U4B snRNA gene enhancer activity requires functional octamer and SPH motifs

Zulkeflie Zamrod and William E. Stumph*

Department of Chemistry and Molecular Biology Institute, San Diego State University, San Diego, CA 92182, USA

Received September 11, 1990; Revised and Accepted November 12, 1990

ABSTRACT

Expression of the chicken U4B small nuclear RNA (snRNA) gene is stimulated by a transcriptional enhancer located approximately 190–227 base pairs upstream of the transcription start site. This enhancer is composed of at least two functional motifs: an octamer (binding site for Oct-1) and an SPH motif. We now report that these two motifs functionally cooperate to stimulate U4B snRNA gene expression, and both are required for the formation of a stable transcription complex. Expression in frog oocytes of 24 different point mutant constructions indicates that the functional SPH motif is at least 15 base pairs in length. It is a recognition site for a sequence specific DNA-binding protein, termed SBF, purified from chicken embryonic nuclear extracts. The ability of the mutant SPH motif constructions to be recognized by SBF *in vitro* correlates with their transcriptional activities, suggesting that SBF mediates the stimulatory effect of the U4B SPH motif. These results are similar to our recent findings on the chicken U1 gene enhancer, which also contains adjacent binding sites for Oct-1 and SBF. These studies, together with evolutionary considerations and sequence comparisons among snRNA gene enhancers, suggest that cooperativity between octamer and SPH motifs could be a widely-employed mechanism for generating vertebrate snRNA gene enhancer activity.

INTRODUCTION

Small nuclear RNAs of the U family (U1, U2, U3, etc.) are abundant and metabolically stable RNAs present in the nuclei of eukaryotic cells. They are well conserved in metazoan evolution and range in size from approximately 60 to 200 nucleotides in length. Furthermore, these snRNAs are distinctively capped with 2,2,7-trimethylguanosine (except U6) and they are not polyadenylated (1). The U1, U2, U4, U5 and U6 snRNAs are components of small nuclear ribonucleoprotein particles (snRNPs) that are involved in splicing pre-messenger RNAs (2).

The genes encoding these snRNAs are unusual in several respects. With the exception of the U6 gene, they are RNA polymerase II transcription units, but they lack introns and many of the regulatory signals commonly associated with protein-coding genes (3, 4). For example, instead of a TATA box, there is a unique element (termed the proximal sequence element, or PSE) located 50 to 60 base pairs upstream of snRNA genes that specifies the transcription initiation site. Moreover, 3' end-formation of the snRNAs is coupled to the initiation of transcription from an snRNA gene promoter (5–7). The PSE is apparently involved in establishing a unique transcription complex capable of recognizing an snRNA-specific 3'-end-formation signal located in the 3'-flanking DNA.

Transcription of snRNAs is also regulated by an enhancer region located about 180–250 base pairs upstream of the transcription initiation site. For several years the so-called octamer sequence (ATGCAAAT and its variants) has been recognized as an important element of snRNA gene enhancers. The octamer motif is a binding site for the transcription factor Oct-1, also known as NF-A1 and OTF-1 (8). Recent studies of frog U2 and U5 RNA genes (9, 10), chicken U1 and U4 RNA genes (11–13), and human U1, U2, and U4 RNA genes (14–17) have provided convincing evidence that additional sequences upstream and/or downstream of the octamer are also involved in snRNA gene enhancer function. However, the detailed sub-structure of these enhancer regions is still not well understood.

We recently reported (12) that an octamer motif and an adjacent SphI Postoctamer Homology (SPH) motif are both essential elements of the chicken U1 gene enhancer. U1 gene constructions with an intact enhancer (containing octamer and SPH motifs) were efficiently expressed in frog oocytes when co-injected with a competing U4B snRNA gene template. In contrast, mutant U1 constructions that lacked either the octamer or the SPH motif failed to be expressed at a detectable level in the competition assay (12). Thus functional octamer and SPH motifs in the U1 enhancer cooperate or synergize in some manner to promote the formation of a stable U1 transcription complex.

Roebuck et al. (12) also demonstrated that the SPH motif is a recognition site for a novel sequence-specific DNA binding protein. This factor was partially purified from chicken embryo nuclear extracts and was termed SBF, for SPH motif Binding

* To whom correspondence should be addressed

Factor. The transcriptional activities of U1 gene templates containing mutations in the SPH motif correlated well with their relative abilities to be recognized *in vitro* by SBF (12).

A region similar in sequence to the U1 gene SPH motif was also noted just downstream of the octamer motif in the chicken U4B RNA gene enhancer (11). Similar to our findings on the U1 gene enhancer, we now report that the SPH and octamer motifs are both required for the functional activity of the U4B gene enhancer. An extensive point mutational analysis of the U4B SPH motif indicates that nucleotides essential for its function are spread over a region at least 15 base pairs in length. In addition, the same factor (SBF) that was isolated based upon its ability to bind to the U1 SPH motif (12) also binds efficiently to the wild type U4B SPH motif, but not to transcriptionally defective templates containing mutations in the SPH motif. These findings suggest that a common mechanism dependent upon both octamer and SPH motifs is involved in enhancing U1 and U4B gene expression. Evolutionary considerations suggest that synergism between octamer and SPH motifs might be a widely-employed mechanism for generating vertebrate snRNA gene enhancer activity.

MATERIALS AND METHODS

Construction of U4B Deletion Mutants

U4B plasmids used in these studies contained the chicken U4B gene (18) cloned into the pUC19 vector. The transcription templates pU4BΔ-227, pU4BΔ-208, pU4BΔ-199, and pU4BΔ-117 were generated by truncating the 5'-flanking DNA at AatII, Eco47III, SphI, and KpnI restriction sites, respectively, such that each clone retained the amount of 5'-flanking DNA indicated by its name. All U4B constructions contained 114 bp of 3'-flanking DNA (truncation at an Eco47III restriction site), except pU4BΔ-199 and pU4BΔ(-200/-194), which retained 283 bp of 3'-flanking DNA. Construction of these last two plasmids was previously described (13).

Construction of Point Mutations in the U4B Enhancer

The pseudo-wild-type template, pU4BΨWT, is similar to pU4BΔ-227 except for two substituted bases at positions -208 and -207 which together introduce a unique BglII restriction site between the octamer and SPH motifs. To construct pU4BΨWT, the synthetic oligonucleotide 5'-GCCAAGCTTCA-CACCTACTTTGCATAGatCTTCCCAGCATGCT-3' (lower case letters denote nucleotides that differ from the wild-type sequence) was made double stranded by mutually-primed synthesis (19, 20). After cutting with HindIII and SphI, the double stranded oligonucleotide was ligated into pU4BΔ-227 that had been previously prepared by digestion with HindIII and SphI (HindIII cuts pU4BΔ-227 in the vector polylinker just upstream of position -227).

Following the construction of pU4BΨWT, it was used as a parental plasmid to generate point mutations within the U4B enhancer SPH motif. This was done by replacing the DNA sequence between the BglII and NotI sites of pU4BΨWT with synthetic DNA oligonucleotides. The series A constructions were generated by annealing and cloning the following oligonucleotides: (A1) 5'-GatCgaaCCAGCATGCCTCGC-3' and 3'-cttGGTCGTACGGAGCGCCGG-5'; (A2) 5'-GatCTTCCC-AaacctaCTCGC-3' and 3'-AAGGGTttggatGAGCGCCGG-5'; (A3 and A4) 5'-GatCTTCCCAGCATGCCTCta-3' and 3'-AA-GGGTCGTACGGAGat(C/a)CGG-5'. The series B, C, D, and

E constructions were prepared using the following four degenerate oligonucleotides for mutually-primed synthesis: (B1-B6) 5'-TTCAGatCTTC(C/a)(C/a)(A/c)GCATGCCTCG-CGGCCGC-3'; (C1-C4) 5'-TTCAGatCTTCCCA(G/t)(C/a)-(A/c)TGCCTCGCGGCCGC-3'; (D1-D5) 5'-TTCAGatCTT-CCCAGCA(T/g)(G/t)(C/a)CTCGCGGCCGC-3'; (E1-E5) 5'-TTCAGatCTTCCCAGCATGC(C/a)(T/g)(C/a)GCGGCCGC-3'. After second strand synthesis, they were digested with BglII and NotI prior to cloning. Plasmid minipreps were sequenced using the dideoxy chain termination method, and clones possessing unique sequences were selected for large scale plasmid preparations. All templates were confirmed by resequencing the enhancer region prior to injection into oocytes.

Construction of the U1 competitor template

The U1 template used as a transcriptional competitor is a chimeric construction that contains the chicken U1-52a gene and its proximal regulatory region, but with its enhancer replaced by the U4B gene enhancer. It was constructed by cloning a 168 bp Asp718 fragment (containing U4B 5'-flanking sequences from position -280 to -117) into the Asp718 site of the pUC19 polylinker upstream of position -90 in the U1 5'-truncation template, pU1Δ-90 (12). In this construction, the U4B enhancer is in its normal orientation and approximately the same distance upstream as found in the wild-type U4B gene. When injected by itself into oocytes, this chimeric plasmid efficiently directs U1 RNA synthesis (21); however, in competition with a wild-type U4B template it is expressed relatively inefficiently, even though it contains the same enhancer. The most likely reason for this is that the two templates differ in the proximal regulatory region, with the proximal region of the U4B gene having a higher level of activity in oocytes. Evidence to support this comes from experiments showing that enhancerless U4B templates that contain the U4B proximal region are expressed fairly well in frog oocytes under non-competitive conditions, whereas comparable U1 constructions fail to be expressed (11, 13).

Oocyte nuclear injections and RNA analysis

Microinjection of templates into *Xenopus laevis* oocytes, isolation of RNA, and analysis of products were carried out as previously described (18). For experiments without U1 competitor, each oocyte was injected with ~30 nl solution containing U4B template DNA (400 μg/ml), a *Xenopus* 5S gene template (1 μg/ml), and [α -³²P]GTP (8 μCi/μl). When U1 competitor plasmid was co-injected, it was included at a concentration of 300 μg/ml and the concentration of U4B template was reduced to 100 μg/ml. Aliquots of total RNA were analyzed in 8% denaturing polyacrylamide gels. Following short autoradiographic exposures, the 5S RNA product was used to normalize for efficiency of injection and recovery of RNA by adjusting the sample volumes on subsequent gels.

When comparing different sets of competition injections, some variability is usually seen in the level of U1 gene expression, even when the same U4B templates are involved. This variability is related to the fact that different batches of oocytes differ in their intrinsic abilities to express the U1 and U4B genes relative to each other. Therefore, the sole role of the co-injected U1 competitor template is to increase the sensitivity of the U4B transcription assay. The level of U1 expression should not be taken as an accurate measure or reflection of the relative transcriptional strengths of the U4B templates.

Isolation and purification of SBF

Purification of SBF was carried out as previously described (12). Briefly, nuclear extract was prepared from 10-day chick embryos and subsequently subjected to heparin-agarose column chromatography. The 250 mM KCl step fraction from the heparin-agarose column was then loaded onto a DNA-affinity column prepared using concatenated oligonucleotides (22) containing the sequence of the chicken U1-52a gene SPH motif: 5'-GATCAAACGCGCGCTGCATGCCGGGAGCACCAC-3' and 3'-TTTGCGCGCGACGTACGGCCCTCGTGCTAG-5' (23). SBF was eluted in high salt, dialyzed and stored in a liquid-nitrogen freezer until used (12).

DNase I footprint analysis

DNase I footprinting experiments were performed as previously described (12), except poly(dA-dT)·poly(dA-dT) was substituted for poly(dI-dC)·poly(dI-dC) as nonspecific DNA. For footprinting the upper strand, a 168 bp Asp718 fragment (extending from position -117 to position -280 in the U4B gene 5'-flanking DNA) was labeled by filling in using Klenow enzyme and [α - 32 P]dGTP. It was then cut with AatII (position -227) and gel-purified. The fragment for the lower strand was prepared by labeling the same Asp718 fragment with T4 polynucleotide kinase and [γ - 32 P]ATP, followed by cutting with BglI (position -266) and gel purification. Binding mixtures, containing increasing amounts of SBF, were subjected to DNase I digestion and analyzed by electrophoresis in 6% sequencing gels as previously described (12).

Gel mobility shift assays

DNA fragments for gel mobility shift assays were isolated by digesting the various U4B constructions with BstEII and HindIII, which cut respectively at position -133 in the U4B 5'-flanking DNA and in the vector just upstream of position -227. In each case this released a 98 bp fragment that was labeled by filling in with Klenow enzyme, [α - 32 P]dGTP, and unlabeled dATP, dTTP and dCTP. Binding reactions with SBF were carried out as previously described (12) except poly(dA-dT)·poly(dA-dT) was used as a nonspecific competitor DNA. The samples were run in 4% nondenaturing polyacrylamide gels containing 25mM Tris, 190mM glycine, 1mM EDTA (pH 8.3) and visualized by autoradiography.

RESULTS

Two Enhancer Motifs Functionally Cooperate to Activate U4B Gene Expression

Analysis using an oocyte expression system indicated that sequences required for U4B gene enhancer activity were located approximately 190–227 bp upstream of the coding region of the chicken U4B RNA gene (13). The sequence of the enhancer region in the ‘wild-type’ template, U4BΔ-227, is shown in Fig. 1. The locations of the octamer and SPH motifs are indicated by boxes above the sequence.

To examine the separate functional contributions of the octamer and SPH motifs in the U4B enhancer, the transcriptional activities of several deletion templates were analyzed by injection into *Xenopus* oocytes (Fig. 1). A plasmid containing a 5S RNA gene was co-injected as an internal control to normalize for injection efficiency and recovery of product. The U4BΔ-117 template, which lacks the entire enhancer region, was expressed in oocytes with about a 3–4 fold reduced efficiency relative to the wild-

type template (Fig. 1-A, compare lanes 1 and 5). Templates lacking an octamer motif (U4BΔ-208) or having a 5 bp deletion within the SPH motif (U4BΔ[−200/−194]) were each expressed at approximately the same level as the enhancerless construction, suggesting that both the octamer and SPH motifs are necessary for enhancer activity.

Because U4B constructions lacking enhancer function were expressed at a relatively high basal level in oocytes (Fig. 1-A), it was desirable to make the transcription assay more sensitive to the presence or absence of a functional U4B enhancer. This was accomplished in subsequent experiments by co-injecting a U1 RNA gene into the oocytes along with the various mutant U4B constructions. Under these conditions of template competition, none of the constructions containing deletions within the U4B enhancer were detectably expressed (Fig. 1-B, lanes

Figure 1. Effects of deletions in the chicken U4B enhancer on U4B gene activity under non-competitive conditions (A) or competitive conditions (B). U4B wild-type, pseudo-wild-type, or deletion mutant templates, as indicated above each lane, were injected into *Xenopus* oocytes together with [α - 32 P]GTP. A 5S gene template was coinjected as an internal control. In the series of injections shown in panel B, the injection mixture contained in addition a chicken U1 gene template as a transcriptional competitor. After 18–24 hours, RNA was isolated from the oocytes, and analyzed in 8% denaturing polyacrylamide gels, followed by autoradiography. The bands corresponding to the U4B, U1, and 5S transcription products are indicated beside the panels. The bottom part of the figure shows the nucleotide sequence of the U4B enhancer region and the extent of the deletion mutation in each template. Solid lines represent vector DNA sequences. Dashed lines indicate internally deleted nucleotides. Lower case letters in the U4B Ψ WT sequence denote base substitutions that create a unique BglIII restriction site. Naturally occurring SphI and NotI restriction sites are also indicated. The positions of the octamer and SPH motifs are indicated by boxes. For comparison, the sequence of the chicken U1 gene SPH motif (12) is shown above the U4B sequence but inverted relative to its natural orientation in the U1 gene. Lower case letters represent nucleotides in the U1 sequence that differ from those in the U4B SPH motif.

1–4). Importantly, neither the octamer motif nor the SPH motif alone was capable of activating U4B gene expression. However, a construction (U4B Ψ WT) with a two-base substitution between the octamer and SPH motifs was an active template (Fig. 1-B,

lane 5). These results indicate that the two motifs are both required for activity in the competition assay, and therefore they functionally cooperate to promote the formation of a stable transcription complex on the U4B template.

Figure 2. Footprinting analysis of SBF binding to the U4B enhancer region. A DNA fragment end-labeled on either the upper or the lower strand was incubated with increasing amounts of affinity-purified SBF (given in μ l above the middle five lanes in each panel) and then partially digested with DNase I prior to loading on a 6% denaturing polyacrylamide gel. C+T and A+G chemical sequencing reactions were loaded as markers in the outside lanes of each panel. The locations of the octamer and the SPH motifs are indicated by the boxes alongside the autoradiograms. The region protected from DNase I digestion, which corresponds to the SPH motif, is delineated by a bracket alongside each panel. Below the autoradiograms, the U4B enhancer sequence is shown and the protected region on each strand is indicated by striped bars. Bars that are farther apart represent partial protection or ambiguity due to lack of DNase I digestion of unbound DNA. The centrally located SphI restriction site and the octamer motif are shown in boldface letters. For comparison, the DNase I protection pattern of SBF on the chicken U1 gene enhancer is also presented (12).

SPH Motifs in the U1 and U4B Enhancers Are Recognized by the Same DNA-Binding Protein

The functionally defined U1 and U4B SPH motifs share extensive sequence similarity (14 identical residues over an 18 base pair region; see sequence comparison in Fig. 1). We recently reported the identification of a novel factor, termed SBF, that binds sequence-specifically to the SPH motif of the U1 enhancer (12). To determine whether this same factor also binds to the U4B SPH motif, SBF was purified by affinity chromatography on a sequence-specific affinity resin prepared using the DNA sequence of the U1 SPH motif (12). The affinity-purified SBF was then used in a footprinting assay with a DNA fragment containing the U4B enhancer region. The upper part of Fig. 2 shows the pattern of protection from DNase I digestion for both the non-template (upper) and template (lower) strands. The region of protection is summarized on the nucleotide sequence shown in the lower part of the figure. For purposes of comparison, the pattern of protection on the U1 enhancer is also depicted (12). In both instances, the SBF footprint is nearly symmetrical about the centrally-located SphI restriction site. Within the limits of the DNase I footprinting assay, the patterns of protection are essentially identical on the U1 and U4B SPH motifs. Electrophoretic mobility shift assays also indicated that the functionally-defined SPH motifs of the U1 and U4B enhancers compete for binding to the same factor, SBF (data not shown).

Point Mutations Indicate that the U4B SPH Motif is Minimally 15 Base Pairs in Length

To identify nucleotides important for function, we subjected the U4B SPH motif to an extensive point mutational analysis. An 18 bp region in the U4B enhancer was chosen for study based upon: 1) sequence homology to the U1 SPH motif; 2) the results of the earlier expression assays (Fig. 1); and 3) the results of the DNase I footprinting experiments (Fig. 2). Specifically, all of the nucleotides from positions -205 to -188 (TTCCC-AGCATGCCTCGCG) were altered either individually or in various combinations by cloning synthetic oligonucleotides between the BglII and NotI sites that flank the SPH motif in the U4B Ψ WT plasmid. To facilitate discussion, nucleotide positions within the region under study will be referred to as positions -9 to $+9$, as shown above the U4B Ψ WT sequence in Fig. 3-F. Five different series of mutations were generated. In the initial set of constructions (series A), the first three, central six, or last three nucleotides in the 18 bp region were altered. Constructions comprising series B, C, D, and E represent mutations at positions -6 to -4 , -3 to -1 , $+1$ to $+3$, and $+4$ to $+6$ respectively (Fig. 3-F).

A total of 24 different point mutants were obtained and assayed for activity by injection into *Xenopus* oocytes. A U1 gene was co-injected as a transcriptional competitor. The results of these expression assays are shown in Fig. 3. Mutation of three bases at the 5' end of the region under study (positions -9 to -7 , template A1) had no apparent effect on activity in the oocyte assay (Fig. 3-A, compare lanes 2 and 6). In contrast, the substitution of six base pairs at the center of the SPH motif (positions -3 to $+3$) led to a complete loss of activity (template A2). The construction (A3) with base substitutions at positions $+7$ and $+8$ was expressed at a reduced but still significant level, whereas transcription from the triple point mutant (A4, mutated at positions $+7$ to $+9$) was even further reduced. These results indicate that nucleotides extending to position $+9$, as well as nucleotides within the SphI restriction site, are important for enhancer activity.

The results of injecting the series B mutants are shown in Fig. 3-B. Nucleotides at positions -6 and -5 are clearly important for SPH motif activity, since the introduction of a single point mutation at either of these positions drastically reduced U4B gene expression (templates B1 and B6). Moreover, transcription of the corresponding double point mutant (B4) was nearly undetectable. (Note that the autoradiogram in Fig. 3-B is overexposed relative to the other panels in the figure, which is evident since the band arising from the Ψ WT template is darker in panel B than in the other panels). The functional importance of the nucleotide at position -4 remained ambiguous from the available data, since a solo point mutation at that position was not obtained. Mutations at that position seemed to have minor variable effects depending upon the other point mutations introduced.

All combinations of mutations at positions -3 to -1 (the series C mutants) had approximately equal effects on expression (Fig. 3-C). In each case, expression was reduced about 2 fold from the pseudo-wild-type level. The series D templates, which contain mutations at positions $+1$ to $+3$, fell into two categories. Two of the templates (D2 and D4) were expressed as well as U4B Ψ WT, whereas three others were expressed at about a 2-fold reduced level. The combined results from the expression assays using the series C and D templates indicate that the SPH motif is surprisingly tolerant of mutations within the SphI restriction site, since even triple mutations in either half of the site caused only about a 2-fold reduction in activity. On the other hand, these nucleotides are clearly important for function since the simultaneous mutation at all six positions (template A2) totally obliterated template activity. However, the exact sequence requirements are apparently not very rigid since many of the bases can be altered individually and in various combinations with only a marginal loss of activity. Apparently, a fairly large number of point mutations in this region is required to produce a pronounced effect.

Fig. 3-E shows the results of the expression assays using the series E constructions (mutations at positions $+4$ to $+6$). All of these templates were expressed at least as well as the parental pseudo wild-type construction. This is not surprising since the alterations at positions $+4$ and $+5$ fortuitously changed those bases to match those found at the respective positions in the fully-functional U1 gene SPH motif (see Fig. 1). The data further indicate that the base at position $+6$ can be altered with no detectable loss of SPH motif activity.

The relative transcriptional activities of the various mutant constructions are summarized in the column at the right of Fig. 3-F. Based upon these data, nucleotides at positions -6 , -5 , $+7$ and/or $+8$, and $+9$ play crucial roles in the activity of the SPH motif. Nucleotides within the central SphI restriction site (-3 to $+3$) are also necessary for activity, although individual mutations and certain combinations of mutations in this region are clearly tolerated. In the transient oocyte expression assay, no effect on expression was observed resulting from mutations at positions -9 to -7 or $+4$ to $+6$.

Transcriptional Activity Correlates with the Ability of Mutant Templates to Bind SBF

The footprinting experiments described earlier (Fig. 2) indicated that the U4B SPH motif is a recognition site for a sequence specific DNA-binding protein termed SBF (12). If the enhancer activity of the SPH motif is mediated by SBF, templates that exhibit a reduced transcriptional activity due to mutations in the SPH motif would be expected to have a reduced binding affinity

Figure 3. Mapping nucleotides in the SPH motif required for chicken U4B enhancer activity in *Xenopus* oocytes. U4B templates containing point mutations in the SPH motif were injected into oocytes together with a U1 competitor template, [α - 32 P]GTP, and a 5S gene internal control plasmid. Total RNA was isolated and analyzed as described in the legend to Fig. 1. The U4B templates injected are indicated above each lane. Bands representing the U1, U4 and 5S transcripts are also indicated. (A)–(E) Transcription results arising from injection of the series A through series E mutant templates respectively. Each panel represents a separate set of injections. (F) A table showing the mutant templates assayed and summary of their relative transcriptional activities. Sequences within the SPH motif of each template are shown. Dots indicate nucleotides identical to those in the U4BΨWT sequence, and lower case letters indicate nucleotide substitutions. The 18 bp region subjected to point mutational analysis is overlined; positions within this region are identified as –9 to +9. Relative expression levels, estimated by visual inspection, are summarized in the column at the right.

for SBF. Therefore, electrophoretic mobility shift assays were carried out using affinity-purified SBF and a representative sample of labeled DNA fragments containing mutations in the SPH motif. The results are shown in Fig. 4.

DNA fragments derived from the pseudo-wild-type template and most of the mutant templates formed complexes that migrated with a slower mobility than the free DNA. Two retarded complexes were predominant (c1 and c2). This pattern is exactly

like that observed for complexes formed between SBF and DNA fragments that contain the U1 gene SPH motif (12).

Significantly, there was a good correlation between transcription levels and the abilities of the various templates to be recognized by SBF. For example, the strength of the complexes formed by the four series A mutants reflected their relative transcriptional levels (A2 < A4 < A3 < A1 ≈ ΨWT). The series C, D, and E constructions were all transcribed fairly

Figure 4. *In vitro* binding of SBF to DNA fragments containing point mutations in the U4B SPH motif. DNA fragments from selected clones were end-labeled with ^{32}P , incubated with affinity-purified SBF, and electrophoresed in nondenaturing polyacrylamide gels. All DNA fragments were identical except for the point mutations in the SPH motif shown for each construction in Figure 3-F. A dash above a lane indicates that no SBF was added to the reaction, and the gradient triangles indicate increasing amounts of SBF. In the upper left panel, 2.5 and 5.0 μl of SBF were employed. In the remaining panels, 3.5 and 7.0 μl of a different preparation of SBF was used. DNA-protein complexes are indicated as c2 and c1, and the letter F indicates the mobility of the free probes.

efficiently in the expression assay, and enhancer fragments derived from them were good substrates for recognition by SBF. On the other hand, fragments derived from the series B templates, which were seriously impaired in transcription, bound very inefficiently to SBF. Taken together, these findings suggest that SBF has a functional role in mediating the enhancer activity of the U4B SPH motif.

DISCUSSION

Transcription of vertebrate snRNA genes is regulated by an enhancer region in the 5'-flanking DNA located 180–250 base pairs upstream of the transcription initiation site (3, 4). However, the mechanism by which these enhancers activate snRNA transcription is not yet understood. Our laboratory recently reported that adjacent (but nonoverlapping) octamer and SPH motifs in the chicken U1 enhancer functionally cooperate to activate U1 RNA gene expression (12). Under conditions of transcriptional competition, deletion of either motif resulted in the loss of detectable gene expression and enhancer activity. The results presented here extend those findings to the U4B gene enhancer.

Octamer motifs are apparently present in the enhancer regions of all vertebrate snRNA genes cloned and sequenced (although in some cases the octamers exist as sequence variants whose activities remain to be demonstrated). Since functional SPH motifs have now been demonstrated in the U1 and U4B gene enhancers, the question arises whether SPH motifs (binding sites for SBF) might also be conserved within the enhancers of other vertebrate snRNA genes. At the present time there is insufficient experimental data to answer this question. However, the enhancer regions of several other vertebrate snRNA genes contain sequences very similar to the chicken U1 and U4B SPH motifs.

		-9	-6	-3	+3	+6	+9		strand
Chicken U4B	-208	CGC	TTC	CCA	GCA	TGC	CTC	GCC	-185 NT
Chicken U1	-183	GTG	CTC	CCG	GCA	TGC	AGC	GGC	-206 T
Chicken U4X	-208	GGG	CTC	CCG	GCA	CAC	CCC	GGC	-185 NT
Human U4C	-219	GAT	TTC	CCA	GCA	TGC	CAA	GGC	-196 NT
X. laevis U5	-187	CAG	CTC	CCG	GCA	ACC	GAC	CCG	-210 T
X. laevis U2	-235	ACC	TTC	CCG	GCA	TGC	CCG	GAC	-258 T
"Consensus" SPH motif									
		---	YTC	CCR	GCR	YNC	NNN	GCG	---
Chicken U2	-195	CGC	CTC	CCG	GCA	TGC	CCC	GGC	-218 T
Human U4'B	-218	TTC	CTC	CCA	TAG	TGC	TTT	GCT	-195 NT
X. laevis U1a	-226	GCG	CTC	CCA	TAG	TTC	ATT	GCT	-249 T
Mouse U1b2	-186	GCG	GAC	CCG	GCA	GAC	CCT	CCG	-209 T
Mouse U6	-256	CTT	CTC	CCA	CAG	TTC	TCT	GGC	-279 T

Figure 5. Sequence comparison of the chicken U1 and U4B SPH motifs with putative SPH motifs from other vertebrate snRNA genes. Each of the six sequences shown at the top of the figure has been functionally implicated in snRNA gene enhancer activity. These six sequences were used to derive the consensus SPH motif presented in the middle of the figure. (If a nucleotide was present in at least five of the six sequences it was judged to form a consensus; R = purine; Y = pyrimidine.) The lower part of the figure shows sequences from the enhancer regions of five additional snRNA genes, but these have not been shown to be required for enhancer activity. Underlines indicate nucleotides that match conserved positions in the consensus sequence. Lower case letters correspond to nucleotides that do not agree with the consensus sequence. T and NT in the column at the right indicate whether the sequence shown for each gene is from the template or non-template strand. References: chicken U1: (12); chicken U4X: (18) and Botros, Zamrod, Miyake, and Stumph, in preparation; human U4C: (16); X. laevis U5: (10); X. laevis U2: (26); chicken U4'B: (27); X. laevis U1a: (28); mouse U1b2: (29); and mouse U6: (30).

A number of these are shown in Fig. 5. The upper part of the figure shows an alignment of six sequences for which there is functional evidence supporting their role in snRNA gene enhancer activity (9, 10, 12, 16). To our knowledge, detailed mutational analyses have been reported only for the chicken U1 and U4B genes. Inspection of the other sequences, however, supports the idea that they may be functionally equivalent to the chicken U1 and U4B SPH motifs, and that their activity may therefore be mediated by SBF. The *Xenopus* U5 sequence shown in the figure is recognized by a factor in unfractionated extracts from Molt-4 cells (10), but to our knowledge that factor has not been further purified or characterized. The human U4C sequence was postulated (16) to contain a binding site for the transcription factor AP-2 where it exhibits strongest similarity to the chicken U4B SPH motif (positions -9 to -3). However, the AP-2 site homology was seriously disrupted in our construction U4BΨWT-A1, and that construction was nevertheless still well-expressed and a good substrate for recognition by SBF. Thus it seems unlikely that the activity of the SPH motif is mediated by AP-2. Moreover, authentic AP-2 binding sites do not compete for binding to SBF in mobility shift assays (our unpublished data).

A consensus SPH motif sequence was derived from the six functionally-defined sequences and is shown in the middle of Fig. 5. The lower section of Fig. 5 lists sequences from the enhancer regions of five additional vertebrate snRNA genes that were not included in deriving the consensus. Their reasonable match to the consensus suggests the possibility that those sequences also may be functional SPH motifs, especially since many of the nucleotides that deviate from the consensus were tolerated in the U4B SPH motif mutants studied (e.g., G, A, T, A, and G at positions -9, -8, -3, -2, and +1 respectively). Because a considerable amount of degeneracy is permitted, it is possible that SPH-like motifs may be important for enhancer function in a wide variety of vertebrate snRNA genes, although they may not be immediately apparent by inspection and sequence comparison.

Nucleotides extending over a region at least 15 bp in length are important for SBF binding and U4B enhancer activity (Figs.

3 and 4). Interestingly, positions -9 to -7 appear to be highly conserved in vertebrate evolution (Fig. 5), yet simultaneous mutation at all three of these positions had no apparent effect in our functional assays (expression or binding). However, this result is in agreement with our mutational and functional analysis of the U1 gene SPH motif (12). In that case, deletion of the nucleotides corresponding to positions -8 to -11 had no detectable effect, but when combined with a set of point mutations at the opposite end of the SPH motif (positions +6 to +10) which by themselves had only a moderate effect, the result was a drastic reduction of U1 gene expression and SBF binding (12). From these results, we believe that the evolutionarily conserved nucleotides at positions -9 to -7 are indeed integral to the SPH motif, but the transient expression assay in oocytes does not effectively measure their contribution unless mutations at those positions are combined with partially disabling base changes in other parts of the SPH motif.

It therefore seems likely that the SPH motif is at least 18 bp in length. This uncommonly long recognition sequence may explain the ability of the SPH motif to tolerate a substantial number of base substitutions without suffering a striking loss of activity. In fact, bases at certain internal positions (e.g., +4 to +6) can apparently be altered and drift freely during evolution with little or no constraint. Other transcription factors have been characterized that recognize unusually long and highly degenerate sequences (e.g., LSF [24] and LF-B1 [25]).

In conclusion, it seems possible that SPH-like motifs may be widely distributed in the enhancers of vertebrate snRNA genes. Functional cooperativity between octamer and SPH motifs may thus be a commonly employed strategy to generate snRNA gene enhancer activity. However, studies on additional snRNA genes will be required to determine how generally applicable this model may be.

ACKNOWLEDGEMENTS

We thank Ken Roebuck for the U1 competitor template, Daniel Szeto and Nancy Fan for affinity-purified SBF, and Kathleen McNamara and Doug Allen for excellent technical assistance. We also thank Sanford Bernstein and Ken Roebuck for critically reading the manuscript prior to submission. This work was supported by Grant GM-33512 from the National Institutes of Health and in part by the California Metabolic Research Foundation. Z.Z. is a fellow of the California Foundation for Biochemical Research.

REFERENCES

- Reddy, R., and Busch, H. (1988) In Birnstein, M.L. (ed.), *Structure and function of major and minor small nuclear ribonucleoprotein particles*. Springer Verlag KG. Heidelberg, pp 1-37.
- Steitz, J.A., Black, D.L., Gerke, V., Parker, K.A., Kramer, A., Frendewey, D., and Keller, W. (1988) In Birnstein, M.L. (ed.), *Structure and function of major and minor small nuclear ribonucleoprotein particles*. Springer Verlag KG. Heidelberg, pp 115-154.
- Dahlberg, J.E., and Lund, E. (1988) In Birnstein, M.L. (ed.), *Structure and function of major and minor small nuclear ribonucleoprotein particles*. Springer Verlag KG. Heidelberg, pp. 38-70.
- Parry, H.D., Scherly, D., and Mattaj, I.W. (1989) *Trends Biochem. Sci.*, **14**, 3633-3644.
- Neuman de Vegvar, H.E., Lund, E., and Dahlberg, J.E. (1986) *Cell* **47**, 259-266.
- Hernandez, N., and Weiner, A.M. (1986) *Cell* **47**, 249-258.
- Hernandez, N., and Lucito, R. (1988) *EMBO J.* **7**, 3125-3134.
- Sturm, R.A., Das, G., and Herr, W. (1988) *Genes Dev.* **2**, 1582-1599.
- Tebb, G., and Mattaj, I.W. (1989) *Mol. Cell. Biol.* **9**, 1682-1690.
- Kazmaier, M., Tebb, G., and Mattaj, I.W. (1987) *EMBO J.* **6**, 3071-3078.
- Roebuck, K.A., Walker, R.J., and Stumph, W.E. (1987) *Mol. Cell. Biol.* **7**, 4185-4193.
- Roebuck, K.A., Szeto, D.P., Green, K.P., Fan, Q.N., and Stumph, W.E. (1990) *Mol. Cell. Biol.* **10**, 341-352.
- McNamara, K.J., Walker, R.J., Roebuck, K.A., and Stumph, W.E. (1987) *Nucleic Acids Res.* **15**, 9239-9254.
- Murphy, J.T., Skuzeski, J.M., Lund, E., Steinberg, T.H., Burgess, R.R., and Dahlberg, J.E. (1987) *J. Biol. Chem.* **262**, 1795-1803.
- Ares, M., Jr., Chung, J.-S., Giglio, L., and Weiner, A.M. (1987) *Genes Dev.* **1**, 808-817.
- Weller, P., Bark, C., Janson, L., and Petterson, U. (1988) *Genes Dev.* **2**, 1389-1399.
- Janson, L., and Petterson, U. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 4732-4736.
- Hoffman, M.L., Korf, G.M., McNamara, K.J., and Stumph, W.E. (1986) *Mol. Cell. Biol.* **6**, 3910-3919.
- Oliphant, A.R., Nussbaum, A.L., and Struhl, K. (1986) *Gene* **44**, 177-183.
- Hill, D.E., Hope, I.A., Macke, J.P., and Struhl, K. (1986) *Science* **234**, 451-457.
- Roebuck, K.A. (1989) Ph.D. Thesis, University of California, San Diego, and San Diego State University.
- Kadonaga, J.T., and Tjian, R. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 5889-5893.
- Earley, J.M., III, Roebuck, K.A., and Stumph, W.E. (1984) *Nucleic Acids Res.* **12**, 7411-7421.
- Huang, H.-C., Sundseth, R., and Hansen, U. (1990) *Genes Dev.* **4**, 287-298.
- Frain, M., Swart, G., Monaci, P., Nicosia, A., Stampfli, S., Frank, R., and Cortese, R. (1989) *Cell* **59**, 145-157.
- Korf, G.M., and Stumph, W.E. (1986) *Biochemistry* **25**, 2041-2047.
- Bark, C., Weller, P., Zabielski, J., and Petterson, U. (1986) *Gene* **50**, 333-344.
- Krol, A., Lund, E., and Dahlberg, J.E. (1985) *EMBO J.* **4**, 1529-1535.
- Howard, E.F., Michael, S.K., Dahlberg J.E., and Lund, E. (1986) *Nucleic Acids Res.* **14**, 9811-9825.
- Das, G., Henning, D., Wright, D., and Reddy, R. (1988) *EMBO J.* **7**, 503-512.