MOLECULAR AND CELLULAR BIOLOGY, Sept. 1988, p. 3906–3917
0270-7306/88/093906-12$02.00/0

Vol. 8, No. 9

Early Retinoic Acid-Induced F9 Teratocarcinoma Stem Cell Gene *ERA-1*: Alternate Splicing Creates Transcripts for a Homeobox-Containing Protein and One Lacking the Homeobox

GREGORY J. LAROSA[1,2] AND LORRAINE J. GUDAS[2,3]*

Program on Cellular and Developmental Biology,[1] and the Department of Biological Chemistry and Molecular Pharmacology,[3] Harvard Medical School, and the Dana-Farber Cancer Institute,[2] 44 Binney Street, Boston, Massachusetts 02115

Received 28 March 1988/Accepted 15 June 1988

Retinoic acid (RA), the natural acidic derivative of vitamin A, can modulate the expression of specific genes and can induce some cell types, such as the murine F9 teratocarcinoma stem cell line, to differentiate in culture. As an initial step toward understanding the molecular mechanism(s) by which RA exerts these effects, we previously isolated cDNA clones for a gene, *ERA-1*, which has the characteristics of an early, direct target for RA. We demonstrated that RA causes a rapid, dose-dependent, and protein synthesis-independent expression of the *ERA-1* gene (G. J. LaRosa and L. J. Gudas, Proc. Natl. Acad. Sci. USA 85:329–333, 1988). We now report the full-length cDNA sequence and the further characterization of this gene. The data indicate that the RA-induced 2.2- to 2.4-kilobase *ERA-1* RNA species that we previously detected consists of two alternately spliced messages. One mRNA encodes a protein with a predicted mass of about 36 kilodaltons (kDa) that possesses the Hox 1.6 homeobox domain. The other mRNA encodes a truncated protein of about 15 kDa which is identical to the 36-kDa protein for 114 amino acids at the amino-terminal end but which lacks the homeobox amino acid sequence. The RA-associated increase in the *ERA-1* mRNA level does not appear to be due to message stabilization, suggesting that the response is at the level of transcription. By Northern (RNA) blot analysis, the usual 2.2- to 2.4-kilobase mRNA species was also rapidly expressed in P19 teratocarcinoma cells during their differentiation to fibroblastic cells in response to RA and was detected in day 10.5 and day 13.5 mouse embryos. This result indicates that the expression of this gene is not limited to the endodermal differentiation of F9 cells.

Understanding the regulation of cellular differentiation and the control of the expression of specific genes during differentiation and development are central problems of developmental biology which have received considerable attention over the past several years. Many murine teratocarcinoma stem cell lines are capable of undergoing a controllable differentiation in culture and thereby provide an excellent in vitro model for molecular and biochemical studies of cellular differentiation and early embryonic development (35). For instance, when the mouse teratocarcinoma stem cell line F9 is grown in monolayer culture in the presence of retinoic acid (RA), the cells undergo a rapid and irreversible differentiation into cells with the biochemical and morphological characteristics of the parietal endoderm of the 3- to 5-day mouse blastocyst (58).

In addition to its effects on the differentiation of teratocarcinoma stem cells, RA, the natural acidic derivative of vitamin A (retinol), can affect the proliferation and differentiation of a variety of cell types (49, 57), suppress carcinogenesis in vivo (6, 49, 57), and affect pattern formation during development (61, 62). The recent work of Thaller and Eichele (60) suggests that RA may be the natural morphogen which forms a gradient for establishing the anterior-posterior pattern formation during chicken limb development. The mechanism by which RA exerts these profound effects on differentiation, growth, and development and on the corresponding alteration in the expression of specific genes is unknown.

In order to further understand the mechanism by which RA influences gene expression and differentiation, specific cellular genes that exhibit changes in their expression during RA-induced differentiation of monolayer F9 mouse teratocarcinoma stem cells have been isolated (7, 12, 15, 20, 31, 37, 55, 64, 65). Initially, these included genes which begin to be expressed in F9 cells relatively late (18 to 24 h) after RA addition, such as those encoding the extracellular matrix proteins laminin B1 and collagen IV (α1) (31, 64), SPARC (37), major histocompatibility antigens (15), and certain homeobox-containing proteins (7, 12).

It has been suggested (11) that RA influences gene expression in a manner analogous in some respects to that of steroid hormones (1, 46). In such a model, the binding of RA to a specific nuclear receptor would cause the rapid induction of specific primary-response (early) genes, some of which may regulate the differentiation process and the expression of secondary (late) genes such as those mentioned above. Strong support for this proposal has come recently from the identification of a specific nuclear RA receptor protein with a putative DNA-binding domain very similar to the DNA-binding domain of steroid receptor proteins (23, 47). Additionally, as part of our studies to further understand the mechanism by which RA induces F9 stem cell differentiation, we recently described the isolation of a cDNA clone for a gene, early retinoic acid 1 (*ERA-1*), which exhibits a rapid, RA concentration-dependent but protein synthesis-independent induction after RA treatment of F9 cells (32). *ERA-1* appears to be a primary target for RA action, with characteristics similar to those of steriod hormone early-response genes (1).

* Corresponding author.

The further characterization of the *ERA-1* gene should provide important information concerning the mechanism of action of RA on the expression of specific genes and the differentiation of F9 cells. Here we report the full-length *ERA-1* cDNA sequence and the partial genomic sequence of the putative promoter region as well as data concerning the splicing and expression of this gene. The results indicate that the *ERA-1* gene contains the Hox 1.6 (3) homeobox sequence and encodes two alternately processed mRNAs which can direct the synthesis of two distinct proteins, one which contains the homeobox and another truncated protein which lacks the homeobox. Additionally, the increase in the steady-state levels of *ERA-1* mRNAs following RA addition appears to be a transcriptional response. Finally, the increase in *ERA-1* mRNAs also occurs in RA-treated P19 teratocarcinoma cells, and *ERA-1* mRNAs are present in mouse embryonic RNA.

The homeobox is a 180-base-pair (bp) DNA sequence that was originally found to be highly conserved in the developmentally important *Drosophila* homeotic genes (39, 41, 51) and has since also been found to be present and highly conserved in mammalian genes (see reference 36 for references). Some murine homeobox-containing genes have been shown to exhibit temporal and spatial restriction of expression during embryonic development (2, 30), suggesting that mammalian proteins containing a homeobox domain may, like the *Drosophila* counterparts, play important roles during murine embryonic development. The homeobox amino acid sequence shows homology to DNA-binding proteins of the helix-turn-helix class (33, 54), and immunolocalization studies indicate that the protein products of several *Drosophila* genes and one mammalian homeobox gene show nuclear localization (8, 17, 27, 66). Fusion proteins containing the homeobox of the engrailed *Drosophila* gene or the homeobox of the Hox 1.5 mouse gene have been shown to have sequence-specific DNA-binding activities in vitro (16, 21), further suggesting that homeobox-containing proteins may directly regulate specific sets of genes important during development. That such genes show regulated expression during RA-induced differentiation of F9 stem cells suggests the possibility that some homeobox-containing genes play a role in some of the striking in vitro and in vivo effects of RA. Thus, the F9 differentiation model provides a useful system for molecular and biochemical studies of the mechanisms of RA action and mammalian differentiation and development as well as for further study of both the function of the putative homeobox-containing proteins and the regulation of homeobox gene expression.

MATERIALS AND METHODS

Cell culture, nucleic acid isolation, and blot hybridizations. F9 and P19 teratocarcinoma stem cells were cultured as described elsewhere (65). Total cellular RNA was isolated by the guanidine-hydrochloride method essentially as described previously (14, 59, 65). Poly(A)$^+$ RNA was purified from total RNA by poly(U)-Sephadex column chromatography (24). Cytoplasmic RNA was isolated as described elsewhere (34), except that nuclei were pelleted through the sucrose step gradient by centrifugation at 1,000 × *g* instead of 10,000 × *g* and then 5.35 g of solid guanidine-hydrochloride and H_2O were added to the cytoplasmic supernatant to a final concentration of 7 M guanidine-hydrochloride in a final volume of 8 ml. The RNA was then purified as described for the isolation of total cellular RNA by the guanidine-hydrochloride method. Day 10.5 and day 13.5 murine embryonic RNA samples were isolated from embryos by the guanidinium isothiocyanate-cesium chloride method (34) and were gifts of M. Mercola (Dana-Farber Cancer Institute, Boston, Mass.). Northern (RNA) blot analysis was done as previously described (32).

Screening for *ERA-1* genomic clones. Duplicate nitrocellulose filter replicas (4) of approximately 200,000 recombinant bacteriophage (20,000/150-mm petri dish) of a mouse spleen genomic DNA library (kindly provided by R. M. Perlmutter, California Institute of Technology, Pasadena, Calif.) were screened by hybridization, as previously described (32), with the 5′ 1.2-kilobase (kb) *Eco*RI fragment of *ERA-1-993*. A single plaque showed a positive hybridization signal and was subsequently plaque purified and named λGERA-1 (genomic *ERA-1*). The mouse DNA was partially mapped by Southern blot analysis and was subcloned in whole and in part into pUC-9 by standard procedures (34).

DNA sequence analysis. Sequences of various subclones of *ERA-1-993*, *ERA-1-399*, and λGERA-1 were determined by the dideoxy-chain termination protocol by using the Sequenase system (U.S. Biochemical Corp.) with supercoiled plasmid miniprep DNA (10). Sequencing primers were either the M13 universal primer (New England BioLabs, Inc.), reverse sequencing primers (New England BioLabs), or specific *ERA-1* 16-mer oligonucleotide primers synthesized in an Applied Biosystems 380A DNA synthesizer at the Dana-Farber Cancer Institute. The entire sequence of the *ERA-1-993* cDNA clone was determined for both strands.

S1 protection analysis. The two RNA species of *ERA-1* were demonstrated by S1 protection analysis (5) by using a 608-bp *Nco*I-*Pst*I fragment of the *ERA-1-993* cDNA previously subcloned into pUC-9 (pERA-1-993$_{Nco\text{-}Pst}$). pERA-1-993$_{Nco\text{-}Pst}$ was digested with *Bgl*II and was 5′ end labeled with [γ-^{32}P]ATP (6,000 Ci/mmol) and polynucleotide kinase in 66 mM Tris chloride [pH 7.5]–10 mM $MgCl_2$–5 mM dithiothreitol–0.1 mM EDTA for 45 min at 37°C. A digestion with *Bgl*I yielded a labeled 579-bp fragment, containing 408 bp of the *ERA-1-993* cDNA sequence and 171 bp of the pUC-9 sequence, which was purified by agarose gel electrophoresis and absorption to and elution from glass beads (63) by using the Geneclean kit (Bio 101, Inc.). For hybridization, approximately 2 fmol of a labeled, double-stranded DNA probe was mixed with 15 μg of total or cytoplasmic RNA or 1 μg of poly(A)$^+$ RNA in 10 μl of 80% formamide–0.4 M NaCl–40 mM piperazine-*N*,*N*′-bis(2-ethanesulfonic acid) (PIPES; pH 6.4)–1 mM EDTA and, after denaturation at 90°C for 3 min, was incubated at 51°C for 12 to 16 h. Hybridization reactions were terminated, and the S1 nuclease digestions were started by the addition of 250 μl of ice-cold buffer containing 30 mM sodium acetate (pH 4.5), 3 mM $ZnSO_4$, 300 mM NaCl, 50 μg each of native and denatured salmon sperm DNA per ml, and 500 U of S1 nuclease (Bethesda Research Laboratories) per ml. Digestions were then carried out at 37°C for 45 min, followed by phenol-chloroform extraction and precipitation of the nucleic acid with ethanol. The pellets were resuspended in 95% formamide–20 mM EDTA–0.05% bromophenol blue–0.05% xylene cyanole FF; the protected fragments were visualized by autoradiography after electrophoresis on denaturing 4% polyacrylamide–8 M urea gels.

To map the transcription start site, S1 protection was carried out by using a probe derived from a subclone, pGERA-1$_{Bam\ 1.7}$, which contains a 1.7-kp *Bam*HI fragment from the λGERA-1 genomic clone. The *Bam*HI fragment was excised with *Bam*HI, 5′ end labeled, and used in an S1 protection assay as described above, except that the pro-

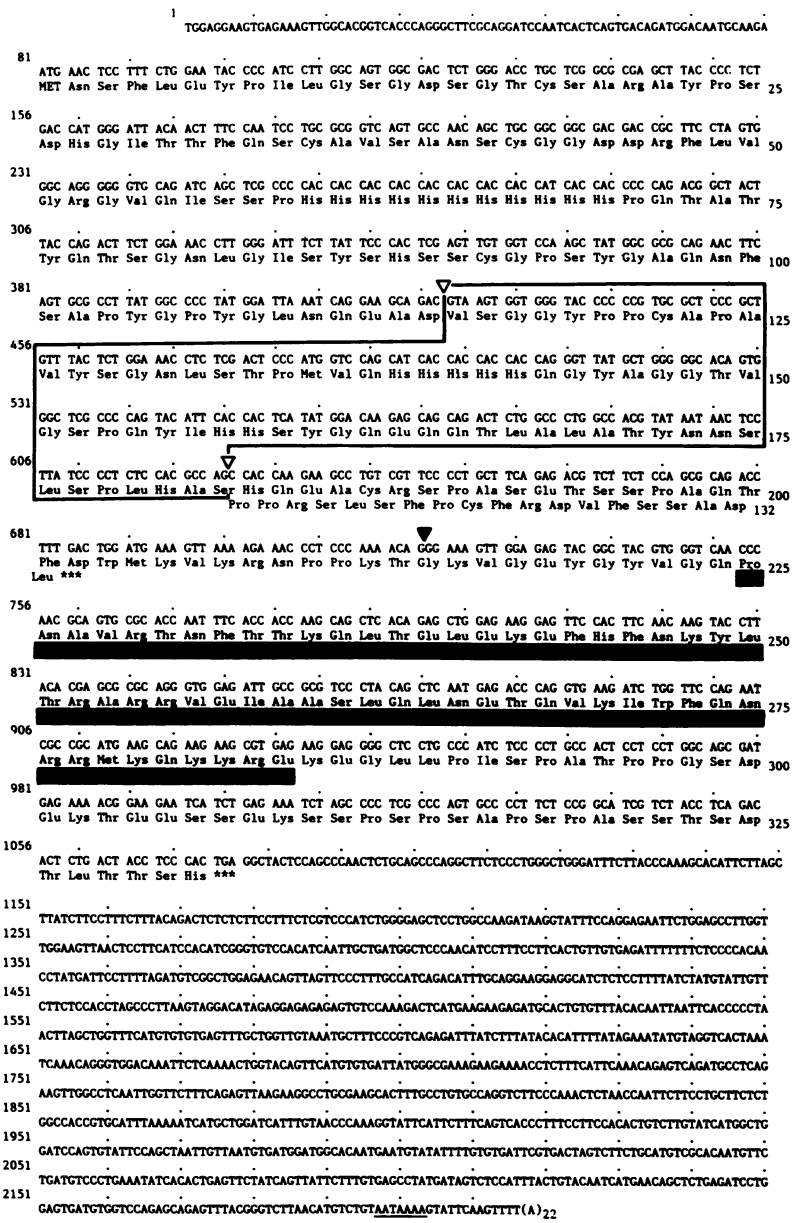

FIG. 1. The complete nucleotide sequence of the *ERA-1-993* cDNA and the predicted amino acid sequences of the *ERA-1-993* and *ERA-1-399* proteins. The nucleotide sequence of *ERA-1-993* is shown with the Hox 1.6 homeobox region underlined (heavy underline). The additional splice acceptor-donor pair that is joined in *ERA-1-399* cDNA is shown at positions 423 and 626 (▽). The sequence of the *ERA-1-993* cDNA which is missing in *ERA-1-399* is shown within the boxed region. The common splice site is shown at position 721 (▼), and the polyadenylation signal is underlined (light underline). The amino acid sequences of the *ERA-1-993* and *ERA-1-399* open reading frames are identical up to Asp-114 (▽). The *ERA-1-993* amino acid sequence continues through the boxed region to the TGA (***) at position 1074. The *ERA-1-399* amino acid sequence resumes after the boxed region at position 626 (second ▽), ending at the TGA (***) at position 683.

tected fragments were visualized with a denaturing gel containing 8% polyacrylamide.

Primer extension analysis. The 5′ ends of the *ERA-1* mRNAs were also mapped by primer extension. An 18-mer oligonucleotide primer, complementary to nucleotide positions 58 to 76 of the *ERA-1-993* cDNA, was synthesized as described above. The oligonucleotide sequence was 5′-GCATTGTCCATCTGTCAC-3′. This primer was 5′ end labeled with polynucleotide kinase as described above, and 2.4 fmol was hybridized to 2 μg of either poly(A)$^+$ RNA from stem or RA-treated F9 cells or to poly(A)$^-$ RNA from RA-treated cells in 5 μl of 10 mM Tris chloride (pH 8.3)–250 mM KCl–1 mM EDTA at 50°C for 6 h. Then, 8 μl of H_2O and 5 μl of 200 mM Tris chloride (pH 8.3)–32 mM $MgCl_2$–40 mM dithiothreitol–2 mM each of dATP, dCTP, dGTP, and dTTP were added. Actinomycin D (Calbiochem-Behring), RNasin (Promega Biotec), and reverse transcriptase (Molecular Genetic Resources) were then added to final concentrations of 50 μg, 300 U, and 1,000 U per ml, respectively. The reaction was incubated at 42°C for 1 h, at which time another 10 U of reverse transcriptase was added and the reaction was incubated at 42°C for another 30 min. EDTA was added to a final concentration of 10 mM, and the nucleic acids were precipitated with 2.5 M ammonium acetate and ethanol. The

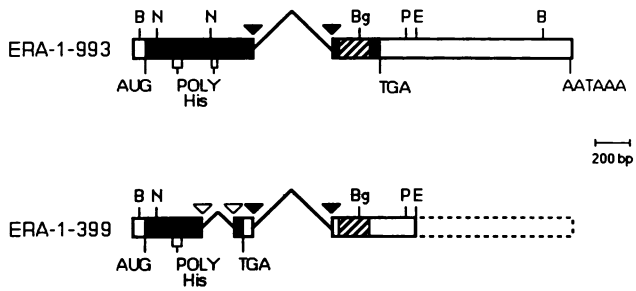

FIG. 2. Schematic illustration of the structures, reading frames, and splicing patterns of *ERA-1-993* and *ERA-1-399* cDNAs. The structure of *ERA-1* cDNA is shown with the open reading frame (▬) bounded by the AUG and TGA codons, the two polyhistidine tracts, the splice site (▼), the homeobox (▨), and the polyadenylation signal. The *ERA-1-339* cDNA is also shown in the same way, with the additional splice site indicated (▽) and the truncated reading frame (▬) bounded by the same AUG but a different TGA codon. Several restriction sites are indicated above *ERA-1-993*: B, *Bam*HI; N, *Nco*I; Bg, *Bgl*II; P, *Pst*I; E, *Eco*RI. There are *Eco*RI sites at both ends of the *ERA-1-993* cDNA and at the 5′ end of *ERA-1-399* that are the result of linker addition during the initial cloning (32) and are not shown. The 3′ 1-kb portion of *ERA-1-399* is shown as a dotted line, since this region was present in the original phage clone but has not been further characterized.

precipitate was resuspended and analyzed as described above by electrophoresis on a denaturing 8% polyacrylamide–urea gel.

RESULTS

Full-length *ERA-1-993* cDNA sequence and predicted amino acid sequence. The isolation of the *ERA-1*cDNA clones has been reported previously (32). Initially, a cDNA clone, which we now call *ERA-1-399*, was isolated but appeared to be shorter than the full-length of the RNA as judged by Northern blot analysis. This *ERA-1-399* clone was used to rescreen the same F9 cDNA library, and we were able to isolate the apparently full-length *ERA-1* cDNA (32), which we now call *ERA-1-993*. This nomenclature, adopted for clarity, indicates the two different lengths of the open reading frames (399 bp versus 993 bp) present in each of these two *ERA-1* cDNAs. This will be discussed further below. Both *ERA-1* cDNAs hybridize, by Northern blot analysis, to a 2.2- to 2.4-kb RNA band which is rapidly induced by RA in F9 cells in a protein synthesis-independent manner, suggesting that increased expression of the *ERA-1* gene is a direct, primary response to RA in these cells (32). The full-length *ERA-1-993* DNA sequence has been determined and is shown in Fig. 1. The structure of the *ERA-1-993* cDNA is shown schematically in Fig. 2. The *ERA-1-993* cDNA sequence from position 1 to the *Eco*RI site at position 1231 is identical, with a few discrepancies, to the partial sequence flanking the Hox 1.6 homeobox that was recently reported (3). It is therefore likely that the *ERA-1-993* cDNA that we have isolated corresponds to the transcription unit containing the Hox 1.6 homeobox located within the Hox 1 cluster on mouse chromosome 6 (3, 19). However, the splicing patterns, reading frame, and encoded proteins we propose for the *ERA-1*/Hox 1.6 gene are substantially different from those previously reported (3). The differences between the results reported here and those of Baron et al. (3) will be further discussed below.

The *ERA-1-993* cDNA is 2,216 bp long with a major open reading frame of 993 bp beginning with a potential AUG initiator codon at nucleotide 81 and ending with a TGA terminator codon at position 1074. There are two possible AUG initiation codons at positions 66 and 81; however, the former, having a pyrimidine (cytidine) at the −3 position, does not conform to the consensus for translation start sequences (9, 29) and is an unlikely initiator. There is a 1,142-bp 3′-untranslated region with an AATAAA polyadenylation signal (48) 14 bp upstream of the 3′-terminal poly(A) tract. There are several isolated copies of the AUUUA pentanucleotide found in messages with short half-lives, but the context in which these are found is not similar to that previously shown (53).

The open reading frame predicts a protein with a molecular mass of about 36 kilodaltons which is rich in serine (12.4%). Beginning at Pro-225 (nucleotide position 753) and ending at Glu-284 (nucleotide position 932) is a homeobox sequence that is identical to the recently published sequence of the mouse Hox 1.6 homeobox (3) at both the nucleic acid and amino acid levels. This homeobox is somewhat more divergent than other *Drosophila* and murine homeoboxes of the Antennapedia class, having only 55 to 63% exact matches at the amino acid level with the *Drosophila* Antennapedia and several murine sequences (Fig. 3). However, it shows an 85% exact match (Fig. 3) at the amino acid level

	1 60	% MATCH WITH ERA-1-993
ANTP	RKRGRQTYTRYQTLELEKEFHFNRYLTRRRRIEIAHALCLTERQIKIWFQNRRMKWKKEN	62 (67)
ERA-1-993	PNAV-TNF-TK-LT---------K----A--V---AS-Q-N-T-V----------Q--RE	
F90-2	NNS--TNF-NK-LT--------------A------NT-Q-N-T-V----------Q--RV	85 (90)
HOX 1.1	----------TL----------H	60 (65)
HOX 1.2	GR----------N----------	62 (67)
HOX 1.3	G--A-TA----------S----------D-	63 (68)
HOX 1.4	PE-S-TA---Q-V----------T---S---V----------D-	67 (73)
HOX 1.5	S----TA---P-LV----------M-P--V-M-NL-N----------Y--DQ	62 (67)

FIG. 3. Alignment comparisons of the *ERA-1-993*/Hox 1.6 homeobox amino acid sequence with other *Drosophila* and mouse Hox 1 homeoboxes. The complete amino acid sequence of the prototypical homeobox from the *Drosophila* Antennapedia (ANTP) gene (39) is shown with the single letter code. The amino acid sequences for *ERA-1-993*, F90-2 (25), Hox 1.1 (12), Hox 1.2 (13), Hox 1.3 (45), Hox 1.4 (50), and Hox 1.5 (40) are aligned with that for the *Drosophila* ANTP gene; only nonidentical amino acids are indicated. The percentages of exact amino acid matches for the *ERA-1-993*/Hox 1.6 homeobox with each of the other sequences are indicated, and the percentages of matches including conservative amino acid substitutions are shown in parentheses.

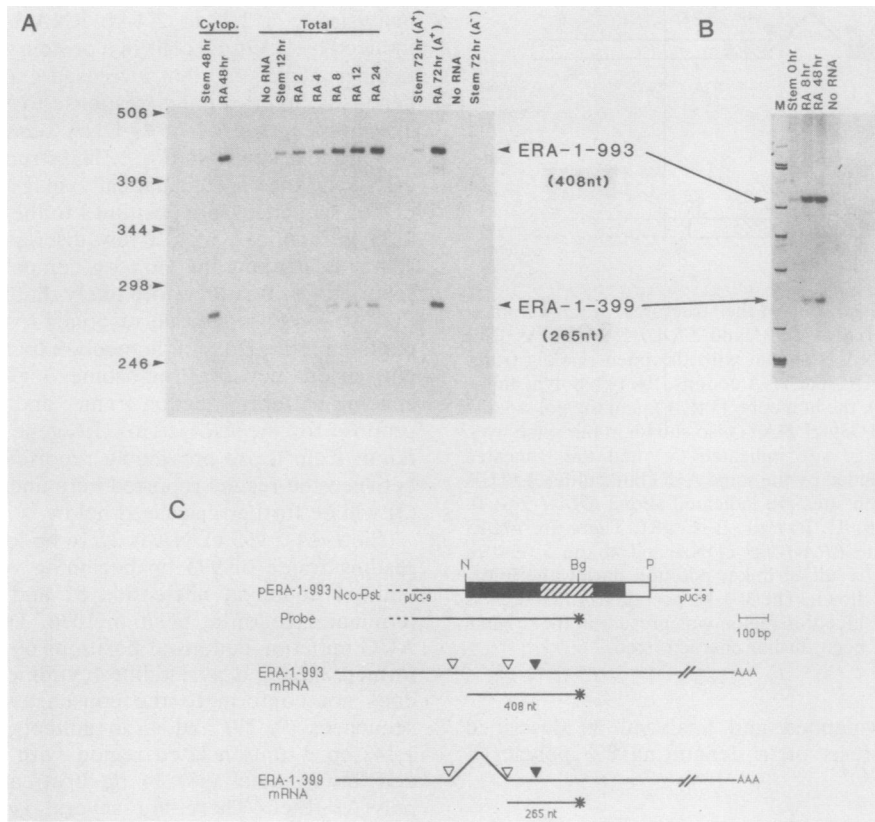

FIG. 4. Detection of both *ERA-1-993* and *ERA-1-399* messages by S1 protection analysis of F9 cell RNA. (A) Fifteen micrograms of either F9 cytoplasmic (Cytop.) or total cellular (Total) RNA or 1 μg of either poly(A)$^+$ or poly(A)$^-$ RNA was hybridized with a probe end labeled with P^{32} (see panel C). The RNA-DNA hybrids were digested with S1 nuclease, and the protected products were visualized by denaturing acrylamide gel electrophoresis. The lanes contain the following samples: cytoplasmic RNA isolated from F9 48-h stem cells or isolated from cells 48 h after the addition of 1 μM RA (lanes 1 and 2); total RNA isolated from 12-h stem cells or from cells treated with 1 μM RA for 2, 4, 8, 12, or 24 h (lanes 4 to 9); poly(A)$^+$ and poly(A)$^-$ RNAs purified from total RNA isolated from 72-h stem cells or isolated 72 h after the addition of 1 μM RA (lanes 10, 11, and 13). Lanes containing the no-RNA controls are indicated (lanes 3 and 12). The *ERA-1-993* protected fragment (408 nucleotides) and the *ERA-1-399* protected fragment (265 nucleotides) are marked. The size markers (in bases) were ^{35}S-labeled pBR322 *Hin*fI-*Pst*I digestion products. (B) S1 nuclease protection analysis of 15 μg of total cellular RNA from 0-h stem cells (lane 1) or 8- and 48-h 1 μM RA-treated cells (lanes 2 and 3). The probe, analysis, and markers (M) are described above. (C) Schematic representation of the S1 probe, both *ERA-1-993* and *ERA-1-399* RNAs, and the predicted fragments which should be protected by hybridization of the probe to each respective RNA. The subclone of *ERA-1-993* cDNA, pERA-1-993$_{Nco\text{-}Pst}$, contains the 608-bp *Nco*I-*Pst*I fragment, which starts at the *Nco*I site at position 482, within the region which is deleted from *ERA-1-399*, and continues through the homeobox to the *Pst*I site at position 1098. The probe was end labeled at the *Bgl*II of pERA-1-993$_{Nco\text{-}Pst}$ within the homeobox region (*ERA-1-993* position 890) and extends through the *Nco*I site and into the flanking pUC-9 vector sequence. The predicted fragments to be protected are shown: a 408-nucleotide fragment should be produced by protection of the probe by hybridization with *ERA-1-993* mRNA; a 265-nucleotide fragment should result from hybridization to *ERA-1-399* mRNA. The common splice junction (▼) and the *ERA-1-399*-specific splice acceptor and donor sites (▽) are indicated. The homeobox (▨) and the open reading frame (▬ are also indicated. The gels were exposed to Kodak XAR film with one intensifying screen for 5 (A) and 4 (B) days. nt, Nucleotides.

with the F90-2 homeobox (25), which is within the *Drosophila* Antennapedia cluster.

Another interesting aspect of the putative *ERA-1-993* protein is the presence of two polyhistidine tracts of 11 and 5 histidines. Homeobox genes often encode proteins with stretches of single amino acids (27 and references therein). However, none to date has been proposed to contain polyhistidine.

***ERA-1-399* cDNA predicts a second, differently spliced message.** An unexpected finding resulted from the sequencing of the shorter *ERA-1-399* cDNA. It was found that this cDNA was not lacking sequences present at the 5′ end of the *ERA-1-993* cDNA, as expected, but was missing 203 bp within the *ERA-1-993* open reading frame (Fig. 1 and 2). In this *ERA-1-399* cDNA, the nucleotides at positions 422 and 626 of *ERA-1-993* are joined, thereby deleting the intervening 203 bp. The effect of this additional splice is to cause a frameshift such that a truncated protein would be translated. The *ERA-1-399* open reading frame would presumably start from the same initiator AUG but would continue only 399 bp to a terminator TGA at *ERA-1-399* position 480 (*ERA-1-993* position 683). This mRNA would encode a protein of about 15 kilodaltons which would be identical to the *ERA-1-993* protein for the first 114 amino acids (to Asp-114) but which would then diverge and end after 19 additional amino acids. Therefore, while the *ERA-1-993* protein would contain the Hox 1.6 homeobox domain, the very similar *ERA-1-399* protein would not.

Detection of both the *ERA-1-993* and *ERA-1-399* RNAs. It is intriguing that the *ERA-1* gene may encode two distinct RNA species capable of directing the synthesis of similar proteins which differ mainly in the presence or absence of the

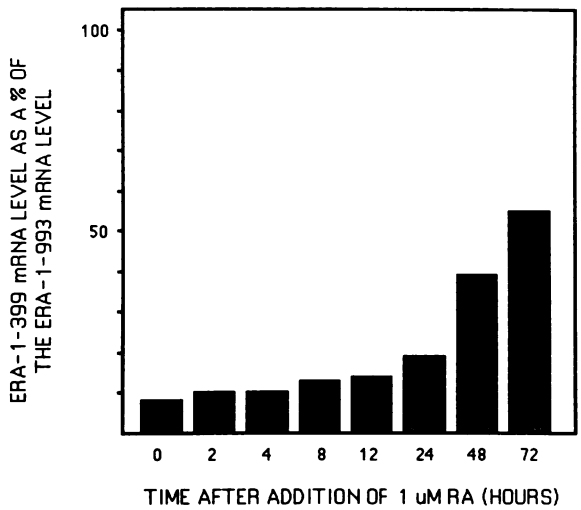

FIG. 5. The *ERA-1-399* RNA is preferentially induced at later times after RA addition. The protected fragments from the S1 protection experiments shown in Fig. 4A and B were quantitated by densitometry. The values for the *ERA-1-399* signal were compared with those for the corresponding signal for *ERA-1-993*. The results indicate the relative levels of *ERA-1-399* mRNA at various times after the addition of RA, expressed as percentages of the level of *ERA-1-993* mRNA.

homeobox domain. To verify this possibility, we attempted to directly demonstrate the existence of both the *ERA-1-993* and *ERA-1-399* RNA species predicted by the sequence data. F9 stem cells were grown in the absence or presence of 1 μM RA, and after various time periods, three RNA classes were isolated: total cellular, total poly(A)$^+$, and cytoplasmic RNAs. These RNA samples were subjected to S1 protection analysis (Fig. 4) by using a probe from the *ERA-1-993* cDNA which would distinguish between the two different species of *ERA-1* RNA. The probe (Fig. 4C) was derived from a 608-bp subclone of *ERA-1-993* cDNA. This fragment begins at the *Nco*I site (position 482) which is within the 203-bp segment that is deleted in the *ERA-1-399* cDNA, spans the homeobox region, and ends at the *Pst*I site (position 1098) downstream of the end of the open reading frame. This subclone was 5′ end labeled at the *Bgl*II (position 890) site within the homeobox, producing an antisense probe which is complementary to the *ERA-1-993* sequences from the *Bgl*II site within the homeobox to the upstream *Nco*I site and which continues into the flanking pUC-9 vector sequence. This probe produces a 408-bp protected fragment if hybridized to *ERA-1-993* mRNA or a 265-bp protected fragment if hybridized to *ERA-1-339* mRNA (Fig. 4C).

Both RNA species are detected in F9 cytoplasmic RNA 48 h after RA addition to the cells, suggesting that the longer *ERA-1-993* RNA is not simply a splicing precursor of *ERA-1-399* RNA (Fig. 4A, lanes 1 and 2). Both *ERA-1-993* and *ERA-1-399* RNAs are enriched in the poly(A)$^+$ but not the poly(A)$^-$ RNA fraction after RA treatment of the cells (Fig. 4A, lanes 10 to 13). Analysis of total cellular RNA isolated at various times after RA addition demonstrates that both RNA species are induced in F9 cells after RA addition with the same initial kinetics (Fig. 4A, lanes 3 to 9). S1 protection analysis with a different probe end labeled at the same *Bgl*II site but extending 890 bp through the 5′ end of the *ERA-1-993* cDNA into the flanking vector sequence produces an 890 bp protected fragment, indicating that this region of the *ERA-1-993* cDNA is present in cytoplasmic RNA and is not simply part of an unspliced precursor (data not shown). The results described are not in agreement with those of Baron et al. (3); these authors were unable to detect an S1-protected fragment when they used the Hox 1.6 cDNA-2 as a probe, and their cDNA-2 probe was essentially the same as *ERA-1-993* from position 1 to 127.

At early time points (<8 h), *ERA-1-399* RNA is present at a level which is approximately 10% of the *ERA-1-993* level. However, at later times, *ERA-1-399* RNA seems to be preferentially induced, since it reaches a level which is approximately 39% of the *ERA-1-993* level by 48 h and 56% of that level by 72 h after RA addition (Fig. 4A and B and Fig. 5). This suggests that there may be a temporally regulated activity which influences the splicing decisions with respect to these RNAs in that the affinity for the additional *ERA-1-399* splice sites is increased after 12 to 24 h

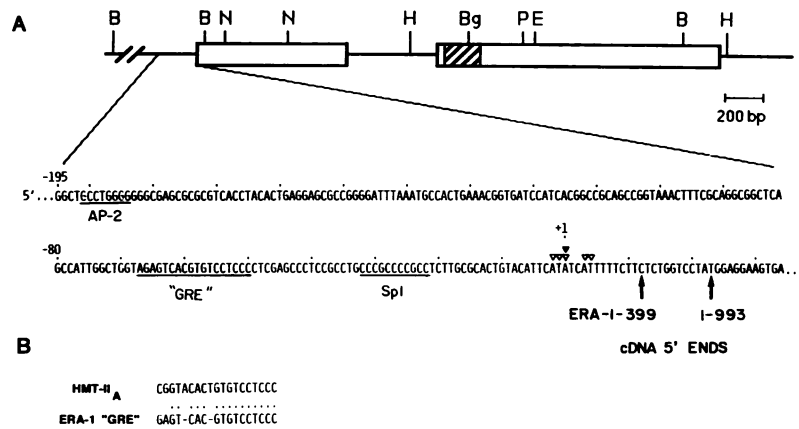

FIG. 6. Schematic representation of the *ERA-1* genomic structure and partial 5′-flanking sequence and a comparison of GREs. (A) The *ERA-1* gene structure is shown with the two exons (▭), the homeobox (▨), and some restriction sites, as indicated: B, *Bam*HI; N, *Nco*I; H, *Hin*dIII; Bg, *Bgl*II; P, *Pst*I; E, *Eco*RI. The DNA sequences of 195 bp of the 5′-flanking genomic region and 34 bp of the transcribed region are also shown. The positions of the transcription start sites, as mapped in Fig. 7 by S1 protection (▼) and primer extension (▽), are indicated. The position mapped by both S1 protection and primer extension (▼ and ▽) is referred to as position +1 for clarity. Potential binding sites for Sp1 and AP-2 (42, 43) and the *ERA-1* GRE-like sequence are indicated. (B) The *ERA-1* GRE-like sequence is aligned for comparison with the GRE region of the human metallothionein II_A promoter (HMT-II_A; 26).

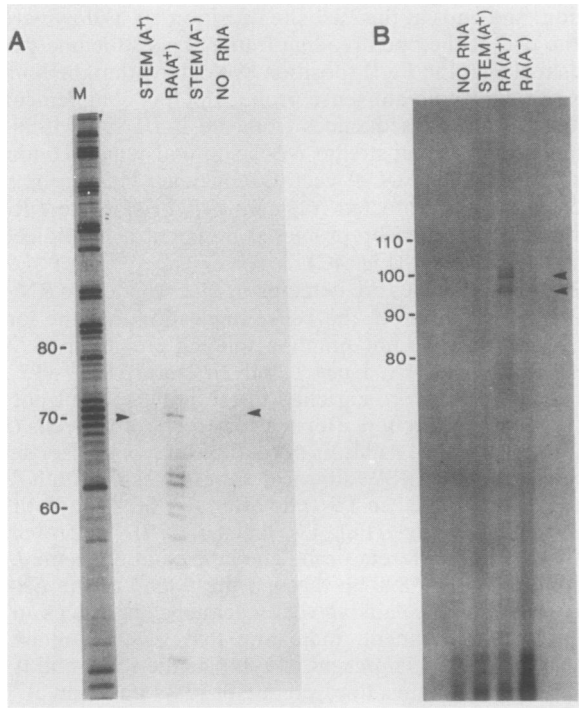

FIG. 7. Mapping the *ERA-1* transcription start site by S1 protection and primer extension analyses. (A) S1 protection was performed with a probe derived from a 1.7-kb *Bam*HI fragment of the λ*GERA-1* genomic clone which overlaps the 5′ end of the *ERA-1-993* and *ERA-1-399* cDNAs and extends approximately 1.7 kb upstream. The probe was made by 5′ end labeling at the *Bam*HI site, which corresponds to position 45 of the *ERA-1-993* cDNA. The probe is complementary to the first 45 sense nucleotides of the *ERA-1-993* cDNA and continues into the flanking sequence. This probe was hybridized with 15 μg of the following F9 RNA samples, and the results were analyzed as in Fig. 4. Lanes: 1, 72-h stem cell poly(A)$^+$; 2, 72-h RA-treated poly(A)$^+$; 3, 72-h stem cell poly(A)$^-$; 4, no-RNA control. (B) Primer extension analysis. A 5′-end-labeled synthetic oligonucleotide primer, complementary to positions 58 to 76 of the *ERA-1-993* cDNA sense strand, was hybridized with 15 μg of RNA and extended with avian myeloblastosis virus reverse transcriptase, and the products were visualized by denaturing acrylamide gel electrophoresis. Lanes: 1, no-RNA control; 2, 96-h stem cell plus dibutyryl cAMP poly(A)$^+$ RNA; 3, 72-h RA-treated poly(A)$^+$ RNA; 4, 72-h RA-treated poly(A)$^-$ RNA. Size markers (in bases) for both panels A and B were a complete set of sequencing reactions (T reaction is shown for A only). The gels were exposed to Kodak XAR film with one intensifying screen for 10 (A) and 8 (B) days.

of RA treatment. Alternatively, the shorter *ERA-1-399* message may become more stable than the longer *ERA-1-993* mRNA at the later time points.

These data indicate that both RNA species are induced in F9 cells after RA addition. Several other lines of evidence suggest that these two different RNA molecules arise because of alternate splicing of the primary transcript. The sequence surrounding the additional presumptive splice sites for *ERA-1-399* RNA (Fig. 1; *ERA-1-993* nucleotides 418 to 425 and 628 to 632) conforms fairly well to the consensus splice site sequences (44). Secondly, the two different RNAs do not arise from the transcription of different alleles. Southern blot analysis of F9 genomic DNA indicates that the gene is a single copy gene (per haploid genome), and the restriction pattern indicates that the *ERA-1* sequence which is missing from the *ERA-1-399* sequence is present in both genomic copies of the gene (32). If there were two distinct *ERA-1-993* and *ERA-1-399* alleles, then hybridization of the 5′ *Eco*RI fragment of *ERA-1-993* cDNA with *Bam*HI-digested F9 genomic DNA should produce two bands, one being 203 bp shorter. However, only one *Bam*HI band is seen (32). Additionally, only a single clone containing the *ERA-1* gene has been isolated (see below) from 200,000 plaques of a mouse spleen genomic library, and this piece of DNA contains the 203-bp sequence of *ERA-1-993* not present in *ERA-1-399* cDNA (data not shown).

Isolation of genomic sequences containing the entire *ERA-1* transcription unit. By using the 1,200-bp 5′ *Eco*RI fragment of *ERA-1-993* cDNA as a probe, a single clone (λGERA-1) containing 12 kb of mouse genomic DNA was isolated. This genomic clone appears to contain the entire *ERA-1* transcription unit along with approximately 7 kb of 5′-flanking sequence and 3 kb of the 3′-flanking sequence. The structure of the *ERA-1* gene was determined by DNA sequence, S1 protection, primer extension, and restriction mapping. The structure, splicing patterns, and 200 bp of the 5′-flanking sequence are shown in Fig. 2 and 6. The gene consists of a 743-bp 5′ exon 1, an approximately 450-bp intron, and a 1,487-bp 3′ exon 2. As stated above, this gene encodes both the *ERA-1-993* RNA of 2,230 nucleotides [without the poly(A) tail], as well as the less abundant *ERA-1-399* message which arises because of an additional splicing event that removes 203 bp within exon 1. The homeobox is located immediately downstream of the splice acceptor site of exon 2. This splice junction, common to both the *ERA-1-993* and *ERA-1-399* RNAs, occurs between *ERA-1-993* positions 721 and 722 (*ERA-1-399* positions 518 to 519) (Fig. 1 and 2).

The transcription start site has been mapped by S1 protection and primer extension analyses (Fig. 7A and B) and is shown in the diagram in Fig. 6A (closed and open triangles). A probe was derived from the 1.7-kb *Bam*HI fragment of the *ERA-1* genomic clone which contains the 5′ end of both of the *ERA-1* cDNAs as well as upstream genomic sequences; this fragment was 5′ end labeled at the *Bam*HI site corresponding to position 45 of the *ERA-1-993* cDNA. In S1 protection analysis, this probe produced a 71-bp protected fragment (the largest protected fragment) that was greatly enriched when the probe was hybridized with RNA from F9 cells treated with RA but not with RNA from untreated cells (Fig. 7A). This analysis maps the 5′ end of the *ERA-1* transcription unit to a point 23 bp further upstream of the 5′ end of the *ERA-1-933* cDNA (Fig. 6A, closed triangle). There are a series of shorter protected fragments which map to base pairs lying downstream of the proposed cap site that are adjacent to A · T base pairs. We believe that these protected fragments do not represent actual RNA species but are artifacts of the S1 analysis due to digestion at less tightly base-paired A · T positions.

The transcription start site was also mapped by primer extension analysis (Fig. 7B). A synthetic oligomer complementary to the *ERA-1-993* sequence from positions 58 to 76 was hybridized to poly(A)$^+$ RNA from F9 cells grown either in the presence or absence of RA and was extended by using reverse transcriptase. Major primer-extended products of 99 and 95 bp were synthesized (with some microheterogeneity) and were also greatly enriched with RNA from RA-treated cells (Fig. 7B). These heterogeneous extended products suggest that there may be several predominantly used cap sites, all close to one another, between 19 and 25 bp further upstream of the 5′ terminus of the *ERA-1-993* cDNA (Fig. 6A, open triangles). However, for further clarity we will refer to the A · T pair at the nucleotide position 23 bp upstream of the 5′ end of the *ERA-1-993* cDNA as the cap

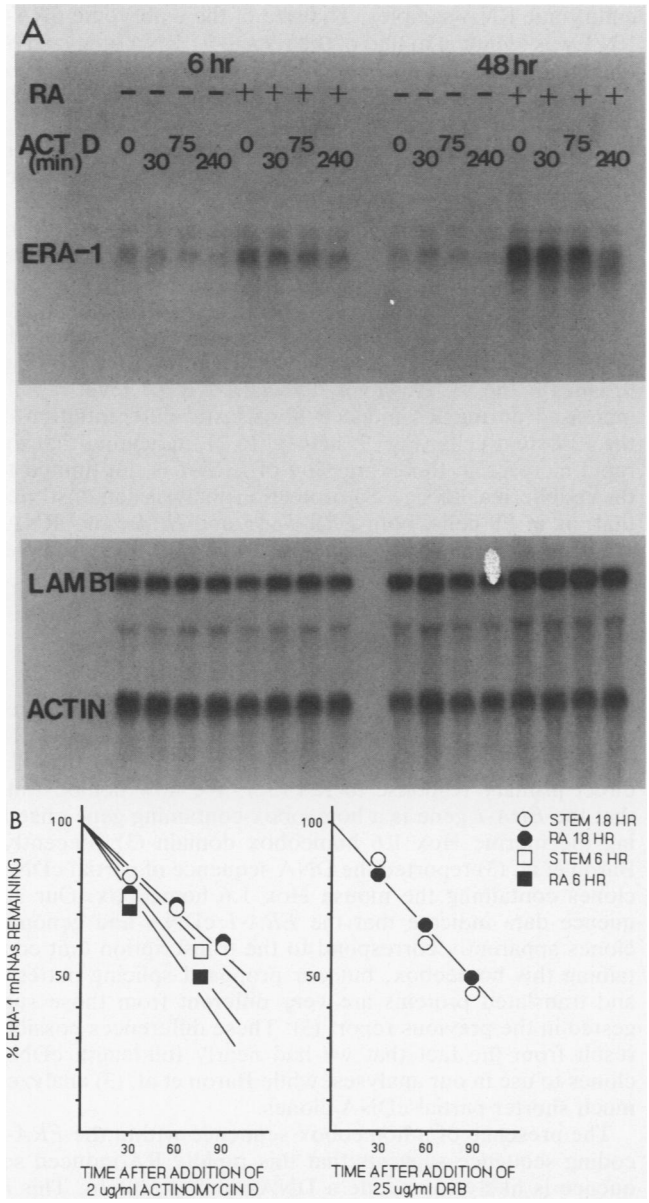

FIG. 8. Northern blot analysis demonstrating the equivalent relative stabilities of *ERA-1* mRNAs before and after RA addition. (A) F9 cells were grown for 6 or 48 h in the absence (−) or presence (+) of 1 μM RA. At the indicated times, RNA was immediately isolated or 2 μg of actinomycin D (ACT D) per ml was directly added to the medium and RNA was isolated after 30, 75, or 240 min. Total RNA (2 μg) was fractionated, blotted, and then hybridized with 6×10^5 cpm of the 1.0-kb 3′ *Eco*RI fragment of *ERA-1-993* cDNA per ml and then subsequently with 2.5×10^5 cpm each of mouse actin (56) and laminin B1 (LAMB) (64) probes per ml. The exposure times were 4.5 days for *ERA-1-993* and 10 h for actin and laminin. (B) Compilation of the densitometric quantitation of two separate RNA stability experiments. A second experiment, similar to the one described above, was performed with actinomycin D and DRB as transcription inhibitors. F9 cells were grown for 18 h with or without added RA (1 μM); RNA was isolated immediately, or either actinomycin D or DRB was added to 2 or 25 μg/ml, respectively, and RNA was isolated after 30, 60, or 90 min. The RNA samples were analyzed by Northern blot hybridization as described above. The Northern blots from both experiments were quantitated by scanning densitometry. The results are shown here as percentages of *ERA-1* RNA remaining at several time points after the addition of the respective inhibitor to 18-h stem (○), 18-h RA-treated (●), 6-h stem (□), or 6-h RA-treated (■) cells. The *ERA-1* hybridization signals were normalized to actin signals to control for amount of RNA loaded per lane. The *ERA-1* signal from cells treated with RA for 18 h and DRB for 90 min was normalized to laminin B1 because of a background problem with the actin signal for this lane.

site (i.e., nucleotide +1), as indicated by S1 protection analysis and by the position of one of the primer-extended products (Fig. 6A, solid and open triangles).

The 250-bp sequence of the 5′ flanking genomic region has been determined, and a portion of this sequence is shown in Fig. 6A. There is no typical TATA box upstream from the cap site, which may contribute to the somewhat heterogeneous start sites we have detected. There is also no obvious CCAAT box. However, there are potential binding sites for known transcription factors; there is a GC box at positions −22 to −31 with potential for Sp1 binding (for a review, see reference 42) and a potential site of AP-2 binding (43) at positions −184 to −191.

Interestingly, there is a sequence element at positions −50 to −67 in which 15 of 19 bp are identical to the region including the glucocorticoid regulatory element (GRE) of the human metallothionein II_A (HMT-II_A) gene (Fig. 6A and B) (26). When this *ERA-1* sequence was aligned with the GRE consensus sequence proposed by Karin et al. (26), 12 of 17 bp matched. This is the same extent of similarity between the proposed estrogen regulatory element of the *Xenopus* vitellogenin A2 gene (28) and this same GRE consensus sequence.

The increase in *ERA-1* RNA is not due to increased message stability. To further characterize the mechanism by which RA causes the induction of the *ERA-1* messages, we were interested in determining whether *ERA-1* mRNA increases resulted from increased transcription of the gene or increased stabilization of the mRNAs. To analyze the relative stability of *ERA-1* RNA before and after RA treatment, F9 stem cells were treated for 6 or 48 h with or without 1 μM RA. At 6 or 48 h, RNA was immediately isolated from some dishes of cells, while 2 μg of actinomycin D per ml was added to other dishes of cells to block transcription and RNA was isolated after an additional 30, 75, or 240 min. The RNA samples were analyzed by Northern blot hybridization (Fig. 8). The results suggest that the stability of *ERA-1* RNA is not significantly affected by RA treatment. In stem cells, the RNA decays with a half-life of approximately 90 to 100 min. After 6 or 48 h of RA treatment, while there is a higher level of RNA present, the RNA decays with approximately the same half-life. This result has been repeated at 18 h of RA treatment by using actinomycin D or 5,6-dichloro-1-β-D-ribofuranosylbenzimidazole (DRB; 52) as RNA synthesis blocking agents. Figure 8B shows a compilation of the data from both sets of Northern blot relative-message-stability experiments. The concentrations of actinomycin D and DRB used, in parallel labeling experiments, were found to inhibit the incorporation of [5,6-^{3}H]uridine by 90 to 95% while only inhibiting the incorporation of L-[^{35}S]methionine by 10 to 30%. These experiments suggest that the increases in *ERA-1* RNA levels do not result from stabilization of the messages, implying that RA treatment increases the transcriptional activity of the *ERA-1* gene.

Using nuclear run-off transcription analysis, we have been able to reproducibly detect only slight increases in the transcriptional activity of the *ERA-1* gene when control stem cell nuclei are compared with nuclei from RA-treated cells (J. Grippo and G. LaRosa, unpublished results). However,

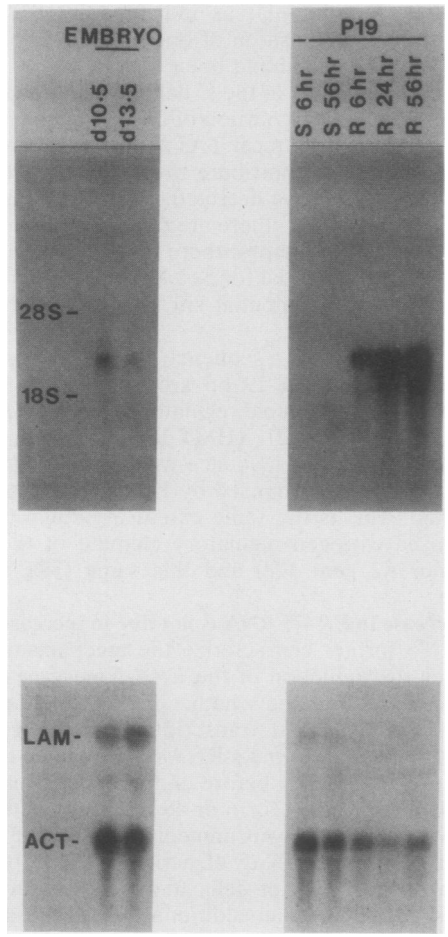

FIG. 9. Northern blot hybridization analysis of *ERA-1* expression in mouse embryos and P19 teratocarcinoma cells. Total RNA (15 μg) from mouse embryos or total RNA (10 μg) from P19 stem or 1 μM RA-treated cells was fractionated, blotted, and hybridized with 2×10^6 cpm of the 5′ 1.2-kb *Eco*RI fragment of *ERA-1-993* cDNA per ml first and then subsequently with 0.25×10^6 cpm each of laminin B1 and actin probes per ml. Lanes: 1, day 10.5 embryo; 2, day 13.5 embryo; 3, 6-h P19 stem cells; 4, 56-h P19 stem cells; 5, 6-h, RA-treated P19 cells; 6, 24-h, RA-treated P19 cells; 7, 56-h, RA-treated P19 cells. The exposure times were 6 days for *ERA-1* and 16 h for laminin and actin. The laminin B1 mRNA level does not show an increase during P19 RA-induced differentiation, as has been seen previously (65).

this inability to detect a larger RA-associated increase in transcription may be attributable to a low level of transcription, reflected by the low abundance of the message (32) combined with the relatively high specific background hybridization to the *ERA-1* probe in this assay. This background hybridization may be caused by the G · C-rich region (encoding the polyhistidine) or the homeobox.

Expression of *ERA-1* RNAs in mouse embryos and P19 teratocarcinoma stem cells. We sought to determine whether the expression of the *ERA-1* gene was a phenomenon unique to RA-treated F9 cells or whether the same RNA was present in other teratocarcinoma stem cells or in the mouse embryo. RNAs extracted from day 10.5 and day 13.5 mouse embryos were assayed for *ERA-1* RNA by Northern blot analysis (Fig. 9, lanes 1 and 2). Using the 5′ 1.2-kb *Eco*RI fragment of *ERA-1-993* cDNA, we found that the typical 2.2- to 2.4-kb RNA band hybridized with the probe in both embryonic RNA samples. The size of the embryonic *ERA-1* RNA was identical to that of the F9 *ERA-1* RNA analyzed on the same gel (data not shown). The level of *ERA-1* RNA decreased between days 10.5 and 13.5 (Fig. 9, lanes 1 and 2). This decrease agrees with that previously reported for Hox 1.6, but, inexplicably, the sizes of the hybridizing RNA in embryos (10 and 4.5 kb) reported previously (3) are very different from what we observed.

The mouse P19 teratocarcinoma stem cell line differentiates into fibroblastic cells when exposed to RA in monolayer cultures (38). P19 cells were grown for 6, 24, or 56 h either with or without 1 μM RA, and RNA was isolated and analyzed by Northern blot hybridization (Fig. 9, lanes 3 to 7). In P19 stem cells, there was very little *ERA-1* RNA (Fig. 9, lanes 3 and 4). However, the *ERA-1* RNA level rapidly increased during RA-induced fibroblastic differentiation of the P19 stem cells (Fig. 9, lanes 5 to 7), indicating that the rapid increase in the expression of *ERA-1* is not limited to the endodermal lineage. S1 protection analysis demonstrated that, as in F9 cells, both *ERA-1-993* and *ERA-1-399* RNAs are present in the embryonic RNA and P19 RNA samples (data not shown).

DISCUSSION

We report here the further characterization of the previously described *ERA-1* gene. This clone, originally isolated on the basis of its rapid induction in F9 stem cells after RA treatment, has been shown to have the characteristics of a direct primary response to RA (32). We now demonstrate that the *ERA-1* gene is a homeobox-containing gene encoding the murine Hox 1.6 homeobox domain (3). Recently, Baron et al. (3) reported the DNA sequence of partial cDNA clones containing the mouse Hox 1.6 homeobox. Our sequence data indicate that the *ERA-1* cDNA and genomic clones apparently correspond to the transcription unit containing this homeobox, but our proposed splicing patterns and translated proteins are very different from those suggested in the previous report (3). These differences possibly result from the fact that we had nearly full-length cDNA clones to use in our analyses, while Baron et al. (3) analyzed much shorter partial cDNA clones.

The presence of a homeobox sequence within the *ERA-1* coding sequence suggests that this rapidly RA-induced sequence is likely to encode a DNA-binding protein. This is consistent with a model for RA action in which a direct primary response to RA is the induction of *ERA-1*, which encodes a DNA-binding protein involved in the regulation of secondary-response genes. Whether the putative protein of *ERA-1* actually regulates any genes which show a slower response to RA remains to be shown. As more promoter sequences of the RA-induced late-response genes are characterized, transient cotransfection experiments with *ERA-1-993* cDNA expression vectors and vector constructs containing the RA-responsive promoter regions of these late genes fused to a reporter gene should make it possible to determine whether the *ERA-1-993* protein interacts with any of these genes. The promoters of the other homeobox sequences, such as Hox 1.1 or Hox 1.2, which are induced at later times after RA addition to F9 cells (7, 12, 13), would be especially likely candidates for regulation by *ERA-1-993*, since it has been shown previously that some homeobox-containing proteins may regulate the expression of others (for a review, see reference 22), while some homeobox proteins have been shown to bind to DNA sequences upstream of their own, as well as other, homeobox regions (16, 21). However, the

experiments to address this possibility with *ERA-1-993* await the characterization of the promoter regions of these other homeobox genes, which are expressed at later times after RA addition to F9 cells.

The *ERA-1* gene gives rise to two differentially processed RNA molecules which would encode two very similar proteins, the approximately 36-kilodalton protein of *ERA-1-993*, which contains the homeobox, and the approximately 15-kilodalton truncated protein of *ERA-1-399*, which does not contain the homeobox (see Fig. 1 and 2). Similarly, the *Xenopus X1Hbox2* gene was found to produce two major transcripts, one which contains the homeobox-coding region and another truncated RNA which does not (67). These two different RNAs arise by differential splicing which results in one RNA molecule with the homeobox region and another RNA without the homeobox. Although the alternate splicing is different from what is seen for *ERA-1*, it would produce the same result, i.e., two proteins which are identical through the amino-terminal portion but which differ primarily in the presence or absence of the homeobox domain. Baron et al. (3) did not detect both of the reading frames reported here, presumably because they did not have a cDNA, like *ERA-1-993*, that contains the region shown boxed in Fig. 1.

This ability to produce two similar proteins, one a homeobox-containing protein and one a homeoboxless protein, raises the potential for some interesting regulatory interactions. While the homeobox domain would presumably direct sequence-specific DNA binding, the amino-terminal portion of the protein may be involved in interaction(s) with one or more other regulatory protein(s). Such an interaction could regulate either the DNA-binding activity or the transcriptional regulatory activity of the *ERA-1-993* homeobox domain protein. The homeoboxless *ERA-1-399* protein could act as a competitive inhibitor of the *ERA-1-993* protein by competing for interaction with such regulatory protein(s) while being unable to bind to DNA. The finding that *ERA-1-399* RNA is initially less abundant than *ERA-1-993* RNA but is preferentially increased at later times after RA addition (Fig. 4 and 5) is consistent with the possibility that the *ERA-1-399* protein may be required later to reduce the initial activity and effects of the *ERA-1-993* homeobox protein. This also suggests the possibility that the affinity of the splicing apparatus for the additional *ERA-1-399* splice sites may be developmentally regulated. Antibodies which should recognize each of the two potential proteins are being generated, and experiments to directly demonstrate the presence of these two proteins in teratocarcinoma cells and in embryonic cells can then be performed.

Analysis of the *ERA-1* promoter region (Fig. 6) reveals that the region lacks typical TATA or CCAAT box homologies but does contain potential binding sites for the transcription factors Sp1 and AP-2 (42, 43). These factors have been shown to bind to and stimulate transcription from both the simian virus 40 early region promoter and the HMT-II_A promoter, and they may activate other cellular promoters (see references 42 and 43 for references). Whether these sites actually function as regulatory elements for the *ERA-1* gene remains to be shown. Additionally, a GRE-like sequence which is very similar to the region containing the GRE in the HMT-II_A 5′-flanking region (26) is located at positions −50 to −67. This is interesting in light of the recent reports that a nuclear RA receptor is a member of the steroid hormone receptor protein family (23, 47). The cDNA clones for this receptor were isolated on the basis of their homology to nuclear hormone receptor DNA-binding domains. The RA receptor DNA-binding domain shows a 43 to 62% exact amino acid match with the DNA-binding domains of several nuclear hormone receptors (23, 47). It is likely that the DNA target sequence for this RA receptor would likewise be similar to the target sequence for the steroid receptors. The proposed target sites for the related glucocorticoid and estrogen receptors show significant similarity (26, 28). This GRE-like sequence upstream of the ERA cap site shows the same extent of similarity to the GRE consensus as does the proposed estrogen regulatory element sequence. This suggests that the *ERA-1* gene may possibly be a target gene for the RA receptor. Experiments involving cotransfection of the RA receptor construct and various *ERA-1* promoter-reporter chimeras will answer this directly.

The *ERA-1* gene is a direct target for RA action in F9 cells, and if it is regulated in a fashion similar to that of steroid hormone-responsive genes, then induction of the *ERA-1* gene should occur at the level of transcriptional initiation. Experiments using RNA synthesis inhibitors to estimate the relative stabilities of the *ERA-1* RNAs indicate that there are no significant differences in their stabilities before versus after RA addition to the cells, with the apparent half-lives of the RNAs being approximately 85 to 105 min (Fig. 8). This half-life value must be qualified by the fact that the drugs themselves may indirectly affect the stability of particular messages (18). However, the same results were obtained with two different inhibitors. The important point is that the stability of the messages under these conditions is not affected by the addition of RA. An in vivo approach to steady-state labeling experiments could provide exact half-life measurements but are impractical because of the low abundance of the *ERA-1* RNAs (32). Additional data indicating a transcriptional response of the *ERA-1* gene has come from preliminary results with transient transfection assays that indicate that 1.7 kb of 5′-flanking *ERA-1* genomic sequences can direct enhanced transcription of the bacterial *CAT* gene in F9 cells after RA treatment (G. Vasios and G. LaRosa, unpublished results).

Finally, the analysis of the pattern of expression of the *ERA-1* gene presented here demonstrates that this expression is not limited to cells differentiating along the endodermal pathway. The *ERA-1* transcripts are expressed in P19 teratocarcinoma stem cells in response to RA as these cells differentiate into fibroblastic cells (Fig. 9). Additionally, we found that the typical 2.2- to 2.4-kb *ERA-1* RNAs were present in mouse embryos at day 10.5 and, at a reduced level, by day 13.5 (Fig. 9). Baron et al. (3) have reported that a Hox 1.6 probe detects 10-kb and 4.5-kb bands along with the 28S and 18S RNAs on Northern blots of RNA from day 9, 10, and 13 mouse embryos. At present, we are unable to explain this difference between their data and our data. The *ERA-1*/Hox 1.6 homeobox shows a high degree of homology with the F90-2 homeobox within the *Drosophila* Antennapedia complex (Fig. 3). The gene containing this homeobox is expressed in both anterior and posterior embryonic structures outside of the regions of segmentation (25). Also, the F90-2 sequence detects two differently sized RNAs which are temporally regulated during embryogenesis (25). The determination of the spatial and temporal expression of the *ERA-1*/Hox 1.6 homeobox gene in early mouse embryos by the use of in situ hybridization of RNA or immunolocalization of the *ERA-1* proteins will be of major interest.

The data presented here on the further characterization of this *ERA-1*/Hox 1.6 gene point to several interesting and important lines of investigation. Future studies concerning this gene will add significant new information pertaining to

the mechanism by which RA can regulate specific gene expression during differentiation and development. Additionally, further work with this gene should contribute to our understanding of the regulation of mammalian homeobox genes and the functions of their protein products.

ACKNOWLEDGMENTS

We thank the members of the Gudas laboratory for helpful suggestions and discussions and Taryn Resnick for help in preparing the manuscript for publication.

This research was supported by Public Health Service grant R01CA39036 from the National Cancer Institute. L.J.G. is an Established Investigator of the American Heart Association; G.J.L. is currently a recipient of a predoctoral fellowship from the National Cancer Center and was formerly supported by Public Health Service grant IT32CA09361 from the National Institutes of Health.

LITERATURE CITED

1. **Ashburner, M., C. Chihara, P. Meltzer, and G. Richards.** 1975. Temporal control of puffing activity in polytene chromosomes. Cold Spring Harbor Symp. Quant. Biol. **38**:655–662.
2. **Awgulewitsch, A., M. F. Utset, C. F. Hart, W. McGinnis, and F. H. Ruddle.** 1986. Spatial restriction in expression of a mouse homeobox locus within the central nervous system. Nature (London) **320**:328–335.
3. **Baron, A., M. S. Featherstone, R. E. Hill, A. Hall, B. Galliot, and D. Buboule.** 1987. Hox-1.6: a mouse homeo-box-containing gene member of the Hox-1 complex. EMBO J. **6**:2977–2986.
4. **Benton, W. D., and R. W. Davis.** 1977. Screening of gt recombinant clones by hybridization to single plaques *in situ*. Science **196**:180–182.
5. **Berk, A. J., and P. A. Sharp.** 1977. Sizing and mapping of early adenovirus mRNAs by gel electrophoresis of S1 endonuclease digested hybrids. Cell **12**:721–732.
6. **Bollag, W.** 1979. Retinoids and cancer. Cancer Chemother. Pharmacol. **3**:207–215.
7. **Brier, G., M. Bucan, U. Francke, A. M. Colberg-Poley, and P. Gruss.** 1986. Sequential expression of murine homeobox genes during F9 EC cell differentiation. EMBO J. **5**:2209–2215.
8. **Carrol, S. B., and M. P. Scott.** 1985. Localization of the fushi tarazu protein during Drosophila embryogenesis. Cell **43**:44–57.
9. **Cavenar, D. R.** 1987. Comparison of the consensus sequence flanking translational start sites in Drosophila and vertebrates. Nucleic Acids Res. **16**:1353–1361.
10. **Chen, E. Y., and P. H. Seeburg.** 1985. Supercoiled sequencing: a fast and simple method for sequencing plasmid DNA. DNA **4**: 165–170.
11. **Chytil, F., and D. E. Ong.** 1978. Cellular vitamin A binding proteins. Vitam. Horm. **36**:1–32.
12. **Colberg-Poley, A., S. Voss, K. Chowdhury, and P. Gruss.** 1985. Structural analysis of murine genes containing homeobox sequences and their expression in embryonal carcinoma cells. Nature (London) **314**:731–738.
13. **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.
14. **Cox, R. A.** 1968. The use of guanidinium chloride in the isolation of nucleic acid. Methods Enzymol. **12**:120–129.
15. **Croce, C. M., A. Linnenbach, K. Huebner, J. R. Parnes, D.H. Margulies, E. Appella, and J. G. Siedman.** 1981. Control of expression of histocompatibility antigens (H-2) and β_2-microglobulin in F9 teratocarcinoma stem cells. Proc. Natl. Acad. Sci. USA **78**:5754–5758.
16. **Desplan, C., J. Theis, and P. H. O'Farrell.** 1985. The Drosophila developmental gene, engrailed, encodes a sequence-specific DNA binding activity. Nature (London) **318**:630–635.
17. **Dinardo, S., J. M. Kuner, J. Theis, and P. H. O'Farrell.** 1985. Development of embryonic pattern in D. melanogaster as revealed by accumulation of the nuclear engrailed protein. Cell **43**: 59–69.
18. **Dreyfuss, G., S. A. Adam, and Y. D. Choi.** 1984. Physical change in cytoplasmic messenger ribonucleoproteins in cells treated with inhibitors of mRNA transcription. Mol. Cell. Biol. **4**:415–423.
19. **Duboule, D., A. Baron, P. Mahl, 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.
20. **Duprey, P., D. Morello, M. Vasseur, C. Babinet, H. Condanine, P. Brulet, and F. Jacob.** 1985. Expression of the cytokeratin endo A gene during early mouse embryogenesis. Proc. Natl. Acad. Sci. USA **82**:8535–8539.
21. **Fainsod, A., L. D. Bogorad, T. Ruusala, M. Lubin, D. M. Crothers, and F. H. Ruddle.** 1986. The homeo domain of a murine protein binds 5′ to its own homeo box. Proc. Natl. Acad. Sci. USA **83**:9532–9536.
22. **Gehring, W. J.** 1987. Homeo boxes in the study of development. Science **236**:1245–1252.
23. **Gigure, V., E. S. Ong, P. Segui, and R. M. Evans.** 1987. Identification of a receptor for the morphogen retinoic acid. Nature (London) **330**:624–629.
24. **Haff, L. A., and L. Bogorad.** 1976. Poly(adenylic acid)-containing RNA from plasmids of maize. Biochemistry **15**:4110–4115.
25. **Hoey, T., H. J. Doyle, K. Harding, C. Wedeen, and M. Levine.** 1986. Homeo box expression in anterior and posterior regions of Drosophila embryo. Proc. Natl. Acad. Sci. USA **83**:4809–4813.
26. **Karin, M., A. Haslinger, H. Holtgreve, R. I. Richards, P. Krauter, H. M. Westphal, and M. Beato.** 1984. Characterization of DNA sequences through which cadmium and glucocorticoid hormones induce human metallothionein-II_A gene. Nature (London) **308**:513–519.
27. **Kessel, M., F. Schulze, M. Fibi, and P. Gruss.** 1987. Primary structure and nuclear localization of a murine homeodomain protein. Proc. Natl. Acad. Sci. USA **84**:5306–5310.
28. **Klein-Hitpab, L., M. Schorpp, U. Wagner, and G. U. Ryffet.** 1986. An estrogen-responsive element derived from the 5′ flanking region of the Xenopus vitellogenin A2 gene functions in transfected human cells. Cell **46**:1053–1061.
29. **Kozak, M.** 1984. Compilation and analysis of sequences from the translational start site in eukaryotic mRNAs. Nucleic Acids Res. **12**:857–872.
30. **Krumlauf, R., P. W. H. Holland, J. H. McVey, and B. L. M. Hogan.** 1987. Developmental and spatial patterns of expression of the mouse homeobox gene, Hox 2.1. Development **9**:603–617.
31. **Kurkinen, M., D. P. Barlow, D. M. Helfman, J. R. Williams, and B. L. M. Hogan.** 1983. Isolation of cDNA clones for basal lamina components: type IV collagen. Nucleic Acids Res. **11**: 6199–6209.
32. **LaRosa, G. J., and L. J. Gudas.** 1988. An early effect of retinoic acid: cloning of an mRNA (ERA-1) exhibiting rapid and protein synthesis-independent induction during teratocarcinoma stem cell differentiation. Proc. Natl. Acad. Sci. USA **85**:329–333.
33. **Laughton, A., and M. P. Scott.** 1984. Sequence of a Drosophila segmentation gene: protein structure homology with DNA-binding proteins. Nature (London) **310**:25–31.
34. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
35. **Martin, G. R.** 1980. Teratocarcinomas and mammalian embryogenesis. Science **209**:768–776.
36. **Martin, G. R.** 1987. Nomenclature for homeobox-containing genes. Nature (London) **325**:21–22.
37. **Mason, I. J., A. Taylor, J. G. Williams, H. Sage, and B. L. M. Hogan.** 1986. Evidence from molecular cloning that SPARC, a major product of mouse embryo parietal endoderm, is related to an endothelial cell "culture shock" glycoprotein of M_r 43,000. EMBO J. **5**:1465–1472.
38. **McBurney, M. W., E. M. Jones-Villeneuve, M. K. Edwards, and P. J. Anderson.** 1982. Control of muscle and neuronal differentiation in a cultures embryonal carcinoma cell line. Nature (London) **299**:165–167.
39. **McGinnis, W., R. L. Garber, J. Wirz, A. Kuroiwa, and W. J. Gehring.** 1984. A homologous protein-coding sequence in Dro-

sophila homeotic genes and its conservation in other metazoans. Cell **37**:403–408.

40. **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.
41. **McGinnis, W., M. S. Levine, E. Hafen, A. Kuroiwa, and W. J. Gehring.** 1984. A conserved DNA sequence in homeotic genes of the Drosophila Antennapedia and bithorax complexes. Nature (London) **308**:428–433.
42. **McKnight, S., and R. Tjian.** 1987. Transcriptional selectivity of viral genes in mammalian cells. Cell **46**:795–805.
43. **Mitchell, P. J., C. Wang, and R. Tjian.** 1987. Positive and negative regulation of transcription *in vitro*: enhancer binding protein AP-2 is inhibited by SV40 T antigen. Cell **50**:847–861.
44. **Mount, S.** 1982. A catalogue of splice junction sequences. Nucleic Acid Res. **10**:459–472.
45. **Odenwald, W. F., C. F. Taylor, F. J. Palmer-Hill, V. Friedrich, Jr., M. Tani, and R. A. Lazzarini.** 1987. Expression of a homeo domain protein in noncontact-inhibited cultured cells and postmitotic neurons. Genes Dev. **1**:482–496.
46. **O'Malley, B. W., and A. L. Means.** 1974. Female steroid hormones and target cell nuclei. Science **183**:610–620.
47. **Petkovich, M., N. J. Brand, A. Krust, and P. Chambon.** 1987. A human retinoic acid receptor which belongs to the family of nuclear receptors. Nature (London) **330**:444–450.
48. **Proudfoot, N. J., and G. G. Brownlee.** 1976. 3′ non-coding region sequences in eukaryotic messenger RNA. Nature (London) **263**:211–213.
49. **Roberts, A. B., and M. B. Sporn.** 1984. Cellular biology and biochemistry of the retinoids, p. 210–286. *In* M. B. Sporn, A. B. Roberts, and D. S. Goodman (ed.), The retinoids, vol. 2. Academic Press, Inc., New York.
50. **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.
51. **Scott, M. P., and A. J. Weiner.** 1984. Structural relationships among genes that control development: sequence homology between the Antennapedia, Ultrabithorax and fushitarazu loci of Drosophila. Proc. Natl. Acad. Sci. USA **81**:4115–4119.
52. **Sehgal, P. B., J. E. Darnell, Jr., and I. Tamm.** 1976. The inhibition of DRB (5,6-dichloro-1-β-D-ribofuranosyl benzimidazole) of hnRNA and mRNA production in HeLa cells. Cell **9**: 473–480.
53. **Shaw, G., and R. A. Kamen.** 1986. A conserved AU sequence from the 3′ untranslated region of GM-CSF mRNA media selective mRNA degradation. Cell **46**:659–667.
54. **Shepard, J. C. W., W. McGinnis, A. E. Carrasco, E. M. DeRobertis, and W. J. Gehring.** 1984. Fly and frog homeo domains show homologies with yeast mating type regulatory proteins. Nature (London) **320**:763–765.
55. **Singer, P. A., K. Trevor, and R. G. Oshima.** 1986. Molecular cloning and characterization of the endo B cytokeratin expressed in preimplantation mouse embryos. J. Biol. Chem. **261**: 538–547.
56. **Spiegelman, B. M., M. Frank, and H. Green.** 1983. Molecular cloning of mRNA from 3T3 adipocytes: regulation of mRNA content for glycophosphate dehydrogenase and other differentiation-dependent proteins during adipocyte development. J. Biol. Chem. **258**:10083–10089.
57. **Sporn, M. B., and A. B. Roberts.** 1983. Role of retinoids in differentiation and carcinogenesis. Cancer Res. **43**:3034–3040.
58. **Strickland, S., and V. Mahdavi.** 1978. The induction of differentiation in teratocarcinoma stem cells by retinoic acid. Cell **15**: 393–403.
59. **Strohman, R. C., P. S. Moss, J. M. Spector, D. Pryzybyla, and B. Paterson.** 1977. Messenger RNA for myosin polypeptides: isolation from single myogenic cell cultures. Cell **10**:265–273.
60. **Thaller, C., and G. Eichele.** 1987. Identification and spatial distribution of retinoids in the developing chick limb bud. Nature (London) **327**:625–628.
61. **Tickle, C., B. Alberts, L. Wolpert, and J. Lee.** 1982. Local application of retinoic acid to the limb bud mimics the action of the polarizing region. Nature (London) **296**:564–565.
62. **Tickle, C., J. Lee, and G. Eichele.** 1985. A quantitative analysis of the effect of all-trans-retinoic acid on the pattern of chick wing development. Dev. Biol. **109**:82–85.
63. **Vogelstein, B., and D. Gillespie.** 1979. Preparative and analytical purification of DNA from agarose. Proc. Natl. Acad. Sci. USA **76**:615–619.
64. **Wang, S. Y., and L. J. Gudas.** 1983. Isolation of cDNA clones specific for collagen IV and laminin from mouse teratocarcinoma cells. Proc. Natl. Acad. Sci. USA **80**:5880–5884.
65. **Wang, S. Y., G. J. LaRosa, and L. J. Gudas.** 1985. Molecular cloning of gene sequences transcriptionally regulated by retinoic acid and dibutyryl cyclic AMP in cultured mouse teratocarcinoma cells. Dev. Biol. **107**:75–86.
66. **White, R. A. H., and M. Wilcox.** 1984. Protein products of the bithorax complex in Drosophila. Cell **39**:163–171.
67. **Wright, C. V. E., K. W. Y. Cho, A. Fritz, T. R. Burglin, and E. M. DeRobertis.** 1987. A *Xenopus laevis* gene encodes both homeobox-containing and homeobox-less transcripts. EMBO J. **6**:4083–4094.