Volume 16 Number 1 1988 Nucleic Acids Research

Organization, sequence and nuclease hypersensitivity of repetitive elements flanking the chicken apoVLDLII gene: extended sequence similarity to elements flanking the chicken vitellogenin gene

Robert J.G.Haché+ and Roger G.Deeley*

Department of Biochemistry, Queen's University, Kingston, Ontario K7L 3N6, Canada

Received September 29, 1987; Revised and Accepted November 20, 1987

ABSTRACT

Analysis of nuclease hypersensitivity of regions flanking the estrogen-dependent, chicken apoVLDLII gene has revealed an hepatic, DNaseI hypersensitive site whose sensitivity is influenced by both the developmental stage and sex of the bird. The site is located 3.0kb upstream from the gene, in a block of middle repetitive elements. Contact hybridization studies indicate that the block consists of contiguous copies of two elements with reiteration frequencies of 500-1000 and 10000-30000 copies per haploid genome. Sequencing of 1.8kb spanning the repeats has revealed that the higher frequency element is a member of the CR1 family. The adjacent lower frequency repeat can also be found next to another member of the CR1 family located in the 3' flanking region of the vitellogenin gene. The hypersensitive site has been mapped to one of the two most highly conserved regions of the CR1 element. This region displays homology with a silencer sequence recently identified in a CR1 element flanking the chicken lysozyme gene.

INTRODUCTION

The avian very low density apolipoprotein II (apoVLDLII) gene specifies a small phospholipid binding protein that constitutes one of two major protein components of the very low density fraction hen serum (1,2). Normally, the gene is expressed only in hen liver at the onset of vitellogenesis, but it can be activated prematurely in birds of either sex by administration of estrogen (3). This can be accomplished in ovo by injection of hormone as early as day 7 of embryogenesis, which results in detectable levels of mature apoVLDLII mRNA by day 9 (4). Expression of the apoVLDLII gene in estrogen treated embryos provides the earliest, currently available indication of the competence of the liver to activate dormant yolk protein genes in response to the hormone.

A region of approximately 25kb that spans the gene and extends 14kb upstream has been cloned and partially characterized, both in terms of primary sequence and the overall transcriptional response of the region to stimulation with estradiol (5-7). Within this region, stimulation with the hormone results in initiation of transcription from the major start sites of the

apoVDLII gene and lower but significant initiation at sites 1105 and 1530 nucleotides upstream (7,8). The limits of the estrogen responsive domain in which the apoVLDLII gene resides have not yet been defined.

In the studies described here, we have used DNase I hypersensitivity as a probe to locate developmentally programmed alterations in chromatin structure in the 5' flanking region of the gene and to investigate the range over which activation of the gene influences this structure. The studies have revealed hypersensitive sites that are influenced both by the developmental stage and sex of the bird, extending as far as 3kb upstream, the most distal of these sites being located in a block of moderately repetitive DNA elements.

The gene itself does not contain repeated sequences but is bracketed by clusters of such elements, several of which are located in the 5' flanking region. With the exception of one block (between -4.6 and -2.8kb) these sequences fall in a region that exhibits extensive haplotypic variation and whose structure is consequently unlikely to play a critical role in determining the regulatory characteristics of the gene (5,6). We have determined the sequence of the repeated cluster between -4.6 and -2.8kb and demonstrated that the hypersensitive site at -3.0kb falls within a member of the CR1 family (9). Members of this family have been found flanking several other avian genes (9). They display a number of unusual properties, including: 1) a predominantly conserved orientation relative to the direction of transcription of the bracketed structural gene, 2) location in areas of transition between nuclease sensitive and resistant regions of chromatin, and 3) the presence of a binding site for a nuclear, non-histone protein (10). Although the origin of the CR1 family has not yet been ascertained, it has been proposed on the basis of their overall structural organization, that dispersion through a reverse transcription mediated mechanism may have contributed to the current interspersed pattern and that the sequences may have originated from the long terminal repeats of a retrovirus (11).

In this manuscript we demonstrate that the block of repetitive elements between -4.6kb and -2.8kb upstream from the apoVLDLII gene in fact consists of two families with reiteration frequences of approximately 1000 and 20000 copies per haploid genome; the more prevalent element being one of the two longest CR1 repeats currently identified. Comparison of the sequence of this region with that of a region bracketing a CR1 element recently identified approximately 2.2kb downstream from the chicken vitellogenin gene (12), has revealed that a portion of the lower frequency repeat is also adjacent to the CR1 element at this second location. A striking feature of the element

upstream from the apoVLDLII gene is that it is marked in the developing embryo by a prominant DNase I hypersensitive site. This site is barely detectable in the adult rooster in which the apoVLDLII gene is dormant, but is enhanced in hen liver where the gene is chronically expressed. It is located in one of the two highly conserved portions of the CR1 element. In day 7 embryos, this DNase I hypersensitive site, or an electrophoretically indistinguishable one, is also detectable in control birds. The hypersensitivity of the region does not appear to be attributable simply to supercoiling dependent alterations in DNA topology, since it displays no preferential sensitivity to S1 nuclease digestion when present in supercoiled, naked DNA.

MATERIALS AND METHODS

Hormone Treatment

Experiments were carried out with White Leghorn embryos, 3-6 month-old cockerels, roosters and hens. Embryos received a single injection of 1.25 mg of 17β-Estradiol in 50µl dimethylsulfoxide on day 5 of embryogenesis. Forty-eight hours later the livers were isolated, frozen in liquid nitrogen and pooled. A typical experiment utilized between 5 and 15 dozen livers. Birds of various ages were given intramuscular injections of 17β-Estradiol in the thigh (20 mg/ml, in propylene glycol, at a dose of 20 mg/kg body weight).

DNase I Digestion of Nuclei

Nuclei were prepared by a modification of the technique of Mirkovitch et al. (13). All steps were performed at 0-4°C. unless otherwise indicated. Fresh adult tissue was passed through a strainer and red blood cells removed by flotation in IS buffer (10 mM Tris HCl, 0.5 mM spermine, 0.125 mM spermidine, 1% thiodyglycol, 10 mM EDTA, 20 mM KCl, 0.5 mM phenylmethylsulfonyl fluoride, 1 mM benzamidine pH 7.2). Tissue was resuspended in IS buffer containing 50% glycerol (v/v) and 1% Nonidet P-40, and was homogenized by 10 strokes of a tight fitting teflon pestle in a Potter-Elvehjem homogenizer. Frozen embryonic tissue was homogenized directly in IS plus 50% glycerol and Nonidet P-40. Nuclei were pelleted at 5000 g for 10 min, then washed once with IS plus glycerol and Nonidet P-40; and twice in IS plus glycerol alone. Purified nuclei were resuspended in a small volume of RS buffer (20 mM Tris-HCl, 100 mM NaCl, 1% thiodyglycol, 0.5 mM phenylmethylsulfonyl fluoride, 1 mM benzamidine, pH 7.2), then diluted to a DNA concentration of 1 mg/ml in digestion buffer (20 mM TrisHCl, 100 mM NaCl, 10 mM $MgCl_2$, 0.1 mM $CaCl_2$, 0.5 mM phenylmethylsulfonyl fluoride, 1 mM benzamidine pH 7.4) and digested with stated amounts of DNase I for 10 min at 37°C. Digested nuclei were lysed with

0.5% sodium dodecyl sulfate (SDS), 0.5 M NaCl, 10 mM EDTA and treated with proteinase K at 37°C overnight. DNA was purified by phenol extraction and ethanol precipitation. Hypersensitive sites were visualized by the indirect end labeling technique of Wu (14), from Southern blots of restricted DNA transferred from 1-1.5% agarose gels following electrophoresis in Tris-acetate buffer, (40 mM Tris, 20 mM NaAc, 2 mM Na_2EDTA, pH 7.2) onto Zeta-Probe nylon membranes (Bio Rad) by direct alkaline transfer (15).

Contact Hybridization

A 1.8kb EcoRI-PstI DNA restriction fragment containing the block of repetitive DNA elements of interest was isolated by agarose gel electrophoresis. This fragment was subjected to further digestion with restriction enzymes and the fragments generated were end labeled with $\gamma\ ^{32}P$ ATP using polynucleotide kinase (16). The restriction ladders were visualized by electrophoresis of the labeled fragments through 2% 1.5 mm thick vertical agarose gels in tris-borate buffer (100 mM Tris, 100 mM borate, 2 mM Na_2EDTA, pH 8.3). Hybridization was performed to sheared denatured rooster liver DNA, fixed to Zeta-Probe, essentially as described by Zasloff and Santos (17).

DNA Sequencing

Sequencing of the 1.8kb EcoRI-PstI DNA fragment was accomplished by subcloning exonuclease Bal31 generated, deletion mutants into M13 and sequencing via the chain termination procedure of Sanger et al. (18), as modified by Anderson (19). Additional dideoxy sequencing was performed on individual restriction fragments subcloned into M13 mp8, mp9 or mp19. All sequencing reactions were primed with a 14 mer sequencing primer developed by Lau et al. (20). Any ambiguities in the sequence were resolved by chemical sequencing, as described by Maxam and Gilbert (21), on selected DNA restriction fragments labeled either by polynucleotide kinase or by fill-in synthesis with the Klenow fragment of DNA polymerase I. DNA sequences were visualized by electrophoresis of the sequencing reactions on 8 and 6% polyacrylaimide gels containing 7M urea, followed by autoradiography.

S1 Analysis of Supercoiled Plasmids

Supercoiled 1.8pBR322 was prepared by isolating the plasmid from E. coli HB101 using an alkaline lysis procedure (22), followed by banding the DNA at least twice on CsCl gradients until the plasmid was greater than 80% supercoiled. Supercoiled 1.8pBR322 was treated with S1 nuclease as described by McKeon et al. (23). The S1 treated DNA was subjected to either direct restriction enzyme digestion and was electrophoresed on 1.5% agarose gels in Tris-acetate buffer; or was labeled at the S1 cleavage sites using poly-

nucleotide kinase and γ ^{32}P ATP prior to restriction enzyme digestion and electrophoresis on DNA sequencing gels.

DNA Hybridizations

Hybridizations were carried out at 42°C for 24-48 hr in 50% formamide, 5X SSC, 4 x Denhardt's solution (0.08% bovine serum albumin, 0.08% polyvinyl pyrrolidone, 0.08% Ficol) 0.2% SDS, 10% dextran sulfate, 200 ug/ml yeast RNA and 200 ug/ml sheared, denatured herring-testes DNA. Blots were washed at 52°C in 0.1X SSC, 0.1% SDS over a one and one-half hour period with at least four changes of buffer.

Computer Analysis

Computer analysis of DNA sequence data was performed using the MicroGenie software package from Beckman Instruments. In addition, open reading frames in the 1.8kb EcoRI-PstI fragment were compared to the Dayhoff protein library for homologous proteins.

RESULTS AND DISCUSSION

The 5' flanking repeated region contains an estrogen inducible DNase I hypersensitive site

A previous report by Kok et al. (24) examined the pattern of DNase I hypersensitive sites in the vicinity of the 5' end of the apoVLDLII gene in adult birds. They were able to demonstrate that upon activation of transcription of the apoVLDLII gene in response to the administration of a bollus of estrogen, two new DNase I hypersensitive sites appeared at the 5' end of the gene in the liver, but not in non-expressing tissues.

In Figure 1, we have extended the analysis by examining the region upstream from -2.5kb and by analysing the profile of sites present at early stages of embryogenesis. Panel B illustrates the profile of sites present in a region extending from -4.6 to +0.8kb in nuclei isolated from rooster liver, hen liver and oviduct. In agreement with the results of Kok et al. (24), we observe a major constitutive site in rooster liver located at -1.8kb. However, we find that the dominant constitutive site present at -1.8 kb is is not detectable in either oviduct or in hen liver. In experiments that will be described elsewhere, we have detected specific protein binding at 3 sites between -1.8 and -2.4kb. One of these sites maps within the hypersensitive region centered at -1.8. Loss of the site at -1.8kb appears to take place gradually over a period of several days following activation of the gene in rooster liver suggesting that the decreased sensitivity of this site may be a consequence of prolonged stimulation of the tissue with estrogen rather than

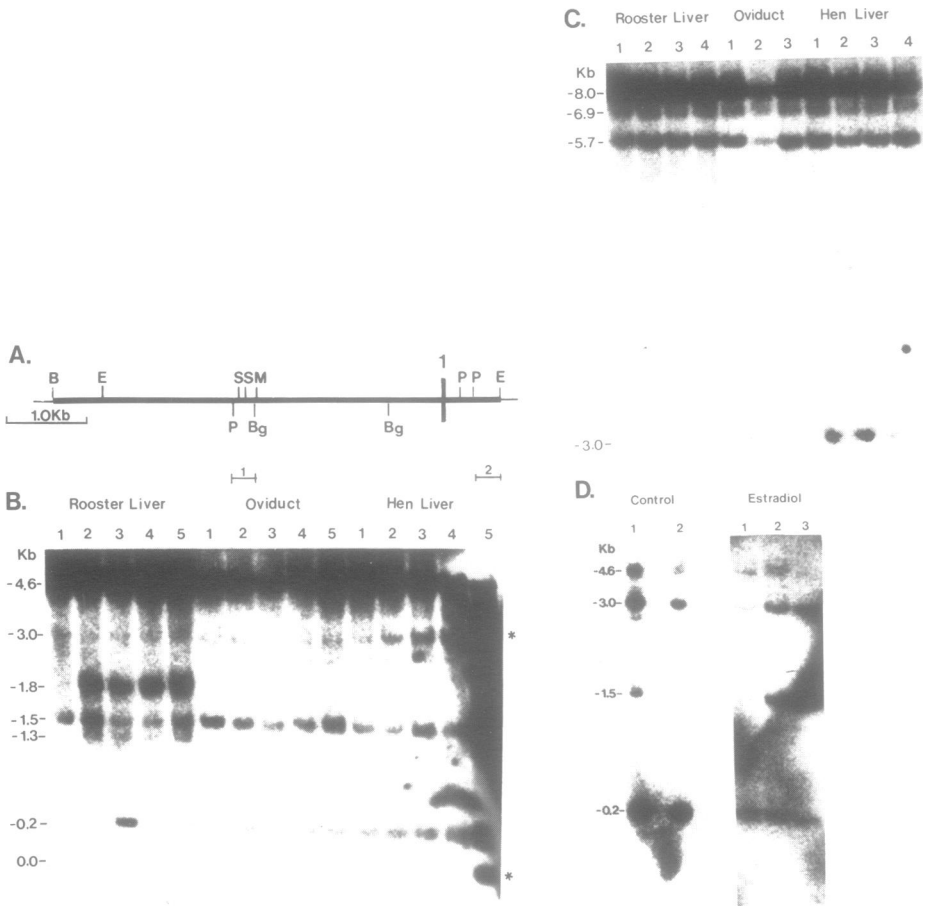

Figure 1: **DNase I hypersensitivity of the 5' flanking region of the apoVLDLII gene.**

A) A partial restriction map covering the regions examined for hypersensitive sites. The map extends from an EcoRI site at +0.8kb to a BamHI site at -5.3kb. The position of the leader exon of the gene is indicated (1). Restriction fragments used to map hypersensitive sites by indirect end-labeling are shown below the map. Probes 1 and 2 abut the BglII site at -2.5kb and the EcoRI site at +0.8kb, respectively. Sites are shown for the following enzymes: BamHI(B), BglII(Bg), EcoRI(E), MspI(M), PstI(P), and SacI(S).

B) Hypersensitive sites between +0.8kb and -4.6kb were analysed in adult birds by preparing nuclei from rooster and hen liver and oviduct according to the protocol described in Materials and Methods. Nuclei were incubated with various concentrations of DNase I at 37°. DNA was then extracted, digested to completion with EcoRI and analysed by Southern blotting using probe 2. The positions of cleavage sites relative to the 5' end of the apoVLDLII gene are indicated at the left hand side of each panel and were

determined by comparison with DNA size markers included in the gels. The intense band at -4.6kb represents the intact fragment produced by EcoRI digestion and is not a hypersensitive site. Lanes 1, DNA from nuclei that were digested with EcoRI immediately after isolation; Lanes 2-5, DNA from nuclei incubated for 10 min with 0, 1, 5, and 10U of DNase I/ml, respectively. Cleavage at some hypersensitive sites was detectable in nuclei without incubation (Lane 1). This cleavage is not observed in DNA prepared by rapid extraction procedures and appears to be attributable to the action of endogenous nucleases during isolation of nuclei. Asterisks indicate fragments produced by cleavage at sites whose sensitivity is enhanced in hen liver.
C) Hypersensitive sites in repeated DNA sequences were analysed as in B) except that the DNA was digested with BglII and hybridized with probe 1. Rooster and hen liver: Lanes 1-4, DNA from nuclei incubated with 0, 1, 5 and 10U DNase I/ml at 37° for 10 min. Oviduct: Lanes 1-3, DNA from nuclei incubated with 0, 5, and 10U DNase I/ml.
D) The hypersensitivity profile in day 7 embryos was analysed in nuclei prepared from normal controls as well as eggs that had been treated with 17β-estradiol on day 5, as described in Materials and Methods. Following treatment with DNase I DNA was extracted and digested with EcoRI followed by Southern blotting with probe 2. Control: Lanes 1 and 2, DNA from nuclei incubated at 37° for 10 min with 0 or 5U DNase I/ml. Estradiol: Lanes 1-3 nuclei incubated with 0, 1 and 5U DNase I/ml.

activation of the gene per se (Haché and Deeley, unpublished, 24). In rooster liver, oviduct and hen liver, we detect a constitutive hypersensitive site located at -1.5kb. In hen liver, where the gene is being expressed, the hypersensitivity profile differs from both oviduct and rooster liver in that there is the typical site located at or very close to the 5' end of the gene also detected by Kok et al. (24) and a site at -3.0kb that is barely detectable in the other two tissues is markedly enhanced. This is shown more clearly in panel C. The end labeled probe used in these experiments was slightly less than 300 nucleotides long and was chosen because it extended from -2.5kb, essentially to the boundary of the repeated region beginning at -2.8kb. Thus we were able to map this hypersensitive site with considerable accuracy to 225-275 nucleotides from the 3' end of the repeated region. Since the site lies within a highly conserved 3'proximal region of a moderately repeated element (see data presented below), the use of additional probes to further improve resolution was precluded. Two other DNase I hypersensitive sites are also apparent in Figure 1C. They map to -5.7 and -6.9kb and are present in all of the tissues examined. The upstream site maps within the region of haplotypic variation and hence may or may not appear in different strains of birds.

During embryogenesis, expression of the apoVLDLII gene in response to exogenously administered estrogen cannot be detected before day 9 of development (4), whereas activation of the vitellogenin gene cannot be detected before day 11. Burch and Weintraub have demonstrated that constitutive hyper-

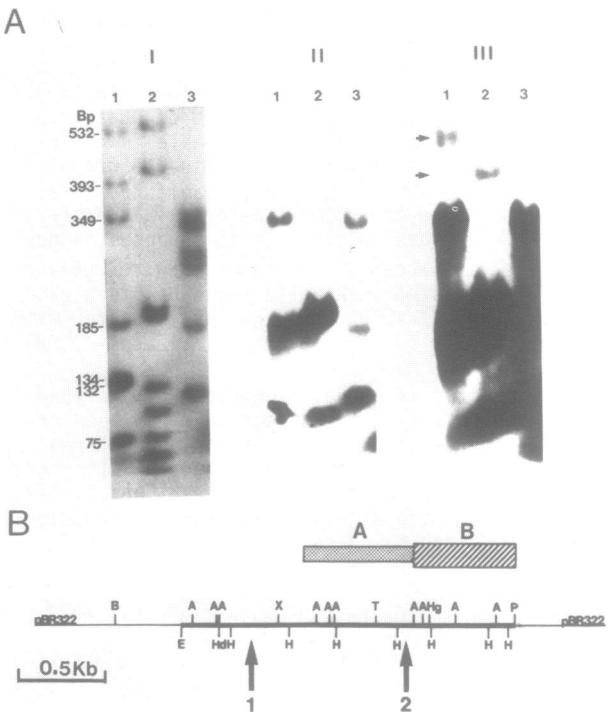

Figure 2: A) **Contact hybridization analysis of repeated elements between -2.8 and -4.6kb.** The 1.8kb EcoRI-PstI fragment was cleaved with various restriction enzymes and the fragments end-labeled with gamma ^{32}P ATP using polynucleotide kinase. Each digest was electrophoresed through a 2% agarose gel, and transferred by passive diffusion under hybridization conditions to Zeta probe nylon membrane that had been presaturated with denatured rooster liver DNA, as described in Materials and Methods. Following washing the blot was subjected to autoradiography. Lanes 1-3, fragments generated by digestion of the 1.8kb fragments with SfaNI/XbaI, AluI or HinfI, respectively. Panel I) Autoradiograph of the gel prior to transfer illustrating the relative efficiencies of labeling sub-fragments. Panel II) Overnight exposure of the contact blot. Panel III) Over-exposure of the blot to reveal poorly labeled fragments of low reiteration frequency. The relative repetition frequencies of each fragment were determined by densitometry of the autoradiograph of the gel prior to transfer and the autoradiographs of the blot.
B) **Schematic representation of repeated elements with two different reiteration frequencies.** Locations of the lower (A) and higher (B) frequency elements are indicated by stippled hatched boxes, respectively. S1 hypersensitive sites are indicated by arrows. Restriction enzyme sites shown: AluI(A), EcoRI(E), HinfI(H), and PstI(P).

sensitive sites around the vitellogenin gene that are present in rooster liver are already detectable in embryos at day 13 (44). However, no analyses were carried out at a stage of development at which the vitellogenin gene

will not respond to estrogen. We have analysed the hypersensitivity profile flanking the apoVLDLII gene as early as day 7 of embryogenesis in both control embryos and embryos that were stimulated with estrogen at day 5. Panel D of Figure 1 profiles the pattern of DNase I hypersensitive sites in the first 4.6kb of apoVLDLII flanking sequence at day 7 of embryogenesis. At this stage of development, the flanking region of the gene is already marked by two of the constitutive sites detected in rooster liver. These are located at -0.2 and -1.5kb. In addition, the pattern of sites in the embryo differs from the pattern in the adult in two respects: 1) The prominent site present at -1.8kb in control rooster liver is not detectable and 2) A site at -3.0kb is very evident in both control and estrogen treated embryos. The location of this site is indistinguishable from one that is enhanced in hen liver but difficult to detect in control rooster liver and oviduct. Consistent with the inability to activate the gene at this stage of development, no alterations were observed upon treatment with estrogen.

It appears likely that nuclease hypersensitive sites in chromatin reflect localized perturbations of the normal nucleosomal array resulting from either alterations in DNA topology and/or interactions with specific DNA binding proteins (25-29). In a number of cases, evidence has been presented that supercoiling dependent, non-B form DNA structures may be involved in the formation of such sites (30-36). It has also been demonstrated that these structures can be generated in naked, supercoiled DNA and detected by digestion with single strand specific nucleases such as S1 (23,34,35,37,41). In experiments that will not be described in detail, we examined the possible contribution of supercoiling dependent alterations in DNA secondary structure in formation of the DNase I hypersensitive site at -3.0kb, by subjecting a supercoiled recombinant plasmid that contained the 1.8kb EcoRI-PstI restriction fragment spanning the repeated sequence region to digestion with the single strand specific nuclease S1. Two S1 senstive, supercoiling dependent sites were detected in the repeated sequence region, the locations of which were determined by analysis on DNA sequencing gels. Cleavage at one site occurred over a 23 nucleotide region centered at nucleotide 360 (see Figure 3) of the insert, and at the second site over a 20 bp region centered around position 1230 (Figure 3). Thus neither site is coincident with the DNase I hypersensitive site at -3.0kb which we estimate lies between nucleotides 1560 and 1620. In the first site a short region of inverted repetition occurs, while a short poly G/poly C stretch of 4 nucleotides occurs at the second. The reason for preferential cleavage occurring at the latter site is not obvious, since there are several

other runs of 4-5 C's or G's in the vicinity of the site that are not cleaved and, based on the complete sequence shown in Figure 3, there are no other features of primary sequence adjacent to the site that would be expected to facilitate cleavage. Conversely, the lack of cleavage at -3.0kb in naked DNA suggests that in this instance the hypersensitivity apparent in chromatin is not dependent solely on DNA sequence but may also involve specific DNA protein interactions. S1 digestions were also carried out with isolated nuclei from adult birds to determine if the sites detectable in supercoiled plasmids were present in chromatin. These experiments did reveal specific S1 sensitive sites in the 5' flanking region of the gene but not within the repetitive region, suggesting that the secondary structures recognized by S1 nuclease in supercoiled plasmids either are not present in chromatin, or are not accessible to the enzyme.

Organization of the repetitive element block

The presence in middle repetitive DNA of a DNase I hypersensitive site whose sensitivity is influenced by the developmental stage of the bird and the state of activity of the apoVLDLII gene in the adult prompted a more detailed analysis of the organization and sequence of the repeated elements between -2.8 and -4.6 kb. The number of elements present, and their relative reiteration frequency was investigated by contact hybridization (42). The results of a typical experiment are displayed in Figure 2A. End labeled restriction fragments of the 1.8kb EcoRI-PstI region containing the repetitive DNA were electrophoresed through an agarose gel (Panel I) and transferred under hybridization conditions to a filter saturated with bound, sheared, denatured genomic DNA. The radioactive blot obtained was washed to remove non-specific hybridization and, then subjected to autoradiography (Panels II and III). Under the conditions used the extent of hybridization is proportional to the relative reiteration frequencies of sequences present in the various restriction fragments, estimates of which were obtained by densitometric analysis of the autoradiograph. These experiments revealed the presence of two regions, located in the 3' 1,100 nucleotides of the 1.8kb fragment, that differed in copy number by a factor of 20-30, the 3' proximal element being the more highly reiterated. The hybridization pattern observed suggests that the two elements are contiguous, since fragments from the different repetition frequencies had common ends in the HinfI and SfaNI/XbaI digests. On the basis of these experiments, the estrogen inducible DNase I hypersensitive site at -3.0kb was centered approximately 250 nucleotides from the 3' end of the higher frequency repeat (element B, Figure 2B).

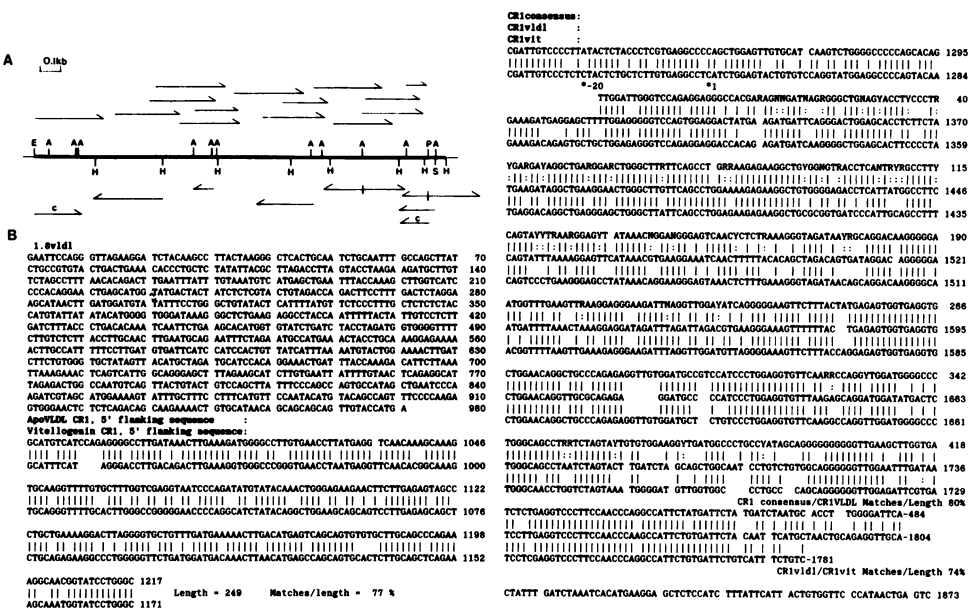

Figure 3. A) **Restriction map and sequencing strategy for the region extending from -2.7 to -4.6kb.** Arrows above the map indicate sequences derived from Bal31 deletions generated as described in Materials and Methods. Arrows below the map indicate sequences derived from dideoxy sequencing of individual restriction fragments. C indicates regions confirmed by chemical sequencing. Enzyme sites shown: AluI(A), EcoRI(E), HinfI(H), PstI(P) and SacI(S).
B) **DNA sequence of the 5' flanking region of the apoVLDLII gene from the EcoRI site at -4.6kb to the SacI site at -2.7kb.** The sequence of the upper strand is shown and is numbered 5' to 3' from the EcoRI site. The degree of similarity between the CR1VLDL and CR1vit flanking, low-frequency, repeated regions is shown as a pairwise comparison on the left side of the figure. On the right hand side of the figure is shown a pairwise comparison of CR1VLDL with the CR1 consensus sequence (11) and the CR1vit sequence. The start of the CR1 consensus sequence is marked by *1 and the 5' flanking sequence extending to *-20 is taken from the sequence of a CR1 element, CR1 CMb, found downstream from the chicken calmodulin gene. Numbering of the apoVLDLII sequence indicates distance from the EcoRI site at -4.6 kb. In order to facilitate comparison of the VLDL and vitellogenin repeated sequences, the vitellogenin sequence has been numbered by aligning the 3' ends of the CR1 VLDL and CR1vit elements.

Sequence of the repetitive elements

The 1.8kb EcoRI-PstI fragment was digested with Bal 31 to generate a series of truncated templates in M13 vectors for sequencing by the dideoxy procedure according to the strategy shown in Figure 3A. The sequence obtained is shown in Figure 3B. Comparison of this sequence with other chicken repeated DNA elements identified the higher frequency repeat (element B in Figure

2B) as a member of the avian CR1 family (9). The reiteration frequency of this family has been estimated to be between 10-30,000 per haploid genome. Using the convention adopted for this element the copy in the apoVLDLII flanking region "points" 5' to 3' towards the gene. This conserved orientation relative to proximal structural genes has been observed for 7 out of 8 other members of this family including those flanking the ovalbumin, calmodulin, and U1 RNA genes (9,11,43). The exception being the most recently characterized element prior to this study, CR1Vtg, which is located downstream from the chicken vitellogenin gene (12). Since the copy number of the higher frequency CR1 repeat was known, we were able to assign a reiteration frequency of 500 to 1000 copies per haploid genome to repeat A.

Figure 3B shows that the CR1 family member flanking the apoVLDLII gene, displays an overall similarity of 80% to the consensus sequence for the CR1 family compiled by Stumph et al. (11). A comparison of CR1VLDL with individual family members indicated that the match extended at least 20 nucleotides beyond the 5' end of the consensus sequence to include the 5' terminal region of the element CR1 CMb which occurs downstream from the chicken calmodulin gene. Because of a lack of available sequence data for CR1 CMb, neither its 5' end nor that of CR1VLDL can be precisely identified from this comparison. However, contact hybridization (Figure 2) placed the 5' transition between CR1VLDL (repeat B) and the lower frequency repeat, between positions 1230 and 1280, or approximately 40 to 90 nucleotides further upstream than could be identified by direct comparison of the two sequences. That the CR1 sequence extends to this location is supported by comparison with the 3' flanking region of the vitellogenin gene containing CR1Vtg (Figure 3B). The match between CR1VLDL and CR1Vtg extends from position 1221 of the apoVLDLII flanking sequence to 1789 with an overall similarity of 74%. This is approximately 120 nucleotides larger than the current CR1 consensus sequence and places one of the S1 sensitive sites detected in plasmids containing this region within less than 10 nucleotides of the junction between low and high frequency repeats. Alignment of the two sequences can be maintained into the region covered by the lower frequency repeat provided that a compensatory shift is made to accommodate two deletions of 16 and 18 nucleotides that appear to have occurred in the apoVLDLII sequence between 1217/1218 and 1220/1221, respectively. The similarity between the sequences extends at the level of 77% from position 1217 to 971, a total length of 818 nucleotides. At the moment, these are the only two CR1 elements that have been found to share these homologous lower frequency flanking sequences. Whether or not their

presence is of functional significance in the coordinate regulation of apoVLDLII and vitellogenin genes, or is related to the evolution of the CR1 family from a more extensive ancestral sequence, remains to be determined.

Examination of the sequence in Figure 3B for open reading frames (ORFs) revealed several extensive open blocks encoding stretches of at least 80 amino acids. All extended open reading frames occur within the repeated DNA elements. Those obtained by direct translation of the sequence shown are organized such that there are 3 overlapping frames extending from nucleotide 679-1533. The introduction of two frame shifts would generate a region capable of coding for a polypeptide in excess of 30000 daltons. However, the sequence displays no significant similarity to ORFs in mammalian LINE families and scanning of the Dayhoff and GenBank data bases has not revealed any significant similarity to any other known polypeptides. Additional computer analysis of the repetitive sequences flanking the apoVLDLII gene failed to reveal any putative promoter sequences for either RNA polymerase II and III, nor have we been able to detect any transcript specific to this locus.

Origin and function of CR1 elements

The origin of the avian CR1 family, its mechanism of dispersion and possible function are currently a matter of speculation. The sequence appears to be restricted to birds and to be located in areas of transitional nuclease sensitivity. The sequence downstream from the vitellogenin gene is located in a region marked by three hypersensitive sites specific to the liver but which are present before the gene is activated (12,44). The precise location of these sites relative to the CR1 element has not been determined. In the case of CR1VLDL, we have mapped a hypersensitive site to one of the two most highly conserved regions of the element between nucleotides 1560-1620 (similarity 90% across family members). However, this site is approximately 200 nucleotides downstream from the other highly conserved region, to which specific protein binding has been detected (10).

The presence of nuclease hypersensitive sites within middle repetitive DNA elements flanking structural genes has not been widely reported. Some examples include constitutive hypersensitive sites in a middle repetitive element flanking the mouse prolactin gene and in a repetitive element within the chicken alpha globin gene cluster (45,46). In only a few instances has the presence of DNase I hypersensitive sites within middle repetitive DNA elements been observed to be influenced by the state of activity of the neighboring gene. Perhaps the most comparable to that described here is a DNase I hypersensitive site mapping within a middle repetitive element 3'

to the chicken lysozyme gene which is only apparent after the gene becomes expressed in the oviduct (47). The middle repetitive element in which this site resides has not been characterized, but it has been shown to map within the region of intermediate DNase I sensitivity 3' to the lysozyme gene.

The most recent hypothesis proposed to explain the origin of the CR1 family is that these elements are the remnants of retroviral long terminal repeats (LTR). This is based on certain structural features of the elements rather than overall sequence similarity, which is low (the maximum overall similarity we have detected with retroviral sequences is approximately 40%). One feature which is displayed by a subset of CR1 elements is a polypurine tract 'AGGGGGAA', occurring at approximately the same location in all cases, which matches the primer binding site preceding the 3' LTR of the avian retrovirus, Rous associated virus 0(RAV-0). The repeated region 1.8VLDL has this sequence located at position 1514-1521 which aligns with position 184-191 of the CR1 consensus (Figure 5B). However, there is no additional similarity to the remaining RAV-0 LTR sequence.

The fact that many CR1 elements have heterogeneous 5' ends while retaining a well defined 3' terminus is consistent with the suggestion that the remainder of the original viral sequence may have been lost as a consequence of recombination. In the majority of cases, this appears to have left only the putative LTR sequence itself and up to 200 nucleotides of 5' flanking sequence. The fact that the homology between elements flanking the apoVLDLII and vitellogenin genes extends more than 300 nucleotides beyond the postulated 5' end of the CR1 consensus and spans a transition between reiteration frequencies suggests that, in a minor population (5-10%), recombination may have left larger segments of the ancestral flanking sequence. It is possible that the S1 hypersensitivity of the boundary between high and low frequency segments of the repeat may have contributed to the loss of these sequences flanking most CR1 elements. The presence of 3 extended open reading frames covering approximately 850 nucleotides immediately upstream from the putative primer binding site of CR1VLDL is also consistent with the possibility of a viral origin for these sequences.

At present, all of the CR1 elements described have been found flanking rather than within genes. However, if these elements are remnants of retroviral insertion events or have amplified perhaps by retroposition as proposed for mammalian Alu sequences, it might be anticipated that they could be present within introns or untranslated regions, providing that their presence did not deleteriously affect expression of the gene involved. Screening of all

avian sequences on file at GenBank revealed evidence for one instance where this may have occurred, involving a sequence within the large intron of the chicken embryonic epsilon globin gene. Sequence similarity was detected at the level of 77% over a stretch of 51 nucleotides between positions 1740 and 1789 of 1.8 VLDL and nucleotides 908 and 957 of the epsilon globin gene. Comparison of the intron sequence with other CR1 elements revealed more extensive matches with elements designated CCM1 (75% match over 68 nucleotides) and CR1 sequence D (70.4% match over 81 nucleotides), supporting the suggestion these sequences may not necessarily be restricted to locations in flanking regions.

At the moment, it is not known whether the hypersensitive site in CR1VLDL is functionally significant as far as expression of the apoVLDLII gene is concerned. The fact that it, like the protein binding site identified previously occurs in a highly conserved region of the element suggests that it may be. Certain rat long interspersed repeats have recently been shown to act as transcriptional "silencers" (48). Very recently, a CR1 element located approximately 1.0kb upstream from the chicken lysozyme gene has been found to contain 'silencer' like elements (49). These are located within a block of 277 bp corresponding to the central region of the CR1 element. This region displays a similarity of 73% to nucleotides 1403-1683 of 1.8VLDL (Figure 3B) and spans the DNase I hypersensitive site we have detected that is located between nucleotides 1560 and 1620. Within the 277 bp region, Baniahmad et al. have identified three elements of 9-10 nucleotides that share a high degree of similarity with other silencer sequences, one of which falls at a position corresponding to the hypersensitive region of 1.8VLDL (49). However, small deletions or insertions have occurred at all sites in 1.8VLDL that correspond to the silencer consensus sequences. Whether or not 1.8VLDL has retained silencer activity despite these alterations is currently under investigation. If CR1 elements evolved from retroviral LTRs they may have, or at least had, the potential to act as enhancers. Thus, it is possible that the 'silencer' activity of some of these elements could be attributable to their ability to form non-functional, enhancer like structures.

ACKNOWLEDGEMENT

We wish to acknowledge the excellent technical assistance of Monika Vasa and Ruth Burch and the suggestions of our colleagues Don Back, Edi Alvarez and Alan Cochrane. We thank Marion Webb for her assistance at all stages of preparation of this manuscript. This research was supported by

grants from the Medical Research Council of Canada to R.G. Deeley and an Ontario Graduate Student Scholarship to R.J.G. Haché.

*To whom correspondence should be addressed

+Present address: Institut für Molekularbiologie und Tumorforschung, Philipps-Universitat, E. Mannkopf-Strasse 1, 3550 Marburg, FRG

REFERENCES

1. Chan, L., Jackson, R.L., O'Malley, B.W. and Means, A.R. (1976) J. Clin. Invest. **58**, 368-379.
2. Jackson, R.L., Lin, H-Y., Chan, L. and Means, A.R. (1977) J. Biol. Chem. **252**, 250-253.
3. Wiskocil, R., Bensky, P., Dower, W., Goldberger, R.F., Gordon, J.I. and Deeley, R.G. (1980) Proc. Natl. Acad. Sci. U.S.A. **77**, 4474-4478.
4. Colgan, V., Elbrecht, A. Goldman, P., Lazier, C.B. and Deeley, R.G. (1982) J. Biol. Chem. **257**, 14453-14460.
5. Wiskocil, R., Goldman, P. and Deeley, R.G. (1981) J. Biol. Chem. **256** 9662-9667.
6. Meijlink, F.C.P.W., Van het Schip, A.D., Arnberg, A.C., Wieringa, B., AB, G. and Gruber, M. (1981) J. Biol. Chem. **256**, 9668-9671.
7. Strijker, R., Blom van Assendelft, G., Dikkeschei, B.D., Gruber, M. and AB, G. (1986) Gene **45**, 27-35.
8. Haché, R.J.G., Wiskocil, R., Vasa, M., Roy, R.N., Lau, P.C.K. and Deeley, R.G. (1983) J. Biol. Chem. **258**, 4556-4564.
9. Stumph, W.E., Kristo, P., Tsai, M-J. and O'Malley, B.W. (1981) Nucl. Acids Res. **9**, 5383-5397.
10. Sanzo, M., Stevens, B., Tsai, M-J. and O'Malley, B.W. (1984) Biochemistry **23**, 6491-6498.
11. Stumph, W.E., Hodgson, C.P., Tsai, M-J. and O'Malley, B.W. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 6667-6671.
12. Van het Schip, F., Samallo, J., Meijlink, F., Gruber, M. and AB, G. (1987) Nucl. Acids Res. **9**, 4193-4201.
13. Mirkovitch, J., Mirault, M-E. and Laemmli, U.K. (1984) Cell **39**, 223-232.
14. Wu, C. (1980) Nature **286**, 854-860.
15. Reed, K.C. and Mann, D.A. (1985) Nucl. Acids Res. **13**, 7207-7221.
16. Richardson, C.C. (1965) Proc. Natl. Acad. Sci. U.S.A. **54**, 158-165.
17. Zasloff, M. and Santos, T. (1980) Proc. Natl. Acad. Sci. U.S.A. **77**, 5668-5672.
18. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463-5467.
19. Anderson, S., Gait, M.J., Mayol, L. and Young I.G. (1980) Nucl. Acids Res. **8**, 1731-1743.
20. Lau, P.C.K. and Spencer, J.H. (1982) Biosci. Rep. **2**, 687-696.
21. Maxam, A.M. and Gilbert, W. (1980) Meth. Enzymol. **65**, 499-560.
22. Birnboim, H.C. and Doly, J. (1979) Nucleic Acids Res. **7**, 1513.
23. McKeon, C., Schmidt, A. and De Crombrugghe, B. (1984) J. Biol. Chem. **259**, 6636-6640.
24. Kok, K., Snippe, L. AB, G. and Gruber, M. (1985) Nucl. Acids Res. **13**, 5189-5202.
25. Dynan, W.S. and Tjian, R. (1985) Nature **316**, 774-778.
26. Butler, P.J.G. (1983) CRC Crit. Rev. Biochem. **15**, 57-91.
27. Reeves, R. (1984) Biochim. Biophys. Acta **782**, 343-393.
28. Yaniv, M. and Cereghini, S. (1986) CRC Crit. Rev. Biochem. **21**, 1-26.
29. Weisbrod, S. (1982) Nature **297**, 289-295.

30. Zhu, J., Allan, M. and Paul, J. (1984) Nucl. Acids Res. **12**, 9191-9204.
31. Selleck, S.B., Elgin, S.C.R. and Cartwright, I.L. (1984) J. Mol. Biol. **178**, 17-33.
32. Margot, J.B. and Hardison, R.C. (1985) J. Mol. Biol. **184**, 195-210.
33. Senear, A.W. and Palmiter, R.D. (1982) Cold Spring Harbor Symp. Quant. Biol. **47**, 539-547.
34. Larsen, A. and Weintraub, H. (1982) Cell **29**, 609-622.
35. Weintraub, H. (1983) Cell **32**, 1191-1203.
36. Nickol, J.M. and Felsenfeld, G. (1983) Cell **35**, 467-477.
37. Htun, H., Lund, E. and Dahlberg, J.E. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 7288-7292.
38. Camilloni, G., Della Seta, F., Negri, R. and Di Mauro, E. (1986) J. Biol. Chem. **261**, 6145-6148.
39. Financsek, I., Tora, L., Kelemen, G. and Hidvegi, E.J. (1986) Nucl. Acids Res. **14**, 3263-3277.
40. Yu, Y-T. and Manley, J.L. (1986) Cell **45**, 743-751.
41. Kilpatrick, M.W., Torri, A., Kang, D.S., Engler, J.A. and Wells, R.D. (1986) J. Biol. Chem. **261**, 11350-11354.
42. Wood, W.I., Nickol, J. and Felsenfeld, G. (1981) J. Biol. Chem. **256**, 1502-1505.
43. Stumph, W.E., Baez, M., Beattie, W.G., Tsai, M-J. and O'Malley, B.W. (1983) Biochemistry **22**, 306-315.
44. Burch, J.B.E. and Weintraub, H. (1983) Cell **33**, 65-76.
45. Weintraub, H., Larsen, A. and Groudine, M. (1981) Cell **24**, 333-344.
46. Durrin, L.K., Weber, J.L. and Gorski, J. (1984) J. Biol. Chem. **259**, 7086-7093.
47. Jantzen, K., Fritton, H.P. and Igo-Kemenes, T. (1986) Nucl. Acids Res. **14**, 6085-6099.
48. Laimins, L., Holmgren-Konig, M. and Khoury, G. (1986) Proc. Natl. Acad. Sci. U.S.A. **83**, 3151-3155.
49. Baniahmad, A., Muller, M., Steiner, Ch. and Renkawitz, R. (1987) EMBO Journal **6**, 2297-2303.