Biomedical Research Foundation, Washington, D.C.
12. **Duboule, D., A. Baron, P. Mähl, and B. Galliot.** 1986. A new homeo-box is present in overlapping cosmid clones which define the mouse HOX-1 locus. EMBO J. **5**:1973–1980.
13. **Edwards, M. K. S., and M. W. McBurney.** 1983. The concentration of retinoic acid determines the differentiated cell types formed by a teratocarcinoma cell line. Dev. Biol. **98**:187–191.
14. **Engelke, D. R., S.-Y. Ng, B. S. Shastry, and G. Roeder.** 1980. Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes. Cell **19**:717–728.
15. **Frankel, A. D., J. M. Berg, and C. O. Pabo.** 1987. Metal-dependent folding of a single zinc finger from transcription factor IIIA. Proc. Natl. Acad. Sci. USA **84**:4841–4845.
16. **Gehring, W. J.** 1985. Homeotic genes, the homeo box, and the genetic control of development. Cold Spring Harbor Symp. Quant. Biol. **50**:243–251.
17. **Ginsberg, A. M., B. A. King, and R. G. Roeder.** 1984. Xenopus 5S gene transcription factor, TFIIIA: characterization of a cDNA clone and measurement of RNA levels throughout development. Cell **39**:479–489.
18. **Hart, C. P., A. Awgulewitsch, A. Fainsod, W. McGinnis, and F. H. Ruddle.** 1985. Homeo box gene complex on mouse chromosome 11: molecular cloning, expression in embryogenesis, and homology to a human homeo box locus. Cell **43**:9–18.
19. **Hartshorne, T. A., H. Blumberg, and E. T. Young.** 1986. Sequence homology of the yeast regulatory protein ADR1 with Xenopus transcription factor TFIIIA. Nature (London) **320**:283–287.
20. **Hauser, C. A., A. L. Joyner, R. D. Klein, T. K. Learned, G. R. Martin, and R. Tjian.** 1985. Expression of homologous homeo-box-containing genes in differentiated human teratocarcinoma cells and mouse embryos. Cell **43**:19–28.
21. **Hofer, E., R. Hofer-Warbinek, and J. E. Darnell.** 1982. Globin RNA transcription: a possible termination site and demonstration of transcriptional control correlated with altered chromatin structure. Cell **29**:887–893.

22. **Jackson, I. J., P. Schofield, and B. Hogan.** 1985. A mouse homoeo box gene is expressed during embryogenesis and in adult kidney. Nature (London) **317**:745–748.
23. **Johnston, M.** 1987. Genetic evidence that zinc is an essential cofactor in the DNA binding domain of GAL4 protein. Nature (London) **328**:353–355.
24. **Joyner, A. L., T. Kornberg, K. G. Coleman, D. R. Cox, and G. R. Martin.** 1985. Expression during embryogenesis of a mouse gene with sequence homology to the *Drosophila* engrailed gene. Cell **43**:29–37.
25. **Kahan, B., and E. D. Adamson.** 1983. A teratocarcinoma-derived bipotential cell line with primitive endoderm properties. Cold Spring Harbor Conf. Cell Proliferation **10**:131–141.
26. **Keeton, W. T.** 1980. Biological science. W. W. Norton & Co., Inc., New York.
27. **Laughon, A., and M. P. Scott.** 1984. Sequence of a *Drosophila* segmentation gene: protein structure homology with DNA-binding proteins. Nature (London) **310**:25–31.
28. **Levine, M., G. M. Rubin, and R. Tjian.** 1984. Human DNA sequences homologous to a protein coding region conserved between homeotic genes of *Drosophila*. Cell **38**:667–674.
29. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
30. **McGinnis, W., R. L. Garber, J. Wirz, A. Kuroiwa, and W. J. Gehring.** 1984. A homologous protein-coding sequence in *Drosophila* homeotic genes and its conservation in other metazoans. Cell **37**:403–408.
31. **McGinnis, W., C. P. Hart, W. J. Gehring, and F. H. Ruddle.** 1984. Molecular cloning and chromosome mapping of a mouse DNA sequence homologous to homeotic genes of *Drosophila*. Cell **38**:675–680.
32. **Messing, J.** 1983. New M13 vectors for cloning. Methods Enzymol. **101**:20–78.
33. **Miller, J., A. D. McLachlan, and A. Klug.** 1985. Repetitive zinc-binding domains in the protein transcription factor IIIA from Xenopus oocytes. EMBO J. **4**:1609–1614.
34. **Pabo, C. O., and R. T. Sauer.** 1984. Protein-DNA recognition. Annu. Rev. Biochem. **53**:293–321.
35. **Preiss, A., U. B. Rosenberg, A. Kienlin, E. Seifert, and H. Jäckle.** 1985. Molecular genetics of *Krüppel*, a gene required for segmentation of the *Drosophila* embryo. Nature (London) **313**: 27–32.
36. **Rhodes, D., and A. Klug.** 1986. An underlying repeat in some transcriptional control sequences corresponding to half a double helical turn of DNA. Cell **46**:123–132.
37. **Rosenberg, U. B., C. Schröder, A. Preiss, A. Keinlin, S. Coté, I. Riede, and H. Jäckle.** 1986. Structural homology of the product of the *Drosophila Krüppel* gene with Xenopus transcription factor IIIA. Nature (London) **319**:336–339.
38. **Rubin, M. R., L. E. Toth, M. D. Patel, P. D'Eustachio, and M. C. Nguyen-Huu.** 1986. A mouse homeo box gene is expressed in spermatocytes and embryos. Science **233**:663–667.
39. **Rudnicki, M. A., and M. W. McBurney.** 1986. Cell culture methods and induction of differentiation of embryonal carcinoma cell lines, p. 19–37. *In* E. J. Robertson (ed.), Teratocarcinomas and embryonic stem cells. A. practical approach. IRL Press, Oxford.
40. **Sakonju, S., D. F. Bogenhagen, and D. D. Brown.** 1980. A control region in the center of the 5S RNA gene directs specific initiation of transcription. I. The 5′ border of the region. Cell **19**:13–25.
41. **Sakonju, S., and D. D. Brown.** 1981. The binding of a transcription factor to deletion mutants of a 5S ribosomal RNA gene. Cell **23**:665–669.
42. **Sakonju, S., and D. D. Brown.** 1982. Contact points between a positive transcription factor and the Xenopus 5S RNA gene. Cell **31**:395–405.
43. **Sanger, F., A. R. Nicklen, and A. Coulson.** 1977. DNA sequencing with chain-terminating inhibitors. Proc. Natl. Acad. Sci. USA **74**:5463–5467.
44. **Schuh, R., W. Aicher, U. Gaul, S. Coté, A. Preiss, D. Maier, E. Seifert, U. Nauber, C. Schröder, R. Kemler, and H. Jäckle.** 1986. A conserved family of nuclear proteins containing structural elements of the finger protein encoded by *Krüppel*, a *Drosophila* segmentation gene. Cell **47**:1025–1032.
45. **Shepherd, J. C. W., W. McGinnis, A. E. Carrasco, E. M. De Robertis, and W. J. Gehring.** 1984. Fly and frog homoeo domains show homologies with yeast mating type regulatory proteins. Nature (London) **310**:70–71.
46. **Southern, E. M.** 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.
47. **Tautz, D., R. Lehmann, H. Schnürch, P. Schuh, E. Seifert, A. Keinlin, J. Jones, and H. Jäckle.** 1987. Finger protein of novel structure encoded by hunchback, a second member of the gap class of *Drosophila* segmentation genes. Nature (London) **327**: 383–389.
48. **Vincent, A.** 1986. TFIIIA and homologous genes. The 'finger' proteins. Nucleic Acids Res. **14**:4385–4391.
49. **Vincent, A., H. V. Colot, and M. Rosbash.** 1985. Sequence and structure of the serendipity locus of *D. melanogaster*. J. Mol. Biol. **186**:149–166.
50. **Wieschaus, E., C. Nusslein-Volhard, and H. Kluding.** 1984. *Krüppel*, a gene whose activity is required early in the zygotic genome for normal embryonic segmentation. Dev. Biol. **104**: 172–186.
51. **Wolgemuth, D. J., E. Engelmyer, R. N. Duggal, E. Gizang-Ginsberg, G. L. Mutter, C. Ponzetto, C. Viviano, and Z. F. Zakeri.** 1986. Isolation of a mouse cDNA coding for a developmentally regulated, testis-specific transcript containing homeo box homology. EMBO J. **5**:1229–1235.