Proc. Natl. Acad. Sci. USA
Vol. 90, pp. 7628–7631, August 1993
Genetics

Six members of the mouse forkhead gene family are developmentally regulated

KLAUS H. KAESTNER, KANG-HUN LEE, JOHANNES SCHLÖNDORFF, HOLGER HIEMISCH, A. PAULA MONAGHAN, AND GÜNTHER SCHÜTZ

Division Molecular Biology of the Cell I, German Cancer Research Center, Im Neuenheimer Feld 280, D-69120 Heidelberg, Germany

Communicated by Wolfgang Beermann, May 20, 1993

ABSTRACT **The 110-aa forkhead domain defines a class of transcription factors that have been shown to be developmentally regulated in *Drosophila melanogaster* and *Xenopus laevis*. The forkhead domain is necessary and sufficient for target DNA binding as shown for the rat hepatic nuclear factor 3 (HNF3) gene family. We have cloned six forkhead gene family members from a mouse genomic library in addition to the mouse equivalents of the genes for HNF3α, -β, and -γ. The six genes, termed *fkh-1* to *fkh-6*, share a high degree of similarity with the *Drosophila* forkhead gene, having 57–67% amino acid identity within the forkhead domain. *fkh-1* seems to be the mammalian homologue of the *Drosophila FD1* gene, as the sequences are 86% identical. *fkh-1* to *fkh-6* show distinct spatial patterns of expression in adult tissues and are expressed during embryogenesis.**

Eukaryotic transcription factors have been divided into several classes according to their characteristic DNA binding domains. These include bZip proteins (containing a basic domain and a leucine zipper), homeobox- and POU-homeodomain-containing proteins, zinc-finger proteins, and the helix–loop–helix proteins (for review, see refs. 1 and 2). The forkhead domain is a highly conserved 110-aa DNA binding region found in a distinct class of transcription factors (3). This domain was named after the protein encoded by the region-specific homeotic *Drosophila* gene forkhead, a gene that is required for the proper formation of the terminal structures of the *Drosophila* embryo (4, 5). The forkhead gene is expressed in ectodermal and endodermal portions of the gut, the yolk nuclei, the salivary glands, and certain cells of the central nervous system. In forkhead mutants, the development of all these tissues is affected, consistent with the proposed role of forkhead as a developmental regulatory gene (5). Over the past 3 years, forkhead-related genes have been described in species ranging from yeast to man (6–15). An example of a developmentally important forkhead gene family member is the *Xenopus laevis* gene *XFD1* [ref. 9; also termed *XFKH1* (10) or pintallavis (11)]. This activin-inducible gene was found to be expressed in the blastopore lip at the onset of gastrulation and was suggested to play a role in the initiation of axis formation.

The functional importance of the forkhead DNA binding region was first delineated through DNA binding assays using deletion mutants of the rat hepatic nuclear factor 3α (HNF3α) (6). A region in the amino-terminal half of the protein was found to be essential for binding of HNF3α to its target site in the transthyretin promoter. This DNA binding region is remarkably well conserved in the *Drosophila* forkhead gene (86% identical amino acids over 110 residues, ref. 3). This fact and the finding that the HNF3α, -β, and -γ transcripts, like those of forkhead, are expressed very early in embryogenesis and are expressed in tissues derived from the primitive gut and neural ectoderm (A.P.M., K.H.K., and G.S., unpublished observations) suggest a role in early embryonic development for the members of the HNF3 family.

The discovery of forkhead gene families in *Drosophila* (12) and *Xenopus* (9) with diverse and developmentally regulated patterns of expression prompted our search for forkhead-related genes in the mouse. Through low-stringency screening of a mouse genomic library, we obtained nine forkhead gene family members, three of which are the mouse homologues of the genes for HNF3α, -β, and -γ, and six of which are distinct and are referred to as *fkh-1* to *fkh-6*.* The transcript distribution of the six *fkh* genes was determined and the genes were found to exhibit distinctive and restricted patterns of expression. As most of these genes are also expressed during early stages of mouse embryogenesis, their further analysis promises to aid our understanding of mammalian development.

EXPERIMENTAL PROCEDURES

Genomic Library Construction and Screening. Genomic DNA was isolated from the murine embryonic stem cell line D3 (16), partially digested with the restriction endonuclease *Sau*3A, size-fractionated (16- to 23-kb fragments), and ligated into λ Dash II (Stratagene) according to Frischauf (17). The 300-bp forkhead domain fragment of HNF3α was amplified from mouse liver cDNA by using PCR (30 cycles; annealing temperature, 52°C) with two oligonucleotide primers derived from the rat HNF3α sequence (ref. 6; 5′-CCAAGACGTTCAAGCGCAGTTACCCTCAC-3′ and 5′-GTAGCAGCCGTTCTCGAACATGTTGCC-3′) and subcloned into pTZ19R (Pharmacia). This probe was used in a low-stringency screen of the D3 genomic library (2×10^6 phages) after labeling by random priming (18). Hybridization and washing of the filters were performed according to Church and Gilbert (19), except that 50 mM NaCl was included in the hybridization and washing solutions and that the hybridization temperature was lowered to 56°C. One set of filters was subsequently washed at 65°C to identify the phages containing HNF3α and -β sequences, which are very closely related (7). Thirty-six positive phages were purified and sorted into six classes by hybridization and sequencing. The forkhead-domain-containing exons of the various λ phages were obtained after *Sau*3A digestion, shotgun subcloning into Bluescript (Stratagene), and colony hybridization using the forkhead domain of HNF3α as a probe. In some cases the forkhead domain was amplified by PCR (5 cycles with annealing at 37°C, followed by 25 cycles with annealing at 45°C) with primers within the forkhead domain [5′-(G/A)CCICCITA(C/T)(A/T)(G/C)ITA(C/T)AT and 5′-(G/A)-TGIC(T/G)(G/A)(ATI(C/G)(T/A)(G/A)TT(C/T)TGCCA]. Sequence analysis was performed with the Heidelberg Unix

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviation: HNF3, hepatic nuclear factor 3.

*The sequences reported in this paper have been deposited in the GenBank data base (accession nos. X71939–X71944).

Sequence Analysis Resources at the German Cancer Research Center.

RNA Isolation and RNase Protection Analysis. Total RNA from a variety of mouse tissues or whole mouse embryos was isolated after homogenization in guanidinium thiocyanate (20). The quality of the RNA preparations was controlled by ethidium bromide staining of the 18S and 28S rRNAs after electrophoretic separation of the RNA in denaturing agarose gels. RNase protection analysis was performed using [α-^{32}P]UTP-labeled antisense RNA probes derived from Bluescript subclones containing 120–500 bp of the genes encoding *fkh-1* to *fkh-6* as described (21). The antisense probes were hybridized against 25–100 μg of total RNA (depending on the abundance of the transcript) at 54°C in 80% (vol/vol) formamide. Specificity of the six probes was shown by the size of the protected probe fragments and the unique expression patterns (see below).

RESULTS

Screening of a mouse genomic library with a 300-nt probe encoding the forkhead domain of the murine HNF3α gene under low-stringency conditions yielded a total of 45 positive signals. Upon plaque purification, 36 phages were isolated. The forkhead-domain-encoding exons of these genomic clones were subcloned and sequenced. Sequence comparison revealed that we had cloned the mouse homologues of the genes for the rat transcription factors HNF3α, -β, and -γ (6, 7) in addition to six forkhead-domain-containing genes. These six genes were termed *fkh-1* to *fkh-6*. The amino acid sequences of the nine mouse genes are depicted in Fig. 1 in comparison to the *Drosophila* forkhead gene. Four conserved subdomains within this sequence that were observed in a comparison of the *Drosophila* forkhead gene family and the rat HNF3α, -β, and -γ sequences (6, 7, 12) are conserved in all mouse sequences as well. Regions A (positions 12–24) and B (positions 44–67) have been proposed to exist as α-helices, whereas region C (positions 72–96) is rich in the "helix breakers" proline and glycine. The fourth sequence near the carboxyl terminus of the forkhead domain (region D, positions 101–110) is rich in basic amino acids and has been proposed to be involved in DNA binding (7, 12).

The overall relationship of the mouse genes to the forkhead gene families of *D. melanogaster* (12) and *X. laevis* (9) is summarized in Table 1. The degree of sequence identity on the amino acid level ranges from 45% to 94% (from 60% to 96% similarity), clearly identifying *fkh-1* to *fkh-6* as members of the forkhead gene family. The extreme degree of similarity between *fkh-1* and *FD1* of *Drosophila* suggests that *fkh-1* is the *FD1* homologue. *FD1* is expressed in the early blastoderm in the position of the precursor cells of the posterior and anterior gut. At later stages, mRNA was also found in cells of the central nervous system (12). At present, we have no information concerning the spatial distribution of the *fkh-1* to *fkh-6* transcripts during the corresponding stages of mouse embryogenesis. fkh-1 mRNA was, however, detected at high levels in total midgestation embryos (see below), leaving the possibility that expression of *fkh-1* could follow a pattern similar to that of *FD1*. *fkh-1* is also closely related to *XFD-4* of *Xenopus* (9) with 91% identical amino acids. A close relationship also exists between *fkh-4* and *fkh-5* and the *Drosophila* gene *FD4* (77% and 81% sequence identity, respectively), all of which are expressed in neuronal tissues (Fig. 2 and ref. 12). *fkh-4* and *fkh-5* also seem to correspond to *XFD-5* (94% and 90% identity). It should be interesting to compare the expression patterns of these genes, once information about the transcript distribution of the *Xenopus* genes becomes available, to ascertain whether they serve similar functions.

We analyzed the expression of fkh-1 to fkh-6 mRNA in a wide range of adult mouse tissues (Fig. 2) and whole mouse embryos from midgestation (day 9.5 postcoitum) to birth (Fig. 3) by RNase protection. All six genes are expressed in a tissue-specific manner, but none is restricted to the derivatives of a single germ layer. fkh-1 mRNA is present in brain, heart, kidney, and fat and to a lesser extent in lung and thymus; its expression is strongest in the midgestation embryo (day 9.5) and declines toward the later stages. *fkh-2* is expressed in the embryo from day 9.5 to 12.5 of gestation, but only in lung and spleen in the adult. *fkh-3* shows strong expression in the lung and gonads but is also found at lower levels in most of the other tissues examined as well as in the embryo starting at around day 15.5. *fkh-4* and *fkh-5*, which are closely related (see Fig. 1), are found in overlapping sets of tissues, both being present in brain and thymus, the former additionally in spleen, ovary, and testes. *fkh-4* is localized to the ventral midbrain/forebrain region at day 9.5 of gestation and is subsequently restricted to distinct regions of the developing midbrain and hindbrain (A.P.M., K.H.K., and G.S., unpublished observations). fkh-6 mRNA is expressed

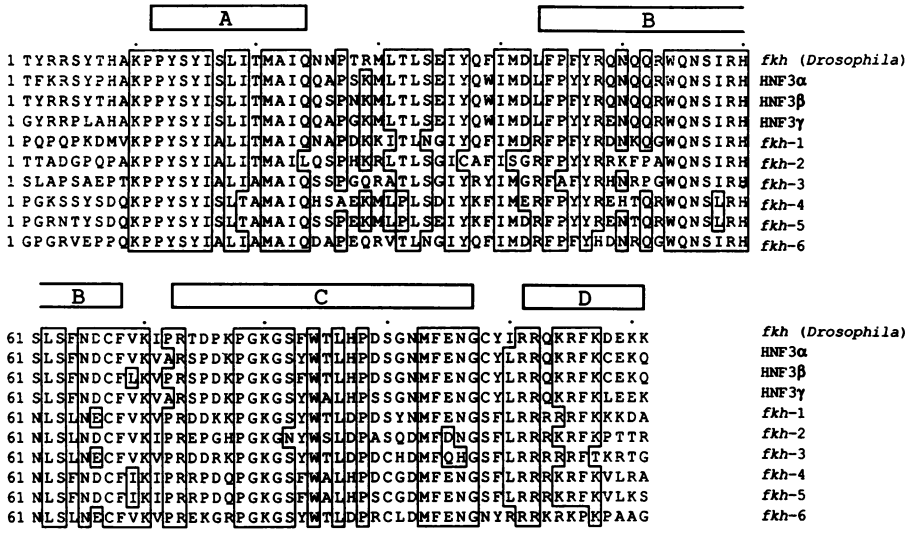


FIG. 1. Forkhead domain sequences encoded by the murine forkhead gene family. The sequences were aligned for maximum overlap for comparison with the *Drosophila melanogaster* forkhead sequence. The numbering refers to the comparison between the forkhead and rat HNF3α proteins (3). Amino acids that are identical in at least 7 of the 10 sequences are boxed. Bars labeled A–D refer to regions mentioned in the text.

Table 1. Sequence comparison of the members of the mouse, *X. laevis*, and *D. melanogaster* forkhead gene families

	% identical amino acids								
	HNF-3α	HNF-3β	HNF-3γ	*fkh-1*	*fkh-2*	*fkh-3*	*fkh-4*	*fkh-5*	*fkh-6*
fkh	84 (92)	88 (93)	80 (87)	65 (76)	57 (69)	56 (71)	63 (78)	67 (76)	58 (70)
FD1	63 (74)	59 (75)	61 (76)	86 (96)	59 (73)	69 (82)	58 (78)	56 (74)	66 (83)
FD2	56 (66)	54 (66)	57 (66)	59 (73)	57 (69)	62 (78)	55 (69)	59 (72)	64 (74)
FD3	57 (73)	58 (74)	55 (72)	62 (77)	76 (84)	59 (76)	60 (73)	60 (70)	59 (76)
FD4	64 (73)	65 (75)	64 (73)	57 (71)	58 (71)	54 (71)	77 (90)	81 (89)	54 (66)
FD5	61 (75)	58 (75)	58 (74)	59 (76)	52 (70)	59 (73)	70 (86)	71 (84)	55 (69)
slp1	48 (60)	47 (62)	49 (64)	57 (68)	53 (65)	58 (66)	48 (64)	50 (65)	58 (70)
slp2	46 (60)	44 (62)	46 (64)	54 (69)	52 (65)	54 (66)	45 (67)	46 (68)	53 (67)
XFD1	87 (95)	87 (95)	86 (90)	65 (77)	58 (69)	53 (69)	61 (80)	66 (79)	59 (75)
XFD2	46 (61)	46 (61)	46 (63)	56 (71)	51 (67)	52 (68)	46 (62)	45 (61)	52 (70)
XFD3	91 (96)	97 (98)	86 (90)	64 (74)	59 (74)	57 (71)	66 (81)	69 (78)	57 (71)
XFD4	64 (75)	63 (75)	65 (77)	91 (95)	61 (74)	67 (81)	58 (74)	59 (72)	67 (80)
XFD5	66 (78)	67 (80)	65 (78)	59 (73)	59 (69)	56 (73)	94 (98)	90 (96)	55 (71)
XFD6	59 (75)	58 (76)	58 (74)	63 (79)	79 (83)	59 (73)	60 (74)	60 (72)	55 (72)

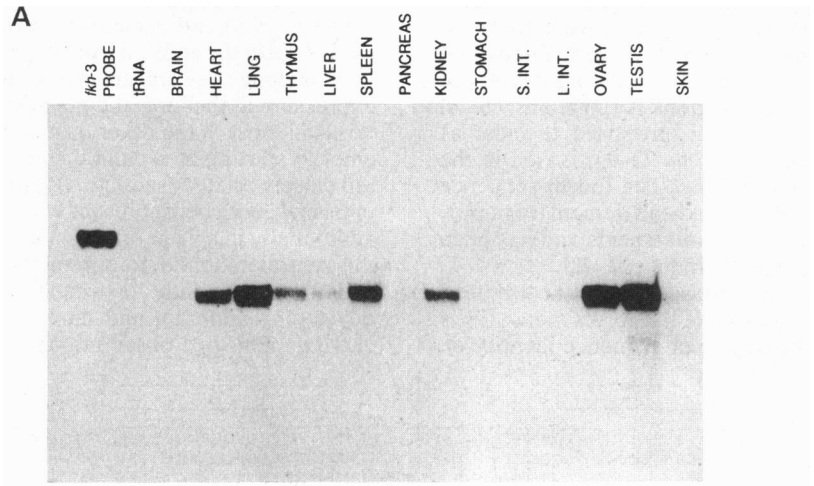
Amino acid sequences of the indicated genes were compared pairwise over the 110-aa stretch of the forkhead domain. Percent similar amino acids is in parentheses.

in lung, kidney, stomach, and intestine and is found in all embryonic stages examined.

DISCUSSION

We have identified nine members of the murine forkhead gene family through a low-stringency hybridization screen of a genomic library. Three of these are the homologues of the rat genes for the transcription factors HNF3α, -β, and -γ (6, 7), while the other six represent genes not previously found in mammals. We envision from these results and from low-stringency genomic Southern blots using the forkhead domain of HNF3α as probe (data not shown) that the mouse forkhead gene family contains several dozen members. This notion is also supported by the fact that for the majority of the forkhead-related genes described in *Drosophila* and *Xenopus* no homologues have yet been found in mice.

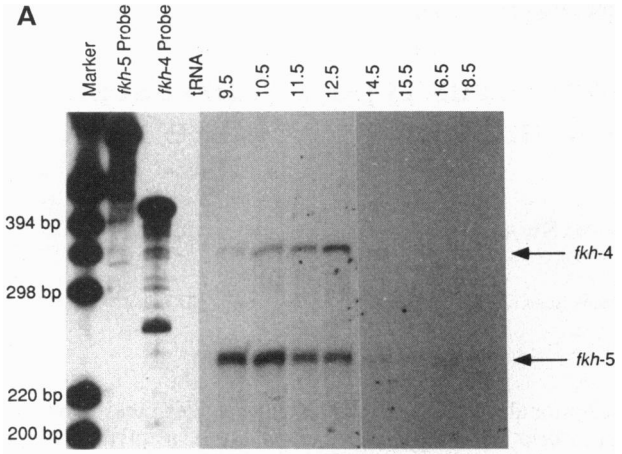
The forkhead gene families of *Drosophila* and *Xenopus* were defined as such by the presence of the conserved 110-aa forkhead domain. The forkhead domain was shown to be



B

fkh	BRAIN	HEART	LUNG	THYMUS	LIVER	SPLEEN	PANCREAS	KIDNEY	STOMACH	S. INT.	L. INT.	OVARY	TESTIS	SKIN	SK. MUS.	FAT
1	+	+	(+)	(+)	-	-	-	+	-	-	-	-	-	-	-	+
2	-	-	(+)	-	-	(+)	-	-	-	-	-	-	-	n.d.	n.d.	n.d.
3	-	+	++	+	(+)	+	-	+	-	-	-	++	++	-	n.d	n.d
4	+	-	-	+	-	+	-	-	-	-	-	+	+	-	-	-
5	+	-	-	+	-	-	-	-	-	-	-	-	-	-	-	-
6	-	-	+	-	-	-	-	+	+	+	+	-	-	-	-	-

FIG. 2. Transcript distribution of *fkh-1* to *fkh-6* in adult mouse tissues. Total RNAs from the mouse tissues indicated were analyzed for the presence of the *fkh-1* to *fkh-6* transcripts by RNase protection analysis. (*A*) Autoradiogram (18-hr exposure) of the RNase protection analysis for *fkh-3* using an antisense probe prepared from a Bluescript subclone containing the forkhead domain of the *fkh-3* gene. (*B*) RNA expression pattern for *fkh-1* to *fkh-6*. Expression: ++, strong; +, medium; (+), weak; −, undetectable. n.d., Not determined; S. INT., small intestine; L. INT., large intestine; SK.MUS., skeletal muscle.



B

fkh	9.5	10.5	11.5	12.5	14.5	15.5	16.5	18.5
1	++	++	+	+	(+)	(+)	(+)	(+)
2	+	+	+	+	(+)	(+)	(+)	-
3	-	-	-	-	(+)	+	+	+
4	(+)	+	+	+	-	-	-	-
5	+	+	+	+	(+)	(+)	(+)	-
6	+	+	+	+	+	+	+	+

FIG. 3. Transcript abundance of *fkh-1* to *fkh-6* during mouse embryogenesis. Total RNA from the gestational ages indicated (in days postcoitum) was analyzed for the presence of the *fkh-1* to *fkh-6* transcripts by RNase protection analysis using gene-specific probes. (*A*) Autoradiogram (60-hr exposure) of the RNase protection analysis for *fkh-4* and *fkh-5*. (*B*) RNA expression pattern for *fkh-1* to *fkh-6*. Expression: ++, strong; +, medium; (+), weak; −, undetectable.

essential for DNA binding of the rat hepatic transcription factor HNF3α (6). HNF3α and two closely related proteins, HNF3β and -γ, were originally purified from rat liver nuclear extracts by their ability to bind specifically to a functionally important cis element in the transthyretin gene, the HNF3 binding site (6, 7). Several conserved sequence elements within the forkhead domain have been suggested to be of importance (7, 12): a carboxyl-terminal domain of basic amino acids reminiscent of similar elements in the helix–loop–helix and leucine-zipper proteins (22, 23), two potential α-helical regions, and a sequence rich in prolines and glycines. Each of these elements is also conserved in *fkh-1* to *fkh-6*, confirming their classification as forkhead gene family members and also suggesting that they too can function as transcription factors. From DNA binding studies performed with the human forkhead-domain-containing genes for human T-cell leukemia virus enhancing factor (15) and interleukin binding factor (14) and those performed with the rat HNF3 and BF1 genes (6, 7, 13), it is clear that the forkhead-related genes can have very different DNA binding specificities. In light of these data, it should be very interesting to determine the DNA binding properties and target preferences of *fkh-1* to *fkh-6*.

The tissue-specific expression of *fkh-1* to *fkh-6* indicates a potential involvement in differentiation and development. All genes except *fkh-3* are expressed at least as early as day 9.5 of gestation, a time at which organogenesis is initiated. Therefore, it can be speculated that these genes are required for the differentiation of specific structures during mammalian development, in analogy to the role of the *Drosophila* forkhead gene in the differentiation of the gut. A precise localization of the transcripts to embryonic tissues by *in situ* hybridization and mutation of these genes through homologous recombination in embryonic stem cells will help to clarify the role of these genes in development.

After completion of the work described above, an independent study on the forkhead gene family in the rat (24) was brought to our attention. Of the nine genes described there only *HFH-6* is the homologue of one of our genes, namely *fkh-2*, providing further evidence of the complexity of the forkhead gene family in mammals.

We thank Drs. G. Kelsey and J. Blendy for critical reading of the manuscript, B. Lupp and H. Kern for technical assistance, and W. Fleischer for oligonucleotide synthesis and photographic work. This work was supported by the Deutsche Forschungsgemeinschaft through Sonderforschungsbereich 229, the Leibniz-Programm, and Fonds der Chemischen Industrie.

1. Johnson, P. F. & McKnight, S. L. (1989) *Annu. Rev. Biochem.* **58**, 799–839.
2. Pabo, C. O. & Sauer, R. T. (1992) *Annu. Rev. Biochem.* **61**, 1053–1095.
3. Weigel, D. & Jäckle, H. (1990) *Cell* **63**, 455–456.
4. Jürgens, G. & Weigel, D. (1988) *Rouxs Arch. Dev. Biol.* **197**, 345–354.
5. Weigel, D., Jürgens, G., Küttner, F., Seifert, E. & Jäckle, H. (1989) *Cell* **57**, 645–658.
6. Lai, E., Prezioso, V. R., Smith, E., Litvin, O., Costa, H. R. & Darnell, J. E., Jr. (1990) *Genes Dev.* **4**, 1427–1436.
7. Lai, E., Prezioso, V. R., Tao, W., Chen, W. S. & Darnell, J. E., Jr. (1991) *Genes Dev.* **5**, 416–427.
8. Oliver, S. G. *et al.* (1992) *Nature (London)* **357**, 38–46.
9. Knöchel, S., Lef, J., Clement, J., Klocke, B., Hille, S., Köster, M. & Knöchel, W. (1992) *Mech. Dev.* **38**, 157–165.
10. Dirksen, M. L. & Jamrich, M. (1992) *Genes Dev.* **6**, 599–608.
11. Ruiz i Altaba, A. & Jessell, T. M. (1992) *Development* **116**, 81–93.
12. Häcker, U., Grossniklaus, U., Gehring, W. J. & Jäckle, H. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 8754–8758.
13. Tao, W. & Lai, E. (1992) *Neuron* **8**, 957–966.
14. Li, C., Lai, C., Sigman, D. S. & Gaynor, R. B. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 7739–7743.
15. Li, C., Lusis, A. J., Sparkes, R., Tran, S.-M. & Gaynor, R. (1992) *Genomics* **13**, 658–664.
16. Doetschman, T. C., Eistetter, H., Katz, M., Schmidt, W. & Kemler, R. (1985) *J. Embryol. Exp. Morph.* **87**, 27–45.
17. Frischauf, A. M. (1987) *Methods Enzymol.* **152**, 190–199.
18. Feinberg, A. P. & Vogelstein, B. (1983) *Anal. Biochem.* **132**, 6–13.
19. Church, G. M. & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1995.
20. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299.
21. Kaestner, K. H., Ntambi, J. M., Kelly, T. J. & Lane, M. D. (1989) *J. Biol. Chem.* **264**, 14755–14761.
22. Murre, C., McCaw, P. S. & Baltimore, D. (1989) *Cell* **56**, 777–783.
23. Landschulz, W. H., Johnson, P. E. & McKnight, S. L. (1988) *Science* **240**, 1759–1764.
24. Clevidence, D. E., Overdier, D. G., Tao, W., Qian, X., Pani, L., Lai, E. & Costa, R. H. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 3948–3952.