Volume 15 Number 16 1987 Nucleic Acids Research

Only two of the four sites of interaction with nuclear factors within the *Xenopus* U2 gene promoter are necessary for efficient transcription

Graham Tebb, Dirk Bohmann[1]* and Iain W.Mattaj

European Molecular Biology Laboratory (EMBL), D-6900 Heidelberg and [1]Division of Molecular Biology, Institute of Cell and Tumour Biology, German Cancer Research Centre, D-6900 Heidelberg, FRG

Received May 11, 1987; Revised and Accepted July 14, 1987

ABSTRACT

An analysis, performed by DNase I footprinting, of the interactions between factors present in Molt-4 nuclear extracts and a *Xenopus* U2 snRNA gene promoter is presented. Four distinct regions of sequence-specific DNA-factor interaction are found. Two of these correspond to the previously identified proximal and distal sequence elements (PSE and DSE) of the promoter. Both of these elements are important in U2 transcription, indicating a functional role for the observed interactions. The other two sites of interaction correspond to a sequence element conserved in many, but not all, vertebrate U snRNA gene promoters (the MSE) and to a region adjacent to the site of transcription initiation (the "cap site"). Site-directed mutants of these latter two elements are constructed which no longer bind nuclear factors. Transcriptional analysis in *Xenopus* oocytes reveals that these mutants are transcribed as efficiently as wild-type U2. Other possible roles for the two factors are discussed.

INTRODUCTION

The U-rich small nuclear RNAs (U snRNAs) represent a family of abundant RNAs found in the form of ribonucleoprotein particles (U snRNPs) at approximately 10^3 to 10^6 copies in the nuclei of higher eukaryotic cells. U6 has recently been shown to be an RNA polymerase III transcript (1, 2), but based on α-amanitin (3, 4) and DRB (5) inhibition studies it is thought that U1-U5 are transcribed by RNA polymerase II. Unlike the majority of polymerase II transcripts the U snRNAs are not polyadenylated, and in addition they have an unusual 2,2,7-trimethylguanosine cap (reviewed in 6) which is generated post-transcriptionally from a monomethyl cap (7). Interest in the study of U snRNAs stems from the observation that U snRNPs can be precipitated by antibodies found in the sera of patients with certain autoimmune disorders (8) and has grown as a result of the discovery that many U snRNPs are involved in the processing of mRNA precursors. There is now

substantial evidence that U1, U2, U5, U7 and a complex of U4 and U6 are required for mRNA processing (for reviews see 9, 10).
Except for U6, U snRNA genes do not possess TATA boxes and their transcription is dependent on the presence of two sequence elements located in the first 300bp of 5′ flanking sequence (4, 11-20). The first of these, termed the proximal sequence element (17), is located between positions -50 and -60 relative to the start site of transcription. It is TATA box-like in function in that it is absolutely required for transcription (11, 15, 17, 18) and that it positions the transcription initiation point (7, 11). The second element is more distal and has enhancer-like properties (11, 15, 17, 19, 21). The promoters of various vertebrate U snRNA genes cloned and characterized thus far have similar structures, and in fact human U1 (4) and U2 (12, 15) genes are transcribed in the *Xenopus* oocyte, although less efficiently than the homologous *Xenopus* genes (18, 22).
In this paper we report an analysis by DNase I footprinting (23) of the promoter region of the *Xenopus* U2 snRNA gene. Using a nuclear extract prepared from the human T-cell line Molt-4 (24) we show the existence of four protected regions. Two of these correspond to the proximal and distal sequence elements, the others to further sequences which we demonstrate are not required for transcription of the gene on microinjection into *Xenopus* oocytes.

MATERIALS AND METHODS

DNase I Footprinting

Nuclear extracts were prepared by the method of Dignam *et al.* (25) with the modifications introduced by Wildeman *et al.* (26). The protein concentration of the extracts was measured by the dye-binding assay (27). For DNase I protection experiments, amounts of extract from 0.25-4.0μl (2-33μg protein) as indicated in the figure legends were preincubated with 25ng of linearized plasmid pUC8 on ice for 15 min. This preincubation has the effect of complexing the non-specific DNA-binding proteins present in the extract. The reactions (10μl) contained 10mM HEPES-KOH (pH 7.9), 20mM KCl, 4mM $MgCl_2$, 4mM spermidine, 0.1mM EDTA, 0.25mM DTT and 10% glycerol. 20,000cpm DNA, end-labelled by means of Klenow DNA polymerase with α-^{32}P-dATP or α-^{32}P-dGTP (Amersham) or using T4 polynucleotide kinase and γ-^{32}P-ATP (Amersham), was added in 2μl water and incubated for an additional 10 min. on ice. 2μl DNase I (Worthington DPFF), freshly diluted in water from a stock solution of 3mg/ml to concentrations between 12.5 and 800 μg/ml was added and digestion was allowed to proceed for 90 seconds at 20°C. Digestion was stopped by the addition of 50μl phenol/chloroform/ isoamylalcohol (25:24:1) and vortexing; the phases were separated by centrifugation and the aqueous phase re-extracted

with 50µl chloroform/isoamylalcohol (24:1) and dialysed on millipore filters against water for 90 mins. The samples were lyophilised, taken up in 2.5µl formamide loading buffer (28) and loaded onto 6% acrylamide-urea sequencing gels together with a G+A sequencing reaction (29) to provide accurate size standards. For the footprinting competition experiment shown in Figure 1B the competitor oligonucleotide was added together with the labelled DNA. The specific competitor was a ds 40mer containing the wild type U2 sequence from -278 to -239; the non-related ds oligonucleotide was a 21mer with the sequence GATCCGTTTAAAGATAGAGAG and its complement.

General Methods

Site-directed mutagenesis was performed essentially as described by Kramer *et al.* (30), and the sequences of the resulting mutants were verified by means of the chain termination method (31). Oocytes were injected, and RNA extracted and analysed as described (32). The construction of mutants -23,-4 -48,-4 and -81,-4 has previously been reported (7); mutants pUC.DSE82+/- were prepared by sub-cloning the Pvu II-Eco RI fragments (blunt-ended by means of Klenow DNA polymerase) of 82+ and 82- respectively (17) into Sma I-cut pUC8, and pUC.DSE242+/- in a similar way from 242+ and 242- (17).

RESULTS

There are four sites of factor-DNA interaction in the U2 promoter

We have previously shown (33) that nuclear extracts of Molt-4 cells represent a rich source of transcription factors. When a ds 39mer oligonucleotide corresponding to the U2 promoter sequence from -78 to -40, which contains the proximal sequence element, is used to probe nuclear extracts prepared from Molt-4 cells or from a *Xenopus* kidney cell line by means of gel retention experiments (34, 35) we see a retarded band of identical mobility, suggesting the binding of similar proteins (M. Kazmaier and H. Parry, unpublished results). When a ds 40mer oligonucleotide containing the U2 distal sequence element is used to examine nuclear extracts of HeLa, Molt-4 and *Xenopus* kidney cells (33, M. Kazmaier and H. Parry, unpublished) a similar gel retention is seen in each case. Furthermore, we have been able to show that the relatively abundant DSE-binding factor(s) present in *Xenopus* extracts produce a DNase I protection pattern identical to that obtained with Molt-4 extracts (data not shown). This evidence suggests that the factors involved in the transcription of U snRNA genes are common to different cell types and species. Given the difficulty of obtaining good extracts from cultured *Xenopus* cells and the low concentrations of transcription factors in these extracts we therefore decided to employ Molt-4 nuclear extracts to characterize the sites of protein binding within the U2 gene promoter.

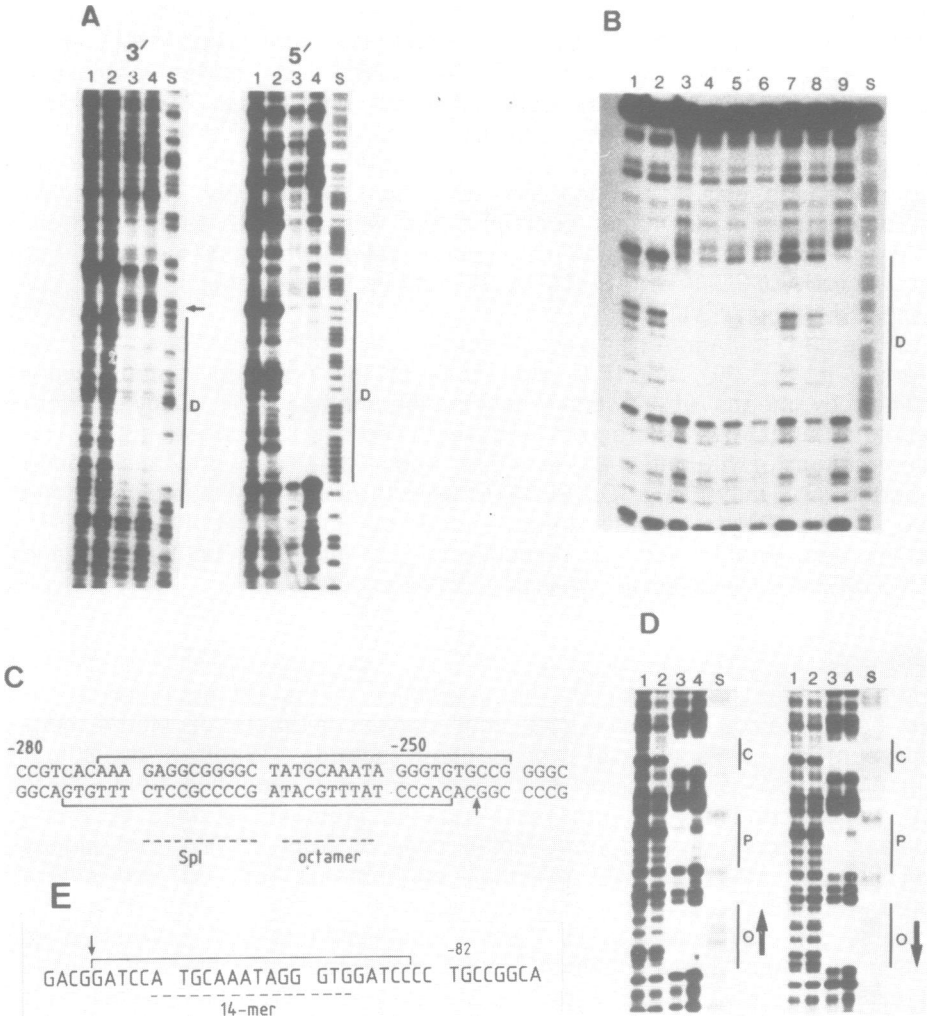


Fig. 1

A: DNase I protection pattern observed over the U2 DSE. Both 3′- (left) and 5′-labelled (right) strands are shown. Lanes 1,2: no extract, 12.5 and 25μg/ml DNase I, respectively; lanes 3,4: 1μl extract (=4μg nuclear protein), 50 and 100μg/ml DNase I; lanes "s": Maxam and Gilbert G+A sequence reactions. The numbers reflect the distance upstream of the start site of transcription. Protected sequences are indicated, and an arrow represents hypersensitivity to DNase I cleavage in the presence of extract. The position of the DSE is indicated by "D".

B: Competition of the protection in A (only the 5′-labelled strand is

shown) with specific and non-related oligonucleotides. Lanes 1,2: no extract, 25 and 100µg/ml DNase I, respectively; lanes 3-9: 0.25µl extract (= 2µg protein), 100µg/ml DNase I. Lane 3: no competitor; lanes 4-8: 50-, 100-, 200-, 350- and 500-fold molar excess of specific competitor; lane 9: 500-fold molar excess of non-related competitor. Lane "s": Maxam and Gilbert G+A sequence reaction. The sequences of the oligonucleotides used are given in the Materials and Methods section.
C: Schematic representation of the protection pattern shown in A. Brackets signify protection, and an arrow indicates a position of hypersensitivity. The consensus Sp1 binding site (36) and the octamer sequence (45, 47) are underlined.
D: Footprints on the 5´-labelled strands of pUC.DSE82- (left) and pUC.DSE82+ (right). Lanes 1,2: no extract, 12.5 and 25µg/ml DNase I, respectively; lanes 3,4: 4µl extract (=33µg protein), 200 and 400µg/ml DNase I. Lanes "s": Maxam and Gilbert G+A sequence reactions. The various protections are indicated as follows: "O" inserted octamer motif-containing oligonucleotide (with arrows to show the orientation of insertion), "P" PSE, "C" "cap site".
E: Schematic representation of the protection pattern shown in D. Only pUC.DSE82+ is shown, since the two constructs give rise to very similar protections. The 14mer oligonucleotide is underlined. Sequences 5´ of position -82 in the gene have been replaced by linker sequences during the construction of this mutant.

As mentioned in the introduction, the distal sequence element (DSE, ref. 17) of the U2 gene has enhancer-like properties, and so it is to be expected that this sequence is able to bind a factor or factors in nuclear extracts. From the pattern of DNase I protection shown in Figure 1 it is clear that this is the case. Comparing Figure 1A lanes 1 and 2 (no extract controls) with lanes 3 and 4, a clear "footprint" is seen over both strands. As demonstrated by the experiment shown in Figure 1B this protection can be removed by the inclusion in the incubation mixture of a 350-fold molar excess of a double-stranded (ds) 40mer oligonucleotide including the DSE (lane 7) but not by a 500-fold molar excess of a non-related ds oligonucleotide (lane 9). The protection pattern over the DSE is represented schematically in Figure 1C. As shown, the extent of protection is approximately 30bp on each strand, and there is a stagger of 4bp between the strands. Included within the protected region are potential binding sites for two previously reported transcription factors, Sp1 (36) and the octamer binding protein (32, 37-40).
It is clear from the experiment shown in Figure 1D that not all of this 30bp region is required for factor binding, since when a ds 14mer oligonucleotide (17) containing the octamer sequence ATGCAAAT, but not the potential Sp1 binding site also found in the DSE, is cloned in

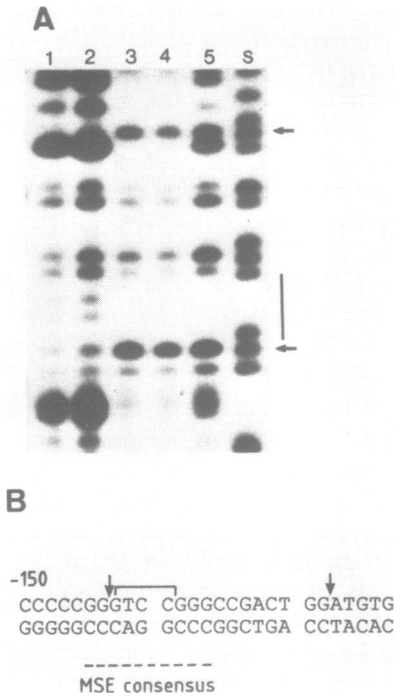


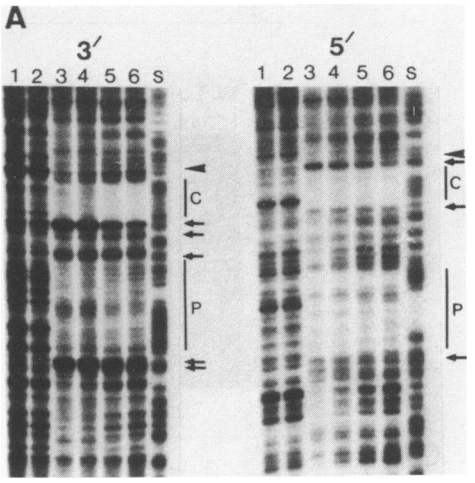
Fig. 2
A: Footprint observed over the MSE 5′-labelled strand. Lanes 1,2: no extract, 12.5 and 25μg/ml DNase I, respectively; lanes 3-5: 4μl extract (=33μg protein), 100, 200 and 400μg/ml DNase I. The protected region and sites of hypersensitivity are indicated as in Figure 1.
B: Schematic representation of the protection in A. The MSE consensus sequence (see Table 1) is underlined.

either orientation into a construct lacking the DSE in clones pUC.DSE82+/- (Figure 1D) and pUC.DSE242+/- (not shown) a protection is observed, although this is shorter than that found over the wild-type DSE (compare Figures 1C and 1E). The protection in these cases is also competable using the specific ds 40mer referred to above, but not using a non-related oligonucleotide (data not shown).
A close inspection of Figure 1A reveals the existence of another region slightly downstream of the indicated protection where the extent of DNase I cleavage is reduced. This is not reproducibly observed (see, for example, Figure 1D) and so is not considered further here.
As described in the Discussion section, there is a sequence common to many vertebrate U snRNA genes located between the distal and proximal sequence elements which we have termed the medial sequence element (MSE). Figure 2 shows the "footprint" observed in this region. Although the changes seen are relatively minor (the disappearance of two bands in the cutting pattern, coupled with the appearance of two hypersensitive sites, one immediately 5′ to the protected region and the other 11bp, or one turn of the DNA helix, 3′ to it) and confined to one strand they have been reproducibly observed in a large number of separate experiments. The protection is sequence-specific, as demonstrated by the observation that when the protected region is mutated (in the mutant ΔMSE

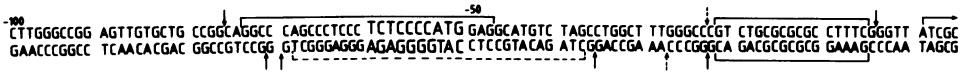
Fig. 3

A: DNase I protection pattern observed over the PSE and the "cap site". Both 3′- (left) and 5′-labelled (right) strands are shown. Lanes 1,2: no extract, 25 and 50μg/ml DNase I, respectively; lanes 3-6: 4μl extract (=33μg nuclear protein), 100, 200, 400 and 800μg/ml DNase I; lanes "s": Maxam and Gilbert G+A sequence reactions. The protections are represented as in Figure 1. The start site of transcription is indicated by an arrowhead.

B: Schematic representation of the protections in A. The conserved region of the PSE (11) is enlarged. The comparatively weak protection over the coding strand of the PSE is indicated by the dashed line. The arrow on the right represents the site of transcription initiation.



B



described in Figure 5) the DNase I cleavage pattern is the same in the presence and absence of extract (data not shown).

The proximal sequence element (PSE, ref. 17) of the U snRNA genes is absolutely required for transcription, and in Figure 3 the protection patterns over this element and the "cap site" are shown. It is noteworthy that the protection on the coding strand of the PSE (Figure 3A, 3′) is considerably weaker than that on the other strand (Figure 3A, 5′). As is clear from the schematic representation, the "cap site" protection does not actually cover the start site of transcription (represented by arrowheads in Figure 3A and an arrow in Figure 3B), but this nomenclature has been adopted to avoid having to rename the PSE. The close proximity of two protected regions can be the result of the presence of two distinct factor-binding sites or alternatively can result from sequence-specific binding at one position followed by the association of a second protein with the DNA-protein complex causing the protection of an additional stretch of DNA in a sequence-independent manner. To distinguish between these two possibilities we performed the experiment shown in Figure 4. The construction of the 'promoter mutants, represented in Figure 4A, has been described previously (7).

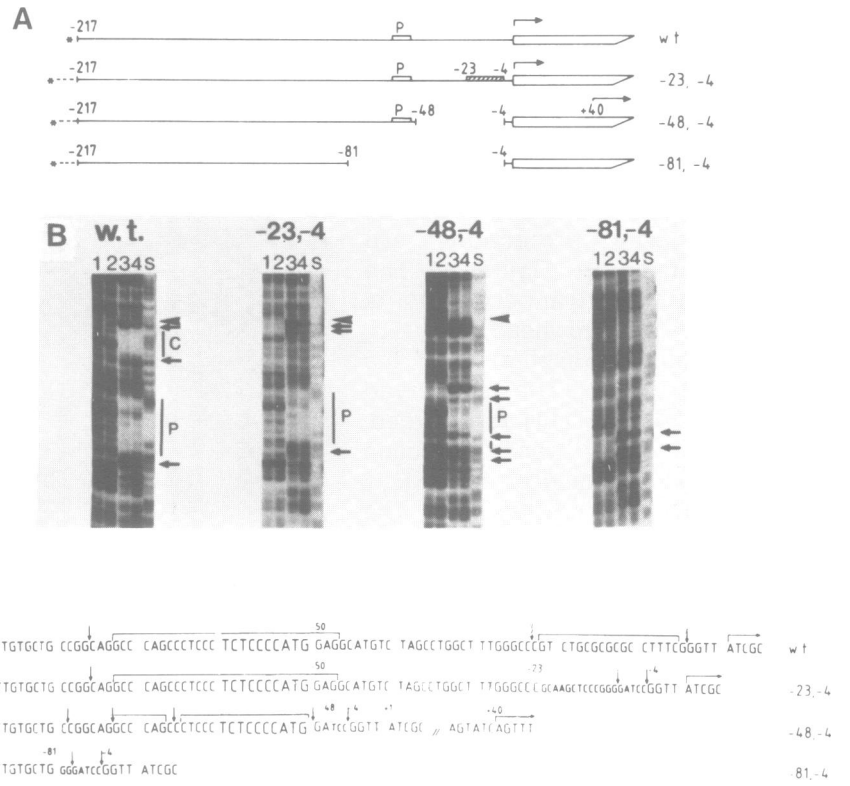


Fig. 4

A: Diagram of the promoter mutants used in this study. Their construction is reported elsewhere (7). Asterisks show the positions of labelling (the dotted lines represent portions of pUC8 vector sequences), the PSE is indicated by "P", a hatched box represents a substitution, the RNA coding sequence is shown by an open box and the start sites of transcription (except for mutant -81,-4 which is not transcribed) are indicated by arrows.

B: Footprints over the 5′-labelled strands of the wild-type U2 promoter and the three mutants. Lanes 1,2: no extract, 25 and 50µg/ml DNase I, respectively; lanes 3,4: 4µl extract (=33µg nuclear protein), 400 and 800µg/ml DNase I; lanes "s": Maxam and Gilbert G+A sequence reactions. The protections are represented as in Figure 1. "P": PSE protection; "C": "cap site" protection.

C: Schematic representations of the protections in B. The conserved region of the PSE (11) is enlarged. Linker sequences introduced during the construction of the mutants are shown in small type. The horizontal arrows on the right indicate the sites of transcription initiation.

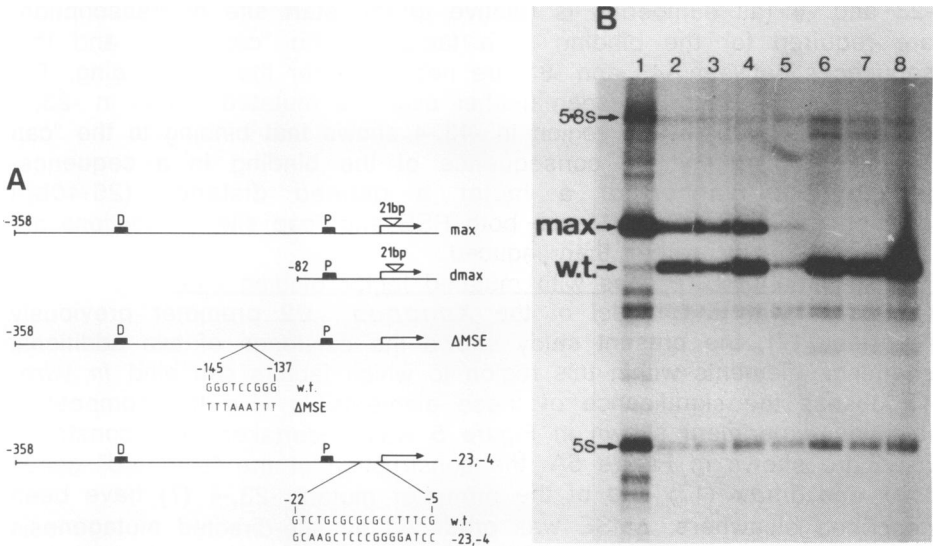


Fig. 5

A: Constructs used for the competition injection experiment. The construction of *max*, d*max* (17) and -23,-4 (7) is described elsewhere, that of ΔMSE in the Materials and Methods section. The DSE ("D") and PSE ("P") are indicated, and substituted sequences are shown.

B: Transcription of the mutants in A in the *Xenopus* oocyte. At least 12 separate oocytes were injected per sample, aiming at the nucleus, with 30-50nl DNA at a concentration of 300μg/ml in water. The oocytes were incubated at 19°C overnight, then RNA was extracted and the amount corresponding to 0.5 oocytes loaded on an 8% acrylamide-7M urea gel. Transcripts originating from the maxigene *max* and the wild-type gene are indicated by arrows, and the positions of 5S and 5.8S RNA are also marked. Lane 1: *max* + U2 with the first 82bp of 5′ flanking sequence only; lane 2: *max* + wild-type U2; lane 3: *max* + -23,-4; lane 4: *max* + ΔMSE; Lane 5: d*max* + U2 with the first 82bp of 5′ flanking sequence only; lane 6: d*max* + wild-type U2; lane 7: d*max* + -23, -4; lane 8: d*max* + ΔMSE.

The mutant -23,-4 is altered in sequence in the region where the "cap site" protection is seen. This mutant initiates transcription at the correct site, while transcripts of the deletion mutant -48,-4 initiate 40 nucleotides downstream of the normal position (arrows in Figure 4A). The longer deletion -81,-4, in which the PSE is removed, results in the loss of transcriptional activity (7). The footprinting data given in Figure 4B and schematically in Figure 4C show that sequences located between

-23 and -4 (all numbering is relative to the start site of transcription) are required for the binding of a factor to the "cap site", and that sequences between -48 and -81 are necessary for the PSE-binding. The fact that protections are seen neither over the mutated region in -23,-4 nor over the +20 to +35 region in -48,-4 shows that binding to the "cap site" is not merely the consequence of the binding in a sequence-independent manner of a factor a defined distance (25-40bp) downstream of the PSE. Hence both PSE and "cap site" protections are sequence-specific rather than induced.

Transcription of U2 genes with mutated factor binding sites

Compared with the model of the *Xenopus* U2 promoter previously reported (17), the present study shows the existence of two additional sequence elements within this region to which factors can bind *in vitro*. To assess the significance of these elements *in vivo* the competition injection experiment shown in Figure 5 was undertaken. The constructs used are shown in Figure 5A: the construction of the "standard" genes *max* and d*max* (17) and of the promoter mutant -23,-4 (7) have been described elsewhere; ΔMSE was produced by site-directed mutagenesis using the method of Kramer *et al.* (30). It is important to note that both -23,-4 and ΔMSE are substitution mutants and as such retain the wild type spacing between regions flanking the mutated sites.

As expected, the injection into the *Xenopus* oocyte of equal amounts of *max* and a U2 mutant deleted to -82 but containing the wild-type coding sequence (Figure 5B, lane 1) results in a very low level of transcription of the wild-type gene. The faint band seen at the wild-type position in this lane (indicated "w.t.") is probably the result of the transcription of endogenous U2 genes, as suggested by the co-injection of wild-type U2 and d*max* (Figure 5B, lane 6), where only the wild-type transcript is produced. Thus sequences upstream of -82 in the U2 gene promoter are required for the ability to compete for transcription in the oocyte. Injection into the oocyte of equal amounts of *max* and wild-type U2 (Figure 5B, lane 2), or of equal amounts of d*max* and the U2 mutant deleted to -82 (Figure 5B, lane 5) leads to equivalent amounts of transcription of these genes. These results confirm previous data (17) and act as controls for the remainder of the injections. Interestingly, as shown in Figure 5B lanes 3 and 4, the two mutants -23,-4 and ΔMSE are also transcribed as efficiently as wild-type U2 in the presence of *max*. Furthermore, the two mutants -23,-4 and ΔMSE can prevent the transcription of d*max*, which lacks both the MSE and the DSE (Figure 5B lanes 7 and 8). Taken in conjunction these findings confirm the result that the DSE is necessary for efficient transcription, and demonstrate that alteration of the sequences required for the binding of nuclear factors to the MSE and "cap site" does not reduce transcriptional efficiency of the U2 gene in the oocyte.

DISCUSSION

The medial and "cap site" sequence elements

The medial sequence element (MSE) reported in this study has previously been identified as occurring in a region of sequence conservation between the *Xenopus* U2 and the human and rat U1 genes (41). In addition, as shown in Table 1, this sequence is also found in a number of other vertebrate U genes in a corresponding position (approximately 130-140bp upstream of the start site). A function for the MSE (in their study termed the "C region") in the the human U1 gene has been proposed by Murphy *et al.* (42), who report that (large) deletions in this region have the effect of reducing transcriptional levels 3- to 5-fold on injection of the genes into *Xenopus* oocytes. In the case of the *Xenopus* U5 gene, however, similar deletions have been shown to have no effect on transcription (20) and for the *Xenopus* U2 gene it has previously been shown (17) that the first 82bp of 5′ non-coding sequence are sufficient for transcription in the oocyte, and neither transcriptional level nor competitive ability (as measured by the type of transcription competition assay shown in Figure 5) is increased by the presence of sequences between -82 and -242 (the DSE is located just upstream of -242). The present study goes further in showing that the region of sequence homology located between -145 and -137 is able to bind a factor in nuclear extracts, but that the substitution of this sequence and the concomitant removal of the ability to bind the factor has no effect on transcription of the *Xenopus* U2 gene in the *Xenopus* oocyte.

Another previously uncharacterized element in the U2 promoter is here termed the "cap site", although as mentioned above this element does

Table 1: A Comparison of Sequences around Position -140 in Vertebrate U snRNA Genes

Gene	Sequence	Position	Ref.
Xenopus U2	GGGTCCGGG	-145/-137	40
Human U1	GGGAGCGGG	-147/-139	4, 55
Rat U1	GGGAGCGCG	-144/-146	56
Mouse U1	GGGAGCGTG	-139/-131	57
Xenopus U1B	GGGTCAGGC	-134/-126	16, 18
Chicken U2	GGGAGCGGG	-146/-138	48
Xenopus U5	CGGTGCGCG	-137/-129	20
Human U2	GGGAACGCC	-198/-190	15
Human U2	GGGAATGGG	-92/-83	15
CONSENSUS	GGG$^{A}_{T}$NCG$^{G}_{C}$G		

not overlap the start site of transcription. As shown in Figure 3A a strong protection is observed over both strands of this region, although footprinting experiments using lower amounts of extract reveal that this protection is the first to be affected by a lowering of the extract concentration (data not shown). The binding to the element is sequence-specific, rather than induced by the prior binding of a transcription factor to the PSE, as shown by the mutant data presented in Figure 4, but the experiment in Figure 5, coupled with the previous report that sequences between -4 and -48 are not required for transcription (7), again suggest that this element has no effect on U2 transcription in the oocyte. Interestingly, Murphy *et al.* (42) report the existence of an "element A" in the human U1 gene which is located immediately adjacent to the coding region, and which is essential for accurate initiation of transcription in the *Xenopus* oocyte and which in addition has a stimulatory effect on transcription in the oocyte. The observations that all sequences in the *Xenopus* U2 gene between -48 and -4 can be removed without affecting the accuracy of transcription initiation (although initiation takes place at position +40 in the coding sequence, in other words at the normal distance downstream of the PSE) and that removal of the binding site between -23 and -4 affects neither the level nor the accuracy of transcription argue strongly against the existence of any essential sequences between the PSE and the start site in this gene. Unfortunately we are unable to assess the transcriptional efficiency of the -48,-4 mutant since the RNA product is different from that of the wild type U2 gene and so might have a differential stability in the oocyte. Measurement of transcription by "run on" experiments in the oocyte is unfeasible, so we are left with no way of quantifying the transcription of this mutant.

From the present study, then, the functions of the medial and "cap site" sequence elements remain unclear, although it could be speculated that factors binding to these sequences may have effects more subtle than we are able to observe in our experimental system, for example they may somehow be involved in co-ordinating the regulation of the production of the protein and RNA moieties that must eventually be assembled into active snRNPs. Other possibilities are that these elements may be involved in developmental or cell-cycle specific regulation.

As mentioned above, Murphy *et al.* (42) are able to observe effects on transcription in the oocyte of mutations within these regions. There are several possible explanations for the discrepancies between their results and ours. First, the two systems under study are different and it is not inconceivable that the regulatory possibilities of the human U1 gene may be different from those of the *Xenopus* U2 gene. Secondly, we report the effect on transcription of alterations which are small in comparison with those characterized by Murphy and co-workers and so it is possible that they are deleting important sequences which we are

leaving untouched. Finally the nature of the mutants used in the two studies is different: we have reported only site-directed mutants while Murphy *et al.* (42) employed deletion mutants. These lead, we believe, to results which are far more difficult to interpret since the spacing and relative positioning on the DNA helix of factors bound to other sequence elements are altered (see, for example, 43) and the possibities for interaction by "looping" or "bending" of DNA sequences (44) between important elements are also affected.

It thus appears that the only sequence elements in the U2 promoter which are important for the transcription of this gene in the oocyte are the previously identified PSE and DSE (17). As reported in the present study, we have been able to demonstrate factor binding *in vitro* to both of these elements, and we should now like to consider each of them in turn.

The proximal sequence element

The proximal sequence element of U genes has been shown to be absolutely required for transcription (11, 15, 17, 18). In the *Xenopus* U2 gene the 5′ boundary of this element has been shown to lie at or downstream of position -82 (17) and the 3′ boundary to be located at or upstream of -48 (7); for the human gene it has been shown that the 5′ border of the region is at or downstream of position -62 (15). Within the sequence between -62 and -48 there is a 10bp stretch which is highly conserved between vertebrate U genes (11), and it is highly likely that this region is the PSE. As shown in Figure 3, this region is able to bind a factor found in the nuclear extracts used in this study.

Since no *in vitro* transcription system has been found which is capable of recognising U snRNA promoters (ref. 4 and unpublished data) it is difficult to be certain that the factor we show to be bound to the PSE is the one responsible for transcribing these genes *in vivo*. The correlation between binding to the promoter mutants *in vitro* and their transcriptional activity *in vivo* suggests that we are observing a physiologically relevant binding, but in order to make this conclusion more certain, and to define more precisely the sequence required for binding, we are currently performing a saturation mutagenesis study of the promoter region between -62 and -48.

The distal sequence element

The distal sequence element of the *Xenopus* U2 gene contains two sequence motifs which have previously been shown to be important in the transcriptional regulation of many different viral and cellular genes. The first of these is the octamer sequence ATGCAAAT, which is essential for the enhanced transcription of U genes (15, 21), and immediately upstream is found a perfect fit to the consensus sequence required for the binding of the cellular transcription factor Sp1 (36, 45, 46). We shall discuss first the importance of the Sp1 motif, and then return to a discussion of the octamer sequence.

Mangin *et al.* (19) have recently reported that a potential Sp1 binding

site, or a highly similar sequence, occurs as a component of the DSE of many different vertebrate U snRNA genes. It might be supposed, then, that this sequence is in some way important for the transcription of these genes, but Mattaj *et al.* (17) have shown that a synthetic oligonucleotide containing the octamer sequence but not the Sp1 binding site is able to enhance transcription of and partially restore competitive ability to deletion mutants of the *Xenopus* U2 gene lacking the distal sequence element. That the octamer is absolutely required for transcriptional enhancement has been demonstrated by the finding that short deletions in this motif, which leave the Sp1 binding site intact, completely destroy the activity of the human U2 (15) and *Xenopus* U1 (21) DSEs. It is clear, therefore, that the Sp1 binding site is neither sufficient nor essential for enhanced transcription of this gene.

From the data shown in Figure 1, we cannot be sure whether Sp1 binds, under the conditions used in the present study, to the U2 DSE (compare the region of protection seen in Figure 1A, where the Sp1 site is present, with the shorter protection seen in Figure 1D, where it is absent). However, we have performed point mutation experiments, in which either the Sp1 binding site or the octamer sequence were mutated to prevent factors binding to them (G.T. and I.W.M., unpublished data), and these demonstrate that both of the sequence elements are required for the full length protection over the DSE. It appears, then, as though the U2 DSE is bipartite, consisting of adjacent Sp1 and octamer sites.

Support for this idea comes from a comparison of the footprints observed over the U2 DSE (Figure 1) and the *Xenopus* U1 DSE (21), which contains a potential Sp1 binding site located 3′ to the octamer. The extent of DNase I protection is the same in each case, but the U1 DSE footprint is "skewed" so that both the octamer and the Sp1 site are covered, as is the case for the U2 where the Sp1 site is located 5′ to the octamer. In addition, the *Xenopus* U5 DSE has recently been shown to have two separate components, an octamer sequence and a G/C-rich region which does not contain a consensus Sp1 binding site, and both of these components can bind factors present in the Molt-4 nuclear extract used in the present study (20).

From the results presented here it is clear that there is a factor present in the Molt-4 nuclear extract which is able to recognize and bind to the octamer sequence of the U2 promoter. This sequence motif is also found in the immunoglobulin enhancer (47, 48), where it occurs in the reverse orientation, in the promoters of immunoglobulin heavy- and κ light-chain genes (47, 49), as a component of various histone H2b promoters (38, 50), in the promoter regions of many different U1 and U2 genes (11, 13, 15-18, 51), in the *Xenopus* U5 gene promoter (20) and, with a single base change, in the SV40 enhancer (49, 52). The importance of the motif has recently been shown by the finding that a single factor present in HeLa cell nuclear extracts can combine with octamer-containing fragments from U1, U2, histone H2b, SV40 and immunoglobulin light and

heavy chain promoters and enhancers (33, 38, 40), and this suggests that the transcription of all of these genes might be stimulated by the binding of a common factor.

We have already reported a partial purification of this factor from HeLa extracts (33, 40), and have shown that the partially purified factor is able to stimulate *in vitro* transcription from a rabbit β-globin gene in an enhancer-dependent manner (33). The gel retention experiments reported in the Results section suggest that the factor is common to many different cell types and conserved throughout evolution. Support for our data comes from the work of Singh *et al.* (37) and Sen and Baltimore (53), who show the existence in HeLa cell nuclear extracts of a protein (IgNF-A) able to bind to the octamer sequence of immunoglobulin enhancers, and this same sequence is also recognized by nuclear factor III (54), a HeLa cell nuclear protein required for optimal *in vitro* adenovirus DNA replication. These results are consistent with the postulated role of the octamer sequence in the activation of many different genes from a variety of species.

ACKNOWLEDGEMENTS

We thank Nina Dathan and Marion Frick for technical assistance, Petra Riedinger for help with the figures and Elizabeth Hardon, Gennaro Ciliberto and Roberto Di Lauro for comments on the manuscript.

*Present address: Department of Biochemistry, University of California, Berkeley, CA 94720, USA

REFERENCES

1. Kunkel, G.R., Maser, R.L., Calvet, J.P. and Pederson, T. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 8575-8579
2. Reddy, R., Henning, D., Das, G., Harless, M. and Wright, D. (1987) *J. Biol. Chem.*, **262**, 75-81
3. Jensen, E.G., Hellung-Larsen, P. and Frederiksen, S. (1979) *Nucleic Acids Res.*, **6**, 321-330
4. Murphy, J.T., Burgess, R.R., Dahlberg, J.E. and Lund, E. (1982) *Cell*, **29**, 265-274
5. Tamm, I., Kikuchi, T., Darnell, J.E. and Salditt-Georgiev, M. (1980) *Biochemistry*, **19**, 2743-2748
6. Busch, H., Reddy, R., Rothblum, L. and Choi, C.Y. (1982) *Ann. Rev. Biochem.*, **51**, 617-654
7. Mattaj, I.W. (1986) *Cell*, **46**, 905-911
8. Lerner, M.R. and Steitz, J.A. (1979) *Proc. Natl. Acad. Sci. USA*, **76**, 5495-5499
9. Maniatis, T. and Reed, R. (1987) *Nature*, **325**, 673-678
10. Sharp, P.A. (1987) *Science*, **235**, 766-771

11. Skuzeski, J.M., Lund, E., Murphy, J.T., Steinberg, T.H., Burgess, R.R. and Dahlberg, J.E. (1984) *J. Biol. Chem.*, **259**, 8345-8352
12. Westin, G., Lund, E., Murphy, J.T., Pettersson, U., and Dahlberg, J.E. (1984) *EMBO J.*, **3**, 3295-3301
13. Zeller, R., Carri, M.-T., Mattaj. I.W. and De Robertis, E.M. (1984) *EMBO J.*, **3**, 1075-1081
14. Lund, E., Dahlberg, J.E. and Forbes, D.J. (1984) *Mol. Cell. Biol.*, **4**, 2580-2586
15. Ares, M., Mangin, M., and Weiner, A.M. (1985) *Mol. Cell. Biol.*, **5**, 1560-1570
16. Krol, A., Lund, E., and Dahlberg, J.E. (1985) *EMBO J.*, **4**, 1529-1535
17. Mattaj, I.W., Leinhard, S., Jiricny, J. and De Robertis, E.M. (1985) *Nature*, **316**, 163-167
18. Ciliberto, G., Buckland. R., Cortese, R. and Philipson, L. (1985) *EMBO J.*, **4**, 1537-1543
19. Mangin, M., Ares, M. Jr. and Weiner, A.M. (1986) *EMBO J.*, **5**, 987-995
20 Kazmaier, M., Tebb, G. and Mattaj, I.W. (1987) *EMBO J.*, in press
21. Ciliberto, G., Palla, F., Tebb, G., Mattaj, I.W. and Philipson, L. (1987) *Nucleic Acids Res.*, **15**, 2403-2416
22. Lund, E., Bostock, C.J. and Dahlberg, J.E. (1987) *Genes and Dev.*, **1**, 47-56
23. Galas, D.J. and Schmitz, A. (1978) *Nucleic Acids Res.*, **5**, 3157-3170
24. Minowada, J., Ohnura, T. and Moore, G.E. (1972) *J. Nat. Cancer Inst.*, **49**, 891-895
25. Dignam, J.D., Lebowitz, R.M. and Roeder, R.G. (1983) *Nucleic Acids Res.*, **11**, 1475-1489
26. Wildeman, A.G., Sassone-Corsi, P., Grundström, T., Zenke, M. and Chambon, P. (1984) *EMBO J.*, **3**, 3129-3133
27. Bradford, M.M. (1976) *Anal. Biochem.*, **72**, 248-254
28. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, New York).
29. Maxam, A. and Gilbert, W. (1980) *Meth. Enzymol.*, **65**, 499-560
30. Kramer, W., Drutsa, V., Jansen, H.-W., Kramer, B., Pflugfelder, M. and Fritz, H.-J. (1984) *Nucleic Acids Res.*, **12**, 9441-9456
31. Sanger, F., Nicklen, S. and Coulsen, A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467
32. Mattaj, I.W. and De Robertis, E.M. (1985) *Cell*, **40**, 111-118
33. Bohmann, D., Keller, W., Dale, T., Schöler, H.R., Tebb, G. and Mattaj, I.W. (1987) *Nature*, **325**, 268-272
34. Fried, M. and Crothers, D.M. (1981) *Nucleic Acids Res.*, **9**, 6505-6525
35. Schneider, R., Gauder, I., Müller, U., Mertz, R. and Winnacker, E.L. (1986) *Nucleic Acids Res.*, **14**, 1303-1317

36. Briggs, M.R., Kadonaga, J.T., Bell, S.P. and Tjian, R. (1986) *Science*, **234**, 47-52
37. Singh, H., Sen, R., Baltimore, D., and Sharp, P.A. (1986) *Nature*, **319**, 154-158
38. Sive, H.L., Heintz, N. and Roeder, R.G. (1986) *Mol. Cell. Biol.*, **6**, 3329-3340
39. Sive, H.L. and Roeder, R.G. (1986) *Proc. Natl. Acad. Sci. USA*, **83**, 6382-6386
40. Tebb, G., Mattaj, I.W., Keller, W. and Bohmann, D. (1986) in "*RNA Polymerase and the Control of Transcription*", Steenbock Symposium, **16**, 421-426
41. Mattaj, I.W. and Zeller, R. (1983) *EMBO J.*, **2**, 1883-1891
42. Murphy, J.T., Skuzeski, J.T., Lund, E., Steinberg, T.H., Burgess, R.R. and Dahlberg, J.E. (1987) *J. Biol. Chem.*, **262**, 1795-1803
43. Takahashi, K., Vigneron, M., Matthes, H., Wildeman, A., Zenke, M. and Chambon, P. (1986) *Nature*, **319**, 121-126
44. Ptashne, M. (1986) *Nature*, **322**, 697-701
45 Gidoni, D., Dynan, W.S. and Tjian, R. (1984) *Nature*, **312**, 409-413
46. Dynan, W.S. and Tjian, R. (1985) *Nature*, **316**, 774-778
47. Parslow, T.G., Blair, D.M., Murphy, W.J. and Granner, D.K. (1984) *Proc. Natl. Acad. Sci. USA*, **81**, 2650-2654
48. Ephrussi, A., Church, G.M., Tonegawa, S. and Gilbert, W. (1985) *Science*, **227**, 134-140
49. Falkner, F.G. and Zachau, H.G. (1984) *Nature*, **310**, 71-74
50. Harvey, R.P., Robins, A.J. and Wells, J.R.E. (1982) *Nucleic Acids Res.*, **10**, 7851-7863
51. Korf, G.M. and Stumph, W.E. (1986) *Biochemistry*, **25**, 2041-2047
52. Gruss, P., Dhar, R. and Khoury, G. (1981) *Proc. Natl. Acad. Sci. USA*, **78**, 943-947
53. Sen, R. and Baltimore, D. (1986) *Cell*, **46**, 705-716
54. Pruijn, G.J.M., van Driel, W. and van der Vliet, P. (1986) *Nature*, **322**, 656-659
55. Manser, T. and Gesteland, R.F. (1982) *Cell*, **29**, 257-264
56. Watanabe-Nagasu, N., Itoh, Y., Tani, T., Okano, K., Koga, N., Okada, N. and Ohshima, Y. (1983) *Nucleic Acids Res.*, **11**, 1791-1801
57. Marzluff, W.F., Brown, D.T., Lobo, S., and Wang, S.-S. (1983) *Nucleic Acids Res.*, **11**, 6255-6270