Volume 9 Number 20 1981 Nucleic Acids Research

A chicken middle-repetitive DNA sequence which shares homology with mammalian ubiquitous repeats

William E.Stumph, Paula Kristo, Ming-Jer Tsai and Bert W.O'Malley

Department of Cell Biology, Baylor College of Medicine, Houston, TX 77030, USA

Received 20 April 1981

ABSTRACT

We have identified and sequenced two members of a chicken middle repetitive DNA sequence family. By reassociation kinetics, members of this family (termed CR1) are estimated to be present in 1500-7000 copies per chicken haploid genome. The first family member sequenced (CR1U1a) is located approximately 2 kb upstream from the previously cloned chicken U1 RNA gene. The second CR1 sequence (CR1OVa) is located approximately 12 kb downstream from the 3' end of the chicken ovalbumin gene. The region of homology between these two sequences extends over a region of approximately 160 base pairs. In each case, the 160 base pair region is flanked by imperfect, but homologous, short direct repeats 10-15 base pairs in length. When the CR1 sequences are compared with mammalian ubiquitous interspersed repetitive DNA sequences (human Alu and Mouse B1 families), several regions of extensive homology are evident. In addition, the short nucleotide sequence CAGCCTGG which is completely conserved in ubiquitous repetitive sequence families from several mammalian species is also conserved at a homologous position in the chicken sequences. These data imply that at least certain aspects of the sequence and structure of these interspersed repeats must predate the avian-mammalian divergence. It seems that the CR1 family may possibly represent an avian counterpart of the mammalian ubiquitous repeats.

INTRODUCTION

For quite some time it has been known that the DNA of higher organisms consists of repetitive DNA sequences present in many copies per haploid genome as well as single copy (unique) DNA sequences. In a majority of organisms studied (1,2), a large percentage of the repetitive DNA has been shown to consist of sequences approximately 300 base pairs (bp) in length which are interspersed among the single copy sequences. The function of these short interspersed repeats is not known, but they have been postulated to be involved in the regulation of gene expression at the level of transcription (3) or RNA processing (4).

Recently, it has been shown that the interspersed repetitive sequences of the human genome are dominated by a single family of repeats which has been termed the Alu family (5,6). Human Alu family sequences are present in

© IRL Press Limited, 1 Falconberg Court, London W1V 5FG, U.K.

about 300,000 copies per genome, and the nucleotide sequence of several individual members as well as a general consensus sequence has been determined (6-9). Specific regions of the Alu sequence share considerable homologies with: interspersed repeated DNA sequences of Chinese hamster cells, sequences found near the origin of replication of several papovaviruses, sequences found in HeLa cell hnRNA, a species of 4.5S RNA from rodent cells (10,11), and a 7S RNA from HeLa cells (12). Several investigators have identified families of interspersed repetitive DNA sequences in the mouse (13-15). The sequence of one such family from the mouse has been determined (13). The members of this family are approximately 130 bp in length, and they exhibit similar homologies to those mentioned above. Equivalent-type sequences have also been reported in other primate and rodent species (16,17,11).

The conservation of these ubiquitous repetitive sequence homologies among several mammalian species prompted us to examine whether these homologies might also be conserved in a non-mammalian species. We have recently reported the identification and characterization of a chicken DNA genomic clone containing a gene for Ul small nuclear RNA (18). In this paper we report the identification and sequence of a chicken repetitive DNA element (CR1Ula) which is located approximately 2 kb upstream from the chicken Ul RNA gene. We have also identified and sequenced a second segment of DNA (CR1OVa) which is located approximately 12 kb downstream from the 3' end of the ovalbumin gene structural sequence. These two sequences are members of a chicken middle repetitive DNA family which we have chosen to term CR1. Sequence homologies are presented which compare the two CR1 sequences to each other and to their mammalian counterparts.

MATERIALS AND METHODS

Recombinant DNA Clones

The initial characterization of the recombinant Charon 4A phage clone containing the chicken Ul RNA gene has been described (18). It was isolated from a gene library obtained from D. Engel, J. Dodgson, R. Axel and T. Maniatis. A 2.5 kb chicken DNA fragment containing the Ul gene (and CR1Ula sequence) was subcloned into the Eco RI site of the plasmid pBR322 (18). The CR1OVa sequence was identified in a phage clone (C1500) which contains approximately the 3' two-thirds of the ovalbumin gene and 13.5 kb of DNA flanking the 3' end of the ovalbumin gene (Fig. 4). A 2.3 kb Eco RI fragment stretching from a natural Eco RI site on the left to an artificial Eco RI site

on the right end (as shown in Fig. 4) was subcloned into pBR322 and used for sequencing. More specific details of this CR10Va-containing clone will be presented in a publication dealing exclusively with repetitive DNA sequences in the ovalbumin gene domain (Stumph *et al.*, in preparation).

Hybridization Probes

The subcloned U1 2.5 kb fragment was purified from plasmid DNA by cutting with Eco RI and separating the fragments by preparative agarose gel electrophoresis. The purified U1 2.5 fragment was then re-cut with Hinf I to produce the smaller fragments shown in Fig. 1B. The Hinf I fragments were subsequently separated on a 4% preparative polyacrylamide gel, extracted from the gel slices, and purified using the method of Maxam and Gilbert (19). Probes for hybridization were prepared by nick translation (20) of either the isolated fragments or, for other experiments, total chicken genomic DNA. The specific activities obtained using α-^{32}P-dCTP and TTP were approximately 5-10 X 10^8 cpm/μg. Using ^{3}H-dCTP and TTP, a specific activity of approximately 1 X 10^8 cpm/μg was obtained for total chicken DNA.

Southern Hybridizations

For Southern transfer (21), 1 μg of Eco RI cut recombinant phage DNA or 12 μg of Eco RI or Bam HI cut total genomic DNA were loaded per lane on 1% agarose gels. After transfer, the nitrocellulose filters were hybridized to ^{32}P probe in 6 x SSC at 68° and washed in 1 x SSC at the same temperature. The Southern filter hybridizations contained 80 x 10^6 cpm of ^{32}P total chicken DNA or 35 x 10^6 cpm of purified fragments 4 or 2 (Fig. 1B).

Solution Hybridizations

For Cot analysis in which the DNA reassociation was to be monitored by hydroxyapatite (HAP) chromatography, reactions were carried out in 0.12 M sodium phosphate (PB) pH 6.8 at 60° at a genomic driver DNA concentration of 8.7 μg/ml (3.33 ml volume), or in 0.4 M PB at 68° at DNA concentrations of 218 μg/ml (1.0 ml volume) or 4.4 mg/ml (40 μl volume). HAP chromatography was carried out at 60° after adjusting all samples to 0.12 M PB. Each reaction contained 5000 cpm of nick translated ^{32}P fragment 4 probe and 200,000 cpm of ^{3}H total genomic DNA as a tracer. Self-reassociation of the labeled probes was monitored in identical reactions substituting *E. coli* for chicken driver DNA. At Cots greater than 5 X 10^2, self-reassociation of the low complexity ^{32}P probe became significant, so those results were unreliable and not plotted. Hybridizations to be assayed by S1 nuclease digestion and subsequent trichloroacetic acid precipitability were carried out at 68° in 0.6 M NaCl, 10 mM HEPES pH 7.0, 2 mM EDTA containing 850 cpm of ^{32}P fragment

4 probe and driver DNA at concentrations ranging between 0.21 μg/ml and 2.1 mg/ml. S1 digestions were carried out as previously described (20). In all cases equivalent Cot values (those that would obtain at 0.18 M NaCl) have been plotted (22).

DNA Sequencing

The DNA sequencing method of Maxam and Gilbert (19) was used. The preparation and separation of end-labeled DNA fragments and the sequencing reactions were performed as described in Catterall et al. (23).

RESULTS

Location and Reassociation Kinetics of the CR1U1a Repetitive Sequence

The U1 genomic clone isolated from the chicken gene library consists of approximately 13.3 kb of chicken DNA inserted into Charon 4A, and contains 165 bp of chicken DNA flanking the 3' end of the U1 RNA coding sequence and about 13,000 bp flanking the 5' end of the U1 RNA gene (unpublished results). Digestion with Eco RI yields four chicken DNA fragments of sizes 4.3, 3.5, 3.0 and 2.5 kb. An Eco RI digest was run on a 1% agarose gel, transferred to nitrocellulose by the method of Southern (21), and probed with ^{32}P nick translated total genomic DNA. Under such conditions, only bands containing DNA sequences which are repeated many times in the genome should yield a signal (24). The result is shown in Fig. 1A. Only the 2.5 kb band, which contains the U1 RNA gene (18), lights up. Since we have previously shown that the U1 gene is repeated only about 6-10 times in the chicken genome (18) it seems unlikely that it has a great enough repetition frequency to be responsible for the signal observed in Fig. 1A. The result therefore suggests that a more highly repeated sequence is present within the 2.5 kb fragment. The result also suggests that none of the other chicken DNA bands in the recombinant phage clone contain repetitive DNA sequences.

In order to better define the location of the repetitive sequence, the purified 2.5 kb fragment was digested with Hinf I to produce the fragments shown in Fig. 1B. The fragments were separated on a preparative polyacrylamide gel, and extracted and purified as previously described (18). Fragments 2 and 4 were then nick translated with ^{32}P nucleotides and individually used to probe Southern blots of total chicken DNA which had been cut with either Eco RI or Bam HI. Fig. 1C shows that fragment 2 hybridized to specific bands in each lane, and thus probably represents a sequence present only once or twice in the genome. In marked contrast, fragment 4 hybridized intensely to DNA fragments of all sizes, which would be expected only if the

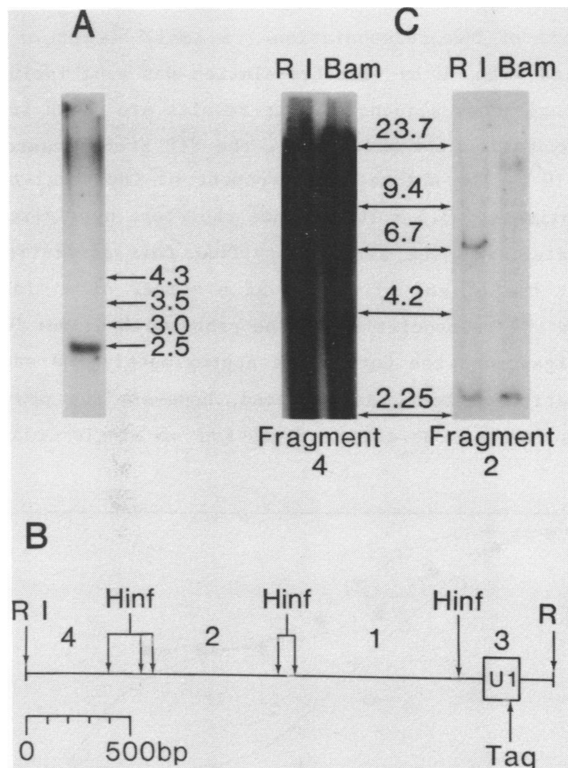

Figure 1: Localization of a Repetitive DNA Sequence ~ 2 kb 5' of the Chicken U1 RNA Gene.

A. Southern blot of Eco RI digested Charon 4A recombinant phage clone containing the U1 gene in a 2.5 kb fragment. The clone also contains three other fragments of the sizes shown which are visible on the ethidium bromide stained gel and all of which are located 5' of the 2.5 kb fragment in the chicken genome. Total chicken genomic DNA labeled with ^{32}P was used as the probe in order to look for fragments containing repetitive DNA sequences.

B. Restriction map of the U1 2.5 fragment. Direction of transcription of U1 RNA is from left to right (5'-3'). The 3' Eco RI site is an artificial site produced during the cloning procedure.

C. Total genomic DNA digested with either Eco RI or Bam HI was run on an agarose gel and transferred to nitrocellulose. Purified fragments 4 and 2 (as shown in B) were labeled with ^{32}P and hybridized separately to different lanes of the blot. Equivalent exposures are shown.

DNA sequence in fragment 4 is repeated many times within the chicken genome.

In order to confirm the above result and to obtain an estimate of the number of copies of the repetitive sequence in the chicken genome, we examined its reassociation kinetics. The ^{32}P-labeled fragment 4 probe was allowed to

reassociate with an excess of sheared, unlabeled total chicken DNA using standard conditions of DNA reassociation. A small amount of total chicken genomic DNA labeled with ^{3}H by nick translation was also included to provide a standard for single-copy sequences. The results are shown in Fig. 2. When assayed by hydroxyapatite chromatography, the ^{32}P probe reassociated with a $Cot_{1/2}$ of 2.7×10^{0}. The unique DNA component of the total chicken DNA had a $Cot_{1/2}$ of approximately 3.5×10^{3}. Since the probe hybridizes approximately 1300 times faster than the single copy DNA, this identifies the DNA sequence present at the 5' end of U1 2.5 as a member of a middle repetitive DNA family. When the reassociation of the probe with driver DNA was assayed by S1 nuclease digestion, the $Cot_{1/2}$ was approximately the same as observed by the hydroxyapatite method. As expected, however, the measured level of reassociation was reduced due to the digestion of single-stranded tails and

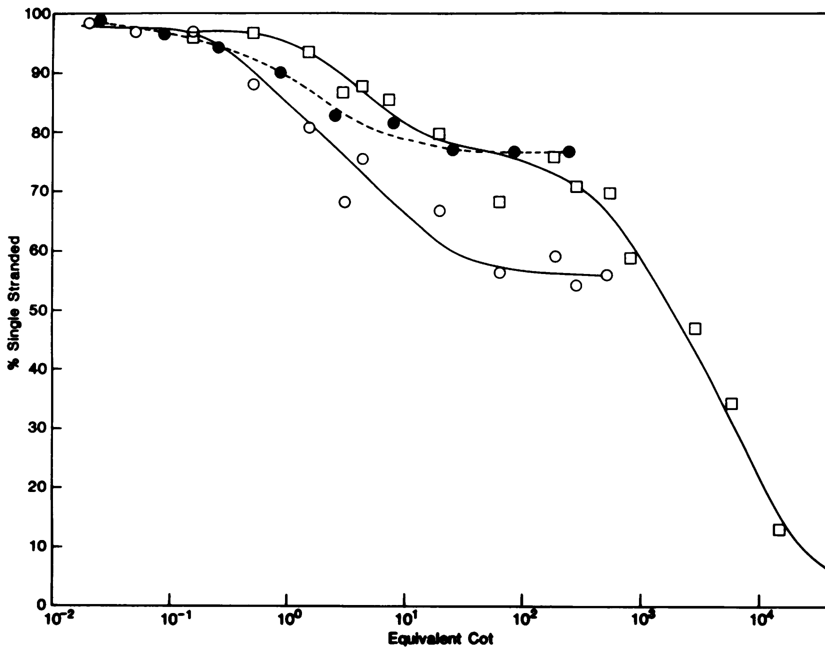

Figure 2: Reassociation Kinetics of the Chicken Repetitive Sequence

—□— Reassociation of driver (total genomic) DNA measured by hydroxyapatite chromatography.

—○— Hybridization of fragment 4 probe with driver DNA measured by hydroxyapatite chromatography.

—●— Hybridization of fragment 4 probe with driver DNA measured by S1 nuclease digestion.

mismatched sequences by the S1 nuclease.

Nucleotide Sequence of Two CR1 Family Members

Fragment 4 from the 5' end of the U1 2.5 clone was sequenced using the method of Maxam and Gilbert (19). The entire sequence of the 390 bp fragment is shown in Fig. 3, and is termed CR1U1a. In order to compare this sequence with that of another related family member, recombinant phage clones containing chicken DNA previously characterized in this laboratory were screened by digesting with Eco R1 and probing Southern blots of these clones with ^{32}P-labeled fragment 4 from the U1 gene clone. One clone which gave a strong hybridization signal was known as C1500. This clone contains about two-thirds of the 3' end of the ovalbumin gene and 13.5 kb of DNA flanking the 3' end of the gene as shown in Fig. 4. The ^{32}P-labeled fragment 4 probe hybridized to the 2.3 kb Eco R1 fragment located at the extreme 3' end of the cloned DNA (data not shown). After subcloning into pBR322, the area of maximum interest in the 2.3 kb fragment was sequenced. This nucleotide sequence is shown in the upper part of Fig. 4, and is designated as CR1OVa.

When the CR1U1a and CR1OVa sequences were compared, the two sequences were found to be homologous over a region of approximately 160 nucleotides. This homology region is shown in Fig. 5A. It is interesting to note that there seem to be two distinct regions which differ in their degree of sequence conservation and which are separated from each other by an insertion/deletion block of 10 nucleotides between residues 274 and 275 of the CR1U1a sequence. The region which lies 3' of the insertion/deletion block, that is CR1U1a

```
        10         20         30         40         50         60
GAATTCAAAA GGTGGAATAA TAAAACAATT TTATCAGTTG TTCGTGCCTG TTCTGTTTGC

        70         80         90        100        110        120
CATCCTTGGT GTGGCAGGAG AATTGCTGCT TCCTTTTTAT TGCGTGCTCA GATCTGGGGT

       130        140        150        160        170        180
TGTGTGATCC GACTGGGGCT TCCCCAAAGG CACTAATGCT TCATTCAACC CTCTGAAAGG

       190        200        210        220        230        240
GTACATAACG GCAGGATAAG GGGAAATCGT TGGAAGTTGG AGGAGGGAAG ATTGAGGTTG

       250        260        270        280        290        300
GACATCAGGG GGAAGTTCTT TACTATGAGA GTGGTGAGGT GCTGGAACAG CTGCCCAGAG

       310        320        330        340        350        360
AGGTTGTGGA TGCCCCGTCC ATCCCTGGAG GTGTTCAAGG CCGGGTTGGA CGTGCCCTGG

       370        380        390
GCAGCCTGGG CTGGTACTGA ATGTGGAGTC
```

Figure 3: Nucleotide Sequence of Fragment #4 (Eco RI-Hinf I, Fig. 1B) Containing the Chicken Middle Repetitive Sequence CR1U1a.

```
        10         20         30         40         50         60
TGTCTTAAGT ATTGCATAGG TGAAGGTATC ACAGGATGCT GTACTAATCA GTTGTGTACC

        70         80         90        100        110        120
CTGCATTTGA GGAAGGCAAA CAGAGAGAGA AAATAAAAGC AAGGGAAGGT GTTAGCATGA

       130        140        150        160        170        180
CTGTAGCTTG ATCATGAGGC ACTGGAACAG GTTGCCCAAA GAGGTTGTGG ATGCCCCATC

       190        200        210        220        230        240
CCTGGAGGCA TTCAAGGCCA GGCTGGATGT GGCTCTGAAC AGCCTGGTCT AGTGGTTGGT

       250        260        270        280        290        300
GACCCTGCAC ATAGCAGGGG GATTGAAACT AGATGATCAT TGTGGTCCTT TTCAACCCAG
```

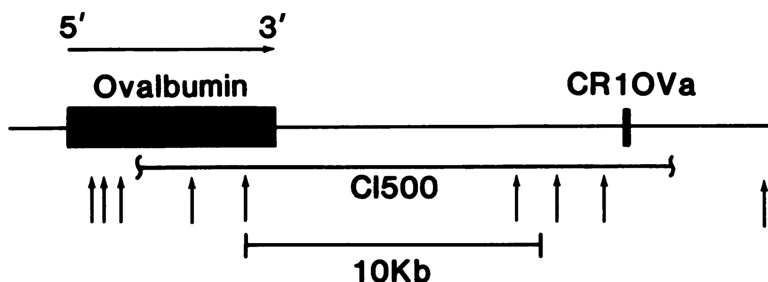

Figure 4: Nucleotide Sequence and Genomic Location of the Chicken CR1 Family Member CR1OVa.

Cl500 is a recombinant phage clone containing chicken DNA from the ovalbumin region of the genome bounded by artificial Eco RI sites and extending over the region shown. Vertical arrows indicate the position of natural Eco RI sites in the chicken genome. Transcription of the ovalbumin gene is from left to right in the diagram. However, the sense of the CR1OVa nucleotide sequence printed at the top reads in a 5' to 3' direction from right to left in the diagram at the bottom of the figure.

residues 275 to 375, exhibits 81% homology to the corresponding region of CR1OVa (83 matches out of 103 positions). The other portion of the sequence, that is, CR1U1a positions 217 to 273, is homologous to CR1OVa at a level of only 60% (37/62). Nevertheless, this still represents a significant degree of homology which is not observed in the flanking DNA sequences. It is also interesting to note that the 160 bp homology region of each CR1 sequence is flanked by a short direct repeat of 10-15 nucleotides. These short repeats are underlined in Fig. 5A and are aligned to show maximum homology in Fig. 5B. It should be pointed out that the short repeats which flank the 160 bp CR1U1a homology region are not homologous to the short repeats which flank the CR1OVa homology region.

A

```
                                       217
CR1U1a          203   GAAATCGTTGGAAG   TTGGAGGAGGGAAGATTGAGGTTGGACATCA--GG--GGGAAGTTCTTTA-CTATGAGAG
                                       ** ***** ** * *  ***   * * ** *  *   ****** * ** * * **  **
CR1OVa           51  GTTGTGTACCCTGCA   TTTGAGGAAGGCAAACAGAGAGAGAAAATAAAAGCAAGGGAAGGTGTT-AGC-AT--GAC
                                       66
                                         274      275
CR1U1a (cont.)                         TGG----------TGAGGTGCTGGAACAGC-TGCCCAGAGAGGTTGTGGATGCCCCGTCC
                                       **           *****  *********  ****** ******************
CR1OVa (cont.)                         TGTAGCTTGATCATGAGGCACTGGAACAGGTTGCCCAAAGAGGTTGTGGATGCCCC----

                                                                                        375
CR1U1a (cont.)                         ATCCCTGGAGGTGTTCAAGGCCGGGTTGGACGTG-CCCTGGGCAGCCTGGGCTGGT ACT GAATGTGGAG       388
                                       ***********  ********* ** **** *** * ***  ******** ** **
CR1OVa (cont.)                         ATCCCTGGAGGCATTCAAGGCCAGGCTGGATGTGGCTCTGAACAGCCTGGTCTAGT G GTTGGTGACCCTGCA   249
                                                                                        233
```

B

```
CR1U1a     GAAATCGTTGGAAG                        CR1OVa     GTTG-TGTACCCTGCA
CR1U1a     GAA-T-GT-GGA-G                        CR1OVa     GTTGGTG-ACCCTGCA
```

Figure 5: Comparison of CR1 Sequences from the U1 Gene and Ovalbumin Gene Regions of the Chicken Genome.

A. The CR1U1a and CR1OVa sequences are aligned to show maximum homology. The underlined areas are short direct repeats which flank the homologous regions of the two sequences.

B. The short direct repeats are aligned to emphasize the homologies.

Comparison of Chicken and Mammalian Repetitive Sequences

Figure 6 shows a comparison of the chicken CR1 sequences to human Alu (7) and mouse B1 (13) repetitive DNA sequences. Several regions of significant homology are observed. The portion of the CR1 sequences used in the comparison shown in Fig. 6A is basically that region of highest homology between the two CR1 sequences (as shown in Fig. 5A). Two adjacent, partially overlapping regions of the human Alu sequence (positions 265-298, and 296-328) share greater than 70% homology with two sections of the CR1U1a sequence in this region (26 matches of 37 positions and 25 matches of 35 positions respectively). Correspondingly, the mouse B1c nucleotides 27 through 85 exhibit 70% homology to the chicken CR1OVa sequence throughout the region shown (47 matches out of 67 positions).

Two portions of the CR1U1a sequence which do not exhibit significant homology to the CR1OVa sequence nevertheless are homologous to segments of the human and mouse repeats. These regions are shown in Fig. 6B and again represent over 70% homology. These data seem to indicate that regions of the CR1U1a sequence which are not shared with CR1OVa nevertheless may still represent related repetitive DNA sequence. The portion of the human Alu sequence which is underlined in Fig. 6B is a 14 bp region which is highly

```
   HUMAN ALU                265   GCTTGAAC--CT-GGGAAGCAGAGGTTGCGCTGA-GCC   298
                                  *** ****  ** *   ** ******** *  ** ***
   CHICKEN CR1U1a           281   GCTGGAACAGCT-GCCCAG-AGAGGTTGTG--GATGCCCCGTCCATCCCT-GG-
                                   ********* * *****  **********  ********    ****** **
   CHICKEN CR1OVa           141   ACTGGAACAGGTTGCCCAA-AGAGGTTGTG--GATGCCCC----ATCCCT-GG-
                                                                   ** ***     * **** **
   MOUSE B1c                                                  27   AT-CCC-----AGCCCTCGGG
A
   HUMAN ALU                            296   GCCGAGATGGCACCATTG--CACTC----CAGCCTGGGC    328
                                              **** * *** **   **  * **    **********
   CHICKEN CR1U1a (cont.)         AGGTGTTCAAGGCCGGGTTGG-AC--GTG--CCCTGGG--CAGCCTGGGCT    372
                                  ***  ********* ** *** *   ***  * ***    ******** **
   CHICKEN CR1OVa (cont.)         AGGCATTCAAGGCCAGGCTGG-AT--GTG-GCTCTGAA--CAGCCTGGTCT    230
                                  *****    **** **** ** **   ** * ** **   ***********
   MOUSE B1c       (cont.)        AGGCA---GAGGC-AGGC-GG-ATTTCTGAGTTC-GAGGCCAGCCTGGTCT     85

                                               C      A
                                               ^      ^
   MOUSE B1c                 -7         TGGTGTACCGGGAGTAGTGGTGCTGCCTTT    25
                                        ******  * **** * ** **** *****
B  CHICKEN CR1U1a            67         TGGTGTGGCAGGAGAATTGCTGCTTCCTTT    87
                                         * ** *********** ***     ***
   HUMAN ALU                242   CGGGAGGCTGAGGCAGGAGAATCGCTTGAACCTGG   276
                                  *** *** * ***  *** ***** * ***  ***
   CHICKEN CR1U1a           189   CGGCAGGATAAGG--GGA-AATCGTTGGAAGTTGG   220
```

Figure 6: Sequence Homologies between the Chicken CR1 Sequences and Mammalian Repetitive Sequences.

A. A region highly conserved between the two chicken CR1 sequences is compared to regions of the human Alu and mouse B1c sequences. The numbers flanking the chicken sequences designate the first and last nucleotides of the series as numbered in Figures 3 and 4. The human Alu and mouse B1c sequences are taken from references 7 and 13 respectively. The numbering of the Alu and B1c nucleotide positions in this figure is that originally used in those respective references.

B. Two additional regions of the CR1U1a sequence are shown which exhibit significant homology to regions in the Alu and B1c sequences. The underlined portion of the human Alu sequence is homologous to sequences found at the origin of replication of several papovaviruses (10).

homologous to sequences found at the origin of replication of several papovaviruses (10). Two regions of the chicken CR1U1a sequence exhibit partial homology to this area; however, the chicken sequences exhibit substantially less homology to the viral sequences than what is observed between the human and viral sequences themselves (not shown).

DISCUSSION

Our data demonstrate the existence of a repetitive DNA sequence, CR1U1a, approximately 2 kb upstream from the previously cloned chicken U1 RNA gene. Weiner (12) and Denison et al. (25) have recently reported the presence of Alu sequences upstream from cloned pseudogenes for human U1 and U2 small nu-

clear RNAs. The upstream Alu sequences are in the same 5'-3' orientation as the small RNA pseudogenes themselves. (The Alu sequence polarity is defined by the direction of transcription of Alu elements in vitro by RNA polymerase III [10,26]). By using the homologies between the CR1 and the Alu sequences shown in Fig. 6, it is possible to assign a polarity to the CR1 family sequences based upon the Alu sequence polarity. The result of this assignment is that the orientation of the CR1U1a sequence is in the same direction as that of the U1 gene, which is identical to the situation found for the Alu sequences located upstream from the human small RNA pseudogenes. The relationship between the CR1OVa sequence and the ovalbumin gene, however, is reversed. In this case the CR1OVa sequence is found downstream from the ovalbumin gene, and the orientation of the CR1OVa sequence is in the opposite direction to that of the ovalbumin gene. Thus, in both of these cases the CR1 sequence "points" toward the structural gene. It is not known whether or not this could be a general phenomenon in the chicken genome.

When the 390 bp Hinf I restriction fragment containing the chicken repeat was labeled with ^{32}P and used as a probe in a DNA reassociation reaction driven by total chicken DNA, a significant fraction of the probe reannealed approximately 1300 times as fast as chicken single copy DNA. This indicates a nominal reiteration frequency of 1300 copies per haploid chicken genome. However, this figure is undoubtedly an underestimate of the true copy number for reasons which have been thoroughly discussed by Houck et al. (5). Factors involved in the retardation of the rate of reassociation of repetitive DNA sequences include sequence mismatching and the resultant lowering of the T_m (27), length effects using short fragments (28,29), and the actual shortness of the repeated sequence region itself. At the criterion used for the hybridizations, a substantial underestimate of the actual copy number would be expected. It has been well documented that Alu sequences are present in at least 300,000 copies per human haploid genome (5,30). However, isolated Alu sequences empirically reassociate with driver DNA only 50,000 to 64,000 times as fast as human single copy DNA (5). Since it appears that the chicken repetitive DNA sequence which we have characterized may be analogous to the human Alu family in many ways, it seems likely that 6800 copies would be a better estimate of the actual reiteration frequency of the chicken sequence (300,000/57,000 x 1300 = 6800). It is possible that additional, more distantly related sequences having a lower degree of homology are not even scored in the Cot assay, yet they may be present in many additional copies per genome. It is not at all surprising that the chicken sequence is less reiterated than

the human Alu family. Most studies using eucaryotic organisms have shown predominant middle repetitive families with kinetic reiteration frequencies in the 1000-2000 fold range (1,5,31), compared to the 50,000 fold kinetic class observed in the human genome. Indeed, the bulk of the middle repetitive sequences of the chicken appear to reassociate at approximately the same rate or slightly slower (Fig. 2) than the CR1 sequence located upstream from the U1 RNA gene.

The cloning of chicken middle-repetitive sequences has also been reported by other investigators (32,33,34). However, these other cloned repetitive sequences appear to be on the order of 2 kb or greater in length, so a close relationship between these cloned repeats and the CR1 sequence reported here is doubtful. It cannot be ruled out, however, that CR1-type sequences could be contained within these longer repeating units.

An important result of the studies presented here is that repetitive DNA sequences which are approximately 80% conserved among different species of mammals (11) appear also to have homologous counterparts in the chicken. Several relatively lengthy regions of the chicken CR1 sequences exhibit greater than 70% homology to the mammalian repeats (Fig. 6). This indicates that at least some of the sequences in these repetitive families must predate the avian-mammalian divergence. Although the chicken repeats do not exhibit the extent of homology to the mammalian repeats as the mammalian sequences do to each other, that is exactly what would be expected from an evolutionary standpoint.

Haynes et al. (11) have pointed out that the nonanucleotide sequence CCAGCCTGG is perfectly conserved at a homologous position in sequences from four different mammalian species. It is very interesting that this same sequence, lacking only the first C residue, is also perfectly conserved at the homologous position in both chicken CR1 sequences (Fig. 6a; CR1U1a 362-369; CR1OVa 220-227). This complete conservation of sequence between the four mammals and a bird argues for an important functional role for this short stretch of nucleotides. It is also interesting to note that the chicken sequences become divergent almost immediately past this stretch of highly conserved nucleotides.

A common structural aspect of mammalian ubiquitous repetitive sequences is that they are often, but not always, flanked by short direct repeats up to 20 bp in length. These short repeats, although highly homologous, are often not perfect copies of each other. This structural aspect of the mammalian repeats also appears to be conserved in the structure of the chicken CR1

sequences. The CR1OVa sequence is flanked by a quite well conserved direct repeat 15-16 bp in length (Fig. 5). The direct repeat flanking the CR1U1a sequence is not so well conserved. However, by analogy to the location of the CR1OVa direct repeats, it seems most likely that the CR1U1a sequences probably are valid direct repeats which have undergone a small number of deletion or insertion events. In the case of the mammalian sequences studied, the short repeats which flank an individual family member are not homologous to the short repeats which flank other family members. This is also the case in the chicken sequences: the CR1U1a short repeats are not homologous to the CR1OVa short repeats.

The partial conservation of sequence homology in these interspersed repeats, together with structural homology predating the divergence of mammals and birds, implies that these repetitive sequence families must play an important role in genomic function. Numerous suggestions have been made in the literature concerning the possible function of these ubiquitous interspersed repeated sequences. These include: origins of DNA replication (10); RNA polymerase III transcription units (10,26,35); boundaries of DNA segments containing genes which are coordinately expressed (7); or involvement in genetic translocation (30). By analogy to the structure of bacterial and yeast transposable elements, the presence of short direct repeats flanking many of the individual repetitive sequences has led to the suggestion that the interspersed repeats could act as mobile genetic elements (11,36,37). Since sequences found in ubiquitous repeats are found also in hnRNA (10,13), in small molecular weight rodent RNAs associated with poly (A) containing RNAs (1,38), in HeLa cell 7S RNA (12), and in polysomal RNA (30), it has also been speculated that these sequences may be involved in mRNA processing or transport, or in other stages of mRNA metabolism. It cannot be ruled out, and is entirely possible, that different sub-families of the ubiquitous repeats could carry out entirely different functions within the cell. The chicken repeat which we have sequenced exhibits only a marginal degree of homology to sequences found at the papovavirus origins of replication. This may argue against the CR1 sequence having a function as an origin of DNA replication. On the other hand, avian replication origins could be divergent from those found in mammals.

In other experiments in our laboratory, we have used the U1 5'-flanking repetitive sequence as a hybridization probe to look for additional related repetitive sequences at other locations within the ovalbumin gene domain. We have examined over 120 kb of this region. No cross-reacting family members

are observed within the 50 kb region of the domain which includes the X, Y and ovalbumin genes. However, in addition to the CR1OVa sequence 12 kb 3' of ovalbumin, the CR1U1a probe also hybridizes moderately to sequences found 18 and 25 kb 5' of the X gene, and 25 kb 3' of the ovalbumin gene. This means that the expressed portion of the domain is itself devoid of, but flanked by, related repetitive DNA sequences. It also appears that there could be a correlation between the location of these sequences and the region of transition of the domain from DNase I sensitivity to DNase I resistance (Lawson et al., manuscript submitted). Although this correlation could be coincidental, further experiments will be performed to test the concept that this chicken repetitive sequence could be involved in specifying the boundaries of functional chromatin domains within the chicken cell nucleus.

ACKNOWLEDGMENTS

We would like to thank Ms. Wanda Beattie for excellent assistance in the sequencing of CR1OVa and Ms. Syndi Herzog for additional technical assistance. Computer analysis of the nucleotide sequences was performed using the NIH-Aim facility of the Stanford MOLGEN Project, and we thank Mr. Albert Ting for help in running the programs on that system.

REFERENCES

1. Davidson, E.H., Galau, G.A., Angerer, R.C. and Britten, R.J. (1975) Chromosoma 51, 253-259.
2. Schmid, C.W. and Deininger, P.L. (1975) Cell 6, 345-358.
3. Britten, R.J. and Davidson, E.H. (1969) Science 165, 349-357.
4. Davidson, E.H. and Britten, R.J. (1979) Science 204, 1052-1059.
5. Houck, C.M., Rinehart, F.P. and Schmid, C.W. (1979) J. Mol. Biol. 132, 289-306.
6. Rubin, C.M., Houck, C.M., Deininger, P.L., Friedmann, T. and Schmid, C.W. (1980) Nature 284, 372-374.
7. Bell, G.I., Pictet, R. and Rutter, W.J. (1980) Nucl. Acids Res. 8, 4091-4109.
8. Deininger, P., Jolly, D. and Friedmann, T. (1980) in ICN-UCLA Symposia on Molecular and Cellular Biology, Alberts, B. Ed., Vol. XIX, pp. 369-378, Academic Press, New York.
9. Pan, J., Elder, J.T., Duncan, C.H. and Weissman, S.M. (1981) Nucl. Acids Res. 9, 1151-1170.
10. Jelinek, W.R., Toomey, T.P., Leinwand, L., Duncan, C.H., Biro, P.A., Choudary, P.V., Weissman, S.M., Rubin, C.M., Houck, C.M., Deininger, P.L. and Schmid, C.W. (1980) Proc. Natl. Acad. Sci., USA 77, 1398-1402.
11. Haynes, S.R., Toomey, T.P., Leinwand, L., and Jelinek, W.R. (1981) Molecular and Cellular Biology 1, 573-583.
12. Weiner, A.M. (1980) Cell 22, 209-218.
13. Krayev, A.S., Kramerov, D.A., Skryabin, K.G., Ryskov, A.P., Bayev, A.A. and Georgiev, G.P. (1980) Nucl. Acids Res. 8, 1201-1215.
14. Cheng, S.-M. and Schildkraut, C.L. (1980) Nucl. Acids Res. 8,4075-4090.

15. Heller, R. and Arnheim, N. (1980) Nucl. Acids Res. 8, 5031-5042.
16. Dhruva, B.R., Shenk, T., and Subramanian, K.N. (1980) Proc. Nat. Acad. Sci., USA 77, 4514-4518.
17. Page, G.S., Smith, S., and Goodman, H.M. (1981) Nucl. Acids Res. 9, 2087-2104.
18. Roop, D.R., Kristo, P., Stumph, W.E., Tsai, M.-J. and O'Malley, B.W. (1981) Cell 23, 671-680.
19. Maxam, A.M. and Gilbert, W. (1977) Proc. Natl. Acad. Sci., USA 74, 560-564.
20. Roop, D.R., Nordstrom, J.L., Tsai, S.Y., Tsai, M.-J. and O'Malley, B.W. (1978) Cell 15, 671-685.
21. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.
22. Britten, R.J., Graham, D.E., and Neufeld, B.R. (1974) in Methods in Enzymology, Grossman, L. and Moldave, K., Eds., Vol. 29, pp. 363-418, Academic Press, New York.
23. Catterall, J.F., Stein, J.P., Kristo, P., Means, A.R. and O'Malley, B.W. (1980) J. Cell Biol. 87, 480-487.
24. Shen, C.-K. and Maniatis, T. (1980) Cell 19, 379-391.
25. Denison, R.A., Van Arsdell, S.W., Bernstein, L.B. and Weiner, A.M. (1981) Proc. Natl. Acad. Sci., USA 78, 810-814.
26. Duncan, C., Biro, P.A., Choudary, P.V., Elder, J.T., Wang, R.R.C., Forget, B.G., deRiel, J.K. and Weissman, S.M. (1979) Proc. Natl. Acad. Sci., USA 76, 5095-5099.
27. Bonner, T.I., Brenner, D.J., Neufeld, B.R. and Britten, R.J. (1973) J. Mol. Biol. 81, 123-135.
28. Wetmur, J.G. and Davidson, N. (1968) J. Mol. Biol. 31, 349-370.
29. Hinnebusch, A.G., Clark, V.E. and Klotz, L.C. (1978) Biochemistry 17, 1521-1529.
30. Tashima, M., Calabretta, B., Torelli, G., Scofield, M., Maizel, A. and Saunders, G.F. (1981) Proc. Natl. Acad. Sci., USA 78, 1508-1512.
31. Eden, F.C. and Hendrick, J.P. (1978) Biochemistry 17, 5838-5844.
32. Eden, F.C., Burns, A.T.H. and Goldberger, R.F. (1980) J. Biol. Chem. 255, 4843-4853.
33. Eden, F.C. (1980) J. Biol. Chem. 255, 4854-4863.
34. Musti, A.M., Sobieski, D.A., Chen, B.B., and Eden, F.C. (1981) Biochemistry 20, 2989-2999.
35. Elder, J.T., Pan, J., Duncan, C.H. and Weissman, S.M. (1981) Nucl. Acids Res. 9, 1171-1189.
36. Duncan, C.H., Jagadeeswaran, P., Wang, R.R.C. and Weissman, S.M. (1981) Gene 13, 185-196.
37. Van Arsdell, S.W., Denison, R.A., Bernstein, L.B., Weiner, A.M., Manser, T., and Gesteland, R.F. (1981) Cell 26, 11-17.
38. Harada, F. and Kato, N. (1980) Nucl. Acids Res. 8, 1273-1285.