Godefroy-Colburn, T., & Thach, R. E. (1981) *J. Biol. Chem. 256*, 11762–11773.
Jen, G., & Thach, R. E. (1982) *J. Virol. 43*, 250–261.
Jenkins, N. A., Kaumeyer, J. F., Young, E. M., & Raff, R. A. (1978) *Dev. Biol. 63*, 279–298.
Kozak, M. (1984a) *Nucleic Acids Res. 12*, 857–872.
Kozak, M. (1984b) *Nature (London) 308*, 241–246.
Lee, G. T.-Y., & Engelhardt, D. L. (1979) *J. Mol. Biol. 129*, 221–233.
Liautard, J. P., & Egly, J. M. (1980) *Nucleic Acids Res. 8*, 1793–1804.
Liebold, E. A., Aziz, N., Brown, A. J. P., & Munro, H. N. (1984) *J. Biol. Chem. 259*, 4327–4334.
Linder, M. C., & Munro, H. N. (1972) *Anal. Biochem. 48*, 266–278.
Lodish, H. F. (1976) *Annu. Rev. Biochem. 45*, 39–72.
Lodish, H. F., & Porter, M. (1980) *J. Virol. 36*, 719–733.
Lowry, O. H., Rosebrough, N. J., Farr, A. L., & Randall, R. J. (1951) *J. Biol. Chem. 193*, 265–275.
Preobrazhensky, A. A., & Spirin, A. S. (1978) *Prog. Nucleic Acid Res. Mol. Biol. 21*, 1–38.
Ray, A., Walden, W. E., Brendler, T., Zenger, V. E., & Thach, R. E. (1985) *Biochemistry 24*, 7525–7532.
Ray, B. K., Brendler, T. G., Adya, S., Daniels-McQueen, S., Miller, J. K., Hershey, J. W. B., Grifo, J. A., Merrick, W. C., & Thach, R. E. (1983) *Proc. Natl. Acad. Sci. U.S.A. 80*, 663–667.
Sargan, D. R., Gregory, S. P., & Butterworth, P. H. W. (1982) *FEBS Lett. 147*, 133–136.
Sarkar, G., Edery, I., Gallo, R., & Sonenberg, N. (1984) *Biochim. Biophys. Acta 783*, 122–129.
Schaefer, F. V., & Theil, E. C. (1981) *J. Biol. Chem. 256*, 1711–1715.
Shull, G. E., & Theil, E. C. (1982) *J. Biol. Chem. 257*, 14187–14191.
Shull, G. E., & Theil, E. C. (1983) *J. Biol. Chem. 258*, 7921–7923.
Walden, W. E., & Thach, R. E. (1982) in *Interaction of Translational and Transcriptional Controls in the Regulation of Gene Expression* (Grunberg-Manago, M., & Safer, B., Eds.) pp 399–416, Elsevier Biomedical, New York.
Walden, W. E., Godefroy-Colburn, T., & Thach, R. E. (1981) *J. Biol. Chem. 256*, 11739–11746.
Watanabe, N., & Drysdale, J. (1981) *Biochem. Biophys. Res. Commun. 103*, 207–212.
Yenofsky, R., Bergmann, I., & Brawermann, G. (1982) *Proc. Natl. Acad. Sci. U.S.A. 79*, 5876–5880.
Zahringer, J., Baliga, B. S., & Munro, H. N. (1976) *Proc. Natl. Acad. Sci. U.S.A. 73*, 857–861.

Chicken U2 and U1 RNA Genes Are Found in Very Different Genomic Environments but Have Similar Promoter Structures†

Gina M. Korf‡ and William E. Stumph*
Department of Chemistry and Molecular Biology Institute, San Diego State University, San Diego, California 92182
Received September 10, 1985; Revised Manuscript Received December 2, 1985

ABSTRACT: We have cloned and analyzed a gene that codes for chicken U2 small nuclear RNA (snRNA). In the haploid chicken genome, there are approximately 35–40 copies of the U2 RNA gene arranged in tandemly repeated units 5.35 kilobase pairs in length. This U2 gene organization contrasts with that of chicken U1 RNA genes, which are found in heterogeneous genomic environments. Although U snRNA genes are transcribed by RNA polymerase II, they lack the usual TATA and CAAT homologies found in the 5′ control regions of most RNA polymerase II transcription units. Nevertheless, a comparison of chicken U2 and U1 RNA gene 5′-flanking DNA sequences reveals two upstream blocks of homology which are also evolutionarily conserved in U2 and U1 RNA genes of other vertebrate species. The first block of conserved sequence is centered around position −55 relative to the RNA cap site, and the other is located near position −200. Interestingly, stretches of sequence with the potential to form Z DNA are located either within or immediately adjacent to both of these two conserved upstream sequence elements, suggesting a possible role for Z DNA in U1/U2 gene expression. Moreover, the chicken U2 and U1 gene promoter regions also contain specific short sequences (i.e., the hexamer GGGCGG and the octamer ATGCAAAT) that have been shown to be required for the expression of a number of mRNA-encoding genes. These findings suggest that the transcription of snRNA genes is controlled by a complex set of factors, some shared with other RNA polymerase II transcription units and others which may be unique to the snRNA genes.

The small nuclear RNAs (snRNAs)[1] of the U family represent a special class of RNA polymerase II transcripts (Eliceiri, 1980; Roop et al., 1981; Murphy et al., 1982; Mattaj & Zeller, 1983) which are metabolically stable and are not polyadenylated. With the exception of U6 RNA, they also contain an unusual 2,2,7-trimethylguanosine cap structure instead of the 7-methylguanosine cap found on mRNA molecules [for a review, see Busch et al. (1982)]. The snRNAs U1–U6, as well as the recently discovered U7–U10 snRNAs (Strub et al., 1984; Reddy et al., 1985), exist in vivo as integral

† This work was supported by grants to W.E.S. from the National Science Foundation (PCM-8310683) and the National Institutes of Health (GM-33512) and in part by the California Metabolic Research Foundation.
* Correspondence should be addressed to this author.
‡ G.M.K. is a predoctoral student in the Department of Biology, San Diego State University.

[1] Abbreviations: snRNA, small nuclear RNA; snRNP, small nuclear ribonucleoprotein particle; bp, base pair(s); kbp, kilobase pair(s).

0006-2960/86/0425-2041$01.50/0

components of small nuclear ribonucleoprotein particles (snRNPs) which are recognized by antibodies from patients with certain autoimmune diseases (Lerner & Steitz, 1979).

Substantial evidence now exists that U1 and U2 snRNPs play a role in the removal of intervening sequences from mRNA precursors (Mount et al., 1983; Padgett et al., 1983; Kramer et al., 1984; DiMaria et al., 1985; Krainer & Maniatis, 1985; Black et al., 1985). The U7 snRNPs have been shown to be involved in the maturation of the 3′ ends of histone mRNA (Strub et al., 1984; Birchmeier et al., 1984). It therefore seems likely that the other U snRNPs may also function in RNA processing events in the cell nucleus.

Our laboratory is interested in elucidating the molecular mechanisms involved in the expression of the genes that code for the snRNAs. We have previously reported the cloning and sequencing of four chicken U1 RNA genes (Roop et al., 1981; Earley et al., 1984). Although these genes lack TATA boxes, there are two evolutionarily conserved regions of high sequence homology in the 5′-flanking DNA. The proximal conserved region consists of the first 60 nucleotides of 5′-flanking DNA, while the distal conserved region is located further upstream at approximately position −200 with respect to the RNA cap site (Earley et al., 1984). The presence of these evolutionarily conserved homologies suggests that DNA sequences in these two distinct regions are essential components of the promoter for chicken U1 RNA gene expression.

This concept is supported by the fact that the 5′-flanking DNAs of cloned U1 and U2 RNA genes of other vertebrate species also contain homologous sequences at similar upstream positions (Marzluff et al., 1983; Watanabe-Nagasu et al., 1983; Tani et al., 1983; Earley et al., 1984; Zeller et al., 1984; Westin et al., 1984a,b; Ares et al., 1985). Moreover, deletions or mutations in the 5′-flanking DNA that remove or disrupt these conserved regions have significant down effects on the transcriptional activity of human and *Xenopus* U1 and U2 RNA genes injected into frog oocytes (Skuzeski et al., 1984; Westin et al., 1984b; Krol et al., 1985; Ciliberto et al., 1985; Mattaj et al., 1985; Ares et al., 1985). Two main conclusions have been drawn from these studies. First, a proximal region (nucleotide positions about −50 to −60) appears to be functionally analogous to a TATA box and is required for the accurate initiation of snRNA transcription (Skuzeski et al., 1984; Ciliberto et al., 1985). Second, a distal region near position −200 has properties similar to (Mattaj et al., 1985), but possibly not identical with (Ares et al., 1985), a classical transcription enhancer element (Gruss, 1984).

In order to study these phenomena more fully in the chicken system, we have now cloned and characterized a gene that codes for chicken U2 RNA. In this paper, we report the genomic organization and nucleotide sequence of this chicken U2 RNA gene, and we compare its structure with that of chicken U1 RNA genes. We find that U2 and U1 RNA genes are organized very differently from each other in the chicken genome but that they have promoter regions that are quite similar in structure, including the two conserved sequence elements mentioned above. Finally, we propose that other specific DNA sequences, and potentially Z-conformation DNA, may also be involved in chicken snRNA gene expression.

Experimental Procedures

Isolation of a Genomic DNA Clone Containing a Chicken U2 RNA Gene. A DNA clone containing a human U2 RNA pseudogene was used as a hybridization probe to screen a chicken genomic DNA library. The human U2 pseudogene clone U2.21.6 (Van Arsdell & Weiner, 1984) was a gift of A. M. Weiner. In this clone, the U2 sequence is contained in a 400 base pair (bp) *Msp*I fragment ligated into the mp8 vector. Replicative form DNA was prepared, and the 400 bp fragment was separated from the vector by *Eco*RI and *Hin*dIII double digestion, followed by purification through agarose and polyacrylamide gels.

This fragment was nick-translated and used to screen a chicken genomic DNA library obtained from C. Hodgson, M.-J. Tsai, and B. W. O'Malley. This library consists of hen oviduct DNA fragments generated by partial cleavage with *Mbo*I and cloned into the polylinker *Bam*HI site of the λ-phage vector EMBL-4. Approximately 100 000 phage plaques were screened by using the following hybridization conditions: 0.9 M NaCl, 0.09 M sodium citrate (pH 7), 0.04% poly(vinylpyrrolidone), 0.04% Ficoll, 0.04% bovine serum albumin, 0.5% sodium dodecyl sulfate, 1 mM ethylenediaminetetraacetate, 64 °C. After three rounds of plaque purification, a single intensely hybridizing clone was obtained and designated λU2-6 (Figure 1).

A 4.0 kilobase pair (kbp) fragment from λU2-6 (defined by a *Bam*HI site at one end and extending to the *Eco*RI site of the polylinker at the other end) was found to hybridize to the U2 probe. This 4.0 kbp fragment (Figure 1) was isolated and subcloned into the *Bam*HI–*Eco*RI sites of pBR322, generating the clone p4.0(λU2-6). The 1.6 kbp *Sst*I fragment was isolated from this plasmid and used for DNA sequencing.

To obtain a U2 coding sequence specific hybridization probe, a 170 base pair *Sau*3A–*Sma*I fragment, containing 163 bp of U2 coding region and only 6 bp of 3′-flanking sequence, was subcloned into the *Bam*HI–*Sma*I sites of the plasmid pSP65 (Promega Biotech). It was reisolated from this plasmid and nick-translated to use as a probe in the genomic DNA Southern blot experiments.

DNA Sequencing. DNA sequencing was carried out by using the chemical degradation method of Maxam and Gilbert (1977). Restriction fragments were 5′ end labeled with T4 polynucleotide kinase and [γ-^{32}P]ATP. Fragments were labeled at the 3′ end by using DNA polymerase I (Klenow fragment) or T4 DNA polymerase and deoxyribonucleotide [α-^{32}P]triphosphates, or, in the case of 3′ overhanging ends, by using terminal deoxynucleotidyl transferase and [α-^{32}P]-ddATP. The sequencing strategy is shown at the bottom of Figure 1.

Analysis of DNA by Southern Blotting. DNA fragments of restriction endonuclease digestions were separated by electrophoresis in 0.75–1.5% agarose gels and transferred to nitrocellulose or nylon filters by the method of Southern (1975). Typically, 1–2 μg of cloned DNA or 10–20 μg of genomic DNA was loaded per gel lane. Genomic DNA was isolated from the liver of a single hen as described in Maniatis et al. (1982). For the copy number blot, 15 μg of chicken genomic DNA was loaded per lane. For use as an internal standard in the copy number blot experiment, a 3.2 kbp *Eco*RI–*Xmn*I fragment was isolated from the plasmid p4.0-(λU2-6). On the basis of a chicken haploid genome size of 1.26 pg of DNA (Maxson et al., 1983), it was calculated that 44 pg would represent a single copy equivalent of a 3.2 kbp fragment in 15 μg of genomic DNA. This amount and multiples of it were loaded, along with 15 μg of *Bam*HI-digested liver DNA, into individual lanes of a gel for use in the copy number blot experiment of Figure 5.

Results

Isolation and Characterization of a Clone Containing a Chicken U2 RNA Gene. Approximately 100 000 plaques from a chicken genomic DNA λ-phage library were screened by

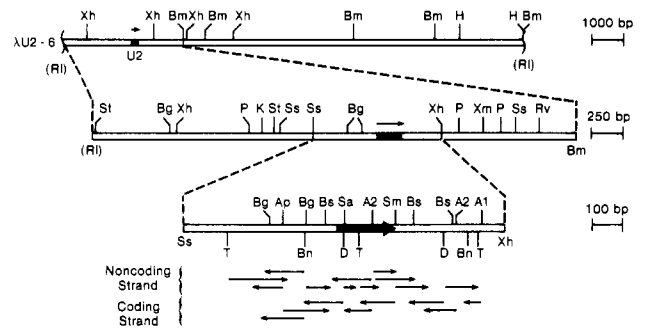

FIGURE 1: Restriction map of a DNA clone containing a chicken U2 RNA gene. The 15 kbp chicken DNA insert in the λ-phage vector EMBL-4 is shown at the top. Wavy lines indicate *Eco*RI sites in the polylinker of the vector. The U2 RNA coding region is illustrated as a bold arrow with the arrowhead indicating transcriptional orientation. An expanded view of the subcloned 4.0 kbp *Eco*RI to *Bam*HI fragment is shown below the phage clone map. A more detailed map of the region sequenced is shown in the third line of the figure. The sequencing strategy is shown at the bottom for the coding and noncoding strands. Restriction enzyme abbreviations: A1, *Aval*I; A2, *Aval*II; Ap, *Apa*I; Bg, *Bgl*I; Bm, *Bam*HI; Bn, *Ban*I; Bs, *Bss*HII; D, *Dde*I; H, *Hin*dIII; K, *Kpn*I; P, *Pst*I; RI, *Eco*RI; RV, *Eco*RV; Sa, *Sau*3A; Sm, *Sma*I; Ss, *Sst*I; St, *Stu*I; T, *Taq*I; Xh, *Xho*I; Xm, *Xmn*I.

using as a hybridization probe a clone that contains human U2 DNA sequences. A single intensely hybridizing clone was obtained. A restriction map of this clone, λU2-6, is shown at the top of Figure 1. Southern blot analysis (not shown) revealed that the U2 sequences were located in a 4.0 kbp *Eco*RI–*Bam*HI fragment positioned at the left end of the phage clone as oriented in Figure 1. This 4.0 kbp fragment was subcloned into pBR322, and finer restriction mapping of the resultant plasmid, p4.0(λU2-6), indicated the presence of a single copy of a U2 RNA gene (Figure 1).

Sequence of the U2 RNA Gene Clone. (A) Coding Region. Figure 2A presents 360 bp of 5′-flanking DNA sequence, 187 bp of U2 coding region, and 293 bp of 3′-flanking sequence determined according to the strategy outlined at the bottom of Figure 1. To our knowledge, only a partial sequence has been published for chicken U2 RNA (Branlant et al., 1982), corresponding to the 3′ portion of the molecule (residues 81 through 187). The DNA sequence in clone λU2-6 and the published RNA sequence are in perfect agreement over this region. Moreover, the chicken DNA clone that we have isolated would code for an RNA that is identical in sequence to the 5′ portion of mammalian U2 RNAs in the region spanned by nucleotides 1–95, a region which is 100% conserved in humans, mice, and rats (Reddy, 1985). For these reasons, we are confident that the clone λU2-6 contains a bona fide chicken U2 RNA gene. This conclusion is independently supported by the presence of conserved transcriptional signals in the flanking DNA and by the results of experiments studying the genomic organization of chicken U2 RNA genes as discussed below.

(B) 5′-Flanking Region. The 5′-flanking DNA of the chicken U2 gene is very GC rich, and as with other snRNA genes does not contain a TATA box characteristic of most genes transcribed by RNA polymerase II. However, when the U2 gene 5′-flanking sequences were compared with four different chicken U1 RNA gene sequences previously determined (Earley et al., 1984), two upstream regions of significant homology could be identified at similar locations. The homologies are shown in Figure 2B. The conserved region most proximal to the RNA cap site is located near position −55. Here, 12 of 15 nucleotides are conserved among the chicken U2 and U1 RNA genes. A second and more distal region is located near position −200. In this region, the U2 and U1

A

```
-360       -350       -340       -330       -320       -310       -300       -290       -280       -270       -260       -250
GCCGGCCGCA ACCCCTTCGG CTCGTCGCCC CGCGCTGACA GAACGGCTCC CGACTGCGGG CCGTTCTGTC GCTGCTGCCG GAGCGGAGCT GCCGCGCCGT GGGAGGGAGG CGGCGGGTGA

-240       -230       -220       -210       -200       -190       -180       -170       -160       -150       -140       -130
GGGGCGGGGC CATGCAAATC GGAGCCGCGG GGGCGGCCGG GAGCGCTCCG GCACGCTGCA CCGGGGCCCG TTGGGGCCGG CACGGGGGCG GAGCGGGAGC GGGAGCGGGA GCGGGGCGGT

-120       -110       -100       -90        -80        -70        -60        -50        -40        -30        -20        -10
GCGGGGCGGG GGGCGGCGCC GTGGCGGCGG GGCCGGGGTG GGTTTGGGCG CGGGGTGCGC GTCCGTGAGC TGTCGGGGGC GCGCGTGGGG GTGGGCGCTG TCCGGGAAGA GCGCGGGTTT

1          11         21         31         41         51         61         71         81         91         101        111
ATCGCTTCTC GGCCTTTTGG CTAAGATCAA GTGTAGTATC TGTTCTTATC AGTTTAATAT CTGATACGTC CTCGATGAGA GGACTTTATA TTAAACGGAT TTTTGGGCGC GGGAGTTGGA

121        131        141        151        161        171        181        +4         +14        +24        +34        +44
CCCGGAGCTT GCTCCCTCCG CTCCGCGCAT CGTCCCGGTA TGGCAGTACC TCCGGGCACG GTGCACCTCC CCCGGGAGGA ATGTGGCGTG GTGAAAGGAG AGAAGGAAGG CGGGGCGGTG

+54        +64        +74        +84        +94        +104       +114       +124       +134       +144       +154       +164
GGGTTGGGCG CGCGGGAGGA GGGTGGCGCT GTGTCCCGCG GGGGGGCGCA GTCGGCTTCG CGGCGCGGCG CGGCGCGGCG CGGCGCGGCG CGGCGCGGCG CGGCGCTGCG TTCGGTTGCG

+174       +184       +194       +204       +214       +224       +234       +244       +254       +264       +274       +284
CTGAGCGCCG TCAGCGGAGC GCGGCCCTGC GCTGTGCGCG CCGCTCGGAC CGCGGGGGAG GGGGCGTTGG CGGCGGCCGC AGCCGCTGGT GGCGCCTCCC TCCGTCTCCC TGCCGGCCGC
```

B

```
                                        "-200" REGION                                          "-55" REGION

CHICKEN U2              -239  GGGCGGGGCCATGCAAATCGGAG-CCGCGGGG-GCG-GCCGGGAGC  -197          -58  CCGTGAGCTGTCGGG  -44
                              ******** ********* * ***** * ** **********                          ***** * ** ****
CHICKEN U1 (CONSENSUS)  -236  GGGYGGGGAYATGCAAATNRANNNCCGCGCGCTGCATGCCGGGAGC  -191          -57  CCGTGCGGTR-CGGG  -44

                                        35/46 = 76% HOMOLOGY                                   12/15 = 80% HOMOLOGY
```

FIGURE 2: Sequence of the chicken U2 RNA gene clone. (A) The sequence of the U2 RNA coding region is shown in large capital letters, corresponding to positions 1 through 187. The 5′-flanking DNA is labeled with negative numbers. The two regions underlined are those that are shown again in part B of the figure and are sequences presumed to be important for expression of the gene. The nucleotides of the 3′-flanking DNA are labeled with numbers preceded by + signs. The noncoding strand is shown. (B) Sequence comparisons showing the homologies present at similar locations upstream of chicken U2 and U1 RNA genes. Asterisks indicate matched bases. R stands for purine and Y for pyrimidine. Dashes have been inserted to maximize alignment. The underlined sequence octamer ATGCAAAT is discussed in the text.

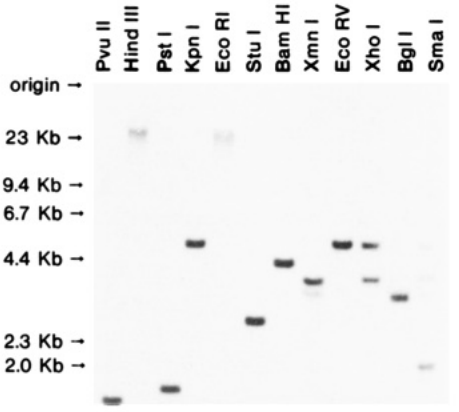

FIGURE 3: Genomic organization of chicken U2 RNA genes analyzed by Southern blot analysis. DNA isolated from the liver of a single hen was digested with the indicated enzymes, electrophoresed on a 0.9% agarose gel, blotted to a nylon filter, and probed with a nick-translated cloned 170 bp fragment specific for U2 coding sequences. The arrows indicate the distances migrated by *Hin*dIII fragments of λ DNA run in adjacent lanes as size markers. The results demonstrate that all detectable chicken U2 RNA genes are present in identical or very similar genomic DNA environments.

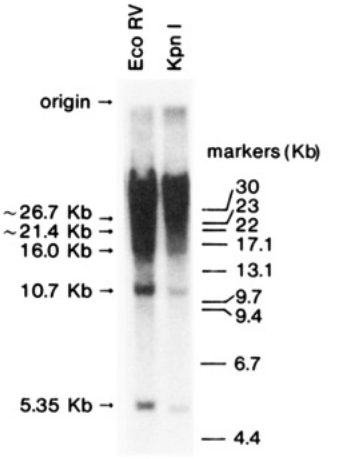

FIGURE 4: Chicken U2 RNA genes are organized as 5.35 kbp tandem repeats. Chicken genomic DNA was partially digested with *Kpn*I or *Eco*RV, electrophoresed on a 0.75% agarose gel, transferred to a nylon filter, and probed with nick-translated U2 coding sequences. The autoradiogram shows that each enzyme produces an identical ladder of partial digestion products corresponding to fragment sizes that are multiples of 5.35 kbp. The lines and numbers on the right side correspond to the distances migrated by *Bgl*II, *Hin*dIII, and *Kpn*I fragments of λ DNA run in adjacent lanes as size markers.

RNA genes share 35 of 46 nucleotides in common (Figure 2B). The conservation of these two blocks of homology strongly suggests that they are functionally important sequences for chicken snRNA gene expression. Of particular interest is the presence of the octamer ATGCAAAT as a conserved component of the "−200" region. This sequence octamer has been shown to be involved in the expression of immunoglobulin genes (Parslow et al., 1984; Falkner & Zachau, 1984; Mason et al., 1985) and, very recently, in the expression of human and *Xenopus* U2 RNA genes (Ares et al., 1985; Mattaj et al., 1985).

(*C*) *3′-Flanking Region.* The 3′-flanking DNA of the chicken U2 RNA gene sequence shown in Figure 2A has a couple of interesting features. First, the sequence AAAGGAGAGA found at positions +27 through +36 resembles an evolutionarily conserved consensus sequence, AAAPuNNAGA, found downstream of vertebrate snRNA genes. This sequence has been shown to be required for 3′ end formation of human U1 and U2 snRNAs (Yuo et al , 1985; Hernandez, 1985). Second, further downstream in the 3′ direction, the pentanucleotide CGCGG is directly repeated 9 times between positions +112 and +156. The significance of this pentanucleotide repeat is not known. Interestingly, however, pentameric repeats of different sequence have also been observed upstream of some chicken U1 RNA true gene and pseudogene loci (Kristo et al., 1984), within the second intron of the human ζ-globin gene (Proudfoot et al., 1982), within the first intron of chicken δ-crystallin genes, and within the switch region of murine immunoglobulin heavy-chain genes (Piatigorsky, 1984). Perhaps such pentameric repeats play a fundamental role in genomic evolution or function.

Genomic Organization of Chicken U2 RNA Genes. To investigate the genomic environment of chicken U2 RNA genes, we performed Southern blot analyses of genomic DNA. DNA prepared from the liver of a single hen was digested with 1 of 12 different restriction enzymes, blotted to nylon filters, and probed with a labeled DNA fragment specific for U2 RNA coding sequences. The results of this experiment are shown in Figure 3. A general pattern of a single strongly hybridizing fragment for each enzyme digestion is observed. Exceptions occur in the case of two enzymes (*Xho*I and *Sma*I) that are sensitive to CpG methylation of their recognition sites; these results are discussed further below. Upon very long exposures, light bands were sometimes observable in other lanes, but these minor bands (including the lower band in the *Xmn*I digestion of Figure 3) were not reproducible from experiment to experiment (data not shown). Although minor heterogeneity cannot be rule out, we can conclude from this experiment that at least most U2 RNA genes in the chicken genome are present in remarkably similar environments.

The results of the experiment shown in Figure 3 are consistent with the possibility that the chicken U2 RNA genes may be organized into an array of tandemly repeated units that contain no recognition sites for *Eco*RI and *Hin*dIII and contain a single recognition site per repeating unit for each of the enzymes *Kpn*I and *Eco*RV. This is suggested by the fact that *Eco*RI and *Hin*dIII produce hybridizing bands that run at the limit mobility of the gel, whereas the next largest bands (*Kpn*I and *Eco*RV lanes) comigrate with a mobility corresponding to a length of 5.35 kbp. Partial digestion products of the same size (5.35 kbp) are also present in the *Xho*I and *Sma*I lanes. These results suggest the possibility that the chicken U2 genes are contained within 5.35 kbp tandemly repeated units. To confirm this arrangement, the experiment shown in Figure 4 was performed. Genomic DNA was digested only partially with *Eco*RV or *Kpn*I and then subjected to Southern blot analysis. The results show that both of these enzymes produce an identical ladder of bands that are multiples of 5.35 kbp in size. This result is expected only if U2 RNA genes are organized in the chicken genome as an array of 5.35 kbp tandemly repeated units. Each unit is recognized once by *Kpn*I and *Eco*RV and not at all by *Eco*RI and *Hin*dIII. The other enzymes used in Figure 3 cut the repeating unit more than once.

In order to determine the number of U2 RNA gene copies in the chicken genome, we performed the experiment shown in Figure 5. Genomic DNA was digested with *Bam*HI which yields a single 4.3 kbp band that hybridizes with a U2 probe on Southern blot analysis (Figure 3). To these genomic DNA digestions, we added increasing amounts of a 3.2 kbp DNA fragment corresponding to 1, 3, 9, 20, 50, or 100 copies of the U2 RNA gene. The autoradiogram shown in Figure 5 reveals that the haploid chicken genome contains between 20 and 50

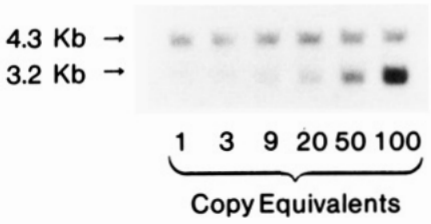

FIGURE 5: Determination of the number of U2 RNA gene copies in the chicken genome. DNA isolated from hen liver was digested with *Bam*HI which yields a single 4.3 kbp band visible upon Southern analysis with a U2 probe (Figure 3). Precisely 15 μg of *Bam*HI-digested genomic DNA was loaded per lane on a 0.9% agarose gel. Also included in each lane were increasing amounts of a cloned 3.2 kbp fragment corresponding to the equivalent of 1, 3, 9, 20, 50, or 100 copies per haploid chicken genome. Densitometer scans of films exposed for various periods of time provided an estimate of 35–40 copies of the U2 RNA gene in the haploid chicken genome.

U2 RNA gene copies. Densitometry of films exposed for various periods of time provided an estimate of 35–40 copies of the U2 RNA gene per haploid chicken genome.

Discussion

Organization of U2 RNA Genes in the Chicken Genome. We have cloned a 15.2 kbp region of chicken DNA that contains a gene coding for U2 RNA. We find that there are about 35–40 copies of the U2 RNA gene present in the haploid chicken genome arranged as tandemly repeated units 5.35 kbp in length. Human and *Xenopus* U2 RNA genes have also been shown to exist in similar tandemly repeated arrays with repeating units of 6.2 and 0.83 kbp, respectively (Van Arsdell & Weiner, 1984; Westin et al., 1984a; Mattaj & Zeller, 1983).

The restriction map of the clone we have isolated (λU2-6) is completely consistent with the genomic mapping data shown in Figures 3 and 4 with the exception of one important feature. Since the clone λU2-6 itself contains no 5.35 kbp internal repetitions, it must contain the region of the chicken genome at the very end of the U2 gene tandemly repeated array. In all other respects, λU2-6 appears to be a faithful representative of the predominant (and perhaps only) class of U2 DNA sequences present in the chicken genome. The isolation of a clone containing an end unit is perhaps not surprising since tandem repeats often do not clone well in λ vectors.

Our results also agree with those of another group (Tani et al., 1983) who probed genomic Southern blots of chicken DNA with a heterologous (rat) U2 DNA probe. They digested chicken DNA with two enzymes, *Eco*RI or *Pst*I, and also observed only a single band of hybridization in each case. The bands in their experiment agree in size with the bands we observe for the same two enzymes in Figure 3.

An interesting phenomenon is revealed by the *Xho*I and *Sma*I digestions in the genomic blot of Figure 3. Although the 35–40 U2 genes are clearly in the same DNA sequence environment, the methylation status of *Xho*I and *Sma*I sites around individual U2 genes can clearly vary. This is indicated by the fact that more than one band is seen upon digestion of genomic DNA with these two enzymes. Since the degree of methylation of a gene is often related to its transcriptional activity (Felsenfeld & McGhee, 1982), these results may be an indication that not all of the U2 genes are equally expressed in the liver, the tissue from which this DNA was isolated. Mangin et al. (1985) have recently observed that the total amount of U1 RNA present in cells is relatively independent of the U1 gene copy number. They suggested that U1 RNA gene expression is subject to a negative feedback control mechanism that regulates the overall synthesis of U1 RNA. The differential methylation of U2 genes that we observe may be a reflection of this sort of feedback control, which could act to alter the expression of a certain proportion of the U2 RNA genes.

Organization of U1 vs. U2 RNA Genes in the Chicken Genome. The organization of chicken U2 RNA genes described above contrasts significantly with the organization of U1 genes in the same organism. Southern blots of chicken genomic DNA reveal multiple bands of hybridization when probed with U1 DNA sequences (Roop et al., 1981). This indicates that chicken U1 genes are found in many different genomic environments. Moreover, at least some U1 RNA genes are closely clustered within 2 kbp of each other in the chicken genome but are not tandem repeats and sometimes have opposing transcriptional orientations (Earley et al., 1984). Finally, the region of sequence identity shared by different chicken U1 RNA genes spans only the coding region and about 59 bp of 5′-flanking DNA (Earley et al., 1984). In contrast, the individual copies of chicken U2 RNA genes are found in apparently identical DNA sequence environments as part of tandemly repeated units 5.35 kbp in length.

These findings indicate that U1 and U2 RNA genes have had a quite different evolutionary history in the chicken. A reasonable interpretation is that U1 RNA genes have remained stable in the chicken genome for an evolutionarily long period of time. As a result, they have maintained their homogeneity only in the regions that are critical for their expression and function. In contrast, the chicken U2 RNA gene family may be derived from a relatively recent event in evolutionary time, possibly from the recent amplification of a single genetic locus.

Features of the Chicken U2 and U1 RNA Gene Promoter Regions. Although chicken U2 and U1 RNA genes are organized very differently, they have nevertheless retained important similar features in their 5′-flanking DNA. The most outstanding feature is the conservation of two specific blocks of sequence homology located approximately 55 and 200 bp upstream of the respective RNA cap sites (Figure 2B). Homologous sequences are also found at similar locations upstream of other vertebrate U1 and U2 RNA genes (Marzluff et al., 1983; Watanabe-Nagasu et al., 1983; Tani et al., 1983; Zeller et al., 1984; Westin et al., 1984a,b; Ares et al., 1985). Sequences in these regions have been demonstrated to be required for the efficient expression of human and frog U1 and U2 RNA genes after injection into *Xenopus* oocytes (Skuzeski et al., 1984; Westin et al., 1984b; Krol et al., 1985; Ciliberto et al., 1985; Mattaj et al., 1985; Ares et al., 1985).

A particularly interesting feature of the distal upstream region is the presence of the sequence octamer ATGCAAAT (Figure 2B). This octamer is perfectly conserved in the chicken U2 RNA gene −200 region and in three of the four chicken U1 RNA genes that we have previously cloned (Earley et al., 1984). In the case of the fourth chicken U1 RNA gene (U1 52b), the related sequence ATGCAGAT is repeated twice within the same region (Earley et al., 1984). The consensus octamer (ATGCAAAT) is found also in the promoter regions of immunoglobulin genes (Parslow et al., 1984; Falkner & Zachau, 1984) and in the heavy-chain enhancer (Ephrussi et al., 1985). Mattaj et al. (1985) have recently demonstrated that this octamer plays a role in *Xenopus* U2 RNA gene expression in a manner analogous to that of a classical enhancer element. It is reasonable to assume that the octamer is similarly involved in the chicken U1 and U2 RNA gene expression.

In Figure 6 we point out some additional potentially important features of the chicken U1 and U2 RNA gene pro-

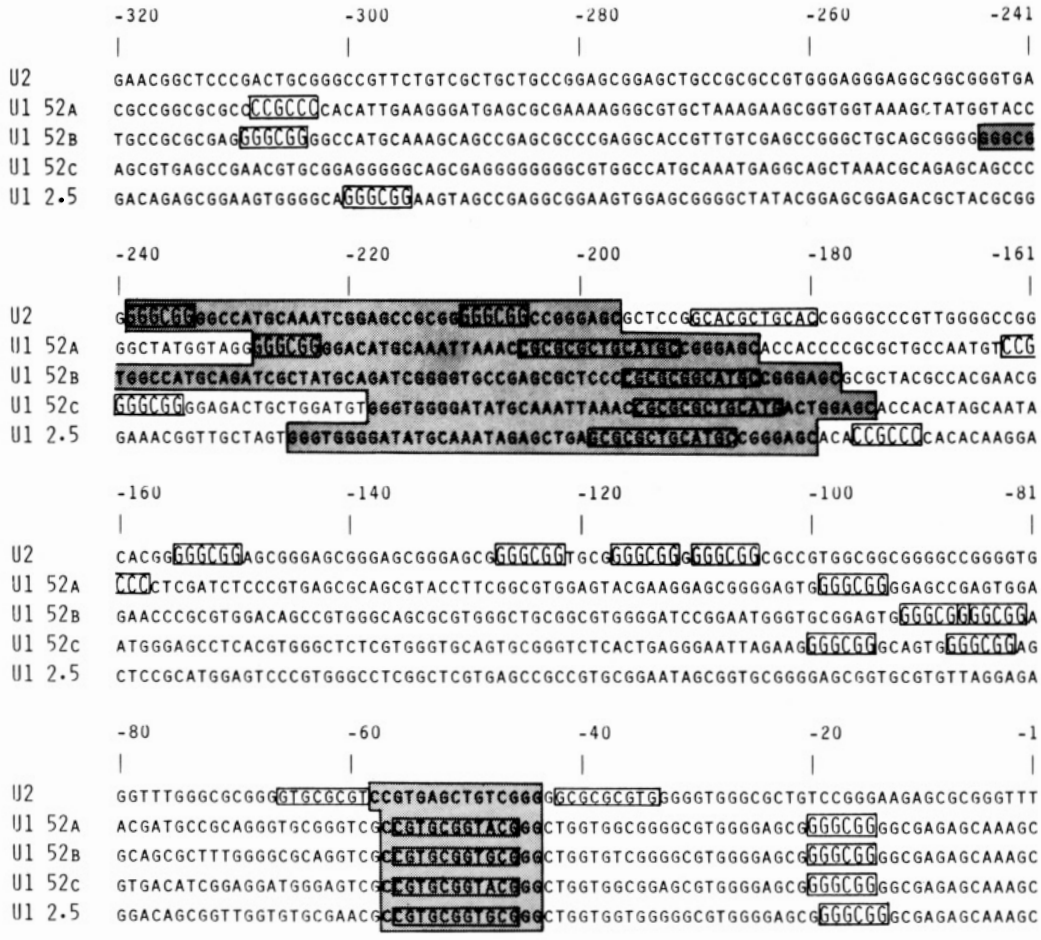

FIGURE 6: Features of the promoter regions of chicken U2 and U1 RNA genes. A total of 320 nucleotides of 5′-flanking DNA is shown for a chicken U2 RNA gene and for four chicken U1 RNA genes. The U1 gene sequences are taken from Earley et al. (1984). The shaded regions correspond to the distal and proximal regions of homology shared by all five genes as defined by the sequence comparisons shown in Figure 2B. [In the U1 52b gene, some of the conserved sequences are duplicated (Earley et al., 1984).] Boxes containing large capital letters indicate the locations of the hexanucleotide GGGCGG or its complement. The boxes around small capital letters indicate sequences that have the potential to exist as Z-conformation DNA. Sequences with a minimum of 8 strictly alternating purines/pyrimidines or with a minimum length of 11 nucleotides with no more than a single out of register nucleotide were assumed capable of adopting the Z conformation (Nordheim & Rich, 1983). The boxed sequences are the only ones in the region shown which meet these criteria.

moter regions. The shaded areas designate the distal −200 and proximal −55 regions of homology as defined by the sequence comparisons shown in Figure 2B. One of the features shown in Figure 6 is the presence of the hexanucleotide sequence GGGCGG at multiple locations in the 5′-flanking DNA. This sequence and its complement are designated by large boxed capital letters. The GGGCGG hexamer is a recognition site for the eucaryotic transcription factor Sp1 (Dynan & Tjian, 1983; Gidoni et al., 1984). There are between three and six copies of this "GC box" upstream of each of the chicken snRNA genes shown, suggesting that the Sp1 factor may play a role in snRNA gene expression.

Finally, sequences with the potential to form Z DNA are shown by boxes around small capital letters in Figure 6. Perhaps the most interesting feature of these potential Z-DNA sequences is their location either within or very near to the conserved −200 and −55 regions. This fact suggests a possible interaction between Z DNA and other DNA sequences required for chicken snRNA gene expression. A possible role of Z DNA in the upstream regions of snRNA genes has previously been discussed (Earley et al., 1984; Krol et al., 1985).

In conclusion, our sequence data indicate that the control of chicken snRNA gene expression may be highly complex. Our findings are consistent with a mosaic model of enhancer and promoter structure and suggest that a combination of discrete transcription factors interacting at multiple regulatory sites may be responsible for promoting specific snRNA gene expression.

ACKNOWLEDGMENTS

We thank Drs. C. Hodgson, M.-J. Tsai, and B. W. O'Malley for the chicken genomic DNA library and A. M. Weiner for the human U2 pseudogene clone. We also thank K. Roebuck for helpful discussions and Dr. S. Bernstein for a critical reading of the manuscript. Finally, we acknowledge Ruth Rice for typing the manuscript.

Registry No. DNA (chicken RNA U2 gene), 101077-51-4.

REFERENCES

Ares, M., Jr., Mangin, M., & Weiner, A. M. (1985) *Mol. Cell. Biol. 5*, 1560–1570.

Birchmeier, C., Schumperli, D., Sconzo, G., & Birnstiel, M. L. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 1057–1061.

Black, D. L., Chabot, B., & Steitz, J. A. (1985) *Cell (Cambridge, Mass.) 42*, 737–750.

Branlant, C., Krol, A., Bel, J.-P., Lazar, E., Haendler, B., & Jacob, M. (1982) *EMBO J. 1*, 1259–1263.

Busch, H., Reddy, R., Rothblum, L., & Choi, Y. C. (1982) *Annu. Rev. Biochem. 51*, 617–654.

Ciliberto, G., Buckland, R., Cortese, R., & Philipson, L. (1985) *EMBO J. 4*, 1537–1543.

DiMaria, P. R., Kaltwasser, G., & Goldenberg, C. J. (1985) *J. Biol. Chem. 260*, 1096–1102.
Dynan, W. S., & Tjian, R. (1983) *Cell (Cambridge, Mass.) 35*, 79–89.
Earley, J. M., Roebuck, K. A., & Stumph, W. E. (1984) *Nucleic Acids Res. 12*, 7411–7421.
Eliceiri, G. L. (1980) *J. Cell. Physiol. 102*, 199–207.
Ephrussi, A., Church, G. M., Tonegawa, S., & Gilbert, W. (1985) *Science (Washington, D.C.) 227*, 134–140.
Falkner, F. G., & Zachau, H. G. (1984) *Nature (London) 310*, 71–74.
Felsenfeld, G., & McGhee, J. (1982) *Nature (London) 296*, 602–603.
Gidoni, D., Dynan, W. S., & Tjian, R. (1984) *Nature (London) 312*, 409–413.
Gruss, P. (1984) *DNA 3*, 1–5.
Hernandez, N. (1985) *EMBO J. 4*, 1827–1837.
Krainer, A. R., & Maniatis, T. (1985) *Cell (Cambridge, Mass.) 42*, 725–736.
Kramer, A., Keller, W., Appel, B., & Luhrmann, R. (1984) *Cell (Cambridge, Mass.) 38*, 299–307.
Kristo, P., Tsai, M.-J., & O'Malley, B. W. (1984) *DNA 3*, 281–286.
Krol, A., Lund, E., & Dahlberg, J. E. (1985) *EMBO J. 4*, 1529–1535.
Lerner, M. R., & Steitz, J. A. (1979) *Proc. Natl. Acad. Sci. U.S.A. 76*, 5495–5499.
Mangin, M., Ares, M., Jr., & Weiner, A. M. (1985) *Science (Washington, D.C.) 229*, 272–275.
Maniatis, T., Fritsch, E. F., & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Marzluff, W. F., Brown, D. T., Lobo, S., & Wang, S.-S. (1983) *Nucleic Acids Res. 11*, 6255–6270.
Mason, J. O., Garath, W. T., & Neuberger, M. S. (1985) *Cell (Cambridge, Mass.) 41*, 479–487.
Mattaj, I. W., & Zeller, R. (1983) *EMBO J. 2*, 1883–1891.
Mattaj, I. W., Lienhard, S., Jiricny, J., & De Robertis, E. M. (1985) *Nature (London) 316*, 163–167.
Maxam, A., & Gilbert, W. (1980) *Methods Enzymol. 65*, 479–560.
Maxson, R., Mohunm, T., & Kedes, L. (1983) in *Eucaryotic Genes* (Maclean, N., Gregory, S. P., & Flavell, R. A., Eds.) pp 277–298, University Press, Cambridge, U.K.
Mount, S. M., Pettersson, I., Hinterberger, M., Karmas, A., & Steitz, J. A. (1983) *Cell (Cambridge, Mass.) 33*, 509–518.
Murphy, J. T., Burgess, R. R., Dahlberg, J. E., & Lund, E. (1982) *Cell (Cambridge, Mass.) 29*, 265–274.
Nordheim, A., & Rich, A. (1983) *Nature (London) 303*, 674–678.
Padgett, R. A., Mount, S. M., Steitz, J. A., & Sharp, P. A. (1983) *Cell (Cambridge, Mass.) 35*, 101–107.
Parslow, T. G., Blair, D. L., Murphy, W. J., & Granner, D. K. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 2650–2654.
Piatigorsky, J. (1984) *Cell (Cambridge, Mass.) 38*, 620–621.
Proudfoot, N. J., Gila, A., & Maniatas, T. (1982) *Cell (Cambridge, Mass.) 31*, 553–563.
Reddy, R. (1985) *Nucleic Acids Res. 13*, r155–r166.
Reddy, R., Henning, D., & Busch, H. (1985) *J. Biol. Chem. 260*, 10930–10935.
Roop, D. R., Kristo, P., Stumph, W. E., Tsai, M.-J., & O'-Malley, B. W. (1981) *Cell (Cambridge, Mass.) 23*, 671–680.
Skuzeski, J. M., Lund, E., Murphy, J. T., Steinberg, T. H., Burgess, R. R., & Dahlberg, J. E. (1984) *J. Biol. Chem. 259*, 8345–8352.
Southern, E. M. (1975) *J. Mol. Biol. 98*, 503–515.
Strub, K., Galli, G., Busslinger, M., & Birnstiel, M. L. (1984) *EMBO J. 3*, 2801–2807.
Tani, T., Watanabe-Nagasu, N., Okada, N., & Ohshima, Y. (1983) *J. Mol. Biol. 168*, 579–594.
Van Arsdell, S. W., & Weiner, A. M. (1984) *Mol. Cell. Biol. 4*, 492–499.
Watanabe-Nagasu, N., Itoh, Y., Tani, T., Okano, K., Koga, N., Okada, N., & Ohshima, Y. (1983) *Nucleic Acids Res. 11*, 1791–1801.
Westin, G., Zabielski, J., Hammarstrom, K., Monstein, H.-J., Bark, C., & Pettersson, U. (1984a) *Proc. Natl. Acad. Sci. U.S.A. 81*, 3811–3815.
Westin, G., Lund, E., Murphy, J. T., Pettersson, U., & Dahlberg, J. E. (1984b) *EMBO J. 3*, 3295–3301.
Yuo, C.-Y., Ares, M., Jr., & Weiner, A. M. (1985) *Cell (Cambridge, Mass.) 42*, 193–202.
Zeller, R., Carri, M.-T., Mattaj, I. W., & De Robertis, E. M. (1984) *EMBO J. 3*, 1075–1081.