MOLECULAR AND CELLULAR BIOLOGY, Jan. 1991, p. 412–424
0270-7306/91/010412-13$02.00/0
Copyright © 1991, American Society for Microbiology
Vol. 11, No. 1

The *Xenopus* B1 Factor Is Closely Related to the Mammalian Activator USF and Is Implicated in the Developmental Regulation of TFIIIA Gene Expression

HILDEGARD KAULEN, PHILIPPE POGNONEC, POLLY D. GREGOR, AND ROBERT G. ROEDER*

Laboratory of Biochemistry and Molecular Biology, The Rockefeller University, 1230 York Avenue, New York, New York 10021-6399

Received 7 August 1990/Accepted 24 October 1990

The *Xenopus laevis* TFIIIA promoter contains a motif that has been implicated in promoter activation in late-stage oocytes and contains the sequence (−269) CACGTG (−264). A cDNA encoding a protein (B1) that binds to this element has been cloned from *X. laevis* and *Xenopus borealis* ovarian cDNA libraries. We show that this protein is a member of the helix-loop-helix family of regulatory proteins and contains 80% sequence identity with the human adenovirus major late transcription factor (MLTF or USF). A survey of B1 protein expression during oogenesis and embryogenesis revealed both oocyte-specific and somatic cell-specific B1 protein-DNA complexes. Immunological data, RNA blot analysis, and proteolytic clipping band shift assays indicated that these complexes most likely represent altered forms of a single B1 polypeptide. Implications for TFIIIA gene regulation during development are discussed.

The regulation of gene expression in eucaryotes involves combinatorial interactions of positive and negative regulatory factors with specific promoter elements. The activities of individual factors may be controlled in turn by posttranslational modifications, by overall abundance, or by interactions with other factors (28).

The positively acting gene-specific transcription factor IIIA (TFIIIA) is required for the expression of 5S RNA genes by RNA polymerase III (4, 16). Expression of the TFIIIA gene during *Xenopus laevis* development provides a model system with which to investigate the molecular mechanisms underlying developmental control at the level of transcription. Both TFIIIA protein and steady-state mRNA levels are highest in early oogenesis, decrease in later stage oocytes, and are drastically reduced during embryogenesis and in somatic cells (18, 36, 42, 47).

The TFIIIA gene has been cloned (35, 45), and a region extending 306 bp upstream of the transcriptional start site has been demonstrated to be sufficient for wild-type expression of the TFIIIA gene in oocytes (27, 35). By analyzing TFIIIA promoter mutants, it was shown that expression of the TFIIIA gene in oocytes is under the control of at least three positively and one negatively acting *cis* elements. The positively acting elements are located from positions −289 to −253, −250 to −173, and −144 to −101. A negative element lies between positions −306 and −289 (35).

Two DNA-binding proteins (B1 and B2) which show specific interactions with two of the positively acting elements were identified in whole ovary extract (35). The B1 binding site includes a 6-bp palindrome (5′CACGTG-3′; positions −269 to −264) previously shown to be part of the recognition sequence for the adenovirus major late transcription factor, USF or MLTF (7, 32). Gel shift competition analysis indicated that the B1 protein shares sequence recognition specificity with human USF (hUSF) (35). Binding of the B2 protein is mediated by promoter sequences from positions −235 to −221. Point mutations in the CACGTG motif of the B1 site or substitutions between positions −239 and −229 of the B2 site interfere both with DNA-protein complex formation in vitro and with transcription of the TFIIIA gene in late-stage oocytes, leading to the hypothesis that the corresponding DNA-binding proteins are regulators of TFIIIA gene expression (35). Hall and Taylor (20) also described a *Xenopus* factor (TDEF, for TFIIIA distal element factor) that binds to the CACGTG element and further reported that this recognition site is required for TFIIIA gene expression in oocytes but in not somatic cells.

The major late transcription factor USF is a ubiquitous protein found in many different cell types. It activates the adenovirus major late promoter (32) and several other cellular genes (8, 10). Purified USF from HeLa cells is composed of 43- and 44-kDa polypeptides that appear to bind independently to DNA (33). The 43-kDa form of hUSF was recently cloned on the basis of amino acid sequence information (19a). It is a member of the Myc-related family of helix-loop-helix regulatory proteins (29).

Since the USF-related *Xenopus* B1 factor was implicated in differential TFIIIA gene expression in oocytes versus somatic cells, we set out to clone this factor by using an hUSF-specific cDNA and to compare the B1 protein-DNA complexes formed during various developmental stages when TFIIIA gene expression is altered. Our studies indicate that the *X. laevis* B1 protein (designated Xl-USF) is highly homologous (80% sequence identity) to the 43-kDa form of hUSF and that the B1 protein-DNA complexes formed with oocyte and somatic cell extracts are distinct and likely result from altered forms of a common B1 polypeptide. Structure-function relationships for the Xl-USF protein are discussed.

MATERIALS AND METHODS

Preparation of extracts. Total ovary extracts were prepared as described by Scotto et al. (35). Staged oocyte extracts were made from manually isolated oocytes, which were segregated into stage 1 through stage 6 according to Dumont (14), after the oocytes had been released from the

* Corresponding author.

ovary following collagenase IA treatment (35). The exact number of oocytes was determined by counting. The oocytes were lysed by vortexing in 2 volumes of a lysis buffer containing 50 mM Tris hydrochloride (pH 8.4), 170 mM KCl, 0.2 mM EDTA, 0.5 mM dithiothreitol (DTT), 0.5 mM phenylmethylsulfonyl fluoride (PMSF), and 20% glycerol. The homogenate was centrifuged at 10,000 × *g* at 4°C for 30 min, and aliquots of the supernatant were quickly frozen and stored at −70°C.

The X58 cell line (derived from disaggregated embryos; courtesy of Raymond Reeves, Washington State University) was grown in monolayer culture in an atmosphere of 5% (vol/vol) CO_2 at 22°C in Dulbecco modified Eagle medium diluted by one-third with distilled water and supplemented with 10% fetal calf serum. The XTC cell line (38) and the SHA cell line (derived from adult kidney; courtesy of D. Shapiro, University of Illinois) were grown in a similar way. The cells were harvested and washed with phosphate-buffered saline. The cell pellet was resuspended in twice the packed cell volume of hypotonic buffer (10 mM *N*-2-hydroxyethylpiperazine-*N'*-2-ethanesulfonic acid [HEPES; pH 7.9], 10 mM KCl, 1.5 mM $MgCl_2$, 0.5 mM DTT, 0.5 mM PMSF). The cells were allowed to swell for 30 min on ice and then disrupted in a motor-driven glass-Teflon homogenizer with tight fitting. To this homogenate was added 0.1 volume of a solution containing 0.3 M HEPES (pH 7.9), 1.5 M KCl, and 0.03 M $MgCl_2$. The lysate was centrifuged at 100,000 × *g* at 4°C for 30 min. Aliquots of the supernatant were quickly frozen and stored at −70°C.

Embryos were generated by fertilizing eggs in vitro according to Krieg and Melton (23). Embryonic stages were collected according to Nieuwkoop and Faber (31). The embryos in each batch were counted. The embryos were completely dejellied by gentle swirling in 2% cysteine hydrochloride (pH 7.8). Dejellied embryos were rinsed several times in TE (10 mM Tris hydrochloride [pH 7.4], 1 mM EDTA) and then homogenized in 2 volumes of a lysis buffer containing 50 mM Tris hydrochloride (pH 8.6), 170 mM KCl, 0.2 mM EDTA, 0.5 mM PMSF, 0.5 mM DTT, and 20% glycerol. Early embryos were disrupted by vortexing; later-stage embryos were disrupted by sonication. The extracts were cleared by spinning at 10,000 × *g* at 4°C. Aliquots were quickly frozen and stored at −70°C. Protein concentrations were determined according to Bradford (5).

Gel shift assays. The gel shift conditions have been described previously (35). Electrophoresis was carried out through a 4 or 6% polyacrylamide gel in 1× TBE (89 mM Tris, 89 mM boric acid, 2 mM disodium EDTA). The B1 site oligonucleotide includes nucleotides −271 to −252 in the TFIIIA promoter. The sequence is as follows:

```
5'-dGATCCATCACGTGCTCCACTAGGACG-3'
      GTAGTGCACGAGGTGATCCTGCCTAG
```

The sequences of the USF and H2B oligonucleotides are given by Scotto et al. (35). The batch of poly(dI-dC) which could completely suppress the nonspecific binding activity from oocyte extract was from Pharmacia (lot AA7880107). When rabbit polyclonal antibodies were used in a gel shift assay, extract and the antiserum or a naive serum control were preincubated for 10 min prior to addition of the labeled oligonucleotide.

Quantitative gel shift measurements were used to derive estimates for the number of molecules of a *trans*-acting factor per oocyte or per cell. The equation used for these calculations was taken from Emerson et al. (15) and Calzone et al. (6).

Serum preparation. Preparation of the various rabbit USF antisera used in this study will be described in detail elsewhere. Briefly, DNA sequences encoding protein domains (e.g., h18–105 indicates that the hUSF cDNA sequence coding for amino acid residues 18 through 105 was used) were subcloned in phase in pATH expression vectors (13). The full-size hUSF cDNA was subcloned in pET-3D, a T7 RNA polymerase-driven expression vector (39). To facilitate cloning of the full-length hUSF cDNA, an *Nco*I site was introduced by using the polymerase chain reaction (PCR). The new sequence, verified by DNA sequencing, is 5'-dACCATGGTGAAG, thus introducing a valine residue between the first and second amino acid residue of the wild-type hUSF protein. The bacterially expressed proteins were then purified by electrophoresis on sodium dodecyl sulfate (SDS)-polyacrylamide gels and injected into rabbits. Sera obtained after the third set of injections were used in the experiment described.

Screening of cDNA libraries. An *X. laevis* cDNA library cloned in λgt10 and prepared from total ovary poly(A)$^+$ mRNA was a generous gift from L. Etkin (Anderson Cancer Center, Houston, Tex.). A *Xenopus borealis* library constructed from total ovary poly(A)$^+$ mRNA in λZAP (Stratagene) was a generous gift of D. D. Brown (Carnegie Institution, Baltimore, Md.). The *X. laevis* cDNA library was screened with a 417-bp *Avr*II-*Apa*I subfragment derived from the hUSF cDNA spanning the DNA binding domain (*Avr*II site at nucleotide 658; *Apa*I site at nucleotide 1075) (19a). The *X. lorealis* library was screened with a 78-bp fragment spanning amino acid residues 43 to 71 from the *X. laevis* cDNA.

Hybridizations were performed at 42°C in a solution containing 20% formamide, 5× Denhardt solution, 0.2% SDS, 50 mM sodium P_i (pH 7.7), 900 mM NaCl, 5 mM EDTA, and 100 μg of denatured salmon sperm DNA per ml. Filters were washed two times at room temperature and then twice at 60 to 65°C in a solution of 1× SSC (0.15 M NaCl plus 0.015 M sodium citrate)–0.1% SDS. Positive plaques were further purified, and the cDNA inserts were analyzed by restriction enzyme digestion and sequencing.

DNA sequencing. The cDNA inserts were subcloned in the vector pBluescript II SK- (Stratagene). Single-stranded DNA was prepared and sequenced by using synthetic primers in a reaction employing the chain termination method and Sequenase reagents (U.S. Biochemical Corp.)

Sequence analysis was performed on an IBM personal computer, using software obtained from DNAstar, Inc. (Madison, Wis.).

mRNA isolation and Northern (RNA) blotting. Poly(A)$^+$ mRNA was isolated from 2 g of mature *X. laevis* ovary tissue and 2 g of liver tissue, using the mRNA isolation kit Fast Track (Invitrogen) according to the supplier's specifications.

For RNA blot analysis, one cycle poly(A)$^+$-selected mRNA was separated on 1.5% agarose–2.2 M formaldehyde gels and transferred to nitrocellulose (44). Blots were hybridized with ^{32}P-labeled DNA fragments under conditions identical to those described for screening of cDNA libraries. The 0.24- to 9.5-kb RNA ladder from BRL Life Technology was used as a size standard.

PCR amplification of the amino-terminal cDNA sequence of Xl-USF. One cycle of poly(A)$^+$-selected mRNA from *X. laevis* ovary was used for cDNA synthesis. The first-strand synthesis was driven by avian myeloblastosis virus reverse transcriptase (Boehringer Mannheim) in a reaction employ-

ing 10 pmol of a gene-specific primer with the sequence 5′-GGCATCATCGCTGGTGAATG-3′ from nucleotides 312 to 292 on the lower strand of the final *X. laevis* cDNA. Excess primer was removed by centrifugation in a Centricon 10 (Amicon Corp.). The second-strand synthesis and a subsequent PCR were carried out by using a second gene-specific primer spanning the sequence 5′-GGAATTCCA GAGAGATGAAGGGGCAACAAAAA-3′ derived from the *X. borealis* cDNA. This primer includes the initiating ATG, the sequence for the first five amino acids, and 7 bp from the 5′ leader which are shared between the hUSF and *X. borealis* USF (Xb-USF) cDNAs. The antisense primer in the PCR reaction was a third Xl-USF gene-specific primer spanning the sequence 5′-GGAATTCCCACAGCCTGGCT CATGGATT-3′ from nucleotides 282 to 262 on the lower strand of the final full-length DNA just upstream of the primer used to prime the first cDNA strand. Both primers used in the PCR reaction included the sequence for an *Eco*RI restriction site at their 5′ ends. The PCR reaction was carried out with a Gene Amp kit from Perkin Elmer. The sample was exposed to 40 cycles of denaturation (95°C, 40 s), annealing (55°C, 1 min), and polymerization (72°C, 3 min). Reaction products were analyzed by gel electrophoresis and subcloned as *Eco*RI fragments in pBluescript SKII- by using standard techniques.

In vitro translation using a rabbit reticulocyte lysate. RNA was generated in vitro by transcribing the Xl- and Xb-USF cDNAs with a T3 or T7 RNA polymerase-driven reaction, using reagents from an RNA capping kit supplied by Stratagene. The cap analog was not used in the reaction. The quality of the RNA was determined by gel electrophoresis. Samples (2 μg) of in vitro RNA were translated in a rabbit reticulocyte lysate (Promega) according to the supplier's specifications. To radiolabel the newly synthesized protein, L-[^{35}S]methionine (ICN) was included in the reaction mixture. To terminate the reaction, the samples were quickly frozen in liquid nitrogen after 20% gycerol were added.

SDS-polyacrylamide gel electrophoresis was performed as described by Laemmli (24).

Proteolytic clipping gel shift assay. Standard gel shift binding reactions were performed as described above. Protein and the labeled B1 site oligonucleotide were preincubated for 10 min at room temperature. Protease was freshly diluted in H_2O added at the indicated concentrations, and the incubation was continued for 10 more min, after which the samples were loaded on a 6% gel shift gel. Endoprotease ArgC was obtained from Boehringer Mannheim Biochemicals. Solutions were made according to the supplier's recommendation.

DNA techniques. All DNA manipulations were done by using standard techniques (26) and according to National Institutes of Health guidelines.

RESULTS

Identification of an oocyte-specific and a somatic cell-specific B1 protein-TFIIIA promoter complex. Electrophoretic mobility shift assays were used to determine whether the USF-related B1 protein implicated in TFIIIA gene transcription in oocytes is also present in somatic cells, in which TFIIIA gene expression is drastically reduced. By using a 26-bp oligonucleotide spanning the B1 protein binding site (B1 site oligonucleotide), one major complex (B1-s) was identified with extracts from the *X. laevis* embryo cell line X58 (Fig. 1A). Formation of the major somatic complex was inhibited by oligonucleotides containing either the B1 protein binding

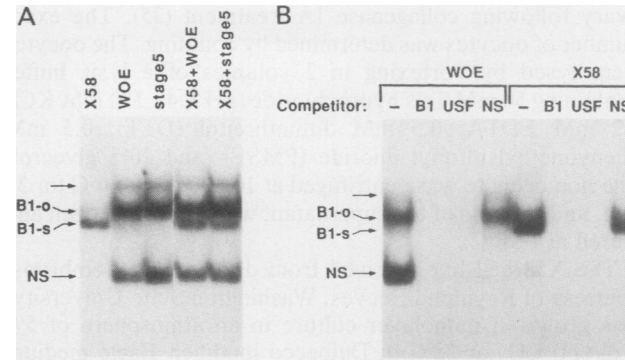


FIG. 1. Detection of oocyte and somatic cell-specific complexes binding to the B1 site in the TFIIIA promoter. (A) A ^{32}P-labeled oligonucleotide spanning from −271 to −252 was incubated with 17 μg of X58 somatic cell extract, 30 μg of whole ovary extract (WOE), 5 oocyte equivalents of stage 5 oocyte extract, or a mixture of two of these extracts. The complexes were resolved on a 4% native polyacrylamide gel and labeled B1-o (oocyte specific), B1-s (somatic cell specific), and NS (nonspecific). The single arrow indicates the labile oocyte-specific complex. The free probe is not shown. (B) Double-stranded oligonucleotides were used as competitors in a gel shift assay employing 30 μg of whole ovary extract (WOE) and 17 μg of X58 somatic cell extract. Competitors were 100 ng (200-fold molar excess) of the B1 site or USF oligonucleotide or a nonspecific H2B oligonucleotide (NS).

site or the USF binding site in the adenovirus major late promoter but not by a nonspecific oligonucleotide containing the binding site for a human H2B octamer-binding protein (17) Fig. 1B). The same picture was observed with other somatic cell extracts from the XTC embryo cell line or the SHA kidney cell line (data not shown).

In contrast, three complexes with different electrophoretic mobilities were resolved upon incubation of the B1 site oligonucleotide with whole ovary extract or with stage 5 oocyte extract (Fig. 1A). However, only two of these complexes (upper arrow and B1-o) were found to be specifically competed for by the B1 site oligonucleotide and not by a nonspecific oligonucleotide (Fig. 1B). Of these two specific complexes, the minor, more slowly migrating oocyte-specific complex (upper arrow in Fig. 1) was found to be labile and could not be resolved in every gel shift assay. The third oocyte-derived complex (NS in Fig. 1B) was found to be nonspecific and during subsequent experiments was found to be completely suppressed with certain batches of the nonspecific competitor poly(dI-dC). An oligonucleotide spanning the hUSF consensus binding site in the adenovirus major late promoter (32) was found to compete as well as the B1 site oligonucleotide for all of the specific complexes in oocyte and somatic cell extracts (Fig. 1B).

A mixing experiment in which an X58 somatic cell extract and either whole ovary extract or stage 5 oocyte extract were preincubated at room temperature before addition of the labeled B1 site probe demonstrated complete additivity (Fig. 1A). This observation excludes the possibility that these complexes are interconvertible and, specifically, that the faster-migrating somatic complex is simply derived from the slower-migrating oocyte complex by the action of a somatic cell-specific protease during extract preparation. Given other considerations of B1 site function during development, it seems more likely that this qualitative difference in complex formation between the oocyte and somatic cell extracts might reflect some more fundamental aspects of

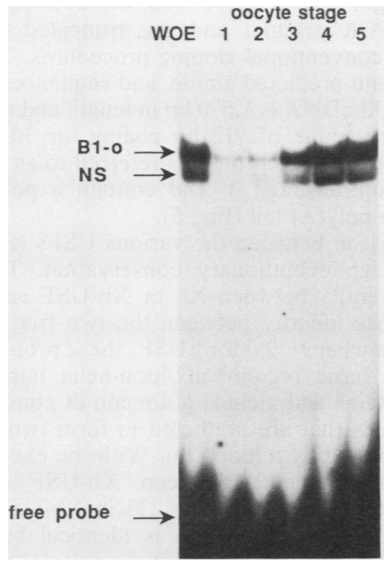


FIG. 2. B1 factor distribution during oogenesis. Staged oocyte extracts were prepared as described in Materials and Methods; 30 μg of whole ovary extract (WOE) or 5 oocyte equivalents of extracts from each stage were assayed, using the ^{32}P-labeled B1 site oligonucleotide as a probe. The complexes were resolved on a 6% polyacrylamide gel. B1-o indicates the oocyte-specific complex. The labile oocyte-specific complex marked with a single arrow in Fig. 1 is not well resolved. NS indicates the nonspecific complex. Competition data showing the specificity of complex B1-o are included in Fig. 3B.

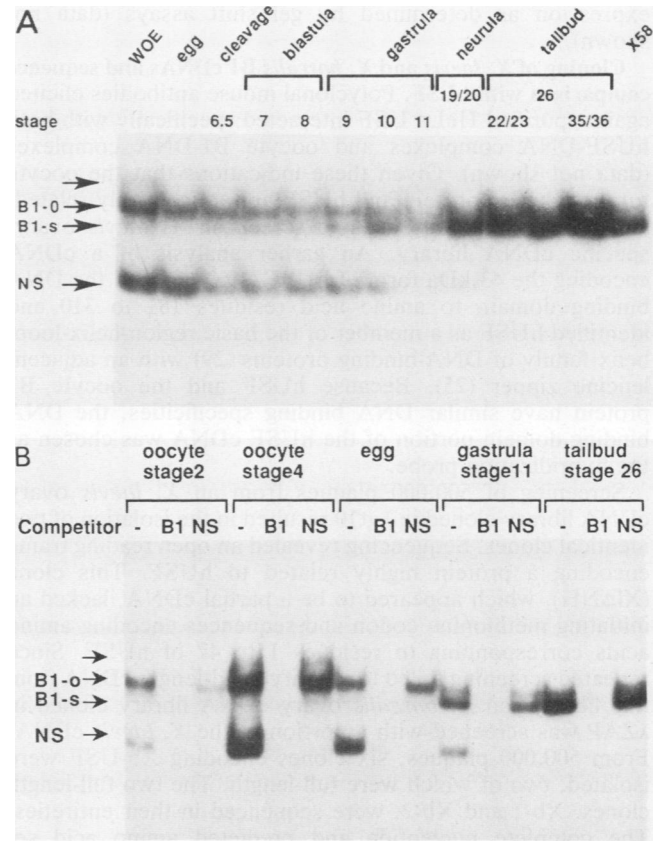


FIG. 3. B1 factor distribution during embryogenesis. (A) Embryonic stages were collected according to Nieuwkoop and Faber (31); the developmental stages are listed above each lane. Samples of 0.5 unfertilized egg or embryo equivalents were incubated with the B1 site oligonucleotide probe. The complexes were resolved on a 4% native polyacrylamide gel under extended running conditions. WOE, 30 μg of whole ovary extract; X58, 17 μg of X58 somatic cell extract. (B) Double-stranded oligonucleotides were used as competitors in a gel shift analysis. Either no competitor (−), 100 ng of B1 site oligonucleotide, or 100 ng of nonspecific H2B oligonucleotide (NS) (corresponding to a 200-fold molar excess) was used. Extracts from 5 oocyte stage 2 or stage 4 equivalents or 0.5 unfertilized egg or embryo equivalents were assayed. The complexes were separated on a 4% native polyacrylamide gel. The labile oocyte complex is resolved and indicated by an arrow.

TFIIIA gene regulation. Hence, we will refer to the proteins responsible for the two oocyte-specific complexes and the somatic cell-specific complex as the oocyte and somatic forms of the B1 protein.

Developmental profile of the oocyte and somatic forms of the B1 protein. To determine the developmental stage at which the capacity for forming the somatic B1 complex replaces the capacity for forming the oocyte B1 complex, we determined the factor distribution during oogenesis and embryogenesis. An analysis of 5 cellular equivalents of staged oocyte (14) extracts revealed that the factor responsible for the oocyte B1 complex accumulates from stage 1 to stage 5/6 during oogenesis (Fig. 2). However, since the volume and the total mass of protein per oocyte are increasing, the ratio of the oocyte B1 protein to the overall cellular protein stays roughly constant. An analysis of constant egg and embryo equivalents indicated that the factor responsible for the oocyte B1 complex persists through gastrula, while the factor responsible for the somatic B1 complex begins to accumulate at early cleavage and by the neurula stage is the dominant, if not the exclusive, form of the B1 protein present (Fig. 3A). The slowly migrating oocyte complex that was seen with total ovary extract and was found to result from a labile interaction (see above) was never observed with unfertilized egg or early embryo extracts. The oligonucleotide competition experiment of Fig. 3B demonstrates the specificity of the complexes. The B1 site oligonucleotide was used as a specific competitor to challenge the complex formation, while an oligonucleotide containing the H2B octamer site was used as a nonspecific competitor.

Quantitative gel shift measurements (6, 15) were used to estimate the minimum number of molecules responsible for the oocyte B1 complex in staged extracts (data not shown). We estimated that this number increased from 10^8 molecules for a stage 1 oocyte to 4×10^9 molecules for a stage 6 oocyte. This increase during oogenesis could reflect either accumulation for maternal use or the need for a constant concentration of the protein during the marked enlargement of the nucleus from stage 1 to stage 6. The somatic form of the B1 protein was estimated to increase from undetectable levels in early embryos to approximately 2×10^4 molecules per cell by the neurula stage. In X58 somatic cells, we estimated the minimum number of molecules for the somatic form of the B1 protein to be 10^4 molecules per cell.

It has been shown previously (18, 36, 47) that the accumulation of TFIIIA protein is developmentally regulated, with highest levels in early oocytes, significantly decreased levels in later-stage oocytes, and drastically reduced levels during embryogenesis and in somatic cells. Our oocyte and embryo extracts reflect this known pattern of TFIIIA

expression as determined by gel shift assays (data not shown).

Cloning of *X. laevis* and *X. borealis* B1 cDNAs and sequence comparison with hUSF. Polyclonal mouse antibodies elicited against purified HeLa USF interacted specifically with both hUSF-DNA complexes and oocyte B1-DNA complexes (data not shown). Given these indications that the oocyte form of the B1 protein and hUSF are antigenically related, an hUSF cDNA was used to screen an *X. laevis* ovary-specific cDNA library. An earlier analysis of a cDNA encoding the 43 kDa form of hUSF (19a) mapped the DNA binding domain to amino acid residues 181 to 310 and identified hUSF as a member of the basic region-helix-loop-helix family of DNA-binding proteins (29) with an adjacent leucine zipper (25). Because hUSF and the oocyte B1 protein have similar DNA binding specificities, the DNA binding domain portion of the hUSF cDNA was chosen as the hybridization probe.

Screening of 500,000 plaques from an *X. laevis* ovary cDNA library cloned in λgt10 resulted in the isolation of two identical clones. Sequencing revealed an open reading frame encoding a protein highly related to hUSF. This clone (XlΔNH), which appeared to be a partial cDNA, lacked an initiating methionine codon and sequences encoding amino acids corresponding to residues 1 to 42 of hUSF. Since repeated screening failed to identify a full-length cDNA from this library, an *X. borealis* ovary cDNA library cloned in λZAP was screened with a portion of the *X. laevis* cDNA. From 500,000 plaques, six clones encoding Xb-USF were isolated, two of which were full length. The two full-length clones, Xb-1 and Xb-2, were sequenced in their entireties. The complete nucleotide and predicted amino acid sequences are given in Fig. 4. The Xb-1 cDNA insert is 2,157 bp in length, while the Xb-2 cDNA insert contains 1,727 bp. These two cDNAs have 5′ leader sequences of different lengths (122 bp for Xb-1 and 73 bp for Xb-2) which are identical in the region of overlap, an identical open reading frame of 924 bp encoding 307 amino acid residues, and 281 bp of identical 3′ nontranslated region before they diverge (Fig. 4). The protein encoded by the cDNAs, Xb-USF, is highly related to hUSF. Only Xb-2 has a polyadenylation signal (AATAAA; underlined in Fig. 4) followed by a poly(A) tail. Interestingly, the longer 5′ leader sequence of Xb-1 includes a second ATG in frame upstream of the ATG at which the similarity to hUSF begins (Fig. 4). If this ATG were used, the protein would be 32 amino acid residues longer at its amino terminus.

By using sequence information obtained from the *X. borealis* cDNAs, the amino terminal portion of an Xl-USF cDNA was cloned by using the PCR technique and oocyte poly(A)$^+$ mRNA. A PCR product of the expected size was obtained and subcloned. Sequencing of several isolates from separate amplification reactions showed that the sequence is very similar but not identical to the corresponding sequence in the Xb-USF and hUSF cDNA clones, and we concluded that it encodes the amino terminus of the Xl-USF cDNA (for details regarding primer sequences, see Materials and Methods).

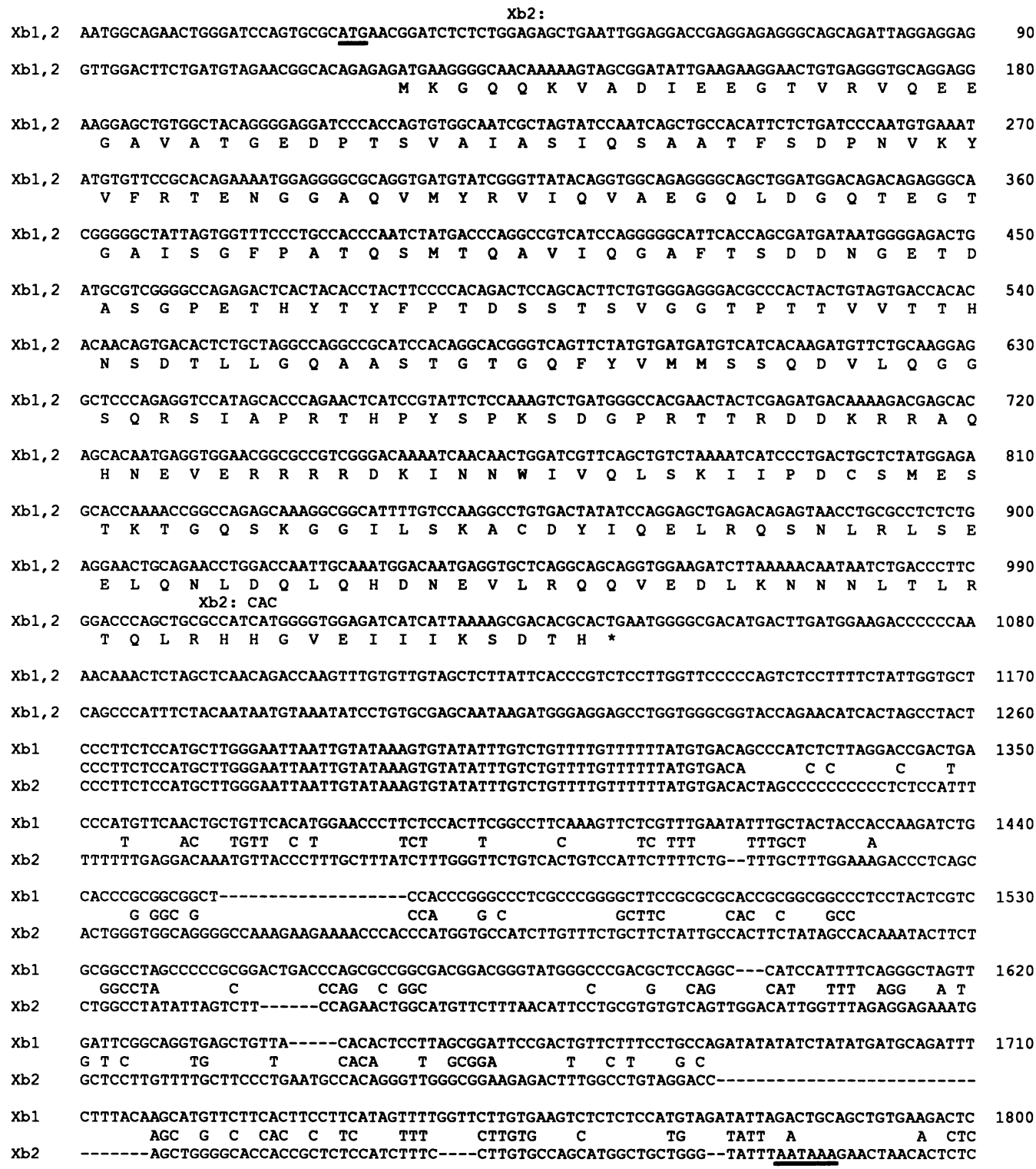
A full-length Xl-USF cDNA (Xl) was constructed from the *X. laevis* PCR product and the truncated cDNA clone XlΔNH by conventional cloning procedures. The complete nucleotide and predicted amino acid sequences are given in Fig. 5. The Xl cDNA is 1,570 bp in length and establishes an open reading frame of 912 bp coding for 303 amino acid residues; the encoded protein is referred to as Xl-USF. The 659 bp of untranslated 3′ end contain a polyadenylation signal and a poly(A) tail (Fig. 5).

A comparison between the various USFs (Fig. 6) clearly reveals a high evolutionary conservation. There is 80% sequence identity between Xl- or Xb-USF and hUSF and 90% sequence identity between the two frog proteins. As reported elsewhere (19a) for hUSF, these proteins are members of the basic region-helix-loop-helix family of DNA-binding proteins and include a domain of conserved hydrophilic residues that are predicted to form two amphipathic helices separated by a loop (29). With the exception of one conservative mismatch between Xb-USF and Xl-USF (serine to threonine at residue 233 in Xb-USF), the basic region-helix-loop-helix domain is identical between these three USF proteins (Fig. 6). All three proteins also contain a conserved leucine repeat motif close to the carboxy terminus (Fig. 6). The first six amino acids (MKGQQK) of the *X. laevis* protein are dictated by the *X. borealis*-derived PCR primer sequence. In addition, the following 110 amino-terminal amino acid residues are also very highly conserved. Interestingly, the first 30 amino acids of Xl-USF, Xb-USF, and hUSF can form an amphipathic alpha helix which might serve as a transcriptional activation domain. Amino acid residues 120 to 160 (Fig. 6) are divergent between all USF proteins, indicating little evolutionary pressure on this part of the protein.

Xl-USF and Xb-USF expressed in vitro show sequence-specific DNA binding. The Xl, Xb-1, and Xb-2 cDNA clones were transcribed with T3 or T7 RNA polymerase, and the resulting RNAs were translated in a rabbit reticulocyte lysate. Each RNA generated a polypeptide which migrated on a 15% SDS-polyacrylamide gel as a 42-kDa species (Fig. 7) even though the open reading frames had predicted sizes of 33 kDa. This aberrant electrophoretic behavior is also seen for hUSF in that a 310-amino-acid protein (molecular weight, 33,517) has an apparent molecular mass of 43 kDa (19a). Translation of in vitro Xb-1 RNA with its larger 5′ leader sequence yielded a second, less abundant polypeptide with an approximate molecular mass of 45 kDa (indicated with an arrow in Fig. 7). This polypeptide may have resulted from initiation from the upstream ATG in frame with the USF open reading frame.

Gel shift analysis with the B1 site probe (Fig. 8) revealed that the Xl, Xb-1, and Xb-2 polypeptides each formed a major complex that ran close to but slightly above the oocyte form of the B1 protein. Control lysates programmed with brome mosaic virus RNA generated only a minor signal which presumably resulted from endogenous rabbit USF. These complexes were all specific, since they were inhibited by specific (B1 site and major late promoter USF site) oligonucleotides but not by the nonspecific H2B oligonucle-

FIG. 4. Nucleotide and predicted amino acid sequences of two *X. borealis* cDNA clones, Xb-1 and Xb-2. Xb-1 has a 122-bp 5′ leader; Xb-2 has 73 bp of identical 5′ leader sequence. Up to nucleotide 1327 in respect to the Xb-1 sequence, the two clones are identical. In the 3′ untranslated region, the two clones diverge completely. The polyadenylation signal AATAAA found in Xb-2 is underlined. A single nucleotide difference (a C instead of a T in Xb-2) between Xb-1 and Xb-2 over the length of the first 1,327 bp is given. This difference does not lead to an amino acid change. The second ATG in frame with the USF open reading frame in Xb-1 is indicated. Nucleotides are numbered on the right. The flanking *Eco*RI sites on each end of the cDNA inserts are not shown.


Xb2:
Xb1,2 AATGGCAGAACTGGGATCCAGTGCGCATGAACGGATCTCTCTGGAGAGCTGAATTGGAGGACCGAGGAGAGGGCAGCAGATTAGGAGGAG 90
Xb1,2 GTTGGACTTCTGATGTAGAACGGCACAGAGAGATGAAGGGGCAACAAAAAGTAGCGGATATTGAAGAAGGAACTGTGAGGGTGCAGGAGG 180
M K G Q Q K V A D I E E G T V R V Q E E
Xb1,2 AAGGAGCTGTGGCTACAGGGGAGGATCCCACCAGTGTGGCAATCGCTAGTATCCAATCAGCTGCCACATTCTCTGATCCCAATGTGAAAT 270
G A V A T G E D P T S V A I A S I Q S A A T F S D P N V K Y
Xb1,2 ATGTGTTCCGCACAGAAAATGGAGGGGCGCAGGTGATGTATCGGGTTATACAGGTGGCAGAGGGGCAGCTGGATGGACAGACAGAGGGCA 360
V F R T E N G G A Q V M Y R V I Q V A E G Q L D G Q T E G T
Xb1,2 CGGGGGCTATTAGTGGTTTCCCTGCCACCCAATCTATGACCCAGGCCGTCATCCAGGGGGCATTCACCAGCGATGATAATGGGGAGACTG 450
G A I S G F P A T Q S M T Q A V I Q G A F T S D D N G E T D
Xb1,2 ATGCGTCGGGGCCAGAGACTCACTACACCTACTTCCCCACAGACTCCAGCACTTCTGTGGGAGGGACGCCCACTACTGTAGTGACCACAC 540
A S G P E T H Y T Y F P T D S S T S V G G T P T T V V T T H
Xb1,2 ACAACAGTGACACTCTGCTAGGCCAGGCCGCATCCACAGGCACGGGTCAGTTCTATGTGATGATGTCATCACAAGATGTTCTGCAAGGAG 630
N S D T L L G Q A A S T G T G Q F Y V M M S S Q D V L Q G G
Xb1,2 GCTCCCAGAGGTCCATAGCACCCAGAACTCATCCGTATTCTCCAAAGTCTGATGGGCCACGAACTACTCGAGATGACAAAAGACGAGCAC 720
S Q R S I A P R T H P Y S P K S D G P R T T R D D K R R A Q
Xb1,2 AGCACAATGAGGTGGAACGGCGCCGTCGGGACAAAATCAACAACTGGATCGTTCAGCTGTCTAAAATCATCCCTGACTGCTCTATGGAGA 810
H N E V E R R R R D K I N N W I V Q L S K I I P D C S M E S
Xb1,2 GCACCAAAACCGGCCAGAGCAAAGGCGGCATTTTGTCCAAGGCCTGTGACTATATCCAGGAGCTGAGACAGAGTAACCTGCGCCTCTCTG 900
T K T G Q S K G G I L S K A C D Y I Q E L R Q S N L R L S E
Xb1,2 AGGAACTGCAGAACCTGGACCAATTGCAAATGGACAATGAGGTGCTCAGGCAGCAGGTGGAAGATCTTAAAAACAATAATCTGACCCTTC 990
E L Q N L D Q L Q H D N E V L R Q Q V E D L K N N N L T L R
Xb2: CAC
Xb1,2 GGACCCAGCTGCGCCATCATGGGGTGGAGATCATCATTAAAAGCGACACGCACTGAATGGGGCGACATGACTTGATGGAAGACCCCCCAA 1080
T Q L R H H G V E I I I K S D T H *
Xb1,2 AACAAACTCTAGCTCAACAGACCAAGTTTGTGTTGTAGCTCTTATTCACCCGTCTCCTTGGTTCCCCCAGTCTCCTTTTCTATTGGTGCT 1170
Xb1,2 CAGCCCATTTCTACAATAATGTAAATATCCTGTGCGAGCAATAAGATGGGAGGAGCCTGGTGGGCGGTACCAGAACATCACTAGCCTACT 1260
Xb1 CCCTTCTCCATGCTTGGGAATTAATTGTATAAAGTGTATATTTGTCTGTTTTGTTTTTTATGTGACAGCCCATCTCTTAGGACCGACTGA 1350
CCCTTCTCCATGCTTGGGAATTAATTGTATAAAGTGTATATTTGTCTGTTTTGTTTTTTATGTGACA C C C T
Xb2 CCCTTCTCCATGCTTGGGAATTAATTGTATAAAGTGTATATTTGTCTGTTTTGTTTTTTATGTGACACTAGCCCCCCCCCCTCTCCATTT
Xb1 CCCATGTTCAACTGCTGTTCACATGGAACCCTTCTCCACTTCGGCCTTCAAAGTTCTCGTTTGAATATTTGCTACTACCACCAAGATCTG 1440
T AC TGTT C T TCT T C TC TTT TTTGCT A
Xb2 TTTTTTGAGGACAAATGTTACCCTTTGCTTTATCTTTGGGTTCTGTCACTGTCCATTCTTTTCTG--TTTGCTTTGGAAAGACCCTCAGC
Xb1 CACCCGCGGCGGCT--------------------CCACCCGGGCCCTCGCCCGGGGCTTCCGCGCGCACCGCGGCGGCCCTCCTACTCGTC 1530
G GGC G CCA G C GCTTC CAC C GCC
Xb2 ACTGGGTGGCAGGGGCCAAAGAAGAAAACCCACCCATGGTGCCATCTTGTTTCTGCTTCTATTGCCACTTCTATAGCCACAAATACTTCT
Xb1 GCGGCCTAGCCCCCGCGGACTGACCCAGCGCCGGCGACGGACGGGTATGGGCCCGACGCTCCAGGC---CATCCATTTTCAGGGCTAGTT 1620
GGCCTA C CCAG C GGC C G CAG CAT TTT AGG A T
Xb2 CTGGCCTATATTAGTCTT------CCAGAACTGGCATGTTCTTTAACATTCCTGCGTGTGTCAGTTGGACATTGGTTTAGAGGAGAAATG
Xb1 GATTCGGCAGGTGAGCTGTTA-----CACACTCCTTAGCGGATTCCGACTGTTCTTTCCTGCCAGATATATATCTATATGATGCAGATTT 1710
G T C TG T CACA T GCGGA T C T G C
Xb2 GCTCCTTGTTTTGCTTCCCTGAATGCCACAGGGTTGGGCGGAAGAGACTTTGGCCTGTAGGACC--------------------------
Xb1 CTTTACAAGCATGTTCTTCACTTCCTTCATAGTTTTGGTTCTTGTGAAGTCTCTCTCCATGTAGATATTAGACTGCAGCTGTGAAGACTC 1800
AGC G C CAC C TC TTT CTTGTG C TG TATT A A CTC
Xb2 -------AGCTGGGGCACCACCGCTCTCCATCTTTC----CTTGTGCCAGCATGGCTGCTGGG--TATTTAATAAAGAACTAACACTCTC
Xb1 TCCAAACTGTTCTAACAAGTACTTGTCTGGAAAAAAATATCCTTTCCTACAGAACCAAAGTACTATATATATTTTTTTCCAGACAAGTAC 1890
Xb2 AAAAAAAAAAAAAAAAAAAAAAA
Xb1 TTGTTAGAACAGTTTGGGCAGGCGAAGCCAGAGGAAACTCTGGTGGAGGTCCGTAGCGGTCCTGACGTGCAAATCGGTCGTCCGACCTGG 1980
Xb1 GTATAGGGGCGAAAGACTAATCGAACCATCTAGTAGCTGGTTCCCTCCGAAGTTTCCCTCAGGATACGTGGCGCTCGTCCGTCGCAGTTT 2070
Xb1 TATCCGGTAAAGCGAATGATTAGAGGTCTTGGGGCCGAAACGATCTCAACCTATTCTCAAACTTTAAATGGGTAAGAAGCCCGGCTCGCT 2160
Xb1 GGCTTGGAGCCGGGCGTGGAATGC 2184

```
   1 ATGAAGGGGCAACAAAAAGTGGCGGATATTGAAGAAGGAAGTGTGCGGGTGCAGGAGGGA
     M  K  G  Q  Q  K  V  A  D  I  E  E  G  S  V  R  V  Q  E  G    20

  61 GCTGTAGCAACAGGGGAAGATCCCACCAGTGTAGCAATCGCCAGTATCCAATCAGCTGCC
     A  V  A  T  G  E  D  P  T  S  V  A  I  A  S  I  Q  S  A  A    40

 121 ACATTCTCTGATCCCAATGTGAAATACGTGTTCCGCACGGAGAATGGGGGGACACAGGTG
     T  F  S  D  P  N  V  K  Y  V  F  R  T  E  N  G  G  T  Q  V    60

 181 ATGTATCGAGTTATACAGCTGGCAGAGGGGCAGCTGGATGGGCAGACAGAGGGCACCGGG
     M  Y  R  V  I  Q  L  A  E  G  Q  L  D  G  Q  T  E  G  T  G    80

 241 GCTATTAGTGGCTTCCCGGCCACCCAATCCATGAGCCAGGCTGTGATCCAGGGGGCATTC
     A  I  S  G  F  P  A  T  Q  S  M  S  Q  A  V  I  Q  G  A  F   100

 301 ACCAGCGATGATGCCGGAGAGACTGATGCGTCGGGGCCAGAGACTCACTACACCTACTTT
     T  S  D  D  A  G  E  T  D  A  S  G  P  E  T  H  Y  T  Y  F   120

 361 CCCACCACAGTCACAGACTCCAGCACTTCCACTGTAGTGACCACCCACACCACTGACACT
     P  T  T  V  T  D  S  S  T  S  T  V  V  T  T  H  T  T  D  T   140

 421 CTCATAGGACAGGCCGGATCCACGGCCCCGGGTCAGTTCTATGTGATGATGTCACCACAA
     L  I  G  Q  A  G  S  T  A  P  G  Q  F  Y  V  M  M  S  P  Q   160

 481 GATGTTCTACAGGGAGGGTCCCAGAGGTCCATAGCACCCAGAACTCATCCATATTCTCCA
     D  V  L  Q  G  G  S  Q  R  S  I  A  P  R  T  H  P  Y  S  P   180

 541 AAGTCCGATGGCCCACGAACTACTCGAGATGACAAAAGACGAGCGCAGCACAATGAGGTG
     K  S  D  G  P  R  T  T  R  D  D  K  R  R  A  Q  H  N  E  V   200

 601 GAACGGCGCCGTCGGGACAAAATCAACAACTGGATCGTTCAGCTGTCTAAAATCATCCCT
     E  R  R  R  R  D  K  I  N  N  W  I  V  Q  L  S  K  I  I  P   220

 661 GACTGCTCTATGGAGAGCACCAAATCCGGGCAGAGCAAAGGCGGCATTTTGTCCAAGGCC
     D  C  S  M  E  S  T  K  S  G  Q  S  K  G  G  I  L  S  K  A   240

 721 TGCGACTATATCCAGGAATTGCGACAGAGCAACCTGCGCCTCTCGGAGGAACTGCAGAAT
     C  D  Y  I  Q  E  L  R  Q  S  N  L  R  L  S  E  E  L  Q  N   260

 781 GTGGACCAATTGCAAATGGACAATGAGTTGCTCAGGCAGCAGGTGGAAGATCTTAAAAAC
     V  D  Q  L  Q  M  D  N  E  L  L  R  Q  Q  V  E  D  L  K  N   280

 841 AAAAATCTGATCCTTCGGACCCAGCTGCGCCATCATGGGGTGGAGATCATCATTAAAAGC
     K  N  L  I  L  R  T  Q  L  R  H  H  G  V  E  I  I  I  K  S   300

 901 GATGGGCGATGAATGGAGCGATGTGAGCCGACGGGAGACCCCAAAACTAACTCTTGCCCA
     D  G  R  *                                                   303

 961 ATTGAACGTCTCCTTGGTTCCCCCAGCTTCCTTTTCTATTGGTGCTTTGCCCATTTCTAC
1021 AATAATGTAAATATTGTGTGTGAGCAATAAGATGGGAGGAGCCTGGTGTTTTGTGGGCGG
1081 TGCAGAACATCACTATCCTTCTCCATGCTTGGGAATGGGTTGTATAATGTGTATATTTGT
1141 CTGTTTTCTATGTCACACTTGCCCCCACCCCCCCCTTTTTTTTTTTGAGGACAAATGTTA
1201 TCCTTCGCTTTATGTTCGGGTTCTGTCCATACTTTTCTGTTTGCTTTGGAAAGACCCTCA
1261 GGGGCCAAAGAAGAAAATCCACCCACGGTGCCATCTACTTTCTGTTGCCAGTTCCATAGC
1321 CACACTGGGACCCTAATCAAATGGACTATTTTAGTCTTCCACAACTTCATTGGCATTCCT
1381 GCAAGTGGCAGTTGTACATTGGGCACCTTGTCTTGCTTCCCTGAAGGCCAGAGTGTTGCC
1441 CAGATGAGACTTGTAGGACCAGCTGGGGCCCCACCACTCTCCTCTCTCCTTGTGCCAGCG
1501 TGGCTGCTGGGTATTTAATAAACAAGTAATACTCTCAAAAAAAAAAAAAAAAAAAAAAAA
1560 AAAAAAAAAA
```

FIG. 5. Nucleotide and predicted amino acid sequences of the full-length Xl cDNA. The sequence encoding the first 6 amino acids MKGQQK is dictated by the primer sequence used to clone the amino-terminal 42 amino acid residues (Materials and Methods). The polyadenylation signal AATAAA preceding the poly(A) tail is underlined. Nucleotides are numbered on the left; amino acids, represented in one-letter code, are numbered on the right. The *Eco*RI linkers at each end of the cDNA sequence are not shown.

otide. We attribute the minor faster-migrating complexes (Fig. 8) to proteolysis products.

RNA blots reveal only one USF mRNA species in oocytes and somatic cells. The data presented above suggest that the *X. laevis* oocyte-derived USF cDNA encodes the oocyte B1 protein or a closely related protein but provides no explanation regarding the molecular difference between the oocyte and somatic B1 proteins. To address this question, *X. laevis* oocyte and somatic (liver) poly(A)$^+$ mRNAs were analyzed by hybridization to different subfragments of the Xl cDNA clone. The first probe tested was a 213-bp DNA fragment spanning the DNA binding domain (the basic region and the helix-loop-helix region) from amino acid residues 189 to 258, on the assumption that this region might be conserved among mRNAs encoding proteins having closely related DNA binding specificities. We observed an mRNA species of 1.7 kb, in close agreement with the size of the Xl clone (Fig. 9, lanes 3 and 4). High-resolution gel electrophoresis (data not shown) failed to detect any difference in size between the oocyte and somatic Xl-USF messages. Given that comparable amounts of poly(A)$^+$ mRNA were analyzed (Fig. 9, lanes 1 and 2), the relative hybridization signals indicate a higher abundance of Xl mRNA in oocytes than in somatic cells.

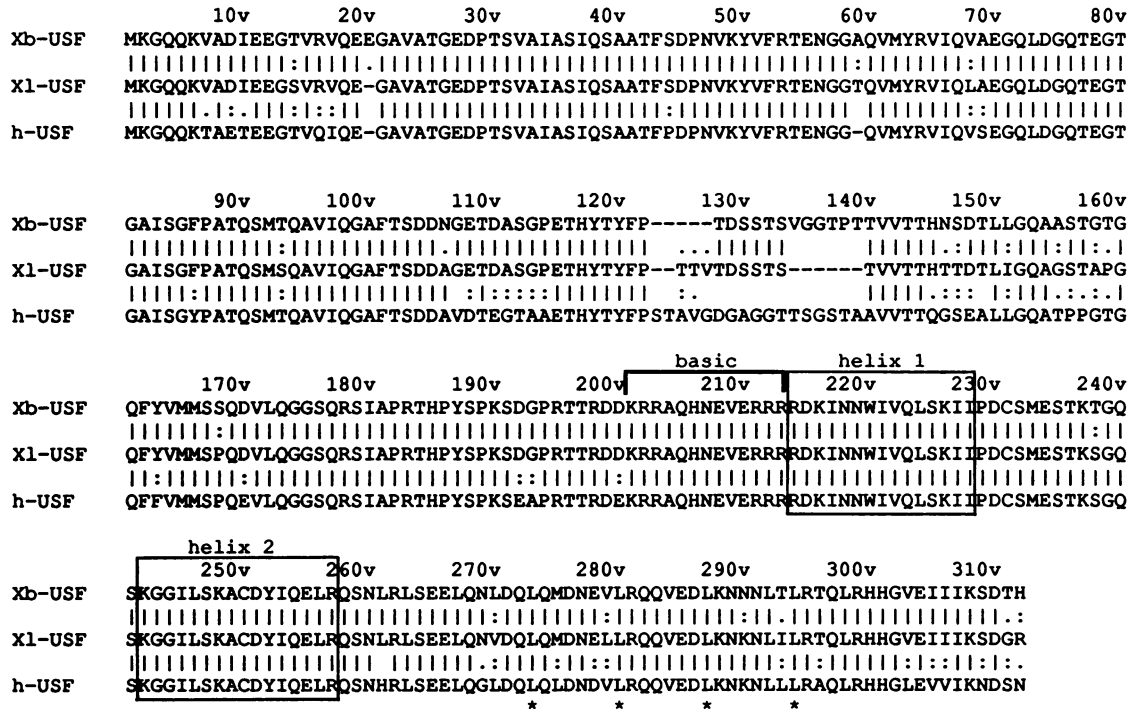


FIG. 6. Amino acid sequence comparison of Xl-USF with Xb-USF and hUSF. Sequences were analyzed with the GAP program, using software from DNAstar, Inc. Vertical bars signify amino acid identities; colons and periods represent conservative amino acid replacements. The basic region, the helix 1-loop-helix 2 region, and the leucine repeat domain are indicated. Amino acid identity is 80.1% between hUSF and Xb-USF, 80.3% between hUSF and Xl-USF, and 90.2% between Xb-USF and Xl-USF.

Reprobing the same RNA blot with cDNA subfragments corresponding to three different parts of the Xl-USF amino terminus gave hybridization patterns that were identical with respect to signal intensities and sizes of the hybridizing mRNA species (data not shown). The Xl cDNA fragments used spanned sequences encoding amino acid residues 43 to 71, 127 to 150, and 43 to 188 of Xl-USF. It is interesting that the fragment spanning amino acids 127 to 150 contains the most divergent portion of the USF proteins and is thus fairly specific for the *X. laevis* gene. These hybridization patterns demonstrate a similarity between the oocyte and liver mRNAs that encompasses a large portion of the coding region of Xl-USF and strongly suggest that the liver and oocyte mRNAs encode the same protein. Altogether, these data point to identical or very similar primary structures for the oocyte and somatic forms of the B1 protein. Thus, it seems highly probable that the oocyte and somatic B1 proteins are derived not from different genes but rather from minor splicing variations or posttranslational modifications.

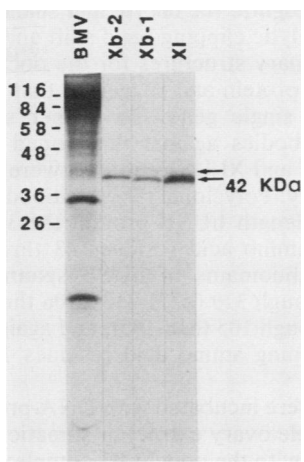


FIG. 7. In vitro-synthesized Xl, Xb-1, and Xb-2 polypeptides. Proteins were synthesized in rabbit reticulocyte lysates as described in Materials and Methods. Translation of brome mosaic virus (BMV) RNA yields four viral proteins of 20, 35, 97, and 110 kDa in size. A 1-μl sample from each 50-μl translation reaction was electrophoresed on a 15% SDS-polyacrylamide gel, dried, and exposed. Size standards (shown on the left): 26.6 kDa, triosephosphate isomerase; 36.5 kDa, lactic dehydrogenase; 48.5 kDa, fumarase; 58 kDa, pyruvate kinase; 84 kDa, fructose-6-phosphate kinase; 116 kDa, β-galactosidase.

A proteolytic clipping band shift assay reveals that the oocyte and somatic forms of the B1 protein are most likely modified or altered versions of the same protein. Protease treatment of DNA-protein complexes followed by electrophoretic analysis provides a convenient method to monitor differences in DNA-binding proteins without purification (34). This method was used to compare the Xl and Xb-2 cDNA-encoded polypeptides and the oocyte and somatic forms of the B1 protein. Each expressed protein or extract was incubated with the ^{32}P-labeled B1 site oligonucleotide under conditions in which the nonspecific binding activity initially observed with oocyte extracts was completely suppressed by a selected batch of poly(dI-dC). After treatment of preformed protein-DNA complexes with varying concentrations of endoprotease ArgC (Fig. 10), electrophoresis revealed several novel, faster-migrating complexes. These represent truncated forms of the proteins bound to DNA, with the fastest-migrating complexes representing limit di-

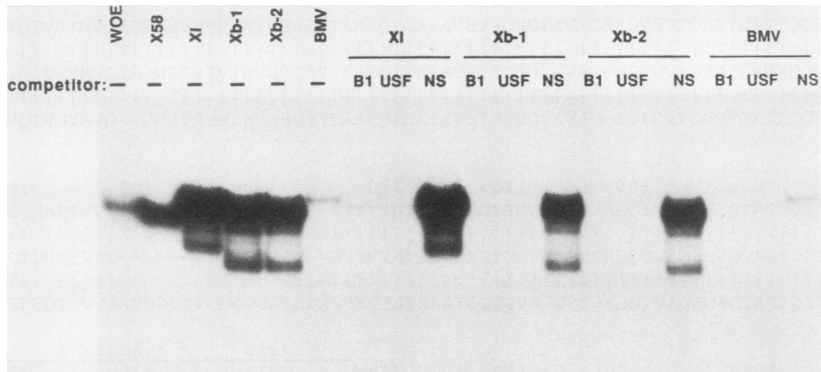


FIG. 8. Demonstration that in vitro-synthesized Xl, Xb-1, and Xb-2 polypeptides bind the B1 site oligonucleotide specifically. A 1-μl sample of a 50-μl in vitro translation reaction was used in a gel shift analysis and reacted with ^{32}P-labeled B1 site oligonucleotide. Standard gel shift reactions using 30 μg of whole ovary extract (WOE) and 17 μg of X58 somatic cell extract are included for reference. Samples (100 ng) of double-stranded B1, USF, or H2B (NS) oligonucleotides (corresponding to a 200-fold molar excess) were included as indicated. The complexes were separated on a 6% native polyacrylamide gel. BMV, Brome mosaic virus.

gests that presumably contain mainly the DNA binding domain portion of the protein.

Notably, the endoprotease ArgC-derived limit digest complexes obtained with the Xb-2 and Xl proteins showed similar but distinct mobilities (Fig. 10). An amino acid sequence comparison of the Xl and Xb-2 proteins revealed 90% sequence identity overall and a conservation of the number and position of arginines, which are the target sites for this particular protease. However, at pH 8 these proteins have a significant intrinsic charge difference (−6.54 for Xl-USF versus −9.51 for Xb-2-USF) which resides mainly in the leucine zipper region of the DNA binding domain. This charge difference can readily explain the different absolute migration behavior of the protease-generated complexes, with the more negatively charged Xb-2 moving more rapidly to the anode.

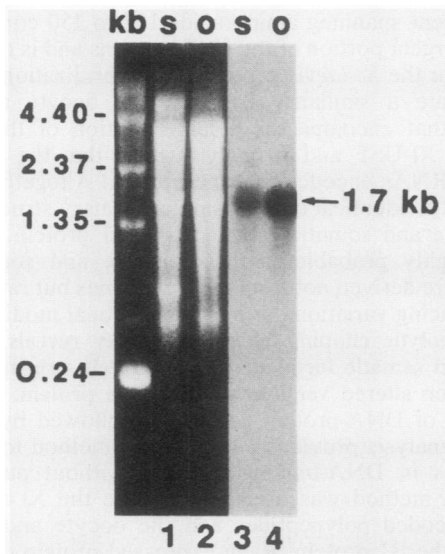


FIG. 9. RNA blot analysis. A ^{32}P-labeled 213-bp fragment spanning the basic region and the helix-loop-helix region of the Xl cDNA was used to probe a RNA blot containing 4 μg of poly(A)$^+$ ovary mRNA or 4 μg of poly(A)$^+$ somatic mRNA from *X. laevis* (lanes 3 and lane 4). The final washing steps were performed at 1× SSC, 0.1% SDS, and 65°C. The ethidium bromide-stained 1.5% agarose gel featuring both mRNAs (lane 1 and lane 2) is shown to demonstrate that equal amounts were loaded. The positions of RNA markers (lane kb) are indicated. o, Oocyte mRNA; s, somatic mRNA.

The oocyte stage 4 and X58 somatic cell extracts (Fig. 10) each revealed the same subset of complexes observed with the Xl and Xb-2 polypeptides, again with small but significant differences in absolute electrophoretic mobility. These could be due to charge differences, as proposed for the Xl and Xb-2 complexes, but could as well reflect other modifications or small differences in primary structure. The fact that the stage 4 complexes run faster than the complex derived from the Xl polypeptide suggests that the B1 protein from oocytes may be more negatively charged than the in vitro-synthesized Xl polypeptide.

Polyclonal rabbit antibodies against bacterially expressed USF proteins recognize the oocyte and somatic forms of B1. RNA and proteolytic clipping band shift analyses pointed to very similar primary structures for the oocyte and somatic forms of the B1 protein and suggested that the proteins are derived from a single gene. To further substantiate this conclusion, antibodies against the cloned and bacterially expressed hUSF and Xl-USF proteins were raised and used in a gel shift assay. Polyclonal rabbit antibodies were elicited against the full-length hUSF protein (h1-310), against Xl-USF spanning amino acid residues 43 through 303 (Xl43-303), against subdomains of hUSF spanning amino acid residues 271 through 310 (h271-310), 106 through 196 (h106-196), and 18 through 105 (h18-105), and against a subdomain of Xl-USF spanning amino acid residues 126 through 150 (Xl126-150).

The antisera were incubated with DNA-protein complexes formed with whole ovary extract or somatic cell extract. All antisera reacted with the oocyte B1 complex and resulted in the formation of specific immunoglobulin-oocyte B1 DNA complexes, whereas only four of six antisera reacted with the somatic B1 complex; incubation with naive serum did not lead to the formation of any immunoglobulin-B1 DNA complexes (Fig. 11). These data clearly show that the oocyte and somatic B1 proteins are highly antigenically related and that the relatedness encompasses a large portion of the protein. The data also show that the proteins are not identical. Altogether, these observations are in agreement

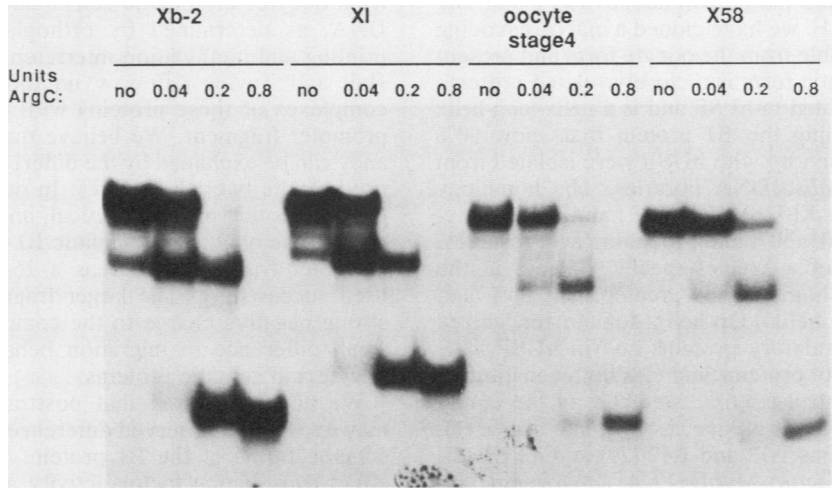


FIG. 10. Proteolytic clipping band shift assay using endoprotease ArgC. Xb-2 and Xl polypeptides (1 μl of a final translation reaction), stage 4 oocyte extracts (3 oocyte equivalents), and X58 somatic cell extract (17 μg) were incubated with ^{32}P-labeled B1 site oligonucleotide under standard gel shift conditions. The preformed protein-DNA complexes were subsequently reacted with increasing units of endoprotease ArgC as indicated to yield a limited proteolysis. The proteolysis-derived complexes were separated on a 6% native polyacrylamide gel.

with our conclusion that the difference between the oocyte and somatic B1 proteins may reside in posttranslational modification, differential mRNA splicing, or interaction with other factors.

DISCUSSION

TFIIIA expression in *X. laevis* is developmentally regulated, with the highest levels of expression in early oogenesis, decreased levels in later-stage oocytes and embryos, and expression of a distinct form in somatic cells (TFIIIA′) (18, 36, 42). The basis of this regulation is believed to be transcriptional, but this has not been proved directly. We have focused on proteins interacting with the B1 binding site, which is required for efficient TFIIIA transcription in oocytes (20, 35) but not in somatic cells (20). Here we describe studies of the oocyte and somatic forms of the B1 protein. Taking advantage of the apparent similarities be-

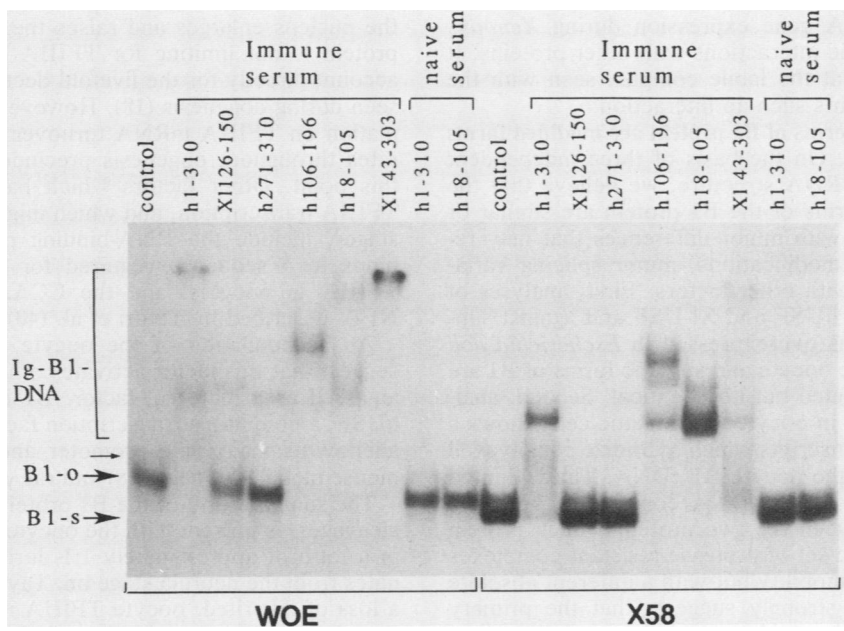


FIG. 11. Interaction of rabbit antibodies raised against bacterially expressed hUSF and Xl-USF with the oocyte and somatic forms of B1. In this assay, 30 μg of whole ovary extract (WOE) or 17 μg of X58 somatic cell extract was incubated with 2 μl of immune serum or naive serum prior to addition of the labeled B1 site oligonucleotide. The control lane contained no serum. Numbers above the lanes indicate the amino acid residues in the hUSF or Xl-USF protein against which the antibodies are directed. B1-o, Oocyte B1 complex; B1-s, somatic B1 complex; Ig, immunoglobulin. The complexes were resolved on a 4% polyacrylamide gel. The free probe is not shown.

tween the B1 protein and the human adenovirus major late transcription factor USF, we have cloned a cDNA encoding a protein indistinguishable from the oocyte form and present evidence that the somatic form is a closely related protein.

The B1 protein is related to hUSF and is a helix-loop-helix protein. cDNAs encoding the B1 protein that show 80% amino acid sequence identity with hUSF were isolated from *X. laevis* and *X. borealis* cDNA libraries. The homology extends throughout the Xl-USF, Xb-USF, and hUSF polypeptides, which have 303, 307, and 310 amino acid residues, respectively. Outside of a species-specific domain in the *Xenopus* protein, the homology is greater than 90% and approaches 100% in a helix-loop-helix domain recognized previously in other regulatory proteins and in hUSF. The helix-loop-helix family of proteins includes both constitutive and developmental control factors. Members of the family include the proto-oncogenes N-*myc*, L-*myc*, and c-*myc* (1), enhancer-binding proteins E12 and E47 (29) and TFE3 (2), and cell determination factors MyoD (12, 41), myogenin (48), the achaete-scute complex (46), daughterless (9), and twist (43). Murre et al. (29, 30) have demonstrated that certain pairs of helix-loop-helix proteins form heterodimers, and Davis et al. (11) showed that these interactions are mediated by the two amphipathic helices of the helix-loop-helix domain, whereas protein-DNA contacts require residues within the basic region. In addition, certain protein-protein interactions of hUSF also require an intact leucine repeat domain (19a).

The potential for heterodimer formation expands the possibilities for transcriptional regulation because it allows combinations of factors that recognize the same DNA sequence in different cell types. For example, E12-MyoD heterodimers induce myogenesis but can be dissociated by competition with a third developmentally regulated helix-loop-helix protein, Id, which thus represses myogenesis (3, 11). Therefore, although the oocyte B1 protein is present at a constant concentration throughout oogenesis while TFIIIA mRNA and protein are decreasing, the B1 protein, like E12, could modulate TFIIIA gene expression during *Xenopus* development by specific interactions with other proteins. It is possible, in fact, that the labile complex seen with the oocyte extract represents such an interaction.

Oocyte and somatic forms of B1 protein are modified forms of a single polypeptide. On the basis of three independent assays of protein and RNA structure, we believe that the oocyte and somatic forms of the B1 protein are similar or identical polypeptides with minor differences that may reflect posttranslational modifications, minor splicing variations, or interactions with other factors. First, analyses of antisera raised against hUSF and Xl-USF and against subdomains of the proteins overexpressed in *Escherichia coli* clearly showed that the oocyte and somatic forms of B1 are highly antigenically related but not identical. Second, analysis of the B1 mRNAs in oocyte and somatic cells shows a single population of transcripts which hybridize equally well to several portions of the oocyte B1 cDNA. Third, limited protease digestions of protein-DNA complexes showed a close relationship between the two proteins. Each protein showed a characteristic set of protease-resistant complexes with the same relative mobility but with a different absolute mobility. This finding strongly suggests that the primary structures are very similar or identical and that modifications might be responsible for the observed differences in absolute migration.

Hall and Taylor (20) have described a protein (TDEF) that binds to the B1 site in the TFIIIA promoter. TDEF isolated from oocytes or somatic cells made identical contacts with DNA, as determined by orthophenanthroline-Cu^{2+} footprinting and methylation interference assays. Furthermore, Hall and Taylor (20) saw no difference in the gel shift complexes of these proteins with use of an 81-bp TFIIIA promoter fragment. We believe that this apparent discrepancy can be explained by the different experimental methods used by the two laboratories. In our hands, the size of the oligonucleotide probe is critical, and we were never able to resolve the oocyte and somatic B1 forms by using a 126-bp promoter fragment, whereas a 26-bp oligonucleotide was used successfully. The larger fragments may contribute a strong negative charge to the complex which obscures the small difference in migration behavior seen between the oocyte and somatic proteins.

We have suggested that posttranslational modifications may explain the observed difference between the oocyte and somatic forms of the B1 protein. Modifications known to effect *trans*-acting factor activity are phosphorylation and addition of O-linked *N*-acetylglucosamine residues. Transcription factors affected by phosphorylation include CREB (19), heat shock factor (37), and TFIIIC (21). RNA polymerase II transcription factors SP1, AP1, and CTF require O-linked *N*-acetylglucosamine residues for full activity (22). Frog USF has a large number of serines and threonines which are potential phosphorylation sites for cellular kinases. Further experiments will examine phosphorylation and glycosylation states of the B1 protein.

Implication of the oocyte and somatic forms of the B1 protein in the regulation of TFIIIA gene expression. The oocyte and somatic forms of the B1 protein exhibit distinct gel shift complexes with the B1 site oligonucleotide. Using this assay, we have determined the developmental profile of these proteins to compare B1 and TFIIIA levels during development. We found that the oocyte form of B1 protein accumulates during oogenesis from 10^8 molecules at stage 1 to 4×10^9 molecules at stage 6. This reflects a constant concentration of oocyte B1 protein throughout oogenesis as the nucleus enlarges and raises the possibility that the B1 protein is not limiting for TFIIIA expression and cannot account directly for the fivefold decrease in TFIIIA mRNA seen during oogenesis (18). However, the absence of information on TFIIIA mRNA turnover and gene transcription rates throughout oogenesis precludes firm conclusions on this point. Other factors which have been implicated in TFIIIA transcription, and which might be limiting at certain stages, include the DNA-binding protein B2 (35), which binds to a sequence required for efficient expression of TFIIIA in oocytes, and the CCAAT box-binding factor NF-Y described by Tafuri et al. (40).

An accumulation of the oocyte B1 protein might also suggest that this factor activates other cellular genes and is amassed as a maternal factor. B1 is the frog homolog of hUSF, a ubiquitous transcription factor which activates the adenovirus major late promoter and the cell-type-specific mouse metallothionein I (8) and rat γ-fibronectin (10) genes.

The somatic form of the B1 protein appears with the first cleavages, is present with the oocyte form of the B1 protein in a ratio of approximately 1:1 during gastrula, and dominates from the neurula stage on. Taylor et al. (42) described a loss of inherited, oocyte TFIIIA mRNA during gastrulation, just at the onset of transcription in *X. laevis* embryos after the mid-blastula-stage transition. This loss and the appearance of a 40-kDa somatic TFIIIA protein (TFIIIA′) (36) correlates with the appearance of the somatic form of the B1 protein. Kim et al. (22a) have recently identified a

somatic TFIIIA transcript which initiates at least 227 bp upstream of the oocyte start site and encodes a somatic TFIIIA protein with 22 additional amino acids at the amino terminus. This finding clearly indicates differential TFIIIA promoter usage in oocytes versus somatic cells, and we suggest that the oocyte and somatic forms of the B1 protein may be involved in the choice between oocyte versus somatic TFIIIA promoters. Since there are no apparent differences in DNA binding properties between the oocyte and somatic forms of the B1 protein (this report; 20), differential promoter activation could be achieved by specific interactions between B1 and other factors which in turn activate the two promoters. The apparent modification of the B1 protein which accompanies the oocyte-somatic switch may regulate these factor interactions.

ACKNOWLEDGMENTS

We thank L. Etkin and D. D. Brown for providing the *X. laevis* and *X. borealis* cDNA libraries. We are indebted to E. Bartnik for assistance in preparation of the figures and for advice in many of the experiments. We are grateful to K. Scotto for advice in the early stages of the project and to the members of the Roeder laboratory for stimulating discussions.

This work was supported by Public Health Service grants CA-42567 and AI 27397 to R.G.R. and by general support from the Pew Charitable Trusts to the Rockefeller University. H.K. was supported by a postdoctoral fellowship from the German Research Foundation, and P.D.G. was supported by fellowship F32AI7696 from the National Institutes of Health and is a Lita A. Hazen fellow; P.P. was supported by a postdoctoral fellowship from the International Agency for Research on Cancer.

REFERENCES

1. **Alt, F. W., R. A. DePhino, K. Zimmerman, E. Legouy, K. Hutton, P. Ferrier, A. Tesfaye, G. D. Yancopoulous, and P. Nisen.** 1986. The human myc family. Cold Spring Harbor Symp. Quant. Biol. **51**:931–941.
2. **Beckmann, H., L. KuoSu, and T. Kadesch.** 1990. TFE3: a helix-loop-helix protein that activates transcription through the immunoglobulin enhancer μE3 motif. Genes Dev. **4**:167–179.
3. **Benezra, R., R. L. Davis, D. Lockshorn, D. L. Turner, and H. Weintraub.** 1990. The protein Id: a negative regulator of helix-loop-helix DNA binding proteins. Cell **61**:49–59.
4. **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.
5. **Bradford, M. M.** 1976. A rapid and sensitive method for quantification of microgram quantities of protein utilizing the principle of protein-dye binding. Anal. Biochem. **72**:248–254.
6. **Calzone, F. J., N. Theze, P. Thiebaud, R. L. Hill, R. J. Britten, and E. H. Davidson.** 1988. Developmental appearance of factors that bind specifically to cis-regulatory sequences of a gene expressed in the sea urchin embryo. Genes Dev. **2**:1074–1088.
7. **Carthew, R. W., L. A. Chodosh, and P. A. Sharp.** 1985. An RNA polymerase II transcription factor binds to an upstream element in the adenovirus major late promoter. Cell **43**:439–448.
8. **Carthew, R. W., L. A. Chodosh, and P. A. Sharp.** 1987. The major late transcription factor binds to and activates the mouse metallothionein I promoter. Genes Dev. **1**:973–980.
9. **Caudy, M., H. Vassin, M. Brand, R. Tuma, L. Y. Jan, and Y. N. Jan.** 1988. daughterless, a *Drosophila* gene essential for both neurogenesis and sex determination has sequence homologies to myc and the achaete-scute complex. Cell **55**:1061–1067.
10. **Chodosh, L. A., R. W. Carthew, J. W. Morgan, G. R. Crabtree, and P. A. Sharp.** 1987. An adenovirus major late transcription factor activates the γ-fibrinogen promoter. Science **238**:684–688.
11. **Davis, R. L., P. F. Cheng, A. B. Lassar, and H. Weintraub.** 1990. The MyoD DNA binding domain contains a recognition code for muscle specific gene activation. Cell **60**:733–746.
12. **Davis, R. L., H. Weintraub, and A. B. Lassar.** 1987. Expression of a single transfected cDNA converts fibroblasts to myoblasts. Cell **51**:987–1000.
13. **Dieckmann, C. L., and A. Tzagoloff.** 1985. Assembly of the mitochondrial membrane system. CBP6, a yeast nuclear gene necessary for synthesis of cytochrome b. J. Biol. Chem. **260**: 1513–1520.
14. **Dumont, J.** 1972. Oogenesis in *Xenopus laevis* (Daudin); stages of oocyte development in laboratory maintained animals. J. Morphol. **136**:153–180.
15. **Emerson, B. M., C. D. Lewis, and G. Felsenfeld.** 1985. Interaction of specific nuclear factors with the nuclease-hypersensitive region in the chicken adult β-globin gene: nature of the DNA binding domain. Cell **41**:21–30.
16. **Engelke, D. R., S.-Y. Ng, B. S. Shastry, and R. G. Roeder.** 1980. Specific interaction of a purified transcription factor with an internal control region of 5S RNA genes. Cell **19**:717–728.
17. **Fletcher, C., N. Heintz, and R. G. Roeder.** 1987. Purification and characterization of OTF1, a transcription factor regulating cell cycle expression of a human H2B gene. Cell **51**:731–741.
18. **Ginsberg, A. M., B. O. King, and R. G. Roeder.** 1984. *Xenopus* 5S gene specific transcription factor TFIIIA: characterization of a cDNA clone and measurements of RNA levels throughout development. Cell **39**:479–489.
19. **Gonzalez, G. A., K. K. Yamamoto, W. H. Fischer, D. Karr, P. Menzel, W. Briggs III, W. W. Vale, and M. R. Montiminy.** 1989. A cluster of phosphorylation sites on the cyclic AMP-regulated nuclear factor CREB predicted by its sequence. Nature (London) **337**:749–752.

19a. **Gregor, P. D., M. Sawadogo, and R. G. Roeder.** 1990. The adenovirus major late transcription factor USF is a member of the helix-loop-helix group of regulatory proteins and binds to DNA as a dimer. Genes Dev. **4**:1730–1740.

20. **Hall, R. K., and W. L. Taylor.** 1989. Transcription factor IIIA gene expression in *Xenopus laevis* oocytes utilizes a transcription factor similar to the major late transcription factor. Mol. Cell. Biol. **9**:5003–5011.
21. **Hoeffler, W. K., R. Kovelman, and R. G. Roeder.** 1988. Activation of transcription factor IIIC by the adenovirus E1A protein. Cell **53**:907–920.
22. **Jackson, S. P., and R. Tjian.** 1988. O-glycosylation of eukaryotic transcription factors: implications for mechanisms of transcriptional regulation. Cell **55**:125–133.

22a. **Kim, S. H., M. K. Darby, K. Joho, and D. D. Brown.** 1990. The characterization of the TFIIIA synthesized in somatic cells of Xenopus laevis. Genes Dev. **4**:1602–1610.

23. **Krieg, P. A., and D. A. Melton.** 1985. Developmental regulation of a gastrula specific gene injected into fertilized *Xenopus* eggs. EMBO J. **4**:3463–3471.
24. **Laemmli, U. K.** 1970. Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature (London) **227**:680–685.
25. **Landschulz, W. H., P. F. Johnson, and S. L. McKnight.** 1988. The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. Science **240**:1759–1764.
26. **Maniatis, T., E. F. Fritsch, and J. Sambrook.** 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
27. **Matsumoto, Y., and L. Korn.** 1988. Upstream sequences required for transcription of the TFIIIA gene in *Xenopus* oocytes. Nucleic Acids Res. **16**:3801–3815.
28. **Mitchell, P. J., and R. Tjian.** 1989. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. Science **245**:371–378.
29. **Murre, C., P. S. McCaw, and D. Baltimore.** 1989. A new DNA binding and dimerization motif in the immunoglobulin enhancer binding protein, daughterless, MyoD and myogenin. Cell **56**: 777–783.
30. **Murre, C., P. S. McCaw, H. Vaessin, M. Caudy, L. Y. Jan, Y. N. Jan, C. V. Cabrera, J. N. Buskin, S. D. Hauschka, A. B. Lasser, H. Weintraub, and D. Baltimore.** 1989. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. Cell

58:537–544.
31. **Nieuwkoop, P. D., and J. Faber.** 1967. Normal table of *Xenopus laevis* (Daudin). North Holland, Amsterdam.
32. **Sawadogo, M., and R. G. Roeder.** 1985. Interaction of a gene-specific transcription factor with the adenovirus major late promoter upstream of the TATA box region. Cell **43**:165–175.
33. **Sawadogo, M., M. W. Van Dyke, P. D. Gregor, and R. G. Roeder.** 1988. Multiple forms of the human transcription factor USF. I. Complete purification and identification of USF from HeLa cell nuclei. J. Biol. Chem. **263**:11985–11993.
34. **Schreiber, E., P. Matthias, M. M. Müller, and W. Schaffner.** 1988. Identification of a novel lymphoid specific octamer binding protein (OTF-2B) by proteolytic clipping bandshift assay (PCBA). EMBO J. **13**:4221–4229.
35. **Scotto, K. W., H. Kaulen, and R. G. Roeder.** 1989. Positive and negative regulation of the gene for transcription factor IIIA in *Xenopus laevis* oocytes. Genes Dev. **3**:651–662.
36. **Shastry, B. S., B. M. Honda, and R. G. Roeder.** 1984. Altered levels of 5S gene specific transcription factor TFIIIA during oogenesis and embryonic development. J. Biol. Chem. **259**: 11373–11382.
37. **Sorger, P. K., M. J. Lewis, and H. R. B. Pelham.** 1987. Heat shock factor is regulated differently in yeast and in Hela cells. Nature (London) **329**:81–84.
38. **Smith, J. C.** 1987. A mesoderm inducing factor is produced by a *Xenopus* cell line. Development **99**:3–14.
39. **Studier, W. F., A. H. Rosenberg, and J. J. Dunn.** 1990. The use of T7 RNA polymerase to direct the expression of cloned genes. Methods Enzymol. **185**:60–85.
40. **Tafuri, S. R., and A. P. Wolffe.** 1990. Developmental regulation of class II transcription factors in *Xenopus laevis*. J. Cell. Biochem. Suppl. **14b**:190.
41. **Tapscott, S. J., R. L. Davis, M. J. Thayer, P. F. Cheng, and H. Weintraub.** 1988. Myo D: a nuclear phosphoprotein requiring a Myc homology region to convert fibroblasts to myoblasts. Science **242**:405–411.
42. **Taylor, W., I. J. Jackson, N. Siegel, A. Kumar, and D. D. Brown.** 1986. The developmental expression of the gene for TFIIIA in *Xenopus laevis*. Nucleic Acids Res. **14**:6185–6195.
43. **Thisse, B., C. Stoetzel, C. Gorostiza-Tisse, and F. Perrin-Schmitt.** 1988. Sequence of the twist gene and nuclear localization of its protein in endomesodermal cells of early *Drosophila* embryos. EMBO J. **7**:2175–2183.
44. **Thomas, P. S.** 1980. Hybridization of denatured RNA and small DNA fragments transferred to nitrocellulose. Proc. Natl. Acad. Sci. USA **77**:5201–5205.
45. **Tso, J. Y., D. J. Van den Berg, and L. Korn.** 1986. Structure of the gene for *Xenopus* transcription factor TFIIIA. Nucleic Acids Res. **8**:5473–5505.
46. **Villares, R., and C. V. Cabrera.** 1987. The achaete-scute gene complex of *D. melanogaster*: conserved domains in a subset of genes required for neurogenesis and their homology to myc. Cell **50**:415–424.
47. **Wormington, W. M., M. Schlissel, and D. D. Brown.** 1982. Developmental regulation of *Xenopus* 5S RNA genes. Cold Spring Harbor Symp. Quant. Biol. **47**:879–884.
48. **Wright, W. E., D. A. Sassoon, and V. K. Lin.** 1989. Myogenin, a factor regulating myogenesis, has a domain homologous to MyoD. Cell **56**:607–617.