MOLECULAR AND CELLULAR BIOLOGY, Mar. 1995, p. 1254–1264 Vol. 15, No. 3
0270-7306/95/$04.00+0

The Embryonic Enhancer-Binding Protein SSAP Contains a Novel DNA-Binding Domain Which Has Homology to Several RNA-Binding Proteins

DANIEL J. DEANGELO,† JEFFERY DEFALCO, LISA RYBACKI, AND GEOFFREY CHILDS*

Department of Molecular Genetics, Albert Einstein College of Medicine, Bronx, New York 10461

Received 9 September 1994/Returned for modification 3 November 1994/Accepted 2 December 1994

Stage-specific activator protein (SSAP) is a 43-kDa polypeptide that binds to an enhancer element of the sea urchin late histone H1 gene. This enhancer element mediates the transcriptional activation of the late histone H1 gene in a temporally specific manner at the mid-blastula stage of embryogenesis. We have cloned cDNAs encoding SSAP by using polyclonal antibodies raised against purified SSAP to screen expression libraries. SSAP is unrelated to previously characterized transcription factors; however, it exhibits striking homology to a large family of proteins involved in RNA processing. The protein is a sequence-specific DNA-binding protein that recognizes both single- and double-stranded DNA. The DNA-binding domain of the protein was localized to the conserved RNA recognition motif (RRM). In addition to tandem copies of this conserved domain, SSAP contains a central domain that is rich in glutamine and glycine and a C-terminal domain that is enriched in serine, threonine, and basic amino acids. Overexpression of SSAP in sea urchin embryos by microinjection of either synthetic mRNA or an SSAP expression vector results in four- to eightfold transactivation of target reporter genes that contain the enhancer sequence. Transactivation occurs beginning only at the mid-blastula stage of development, suggesting that SSAP must be modified in a stage-specific manner in order to activate transcription. In addition, there are a number of other RRM-containing proteins that contain glutamine-rich regions which are postulated to function in the regulation of RNA processing. Instead, we suggest that SSAP is a member of a family of glutamine-rich RRM proteins which constitute a novel class of transcription factors.

Normal embryological development mandates that specific sets of genes be expressed with distinct temporal patterns. The sea urchin contains two different developmentally regulated histone gene families (for a review, see reference 33). The early histone gene family begins to be transcribed soon after fertilization, peaks at about 12 h postfertilization, and then undergoes a precipitous decline. Members of the late histone gene family are transcribed at low levels until about 12 h postfertilization, when they are activated, and reach their peak transcriptional levels in 24 h late-blastula-stage embryos (4, 21, 23, 25, 27). This late histone gene family constitutes the somatic set of histone genes.

We have previously demonstrated that a segment of DNA between positions −288 and −317 (upstream sequence element [USE] IV) in the late histone H1-β gene is responsible for the transcriptional activation during the blastula stage of embryogenesis (29). This highly conserved sequence is found in the same position in the *Strongylocentrotus purpuratus* late H1γ and *Lytechinus pictus* late H1γ genes (24, 29), and we have demonstrated that it is capable of acting as an embryonic enhancer element, activating target genes in a stage-specific manner (28). Nuclear extracts prepared from developmental stages before and after the activation of the late histone gene all contain a factor which specifically binds to the temporal enhancer (11). A 43-kDa polypeptide which binds to and footprints the USE IV enhancer element has been purified. This protein is referred to as stage-specific activator protein (SSAP). SSAP is the only detectable USE IV sequence-specific DNA-binding activity found in nuclear extracts of sea urchin embryos. In addition, multimers of this binding site are capable of functioning as a temporal enhancer element (29, 42a). Immunoblot analysis demonstrates that SSAP is present in low but detectable amounts early in development and then increases as development proceeds. Although SSAP DNA-binding activity is present in the embryo before the enhancer is activated, a change in the molecular weight of SSAP precisely parallels the increase in H1-β gene transcription. Early in development, SSAP appears as a 43-kDa monomer, but it undergoes a posttranslational modification beginning at about 12 h postfertilization (early blastula). Modified SSAP is a dimer that contains at least one 43-kDa subunit and has an apparent molecular mass of approximately 80 to 100 kDa. Thus, it is the disappearance of the 43-kDa species and the appearance of the 80- to 100-kDa species which coincide with H1-β gene activation (11). The correlation between the presence of the SSAP dimer and the pattern of H1-β gene expression strongly suggests that it is this dimeric species which is directly responsible for the blastula stage-specific transcriptional activation of the late H1 gene.

Most well-characterized DNA-binding motifs interact with the major groove of double-stranded DNA. These motifs include families of proteins containing the helix-turn-helix, homeodomain, zinc finger, leucine zipper, and helix-loop-helix structures as well as others (reviewed in reference 40). However, alternate DNA conformations may also be involved in the regulation of gene transcription. S1 nuclease and bromoacetaldehyde have been used to detect altered DNA structures at specific chromosomal locations within the 5′-flanking regions of actively transcribed genes (30, 48). These altered DNA sites are also present in supercoiled recombinant plasmids but not relaxed or linear constructs, suggesting that the DNA secondary structure is dependent on both specific nucleotide se-

* Corresponding author. Mailing address: Department of Molecular Genetics, Albert Einstein College of Medicine, 1300 Morris Park Ave., Bronx, NY 10461. Phone: (718) 430-3569. Fax: (718) 430-8778.

† Present address: Department of Medicine, Massachusetts General Hospital, Boston, MA 02114.

quences and the free energy derived from superhelicity. Non-B-form DNA structures can result from either DNA bending (58), transitions to Z-form DNA (38), DNA strand slippage (19), or double-helical strand melting which may be present de novo or induced by either helicases or other DNA-binding proteins. There have been numerous reports of nuclear proteins which recognize these altered DNA structures (15, 16, 44, 52, 54), yet the mechanism by which they exert a regulatory role in gene transcription has not been elucidated.

In this study, we have used antisera generated against purified SSAP to isolate a set of cDNA clones from a sea urchin embryonic library. We have demonstrated that these cDNA clones encode the embryonic activator protein SSAP and are capable of transactivating the sea urchin late H1 histone gene promoter in vivo. Sequence analysis revealed the presence of two RNA recognition motifs (RRMs) which are common to many of the single-stranded-DNA- and RNA-binding proteins (14, 22a, 41). Upon further investigation, we determined that SSAP is capable of binding to both double- and single-stranded DNA in a sequence-specific manner. Furthermore, we have demonstrated that the RRM contains the sequence-specific DNA recognition motif of a protein involved in transcription activation.

MATERIALS AND METHODS

Microinjection of sea urchin one-cell zygotes. The procedure used to inject the *L. pictus* zygotes was essentially that of McMahon et al. (35) and Colin (6) and exactly as described by Lai et al. (29). Chloramphenicol acetyltransferase (CAT) assays were performed as previously described (31).

Mobility shift assays. Mobility shift assays were carried out as previously described (13, 28). The oligonucleotides used in the studies were USE IV coding strand (5′-GATCGACAAGTTTTAAATCTGATTCTG-3′), USE IV noncoding strand (5′-GATCCAGAATCAGATTTAAAACTTGTC-3′), nonspecific coding strand (5′-GATCTAAGAGAAGTTTGA-3′), and nonspecific noncoding strand (5′-GATCCTCAAACTTCTCTTA-3′).

SDS-PAGE and Western blotting (immunoblotting). For Western blots, samples were first subjected to sodium dodecyl sulfate (SDS)-polyacrylamide gel electrophoresis (PAGE) and then electrophoretically transferred onto nitrocellulose (55). To detect bound antibody, the blots were incubated with alkaline phosphatase-conjugated goat anti-rabbit antibodies (Stratagene) and visualized with the substrates nitroblue tetrazolium and 5-bromo-4-chloro-3-indolyl phosphate (BCIP), both of which were obtained from Bio-Rad. For Western blotting using chemiluminescence detection (Amersham), blots were washed in TBSN (20 mM Tris-HCl [pH 7.4], 150 mM NaCl, 0.05% Nonidet P-40) following transfer. Blots were developed as recommended by the manufacturer.

Isolation of SSAP cDNA clones. A λZAP cDNA library was made from poly(A)$^+$ RNA isolated from 4-h *S. purpuratus* embryos (kindly provided by E. Davidson). The library was plated by using BB4 cells as described by Sambrook et al. (43). The fusion peptide was induced by preincubating the nitrocellulose filters with 10 mM IPTG (isopropyl-β-thiogalactopyranoside). The blots were screened with a 1:200 dilution of the polyclonal antisera (59). To visualize positive clones, the blots were incubated with alkaline phosphatase-conjugated goat anti-rabbit antibodies (Stratagene) and visualized with the substrates nitroblue tetrazolium and BCIP (Bio-Rad).

Sequencing methods. The nucleotide sequence was determined from both strands, using the Sequenase version 2.0 kit (U.S. Biochemical) on either CsCl- or polyethylene glycol-precipitated miniprep DNA.

Affinity purification of antibodies. Nitrocellulose filters with immobilized protein were prepared by overlaying approximately 2,000 immunopositive phage plaques (57). A 1:50 dilution of the polyclonal antisera in 1× Tris-buffered saline–0.1% Tween 20–3% bovine serum albumin (BSA) was incubated with the filters for 2.5 h at room temperature. The filters were washed three times with 1× Tris-buffered saline–0.1% Tween 20. The antibody was eluted with three 1-min washes in 5 mM glycine-HCl (pH 2.3)–150 mM NaCl–0.1% Tween 20–100 μg of BSA per ml. The combined washes were immediately neutralized with 50 mM Tris-HCl (pH 7.4). These antibodies were reacted with Western blots as described previously except that the blots were visualized by using the Western-Light chemiluminescence detection system (Tropix).

In vitro transcription and translation of SSAP cDNA. Plasmid pSSAP7.1 was linearized by using various restriction endonucleases and transcribed in vitro with T3 RNA polymerase. The N-terminal deletions were obtained by using the RNA amplification with transcript sequencing method (45). The appropriate region of the cDNA insert was amplified by PCR. The upstream oligonucleotide contained a 29-base T7 promoter sequence followed by an 8-base translation initiation signal and a 17-base sequence complementary to the cDNA insert. The first two regions of the oligonucleotide contain the sequence GGATCCTAATACGACT CACTATAGGGAGACCACCATG. The transcript produced an 11-nucleotide leader (GGGAGACCACC) followed by an ATG initiation codon and the remainder of the coding region. Approximately 1 μg of RNA was added to a rabbit reticulocyte lysate (Promega). DNA affinity chromatography was performed as previously described (11), with the following modifications: one-half of the reaction buffer (25 μl) was diluted 1:1 with *N*-2-hydroxyethylpiperazine-*N*′-2-ethanesulfonic acid (HEPES) buffer (20 mM HEPES-KOH [pH 7.9], 0.1 mM EDTA, 1 mM dithiothreitol [DTT], 1 mM phenylmethylsulfonyl fluoride, 0.1% Nonidet P-40, 20% glycerol) containing no KCl, and the protein was loaded onto the column with 0.5 μg of poly(dI-dC) as a nonspecific competitor. The column contained approximately 50 μl of packed resin and had a void volume of approximately 50 μl, and fractions corresponding to this size were collected. The column was washed with low-salt HEPES buffer containing 0.15 and 0.25 M KCl, and then protein was eluted in 0.6 M KCl. The fractions from the wild-type SSAP, ΔC389, and ΔN90 constructs were resolved on SDS–10% polyacrylamide gels, while the fractions from the ΔN178 and ΔC202 constructs were separated on 12% gels.

SSAP mRNA for microinjection was synthesized in vitro from linearized pSSAP7.1 with T3 RNA polymerase. Capped, full-length transcripts were obtained by using a Megascript T3 RNA polymerase kit (Ambion). RNA was diluted to appropriate concentrations in sterile diethylpyrocarbonate-treated H_2O.

Expression and purification of recombinant SSAP. Full-length SSAP protein was produced in *Escherichia coli* by using the pET expression system (50). An *Nco*I site was introduced at the first ATG codon of the SSAP cDNA and a *Bam*HI site was placed at the 3′ end of the SSAP cDNA through PCR using the appropriate oligomers. The resulting fragment was then subcloned into pET3d to yield pET-SSAP1. Recombinant SSAP was expressed in *E. coli* BL21(DE3), and SSAP protein was recovered from the insoluble fraction by the method of Hoey and Levine (20). Bacterium-expressed SSAP (bSSAP) was further purified by preparative SDS-PAGE. Renatured insoluble fraction from induced cells was resolved on an SDS–10% reducing polyacrylamide gel. Following electrophoresis, the gel was washed in double-distilled H_2O and stained in 0.25 M KCl–1 mM DTT for 5 min at 4°C. A prominent band at 44 kDa representing bSSAP was excised and washed in cold 1 mM DTT. The gel slice was crushed, and the protein was eluted in 100 mM KCl–25 mM HEPES (pH 7.9)–20% glycerol–12.5 mM $MgCl_2$–0.1% Nonidet P-40–2 mM benzamidine–0.5 mM phenylmethylsulfonyl fluoride–5 μg of pepstatin A per ml–5 μg of leupeptin per ml. Eluted protein was then acetone precipitated, and the resulting pellet was resuspended in 25 mM HEPES–0.1 M KCl–1 mM EDTA–1 mM DTT–20% glycerol–6 M guanidine. Denaturant was removed by dialysis against the same buffer without guanidine-HCl. Active protein was then concentrated by passing this gel-purified material over a double-stranded calf thymus DNA-cellulose column. The column was washed with 0.13 M KCl buffer, and bound bSSAP was eluted with 0.6 M KCl buffer.

Production of defined RNA oligomers with USE IV sequences. RNA oligomers containing the USE IV element sequences were transcribed in vitro, using T7 RNA polymerase from a double-stranded oligonucleotide template containing a T7 promoter immediately upstream from either the positive coding strand or negative noncoding strand of the USE IV oligomers. The runoff transcripts containing the same sequences as the 27-mer single strands of the USE IV enhancer were quantitated and used as competitors in mobility shift assays.

Nucleotide sequence accession number. The SSAP nucleotide sequence has been deposited in GenBank and can be found under accession number L15365.

RESULTS

Cloning of SSAP. To isolate a cDNA clone of SSAP, we screened a 4-h λZAP *S. purpuratus* library with polyclonal antisera raised against affinity-purified SSAP monomer (11). The 4-h sea urchin library is composed mainly of maternally derived transcripts, since zygotic transcription does not become significant until 6 h postfertilization (10). Both mobility shift assays (11) and Western blot analysis (Fig. 1A) have demonstrated that SSAP is present by 4 h postfertilization, albeit in small quantities. The first detectable expression of SSAP must therefore arise from translation of the maternal transcripts.

Twelve independent immunopositive clones were obtained from screening approximately 1.8×10^6 clones. Since the polyclonal antisera cross-react with a few other sea urchin nuclear proteins (Fig. 1A), each clone was tested for its ability to affinity purify anti-SSAP immunoglobulin. This type of affinity purification has been used successfully to isolate different subclasses of the glucocorticoid receptors (57). Antibody affinity purified from each fusion protein was tested for its ability to

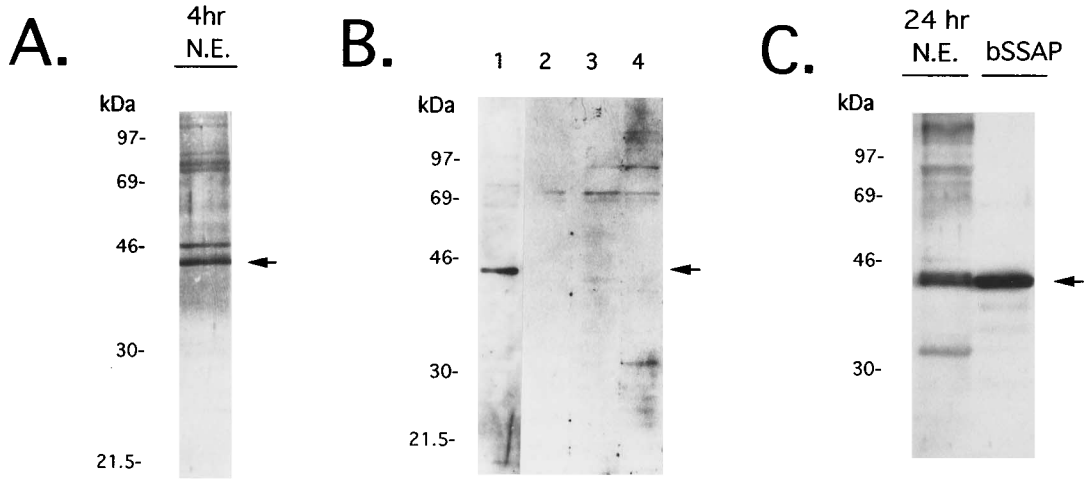

FIG. 1. Identification of recombinant clones by epitope selection. (A) Western blot of 4-h nuclear extract (N.E.). Crude nuclear extract (50 μg) from 16-cell embryos was fractionated by SDS-PAGE, transferred to nitrocellulose, and probed with crude antisera generated against purified SSAP (11) as described in Materials and Methods. (B) Immunopositive recombinant clones isolated from a 4-h *S. purpuratus* library were used to affinity purify immunoglobulin from crude polyclonal antisera generated against purified SSAP. Fusion protein from approximately 2,000 plaques was immobilized onto nitrocellulose and incubated with crude antisera for 2.5 h. The selected antibody was eluted at low pH, neutralized, and used to probe a Western blot of 12-h crude nuclear extract. Lane 1, the antibody used was selected against clone pSSAP7.1, a seropositive clone; lanes 2 to 4, the antibody used was selected against clones pSSAP.1A, pSSAP.1B, and pSSAP.3-1, all seronegative clones. SSAP migrates as a 43-kDa polypeptide. (C) Comparison of the mobilities of authentic SSAP and bSSAP (encoded by clone pSSAP7.1) on SDS-polyacrylamide gels. Crude nuclear extract (1.25 μg) and purified bSSAP (50 ng) were fractionated by SDS-PAGE and processed for Western blot analysis as described in Materials and Methods. The Western blot was probed with antisera against authentic sea urchin SSAP. The blot in panel A was processed by using alkaline phosphatase-conjugated goat anti-rabbit antibodies (Stratagene); the Amersham chemiluminescence detection system was used for panels B and C.

cross-react with the 43-kDa polypeptide, SSAP, by immunoblot analysis. Of the original 12 clones that were isolated, only 4 tested positive; 1 of these positive clones is shown in Fig. 1B. The antiserum that was used to screen the cDNA library was also used to probe a Western blot containing both crude nuclear extract from 24-h late-blastula-stage embryos and bSSAP (Fig. 1C). The bacterial protein was expressed from the putative initiation codon of the pSSAP7.1 cDNA insert that was subcloned into the pET3 expression vector (see below). Authentic sea urchin SSAP and bSSAP have identical mobilities in these denaturing polyacrylamide gels.

All four of the positive clones cross-hybridize with one another, and two contained identical cDNA inserts. The nucleotide sequence of each of the clones predicts an open reading frame of 1,212 nucleotides, corresponding to a protein with an approximate molecular mass of 44 kDa (Fig. 2). This value is nearly identical to the apparent molecular mass of SSAP determined by SDS-PAGE and glycerol gradient sedimentation (11). The open reading frames of each of the four clones were identical; the only differences lie in the 3′ untranslated region.

Approximately 3 nmol of affinity-purified authentic 43-kDa SSAP was subjected to acid hydrolysis for determination of its amino acid composition. The composition of the purified protein was compared with that of the predicted coding sequence of the cDNA clone. The results revealed that the compositions of authentic SSAP and the open reading frame in pSSAP7.1 are nearly identical and provide strong evidence that pSSAP7.1 encodes SSAP (data not shown).

Northern (RNA) analysis of poly(A)$^+$ RNA, using the 1.8-kb pSSAP7.1 cDNA as a probe, revealed a single major mRNA species of about 3.8 kb (Fig. 3). Upon longer exposure of the Northern blot, other, less abundant transcripts are detected (data not shown), probably resulting either from alternatively processed transcripts or from cross-hybridization with related genes. The SSAP mRNA is present in low but detectable amounts in the egg and in 4- and 16-cell embryos corresponding to 2 and 4 h postfertilization. SSAP transcripts begin to accumulate between 4 and 8 h postfertilization and are present throughout embryogenesis, paralleling the development profile of the SSAP levels (11). SSAP seems to be represented by a relatively rare transcript.

Common structural motifs in SSAP. A computer search of the databases by using the SSAP amino acid sequence failed to detect any known sequence-specific DNA-binding motifs. Unexpectedly, within the first 180 amino acids, SSAP shows significant homology to the ribonuclear proteins (RNPs) (Fig. 4). SSAP contains two 90-amino-acid repeats which are homologous to the consensus sequence of the RNA-binding domain or the RRM (14). The RRM is characterized by two aromatic amino acid-rich sequences. Each 90-amino-acid repeat contains two conserved oligopeptide sequences designated RNP-1 and RNP-2. Both of these regions show an excellent match to the consensus sequences of RGFGFVTF for RNP-1 and LFVGGL for RNP-2 (14, 41). Most RNA-binding proteins contain one or more copies of the RRM domain. Some of the RRM-containing proteins have been shown to bind to single- and double-stranded DNA as well as RNA stem-loops. Although many of these proteins have not been shown to be sequence-specific DNA-binding proteins, some show preferential binding to A+T-rich DNA stretches (39). A second notable feature of the SSAP primary structure is the presence of a glycine- and glutamine-rich domain immediately adjacent to the RRM. The region from residues 181 to 290 is composed of 43% glycine and contains no discernible α-helical or β-sheet structures. Overlapping this glycine-rich domain is a region from residues 225 to 290 which is 37% glutamine. Glutamine-rich regions have been identified in many transcription factors (7, 8). The C-terminal 114 residues show no identity to any of the proteins contained in the database. This region, however, is both serine and threonine rich and very basic, with an isoelectric point of 11.3.

Characterization of the SSAP DNA-binding domain. The DNA-binding properties of the in vitro-translated SSAP were assayed by analytical DNA affinity column chromatography

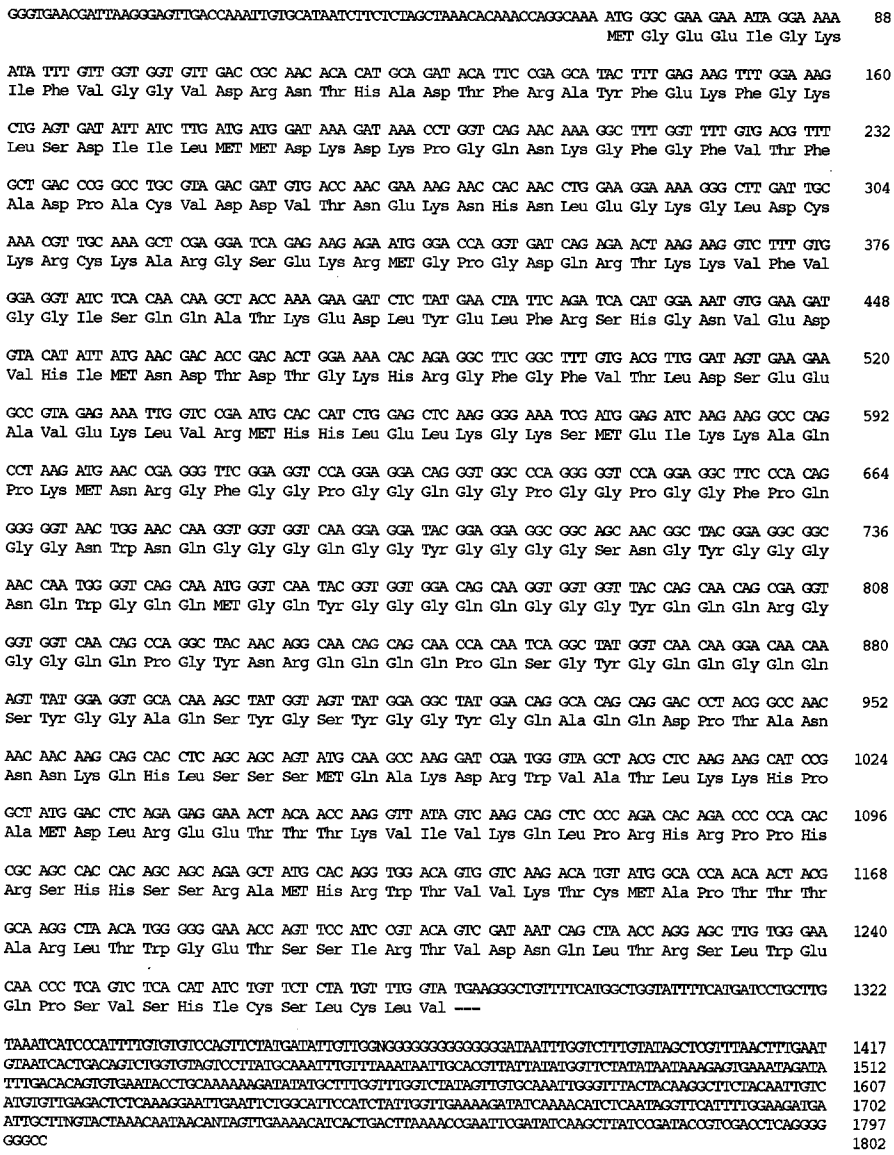

FIG. 2. Nucleotide sequence of the SSAP cDNA clone. The four positive clones identified by epitope selection were subcloned and sequenced on both strands, using the dideoxy method. All four clones contained the same open reading frame. The sequence depicted is from clone pSSAP7.1.

using the SSAP recognition sequence, USE IV. This is the same DNA affinity resin which was used to purify native sea urchin SSAP (11). This assay has been previously applied to the characterization of the Fos-Jun complexes as well as the RNA polymerase I transcription factor UBF (22, 56). The protein was generated by in vitro translation of synthetic SSAP mRNA templates. In vitro-translated proteins which contain a DNA-binding domain will be able to bind to the affinity column in the presence of poly(dI-dC) and be eluted with high salt concentrations. Proteins lacking the DNA-binding domain will appear in both the flowthrough and fractions from the low-salt wash. Essentially wild-type levels of DNA-binding activity were observed with both of the C-terminal deletions (Fig. 5). Construct ΔC380 removed only part of the C-terminal domain, while construct ΔC202 removed both the C-terminal domain and most of the glycine-glutamine-rich domain. Construct ΔN178, which entirely removed both of the RRM repeats, failed to bind to the column.

These data demonstrate that the SSAP DNA-binding domain lies within the RRM repeats and that neither the C-terminal tail nor the glycine-glutamine-rich regions significantly influence the DNA-binding efficiency. Interestingly, construct ΔN90, which removes only the first N-terminal RRM repeat, is still able to bind to the USE IV affinity column with little decrease in efficiency (Fig. 5). The equivalence of the RRM repeats for their DNA-binding activity has not yet been determined, but it is clear that at least part of the RRM domain is necessary for DNA-binding activity.

pSSAP7.1 encodes a sequence-specific DNA-binding protein that recognizes the USE IV enhancer. To characterize the DNA-binding activity of the protein encoded by the isolated SSAP clone, electrophoretic mobility shift experiments were carried out with bSSAP and a 27-bp double-stranded oligonucleotide containing the USE IV enhancer element. In the mobility shift assay, purified bSSAP exhibits a mobility slightly slower than that of authentic SSAP monomer (see Fig. 7A,

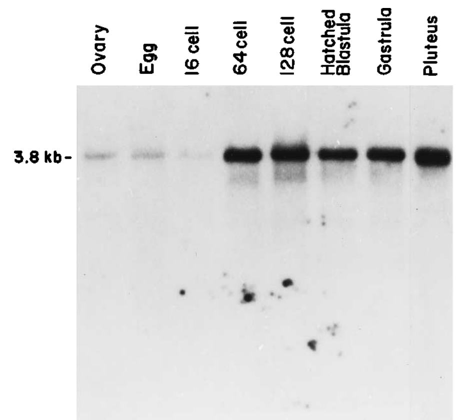

FIG. 3. Developmental Northern blot analysis of the sea urchin SSAP transcripts. Poly(A)$^+$ RNA (10 μg) from each stage of development indicated was separated on a denaturing agarose gel and transferred to nitrocellulose. The membrane was UV cross-linked and probed with a random-primed DNA fragment containing the entire SSAP coding region. The size of the transcript is indicated at the left.

lanes 8 and 9). Since bSSAP and authentic SSAP migrate identically on SDS-PAGE (Fig. 1C), the difference in mobility in gel shift assays may reflect the possibility that bSSAP has not fully assumed its native conformation. Such a discrepancy in gel retardation assays between prokaryotically expressed and authentic DNA-binding proteins has been noted for several cloned transcription factors, for example, Oct-1 (51). The interaction of bSSAP with the USE IV probe is specific since it can be competed for with double-stranded specific DNA (unlabeled probe) but not double-stranded nonspecific DNA (Fig. 6A, lanes 2 and 3). Given the extensive homology of SSAP with RRM-containing proteins, we sought to determine whether bSSAP would exhibit affinity for single-stranded nucleic acid. Mobility shift competition experiments were performed with the individual strands of the USE IV oligomer or nonspecific oligonucleotides as competitors (Fig. 6A, lanes 4 to 7). Both individual strands of the USE IV enhancer element are capable of competing for bSSAP-binding activity with the double-stranded oligonucleotide probe; however, nonspecific single-stranded DNA failed to compete for bSSAP binding. Examination of the USE IV-binding site revealed that it is an imperfect palindrome with the two strands having very similar sequences (9 of 13 bases identical). This could explain the competition using each of the strands of the enhancer element.

More direct evidence of SSAP's ability to interact with the single strands of the enhancer was obtained by using each strand of the USE IV element as a probe in a mobility shift assay. bSSAP binds to both the coding and noncoding strands of the USE IV element (Fig. 6B, lanes 1 and 2), but it fails to recognize either of the control, nonspecific, single-stranded probes (lanes 3 and 4). bSSAP exhibits slightly greater affinity for the coding strand of the USE IV enhancer element than for the noncoding strand. Mobility shift competition experiments using the individual strands of the USE IV oligomer as a probe or nonspecific oligonucleotides as competitors confirmed the specificity of the binding (data not shown).

```
                              RNP-2                                   RNP-1

                              beta-1     alpha-1       beta-2      beta-3    alpha-2         beta-4
Human U1 snRNP A #1           TIYINNLNEKIKKDELKKSLYAIFSQFGQILDILVSR-SLKMRGQAFVIFKEVSSATNALRSMQGFPFYDKPMRIQYAK
Human U1 snRNP A #2           ILFLTNLPEE-TNELMLSMLFNQFPGFKEVRLVP-----GR-HDIAFVEFDNEVQAGAARDALQGFKITQNNAMKISFA
Human hnRNP A1 #1             KLFIGGLSFE-TTDESLRSHFEQWGTLTDCVVMRDPN-TKRSRGFGFVTYATVEEVDAAMNAR-PHKVDGRVVEPKRAV
Human hnRNP A1 #2             KIFVGGIKED-TEEHHLRDYFEQYGKIEVIEIMTDRG-SGKKRGFAFVTFDDHDSVDKIVIQK-YHTVNGHMCEVRKAL
Human U1 snRNP 70K            TLFVARVNYD-TTESKLRRAFEVYGPIKRIHMVYSKR-SGKPRGYAFIEYEHERDMHSAYKHADGKKIDGRRVLVDVER
Human poly(A) bp#1            SLYVGDLHPD-VTEAMLYEKFSPAGPILSIRVCRDMI-TRRSLGYAYVNFQQPADAERALDTMNFDVIKGKPVRINWSQ
Human poly(A) bp#2            NIFIKMLDKS-IDNKALYDTFSAFGNILSCKVVCDEN----SKGYGFVHFETQGAAERAIEKMMGMLLNDRKVFVGRFK
Human poly(A) bp#3            NVYIKNFGED-MDDERLKDLFGPALSVKVMTDES-----GKSKGFGFVSFERHEDAQKAVDEMNGKELNGKQIYVGRAQ
Human poly(A) bp#4            NLYVKNLDDG-IDDERLRKEFSPFGTITSAKVNMEG---GRSKGFGFVCFSSPEEATKAVTEMNGRIVATKPLYVALAQ
Human SF2                     RIYVGNLPPDIRTKD-IEDVFYKYGAIRDIDLKNRRG--G--PPFAFVEFEDPRDAEDAVYGRDGYDYDGYRLRVEFPR
Fly sxl #1                    NLIVNYLPQD-MTDRDVYALFRAIGPINTCRINRDYK-TGYSFGYAFVDVTSEMDSQRAIKVLNGITVRNKRLKVSVAR
Fly sxl #2                    NLYVTNLPRT-ITDDQLDTIFGKYGSIVQKNILRDKL-TGRPRGVAFVRYNKREEAQEAISALNMVIPEGGSQPLSVAL
Fly tra2                      GVFGLNTN---TSQHKVRELFNKYGPIERIQMVIDAQ-TQRSRGFCFIYFEKLSDARAAKDSCSGIEVDGRRIRVDVSI
Fly hdp #1                    KLFIGGLDYR-TTDENLKAHFEKWGNIVDVVVMKDPR-TKRSRGFGFITYSHSSMIDEAQSR--PHKIDGRVVEPKRAV
Fly hdp #2                    KLFVGALKDD-HDEQSIRDYFQHFGNIVDINIVIDKE-TGKKRGFAFVEFDDYDPVDKVVLQK-QHQLNGKMVDVKCOW
Rat hdp                       RKLFIGGLSFETTDESLRSHFEQWGTLTDCVVMRDPN-TKRSRGFGFVTYATVEEVDAAMNAR-PHKVDGRVVEPKRAV
Hamster nucleolin C/23        KTLVLSNLSYSATEETLQEVFEK-----ATFIKVPQNQQGKSKGYAFIEFASFEDAKEALNSCNKMEIEGRTI---RLE
Mus-CArG A                    GKMFVGGLSWDTSKKDLKDYFTKFGEVVDCTIKMDPM-TGRSRGFGFILFKDSSSVEKVLDQK-EHRLDGRVIDPKKAM
EWS (362-447)                 AIYVQGINDS-VTLDDLADFFKQCGVVKM (+18 AAs) KPKGDATVSYEDPPTAKAAVEWFDGKDFQGSKLKVSLAR
TLS (287-372)                 TIFVQGLGEN-VTIESVADYFKQIGIIKT (+18 AAs) KLKGEATVSFDDPPSAKAAIDWFDGKEFSGNPIKVSFAT
E.coli CspA (16-53)                                                   KGFGFITPDDGSKDVFVHFSAIQNDGYKSLDEGQKVSF
Xenopus FRGY1 (48-85)                                                 NGYGFINRNDTKEDVFVHQTAIKKNNPRKYLRSVGDGE

SSAP  Repeat #1               GKIFVGGVDRNTHADTFRAYFEKFGKLSVIILMMDKDKPGQNKGFGFVTFADPACVDDVTNEK-NHLLEGKGLDCKRCK
      Repeat #2               KKVFVGGISQQATKEDLYELFRSHGNVEDVHIMNDTD-TGKHRGFGFVTLDSEEAVEKLVRMH-HLELKGKSMEIKKAQ
```

FIG. 4. Sequence alignment of selected RRMs. The RNP-1 and RNP-2 elements are labeled, and eight other highly conserved positions are shown in boldface. The other, less conserved positions are not highlighted. The regions of β sheet and α helix indicated at the top are based on the predictions of Kenan et al. (22a) determined from the nuclear magnetic resonance spectra and crystal structures of U1 small nuclear RNP-A (snRNP A). A number of slightly different alignments are possible by the introduction of different or additional gaps. The EWS and TLS proteins contain the indicated 18-amino-acid insertion. Poly(A) bp, polyadenylate-binding protein; hdp, helix-destabilizing protein. The numbers following the name of a protein refer to each individual tandem repeat of the RRM domain found within the entire protein sequence. The sequences were compiled from references 18 and 26 and additional references cited in the text. hnRNP, heterogeneous nuclear RNP.

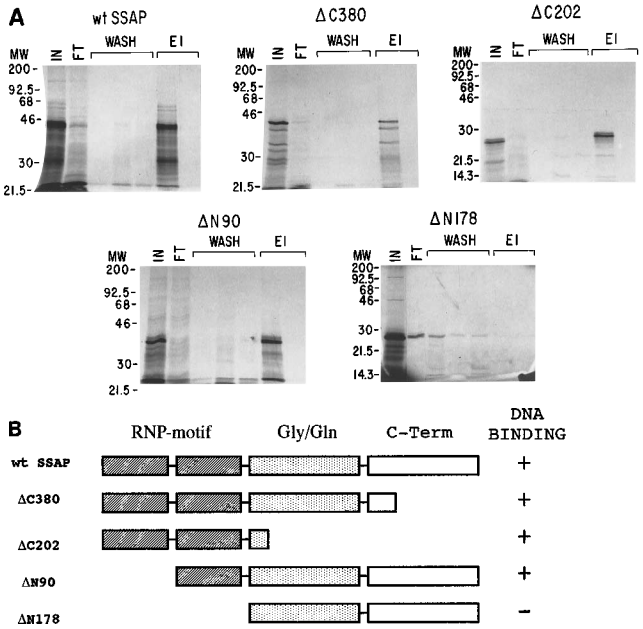

FIG. 5. (A) Identification of the DNA-binding domain of SSAP. Wild-type (wt) and truncated deletions of SSAP were transcribed and translated in vitro and passed over small DNA affinity columns containing the SSAP recognition sequence in the presence of a nonspecific competitor. The columns were washed at a low salt concentration, and bound protein was eluted with a high-salt buffer. Aliquots of the fractions were separated by SDS-PAGE and autoradiographed. Lanes: M, molecular weight (mw) markers (indicated in thousands); IN, input material; FT, flowthrough; WASH, low-salt wash fractions (0.15 and 0.25 M KCl); El, high-salt eluent fractions (0.6 and 1.0 M KCl). (B) Structures of the deleted constructs. Darkened boxes represent RNP repeats, the stippled boxes represent Gly-Glu-rich domains, and open boxes represent the basic C-terminal tail of SSAP.

Authentic SSAP monomer purified from sea urchin nuclei (11) was tested for the ability to bind to single-stranded DNA. The experiments shown in Fig. 6 were repeated except for the substitution of authentic SSAP monomer for bSSAP. The results conclusively demonstrate that native SSAP is a single-stranded-DNA-binding protein (Fig. 7). In fact, SSAP has a greater affinity for the single strands than for specific double-stranded DNA. The native SSAP protein prefers the coding strand of the 27-bp USE IV enhancer to a much greater degree than does the bSSAP protein. This difference between SSAP and bSSAP could be the result of posttranslational modification of SSAP.

A number of RRM proteins are capable of binding to single-stranded DNA, with preferences for either polypyrimidine or polypurine tracts (14). Although purified SSAP is capable of footprinting large double-stranded DNA (11), the A+T-rich nature of the SSAP-binding site suggested that like some other RRM proteins, SSAP might not bind with the specificity required of a transcriptional activator. Three mutant oligonucleotides were synthesized to test the sequence specificity of SSAP binding (Fig. 8A). We compared the abilities of wild-type and mutant oligonucleotides to compete for binding in a mobility shift assay (Fig. 8B). None of the three mutant oligonucleotides were effective competitors compared with the wild-type sequence. The mutant sequences included alterations that substituted a G+C-rich region within the binding site (mutant 3) as well as two subtle alterations that maintained the same A+T content (mutants 1 and 2) (Fig. 8A). We also measured equilibrium dissociation constants of purified native SSAP for double-stranded USE IV oligonucleotide ($K_d = 3 \times 10^{-9}$ M), coding strand ($K_d = 1.8 \times 10^{-10}$ M), noncoding strand ($K_d = 3.2 \times 10^{-10}$ M), mutant 1 ($K_d = 1.6 \times 10^{-9}$ M), mutant 2 ($K_d = 1.6 \times 10^{-9}$ M), and mutant 3 ($K_d = 3.6 \times 10^{-8}$ M) by mobility shift experiments and Scatchard plot analysis (data not shown). These quantitative measurements showed that SSAP has a 10- to 30-fold-higher affinity for single-stranded USE IV target sequence and that the mutations in the single-stranded probes reduced the affinity by 10- to 50-fold.

The sequence homology between SSAP and a wide variety of RNA-binding and putative RNA-binding proteins suggested that SSAP may potentially interact with RNA. We produced two 27-nucleotide RNA molecules with the sequences found in each strand of the USE IV enhancer oligonucleotides by plac-

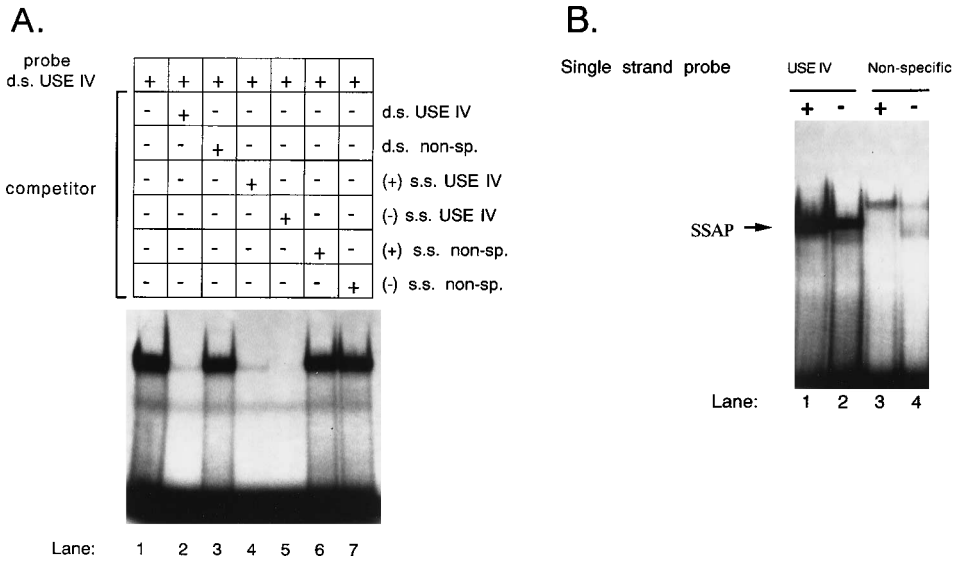

FIG. 6. bSSAP is a sequence-specific DNA-binding protein. (A) Mobility shift assays were performed with 3′-end-labeled 27-bp double-stranded (d.s.) USE IV probe and purified bSSAP protein; 250 ng of the indicated unlabeled double- and single-stranded (s.s.) competitor oligonucleotides was added prior to the addition of probe. (B) Mobility shift assays were performed with bSSAP and the indicated 5′-end-labeled single-stranded probes. non-sp., nonspecific.

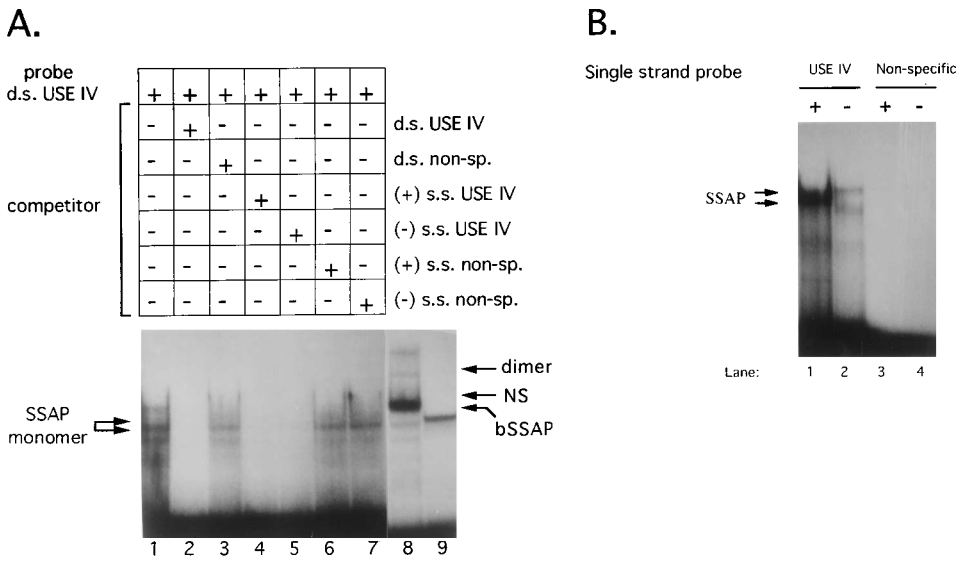

FIG. 7. Authentic SSAP is a single-stranded-DNA-binding protein. (A) Mobility shift assays were performed with 3′-end-labeled 27-bp double-stranded (d.s.) USE IV probe and purified sea urchin SSAP (12) (lanes 1 to 7); 250 ng of the indicated unlabeled double- and single-stranded (s.s.) competitor oligonucleotides was added prior to the addition of probe. Lane 8 is the same as lane 1 except that crude nuclear extract from 8-h morula-stage embryos was used. The dark band in this lane is nonspecific (non-sp.; NS), and the band above that is the dimer form of SSAP (11). Lane 9 is the same as lane 1 except that bSSAP was used as a source of protein. bSSAP migrates more slowly than native SSAP monomer. (B) Mobility shift assays were performed with purified SSAP and the indicated 5′-end-labeled single-stranded probes.

ing a T7 RNA polymerase promoter immediately adjacent to both orientations of the USE IV element. Transcription reactions using T7 RNA polymerase produced intact 27-base RNA molecules that were used as competitors in mobility shift assays using the double-stranded USE IV DNA probe (Fig. 9). While identical sequences in DNA oligonucleotides are effective competitors of bSSAP binding (Fig. 6A, lanes 4 and 5), the RNAs compete very poorly. bSSAP will bind weakly to a radioactive USE IV RNA probe (data not shown), but it is clear that bSSAP prefers DNA as a substrate for binding.

SSAP is able to transactivate target genes in vivo. The biological activity of SSAP in vivo was examined by using a system which allows the introduction of exogenous material into fertilized sea urchin eggs. This technique involves microinjection of nucleic acids into eggs or one-cell zygotes and has been used successfully for the expression of cloned DNAs as well as RNA (6, 35). To test the ability of SSAP to transactivate a target gene, pGC364 was used as the reporter gene construct (Fig. 10A). This plasmid contains the basal promoter of the H1-β gene (−106 to +8) subcloned directly upstream of the bacterial CAT gene (29). In addition, the nucleotide sequence corresponding to the embryonic enhancer element, USE IV, was subcloned upstream of the basal promoter, and three tandem copies of an oligonucleotide containing the USE IV sequence were positioned downstream of the CAT gene (28).

pGC364 was injected alone or coinjected with capped SSAP mRNA, and the injected embryos were collected at different developmental stages and assayed for CAT enzyme activity. When injected alone, pGC364 is expressed with a temporal pattern which is identical to that of the late H1 histone gene (27). Expression is low in 8-h (early-blastula-stage) embryos and increases significantly by the hatching blastula stage at 22 h postfertilization (Fig. 10B). Figure 10B also shows the results of the coinjection of in vitro-synthesized capped SSAP mRNA (0.6 pg) with pGC364 (0.1 pg). These concentrations of mRNA and target reporter appear to be saturating, as determined from a dose-response experiment (data not shown). At the early blastula stage, 8 h postfertilization, when the enhancer is inactive, introduction of exogenous SSAP mRNA did not transactivate pGC364. However, at all subsequent stages tested, when the enhancer is functional and the late histone genes are maximally activated, there was a three- to fourfold transactivation of the pGC364 reporter gene (Fig. 10B). Therefore, although overexpressing SSAP causes the transactivation of the target reporter gene through the USE IV sequence, it can do so only at a specific time during development. These results indicate that the translated products of the injected SSAP RNAs are subject to the same temporal constraints that act upon the endogenous transcription factor.

A number of control experiments were performed to demonstrate that transactivation involves transcriptional regulation by SSAP acting on the target *cis*-acting DNA-binding sequence. We tested an SSAP expression vector, which includes a 1.8-kb insert of the SSAP cDNA downstream of the sea urchin blastula-stage promoter ΔSpec-1 (17). When coinjected with pGC364, an approximate fivefold transactivation of CAT activity was seen at 22 h postfertilization (data not shown). To further test the hypothesis that transactivation by SSAP can be mediated only through a functional enhancer, we cloned the three tandem wild-type SSAP-binding sites and three mutant SSAP-binding sites downstream of the CAT gene of pGC355 (constructs shown in Fig. 11A) and tested these constructs for the ability to be transactivated by synthetic SSAP mRNA. The wild-type enhancer construct showed about a fourfold increase in its level of activation, but the mutant enhancer element failed to respond to coinjected SSAP mRNA (Fig. 11B). In numerous experiments, we routinely observe three- to eightfold transactivation of reporter constructs with functional SSAP-binding sites, but we have never observed transactivation of reporter genes lacking a functional enhancer.

Constructs pGC296 and pGC296A (Fig. 11A), which contain the early H1 promoter upstream of the CAT gene, were also microinjected to rule out the possibility that SSAP acts posttranscriptionally to stabilize primary transcripts containing

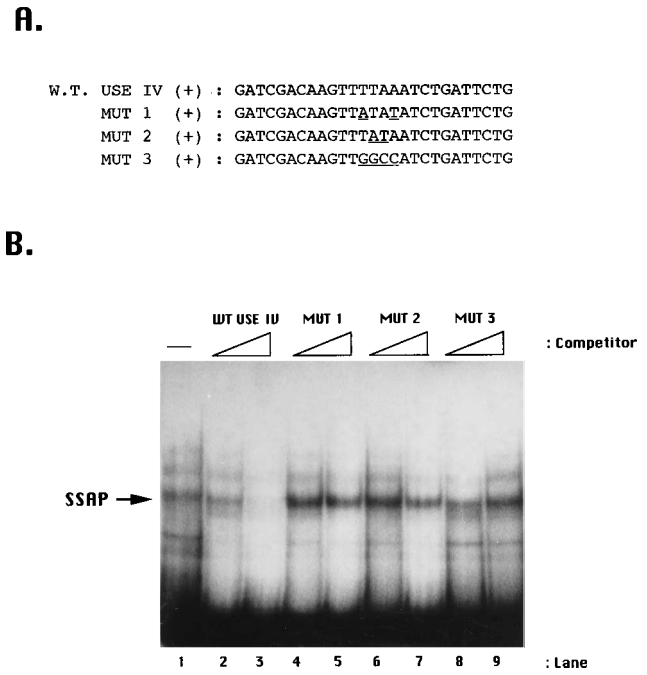

FIG. 8. SSAP binds to the coding strand of USE IV in a sequence-specific manner. (A) Nucleotide sequences of wild-type (W.T.) and mutant (MUT) USE IV coding-strand oligomers. Underlined letters indicate base substitutions. (B) SSAP binding competition between wild-type USE IV probe and mutant USE IV competitor oligomers. Shown is an autoradiogram of an electrophetic mobility shift assay using purified authentic SSAP monomer and end-labeled wild-type USE IV coding-strand oligomer as a probe. The lanes corresponding to reaction mixtures containing 25- and 100-fold excess levels of single-stranded competitor oligomers are indicated above the autoradiogram. The arrow indicates the position of SSAP-probe complex.

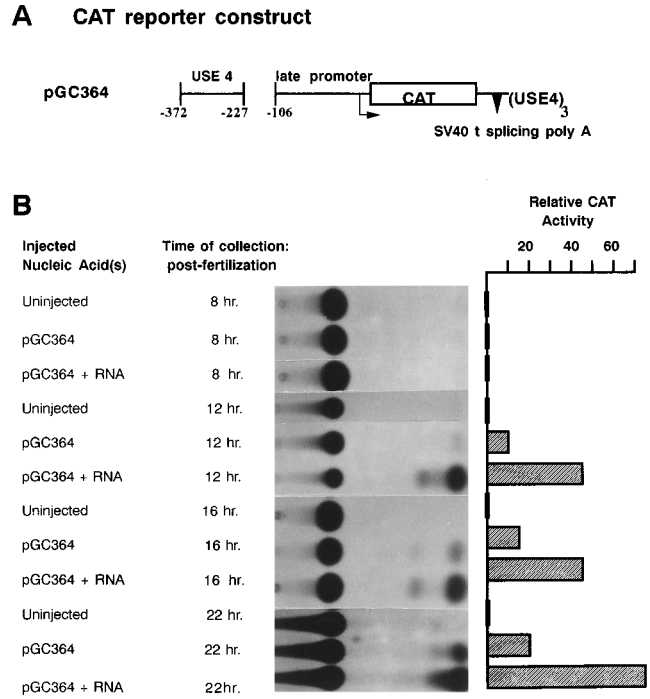

FIG. 10. Developmental time course of transactivation of target genes by overexpression of SSAP in vivo. (A) Schematic representation of linearized microinjected DNA. pGC364 is isogenic with pGC355 (see Fig. 11) except for a 145-bp partial enhancer containing sequences from −227 to −372 placed upstream of the basal promoter (this fragment has two SSAP-binding sites) as well as three tandem copies of the 27-bp USE IV oligonucleotide cloned in the *Bam*HI site downstream of the CAT gene. SV40, simian virus 40. (B) Fertilized eggs were injected with pGC364 alone or coinjected with capped synthetic SSAP mRNA. Injected embryos were collected at the indicated developmental stage, and CAT assays were performed as described in the text. Bar graphs represent levels of activation above that of uninjected control embryos.

the USE IV sequence. The late H1 enhancer does not function in combination with the early H1 promoter (Fig. 11C and reference 42a), but primary transcripts of pGC296A will contain the USE IV sequence at their 3′ ends. Transcripts derived from these early H1 promoter constructs are identical to those of the late H1 promoter-driven reporter genes except for a few nucleotides in the 5′ untranslated region. Injected embryos were collected 12 h postfertilization, a developmental stage when there is overlap in the activities of the early promoter and the late gene enhancer (24). If SSAP was an RNA-binding protein, it should have been able to act on constructs pGC296 and pGC296A in a manner which would produce effects similar to those seen with both pGC364 and pGC355A. The inability of SSAP to transactivate pGC296A (Fig. 11C) not only demonstrates the specificity of the USE IV enhancer in the context of its own promoter but adds supportive evidence for the role of the SSAP protein as an authentic transcriptional activator rather than a stabilizer of RNA molecules.

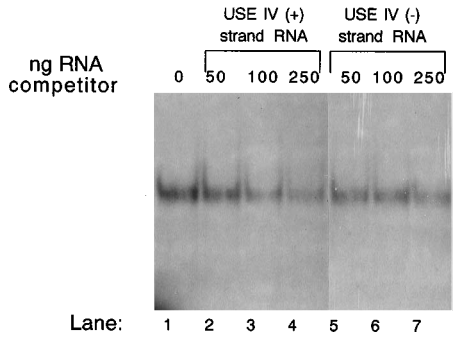

FIG. 9. Electrophoretic mobility shift assays using bSSAP and 50, 100, and 250 ng of RNA oligomers as competitors. The probe is a 3′-end-labeled 27-bp double-stranded USE IV oligonucleotide. The competitor RNAs are 27-bp transcripts representing either the coding or the noncoding strand of the SSAP-binding site (USE IV). Single-stranded RNA competitors were synthesized by in vitro transcription of the double-stranded oligonucleotide template as described in Materials and Methods.

DISCUSSION

To elucidate the mechanisms which regulate temporal gene expression during embryogenesis, we have isolated and begun to analyze the structure and function of the cDNA encoding SSAP. This protein is the only sequence-specific late histone H1 enhancer DNA-binding activity found in sea urchin embryo nuclei (11). In addition, multimers of the binding site for this protein, USE IV, are capable of functioning as a stage-specific embryonic enhancer. The cDNA clones which were isolated by immunological cross-reactivity to the native sea urchin protein were reconfirmed by epitope selection. Surprisingly, SSAP contains two 90-amino-acid repeats which show striking similarity to a number of proteins which bind RNA and single-stranded DNA (14). Furthermore, a 6- and an 8-amino-acid peptide sequence within each 90-amino-acid repeat show homology to the conserved sequences designated RNP-1 and RNP-2, which constitute part of the RRM (Fig. 4). These

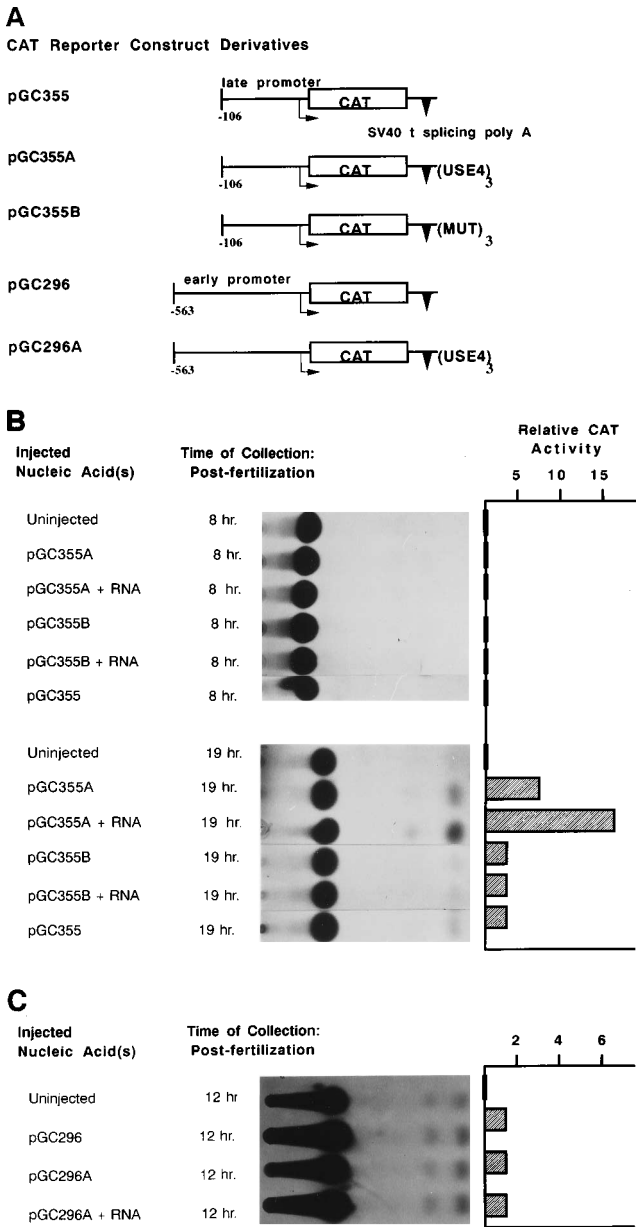

FIG. 11. In vivo transactivation of target genes. (A) Schematic representation of linearized microinjected DNA constructs. pGC355 is the control reporter construct driven by the basal promoter (−106 to +8) of the late H1 gene. pGC355A and pGC355B are isogenic with pGC355 except for three tandem copies of the 27-bp USE IV oligonucleotide or three tandem copies of mutant 3 USE IV oligonucleotide cloned in the *Bam*HI site downstream of the CAT gene. pGC296A is isogenic with pGC296 except for three tandem copies of the 27-bp USE IV oligonucleotide cloned in the *Bam*HI site downstream of the CAT gene. SV40, simian virus 40. (B) Fertilized eggs were injected with the indicated DNA construct alone or coinjected with capped synthetic SSAP mRNA. Injected embryos were collected at the indicated developmental stage, and CAT assays were performed as described in the text. Bar graphs represent levels of activation above that of uninjected control embryos. (C) pGC296, pGC296A, and pGC296A plus SSAP mRNA were injected as additional controls, and embryos were collected 12 h postfertilization. All injections from panels B and C were performed on sibling zygotes, and expression levels from each of the individual constructs can therefore be directly compared.

peptide sequences have been demonstrated to be absolutely required for the binding of these proteins to nucleic acids (36, 46). SSAP, therefore, represents a class of transcription factor in that it is able to specifically bind to alternate conformations of DNA. This property was demonstrated by the formation of sequence-specific protein-DNA complexes with single-stranded DNA probes in the mobility shift assay. SSAP is capable of binding to both the coding and noncoding strands; however, it has a higher affinity for the coding strand (Fig. 6 and 7). This property may seem somewhat troubling until one compares the nucleotide sequence of the USE IV enhancer element that purified SSAP protects in a DNase I footprinting experiment (11). The USE IV coding (5′-AGTTTTAAATCT-3′) and noncoding (5′-AGATTTAAAAGT-3′) strands contain a 9-of-12-base homology stretch. This imperfect palindrome, however, cannot form a stable double-stranded structure within each individual oligonucleotide strand. In addition to the specific binding to single-stranded oligonucleotide probes, purified SSAP is capable of footprinting a large 160-bp double-stranded DNA probe containing the enhancer-binding site (11).

The SSAP polypeptide can be divided into three separate domains with different structural characteristics (Fig. 5). Deletion analysis revealed that the DNA-binding domain is localized to the RRM repeats. In fact, only one repeat is necessary for the accurate and efficient binding of SSAP to the USE IV recognition sequence. The RRM was initially identified in the yeast poly(A)-binding protein, mammalian heteronuclear RNP proteins, and a number of proteins involved in RNA splicing (2, 14, 41). Unlike transformer, doublesex, alternative splicing factor, and other RRM-containing proteins which affect RNA splicing, SSAP does not contain a serine-arginine-rich region. These regions not only are Ser-Arg rich but also contain high proportions of Ser-Arg and Arg-Ser dipeptides (32). Instead of this serine-arginine-rich region, SSAP contains a central domain that is highly enriched in glutamine and glycine residues. The third domain of SSAP is enriched in serine, threonine, and basic amino acids. Recent results of studies of expression of truncated SSAP proteins in sea urchins and GAL4-SSAP fusions in mammalian cells demonstrated that these domains together constitute an extremely potent transcription activation domain (unpublished data). Apparently, in sea urchins this activation domain functions only beginning at the mid-blastula stage of development when SSAP dimers accumulate (11). Overexpression of SSAP monomer at earlier stages of development cannot ectopically transactivate reporter constructs (Fig. 10). Injected mRNAs are globally distributed in the early sea urchin embryo (2a), and they are effectively translated immediately after fertilization (2a, 6). Either this domain requires stage-specific posttranslational modification to function or it must dimerize to form a functional transcriptional activation domain. SSAP is phosphorylated beginning at the blastula stage when the protein dimerizes and the enhancer begins to function (10a). Importantly, overexpression of SSAP effectively transactivates reporter constructs beginning at the blastula stage of development, and control experiments ruled out posttranscriptional models for SSAP function (Fig. 10 and 11). These experiments provide strong support for SSAP functioning as a transcriptional activator.

Recently, there have been several reports of RRM-containing proteins whose function is unknown but that contain additional glutamine-rich domains. In light of the evidence that SSAP is a transcriptional regulator, we suggest that it is equally plausible that these proteins are involved in transcription. The *Drosophila couch potato* gene, which is necessary for the differentiation of embryonic peripheral nervous system cells, contains both an RRM and glutamine-rich domains (1). Additional examples of potential members of the RRM transcription factor family are the Ewing's sarcoma gene (EWS) (12) and the myxoid liposarcoma gene (TLS) (9). The EWS gene, located on chromosome 22, also contains a glu-

tamine-rich and serine/threonine-rich region as well as an RRM domain. Transformed cells from Ewing's sarcoma and other neuroectodermal tumors share a specific reciprocal translocation between chromosomes 11 and 22. This translocation creates a fusion protein in which the RRM domain of the EWS protein is replaced by an *ets* DNA-binding motif (FLI-1). This creates a novel fusion protein between the amino-terminal serine/threonine, glutamine-rich domain of EWS and the *ets* DNA-binding domain of FLI-1. TLS is an RRM protein that is translocated in liposarcomas, generating a fusion protein with a member of the c/EBP family, CHOP. We favor the hypothesis that the novel fusion proteins are able to ectopically activate *ets*- or CHOP-targeted gene transcription via the EWS or CHOP serine/threonine-, glutamine-rich transcriptional activation domains. This interpretation is supported by the recent observation that the EWS/FLI-1 fusion protein is a more potent transcriptional activator than FLI-1 alone (34). Thus, RRM-containing proteins seem to play a fundamental role in development and represent another important class of DNA-binding motifs. Consistent with this idea, we have shown that SSAP is conserved throughout evolution and that a human homolog both footprints the USE IV enhancer and is immunologically related to the sea urchin protein (10a).

A number of single-stranded DNA-binding proteins that utilize different DNA-binding motifs have been shown to be involved in transcriptional regulation. The prokaryotic cold shock proteins have an antiparallel β-sheet structural motif as their nucleic acid-binding motif (47). The proteins contain the RNP-1 motif (Fig. 4), and the structure resembles the RRM antiparallel β-sheet motif. Members of the cold shock protein family include the *E. coli* nucleoid protein H-NS, which binds specifically to single-stranded DNA containing a CCAAT sequence and stimulates transcription from promoters containing this binding site (31). The *Xenopus* Y-box transcription factor is a member of this family of single-stranded DNA-binding proteins (53). YB-1 binds to the CCAAT sequence and stimulates transcription when assayed in vitro (53) or with a transactivation assay (42). It has been suggested that the Y-box-binding proteins mediate transcriptional activation by alterations in the structure of the promoter DNA (42). The Y-box-binding proteins contain no obvious transcription activation domain, and they bind to sequences immediately upstream of the transcription start site. Bending or unwinding of the CCAAT sequence near the RNA start site could provide a more efficient substrate for general transcription factors and RNA polymerase II. SSAP, on the other hand, binds to an enhancer element about 300 bp upstream of the histone H1 coding sequence (11, 28). While it is possible that SSAP functions as a transcriptional activator by altering DNA structure, it has a potent transcription activation domain (unpublished data). The human c-*myc* promoter and far-upstream elements are activated by two different single-stranded DNA-binding proteins. Heterogeneous nuclear RNP protein K binds to the pyrimidine-rich element strand of the CT element of the c-*myc* promoter in a sequence-specific manner and it is capable of transactivating wild-type but not CT mutant promoter constructs (54). Similarly, the far-upstream element of the c-*myc* gene is activated by another sequence-specific single-stranded-DNA-binding protein, FBP (15).

With the use of S1 nuclease and bromoacetaldehyde, Larsen and Weintraub were able to demonstrate alternate DNA conformations at specific locations within the chromatin (30). S1 nuclease-sensitive sequences were located in the 5′ region of genes, and in the specific case of the β-globin gene, the cleavages were detected only in tissues actively transcribing those genes. Alternate conformations of DNA can arise from different nucleotide sequences (for a review, see reference 4); for example, palindromic sequences can rearrange to form cruciform structures. Dinucleotide repeats (19) and stretches of deoxyguanosine (37) can be sensitive to S1 nuclease digestion. In addition, sequences with a low denaturation temperature may exhibit some single-strand character. It is interesting that the USE IV sequence is surrounded by many ATTTTA sequences which are thought to be involved in bending the DNA double helix (25). It is possible that this bending produces an increase in torsional strain which can be relieved by small localized regions of DNA melting. Thus, there is ample evidence for regions of single-stranded DNA in chromatin that may very well be *cis*-acting regulatory sequences.

It is unclear whether SSAP recognizes its binding site in a double-stranded form and then induces a conformational change or recognizes an already preexisting non-B form of DNA. During sea urchin development, there is a switch in the histone protein variants that are used for nucleosome formation (5). The internucleosomal repeat increases from a distance of 200 nucleotides in cleavage-stage embryos to an average of between 210 and 215 nucleotides during the blastula stage of development, and nucleosome core particles isolated from late-stage embryos are more resistant to DNase cleavage and dissociation by salt (49). The change in histone subtypes correlates with enhancer activity and may create functionally different classes of nucleosomes that mediate a conformational change in the enhancer from a B to a non-B form of DNA. We are currently attempting to distinguish between these alternative models of SSAP activity.

ACKNOWLEDGMENTS

This work was supported by NIH grant GM30333. D.J.D., J.D., and L.R. are supported in part by NIH training grants T32 GM07288, T32 GM07128, and T32 GM07491 from NIGMS.

We gratefully acknowledge the extremely useful review of this manuscript by Eric Davidson. We also thank Eric Davidson for the cDNA expression library.

REFERENCES

1. **Bellen, H. J., S. Kooyer, D. D'Evelyn, and J. Pearlman.** 1992. The *Drosophila* couch potato protein is expressed in nuclei of peripheral neuronal precursors and shows homology to RNA-binding proteins. Genes Dev. **6:**2125–2136.
2. **Boggs, R. T., P. Gregor, S. Idriss, J. M. Belote, and M. McKeown.** 1987. Regulation of sexual differentiation in *D. melanogaster* via alternative splicing of RNA from the *transformer* gene. Cell **50:**739–747.

2a. **Cameron, R. A., L. C. Smith, R. J. Britten, and E. H. Davidson.** 1994. Ligand-dependent stimulation of introduced mammalian brain receptors alters spicule symmetry and other morphogenic events in sea urchin embryos. Mech. Dev. **45:**31–47.

3. **Cantor, C. R.** 1981. DNA choreography. Cell **25:**293–295.
4. **Childs, G., C. Nocente-McGrath, T. Leiber, C. Holt, and J. Knowles.** 1982. Sea urchin (*Lytechinus pictus*) late-stage histone *H3* and *H4* genes: characterization and mapping of a clustered but nontandemly linked multigene family. Cell **31:**383–393.
5. **Cohen, L. H., K. M. Newrock, and A. Zweidler.** 1975. Stage-specific switches in histone synthesis during embryogenesis of the sea urchin. Science **190:** 994–997.
6. **Colin, A. M.** 1986. Rapid and repetitive microinjection. Methods Cell Biol. **27:**395–406.
7. **Courey, A. J., D. A. Holtzman, S. R. Jackson, and R. Tjian.** 1989. Synergistic activation by the glutamine-rich domains of human transcription factor Sp1. Cell **59:**827–836.
8. **Courey, A. J., and R. Tjian.** 1988. Analysis of Sp1 in vivo reveals multiple transcriptional domains, including a novel glutamine-rich activation motif. Cell **55:**887–898.
9. **Crozart, A., P. Aman, N. Mandahl, and D. Ron.** 1993. Fusion of CHOP to a novel RNA-binding protein in human myxoid liposarcoma. Nature (London) **363:**640–644.
10. **Davidson, E. H.** 1986. Gene activity in early development, 3rd ed. Academic Press, Orlando, Fla.

10a. **DeAngelo, D. J.** 1993. Ph.D. thesis. Albert Einstein College of Medicine, Bronx, N.Y.

11. **DeAngelo, D. J., J. DeFalco, and G. Childs.** 1993. Purification and charac-

terization of the stage-specific embryonic enhancer-binding protein SSAP-1. Mol. Cell. Biol. **13:**1746–1758.
12. **Delattre, O., J. Zucman, B. Plougastel, C. Desmaze, T. Melot, M. Peter, H. Kovar, I. Joubert, P. Jong, G. Rouleau, A. Aurias, and G. Thomas.** 1992. Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. Nature (London) **359:**162–165.
13. **DiLiberto, M., Z.-C. Lai, H. Fei, and G. Childs.** 1989. Development control of promoter-specific factors responsible for the embryonic activation and inactivation of the sea urchin early histone H3 gene. Genes Dev. **3:**973–985.
14. **Dreyfuss, G., M. J. Matunis, S. P. Roma, and C. G. Burd.** 1993. hnRNP proteins and the biogenesis of mRNA. Annu. Rev. Biochem. **62:**289–321.
15. **Duncan, R., L. Bazar, G. Michelotti, T. Tomonaga, H. Krutzsch, M. Avigan, and D. Levens.** 1994. A sequence-specific single-strand binding protein activates the far upstream element of c-*myc* and defines a new DNA-binding motif. Genes Dev. **8:**465–480.
16. **Gaillard, C., and F. Strauss.** 1990. Sequence-specific single-strand-binding protein for the simian virus 40 early promoter stimulates transcription in vitro. J. Mol. Biol. **215:**245–255.
17. **Gan, L., G. M. Wessel, and W. H. Klein.** 1990. Regulatory elements from the related Spec genes of *Strongylocentrotus purpuratus* yield different spatial patterns with a lac Z reporter gene. Dev. Biol. **142:**346–359.
18. **Grange, T., C. Martin de Sa, J. Oddos, and R. Pictet.** 1987. Human mRNA polyadenylate binding protein: evolutionary conservation of a nucleic acid binding motif. Nucleic Acids Res. **15:**4771–4787.
19. **Hentschel, C. C.** 1982. Homopolymer sequences in the spacer of a sea urchin histone gene repeat are sensitive to S1 nuclease. Nature (London) **295:**714–716.
20. **Hoey, T., and M. Levine.** 1988. Divergent homeo box proteins recognize similar DNA sequences in Drosophila. Nature (London) **332:**858–861.
21. **Ito, M., J. Bell, G. Lyons, and R. Maxson.** 1988. Synthesis and turnover of late H2B histone mRNA in developing embryos of the sea urchin, *Strongylocentrotus purpuratus*. Dev. Biol. **129:**147–158.
22. **Jantzen, H.-M., A. Admon, S. P. Bell, and R. Tjian.** 1990. Nucleolar transcription factor hUBF contains a DNA-binding motif with homology to HMG proteins. Nature (London) **344:**830–836.
22a. **Kenan, D. J., C. C. Query, and J. D. Keene.** 1991. RNA recognition: towards identifying determinants of specificity. Trends Biochem. Sci. **16:**214–220.
23. **Knowles, J. A., and G. Childs.** 1984. Temporal expression of late histone messenger RNA in the sea urchin *Lytechinus pictus*. Proc. Natl. Acad. Sci. USA **81:**2411–2415.
24. **Knowles, J. A., and G. Childs.** 1986. Comparison of the late H1 histone genes of the sea urchins *Lytechinus pictus* and *Strongylocentrotus purpuratus*. Nucleic Acids Res. **14:**8121–8133.
25. **Knowles, J. A., Z.-C. Lai, and G. Childs.** 1987. Isolation, characterization and expression of the gene encoding the late histone subtype H1-gamma of the sea urchin *Strongylocentrotus purpuratus*. Mol. Cell. Biol. **7:**478–485.
26. **Krainer, A. R., A. Mayeda, D. Kozak, and G. Binns.** 1991. Functional expression of cloned human splicing factor SF2: homology of RNA-binding proteins, U1 70K, and *Drosophila* splicing regulators. Cell **66:**383–394.
27. **Lai, Z.-C., and G. Childs.** 1988. Characterization of the structure and transcriptional patterns of the gene encoding the late histone subtype H1-β of the sea urchin *Strongylocentrotus purpuratus*. Mol. Cell. Biol. **8:**1842–1844.
28. **Lai, Z.-C., D. J. DeAngelo, M. DiLiberto, and G. Childs.** 1989. An embryonic enhancer determines the temporal activation of a sea urchin late H1 gene. Mol. Cell. Biol. **9:**2315–2321.
29. **Lai, Z.-C., R. Maxson, and G. Childs.** 1988. Both basal and ontogenic promoter elements affect the timing and the level of expression of a sea urchin H1 gene during early embryogenesis. Genes Dev. **2:**173–183.
30. **Larsen, A., and H. Weintraub.** 1982. An altered DNA conformation detection by S1 nuclease occurs at specific regions in active chick globin chromatin. Cell **29:**609–622.
31. **La Teana, A., A. Brandi, M. Falconi, R. Spurio, C. Pon, and C. O. Gualerzi.** 1991. Identification of a cold shock transcriptional enhancer of the *Escherichia coli* gene encoding nucleoid protein H-NS. Proc. Natl. Acad. Sci. USA **88:**10907–10911.
32. **Mancebo, R., P. C. H. Lo, and S. M. Mount.** 1990. Structure and expression of the *Drosophila melanogaster* gene for the U1 small nuclear ribonucleoprotein particle 70K protein. Mol. Cell. Biol. **10:**2492–2502.
33. **Maxson, R. E., T. J. Mohon, R. Cohn, and L. Kedes.** 1983. Expression and organization of histone genes. Annu. Rev. Genet. **17:**239–277.
34. **May, W. A., S. L. Lessnick, B. S. Braun, M. Klemsz, B. C. Lewis, L. B. Lunsford, R. Hromas, and C. T. Denny.** 1993. The Ewing's sarcoma *EWS*/*FLI-1* fusion gene encodes a more potent transcriptional activator and is a more powerful transforming gene than *FLI-1*. Mol. Cell. Biol. **13:**7393–7398.
35. **McMahon, A. P., T. J. Novak, R. J. Britten, and E. H. Davidson.** 1984. Inducible expression of a cloned heat shock fusion gene in sea urchin embryos. Proc. Natl. Acad. Sci. USA **81:**7490–7494.
36. **Merrill, B. M., K. L. Stone, F. Cobianchi, S. H. Wilson, and K. R. Williams.** 1988. Phenylalanines that are conserved among several RNA-binding proteins form part of a nucleic acid-binding pocket in the A1 heterogeneous nuclear ribonucleoprotein. J. Biol. Chem. **263:**3307–3313.
37. **Nickol, J. M., and G. Felsenfeld.** 1983. DNA conformation at the 5′ end of the chicken adult b-globin gene. Cell **35:**467–477.
38. **Nordheim, A., and A. Rich.** 1983. Negatively supercoiled simian virus 40 DNA contains Z-DNA segments within transcriptional enhancer sequences. Nature (London) **303:**674–679.
39. **Olson, M. O., Z. M. Rivers, B. A. Thompson, W. Y. Kao, and S. T. Case.** 1983. Interaction of nucleolar phosphoprotein C23 with cloned segments of ribosomal DNA. Biochemistry **22:**3345–3351.
40. **Pabo, C. O., and R. T. Sauer.** 1992. Transcription factors: structural families and principles of DNA recognition. Annu. Rev. Biochem. **61:**1053–1095.
41. **Query, C. C., R. C. Bentley, and J. D. Keene.** 1989. A common RNA recognition motif identified within a defined U1 RNA binding domain of the 70K U1 snRNP protein. Cell **57:**89–101.
42. **Ranjan, M., S. R. Tafuri, and A. Wolffe.** 1993. Masking mRNA from translation in somatic cells. Genes Dev. **7:**1725–1736.
42a. **Rybacki, L.** Unpublished data.
43. **Sambrook, J., E. F. Fritsch, and T. Maniatis.** 1989. Molecular cloning: a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
44. **Santoro, I. M., T. M. Yi, and K. Walsh.** 1991. Identification of single-stranded-DNA binding proteins that interact with muscle gene elements. Mol. Cell. Biol. **11:**1944–1953.
45. **Sarkar, G., and S. S. Sommer.** 1989. Access to a messenger RNA sequence or its protein product is not limited by tissue or species specificity. Science **244:**331–334.
46. **Scherly, D., W. Boelens, N. A. Dathan, W. J. van Venrooij, and I. W. Mattaj.** 1990. Major determinants of the specificity of interaction between small nuclear ribonucleoproteins U1A and U2B″ and their cognate RNAs. Nature (London) **345:**502–506.
47. **Schindelin, H., M. A. Marahlel, and U. Heinemann.** 1993. Universal nucleic acid-binding domain revealed by crystal structure of the *B. subtilis* major cold shock protein. Nature (London) **364:**164–168.
48. **Schon, E., T. Evans, J. Welsh, and A. Efstratiadis.** 1983. Conformation of promoter DNA: fine mapping of S1-hypersensitive sites. Cell **35:**837–848.
49. **Simpson, R. T.** 1981. Modulation of nucleosome structure by histone subtypes in sea urchin embryos. Proc. Natl. Acad. Sci. USA **78:**6803–6807.
50. **Studier, F. W., A. H. Rosenberg, J. J. Dunn, and J. W. Dubendorff.** 1990. Use of T7 RNA polymerase to direct expression of cloned genes. Methods Enzymol. **185:**60–89.
51. **Sturm, R. A., G. Das, and W. Herr.** 1988. The ubiquitous octamer-binding protein Oct-1 contains a POU domain with a homeo box subdomain. Genes Dev. **2:**1582–1599.
52. **Tada, H., and K. Khalili.** 1992. A novel sequence-specific DNA-binding protein, LCP-1, interacts with single-stranded DNA and differentially regulates early gene expression of the human neurotropic JC virus. J. Virol. **66:**6885–6892.
53. **Tafuri, S. R., and A. P. Wolffe.** 1990. Xenopus Y-box transcription factors: molecular cloning, functional analysis and developmental regulation. Proc. Natl. Acad. Sci. USA **87:**9028–9032.
54. **Takimoto, M., T. Tomonaga, M. Matunis, M. Avigan, H. Krutzsch, G. Dryfuss, and D. Levens.** 1993. Specific binding of heterogeneous ribonucleoprotein particle protein K to the human c-*myc* promoter in vitro. J. Biol. Chem. **268:**18249–18258.
55. **Towbin, H., T. Staehelin, and J. Gordon.** 1979. Electrophoretic transfer of proteins from polyacrylamide gels to nitrocellulose sheets: procedures and some applications. Proc. Natl. Acad. Sci. USA **76:**4350–4354.
56. **Turner, R., and R. Tjian.** 1989. Leucine repeats and an adjacent DNA binding domain mediate the formation of functional *cFos-cJun* heterodimers. Science **243:**1689–1694.
57. **Weinberger, C., S. M. Hollenberg, J. M. Harmon, S. T. Brower, J. Cidlowski, E. B. Thompson, M. G. Rosenfeld, and R. M. Evans.** 1985. Identification of human glucocorticoid receptor complementary DNA clones by epitope selection. Science **228:**740–742.
58. **Wu, H., and D. M. Crothers.** 1984. The locus of sequence-directed and protein-induced DNA bending. Nature (London) **308:**509–513.
59. **Young, R. A., and R. W. Davis.** 1983. Efficient isolation of genes by using antibody probes. Proc. Natl. Acad. Sci. USA **80:**1194–1198.