Organization of the 3′-boundary of the chicken α globin gene domain and characterization of a CR 1-specific protein binding site

Gabriel Farache, Sergey V.Razin, Félix Recillas Targa and Klaus Scherrer*
Institut Jacques Monod, Université Paris 7, 2 Place Jussieu, 75251 Paris Cedex 05, France

Received December 15, 1989; Accepted January 10, 1990 EMBL accession no. X51600

ABSTRACT

The sequence of a DNA fragment about 1 Kbp long located at the 3′ boundary of the chicken α globin gene domain, including the 3′-side matrix attachment point and the site of transcription termination, was determined. It contains a repetitive DNA element and the AT-rich (easily denaturable) DNA segment conserved at the same position in the duck genome. The repetitive sequence was identified by computer analyzis as being a member of the CR1 family. Within the non-repetitive part of the AT-rich DNA fragment, four topoisomerase II recognition sites were found which might be indicative of matrix attachment. Furthermore, two distinct regions were identified, possessing strong homology to a number of non-coding consensus sequences, one of them to a limited part of the LTR of HTLVIII, and the other to the replication origin of Polyoma virus JC. DNA shift experiments showed that the CR1 repeat binds specifically an abundant nuclear protein factor. The binding site for this factor was identified by footprinting and turned out to be closely related to the previously described recognition site for the TGGCA-binding protein, the chicken analog of nuclear factor 1 (NF-1). Finally, the CR1 repeats within the chicken α and β globin gene domains were mapped. All these observations are discussed in terms of the organization of the 5′ and 3′ boundaries of the functional genomic domains forming a chromatin loop including all avian α type globin genes.

INTRODUCTION

The subdivision of the eukaryotic genome into structural domains of functional significance was long indicated by cytogenetic data and in particular the banding pattern of polytene chromosomes (1−4). Later on large DNA loops anchored to some protein structures were found in deproteinized metaphasic chromosomes and interphase nuclei (5, 6).

It was proposed that such DNA loops represent not only domains of chromosome organisation but also potentially functional units (7). Most extensive results suggesting this idea stem from studies of the chromatin structure of actively transcribed gene domains. When genes are inactive, the chromatin is found in a highly compact state, and the DNA is inaccessible to molecular probes such as DNAse I. Transcriptional activation is accompanied by cooperative unraveling of the chromatin including the coding sequences; genomic domains as long as 100 kilobase pairs (Kbp) which can be preferentially digested by DNAse I were defined in this manner (8). Some of these DNA domains are flanked by regions where a sharp change of chromatin structure takes place, passing from a resistant conformation to a DNAse sensitive state (9−11). This suggests that specific DNA sequences localized there mark the boundaries of the gene domains and could, hence, be involved in such changes.

Although the model of eukaryotic genome organization in segregated domains seems to be largely accepted, and the mechanism of decondensation of specific chromatin domains is considered as a primary step in the control of gene expression (7), little is known about the DNA sequences responsible for defining such genomic domains.

The discovery of a systematic punctuation of eukaryotic DNA by easily denaturable AT-rich regions (12) has suggested that DNA sequences exhibiting such physico-chemical properties could be involved in the definition of structural and functional genomic domains in the eukaryotic genome, as shown for the avian globin genes (13). On the other hand, in the chicken ovalbumin gene domain, members of a family of repetitive sequence, termed the 'CR1 family' (14), have been localized at the boundaries of the genomic domain of sensitivity to DNAse I and of undermethylated DNA (15); the authors propose that CR1 sequences may be involved in defining boundaries of some active chromosomal domains in the chicken genome. Indeed, elements of this family are present in about 7000 copies per genome (14) and some of them are found in specific genomic locations and in orientations which suggest that they might play a role in establishing boundaries of domains of gene activity in chromosomes (9, 15). If this is the case it seems reasonable to expect that certain segments present in the CR1 sequence might be recognised by nuclear factor(s) involved in the control of chromatin structure.

We have taken advantage of available information concerning

* To whom correspondence should be addressed

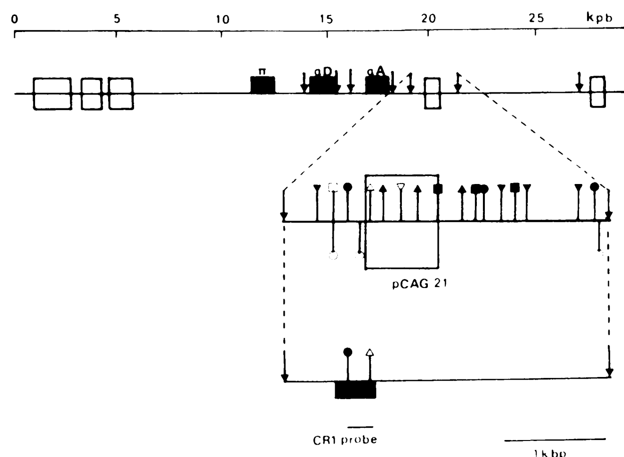

Figure 1—The chicken α globin gene domain—On the top line, black boxes represent the position of the embryonic π, and the adult α^D and α^A globin genes, whereas white boxes represent positions of A+T rich regions. The Bam HI restriction fragment of 3.2 Kbp containing the A+T rich region on the 3′ side of the gene cluster was subcloned in the Bam HI site of the pBR 322 plasmid; the bottom line shows the restriction map of the inserted fragment (pCAG21). In the lower left, the CR1 probe Ava II to Dra I) is shown. Symbols for restriction enzymes are: Ava II, Bal I, Bam HI, Bgl I, Dra I, Hinc II, Hpa II, Sau 3AI, Pvu II.

the structural and functional features of the chicken α globin gene domain to address the question of the structure and possible functional role of specific DNA sequences localised at the boundaries of this gene domain in respect to the definition of active chromosomal regions. The chicken α globin gene cluster is flanked at both sides by easily denaturable AT-rich segments (13). A group of 3 AT-rich segments is situated at the 5′ end of the domain, about 8 Kbp upstream of the gene cluster and, at the 3′-end, an isolated AT-rich 'linker' (ATRL) maps 2 Kbp downstream of the α^A gene. The AT-rich segment at the 3′-end is linked to a repetitive DNA element; it coincides with the end of the domain of gene transcription, of sensitivity to DNAse I and of undermethylated DNA (16, 17); it includes also a site of attachment to the nuclear matrix in erythroid cells (18, 19).

In the work presented here, we undertook the characterization of the DNA segment at the 3′ end of the chicken α globin gene domain. We found that the repetitive DNA element and the AT-rich segment (13, 17, 18) are very closely linked but hardly overlap. Sequencing of the DNA segment led to the observation that the repetitive sequence belongs to the CR1 family. This CR1 element binds specifically to nuclear factor(s) present in bone marrow cells of anemic chicken. Interestingly, the analysis of the sequence in the binding site suggested that this factor is related to both, a nuclear protein specifically binding to the DNAse hypersensitive region upstream of the chicken β^A globin gene (20), and also the ubiquitous TGGCA binding protein (21) which seem to be involved in the control of chromatin structure in chicken cells.

MATERIALS AND METHODS

Recombinant DNA clones

The initial characterization of the recombinant Charon 4A phage clones, containing the chicken α and β globin gene regions has been described (22). A 3.2 Kbp restriction fragment from the α 5 clone (λCαG5), containing the A+T rich region and the repetitive sequence on the 3′-side of the gene cluster, was subcloned in the Bam HI site of the plasmid pBR322 (see Fig. 1) according to (23). The resulting recombinant plasmid (pCAG 21) was used to obtain the DNA probes and restriction fragments for sequencing.

Hybridization probes

Three types of hybridization probes were used in this work. The first, liable to identify repetitive DNA segments, consisted of genomic chicken DNA prepared from erythrocytes and fragmented to about 600 bp by sonication. The second, the CR1 probe was a 230 bp fragment, obtained from the pCAG21 insertion after digestion with HpaII. The third, the AT-probe consisted of a 1600 bp fragment derived from digestion of the pCAG21 insert fragment with the HincII restriction enzyme. Probes for hybridization were labeled by nick-translation (23).

Southern-blot hybridization

For Southern transfers, 1−2 μg per lane of digested recombinant phage or plasmid DNA were loaded on agarose gels. After electrophoresis, DNA fragments were transferred to Gene Screen membranes (NEN) according to the suppliers instruction manual and hybridized to (^{32}P)-labelled probes in 6×SSC at 65°C for 16 hours. Membranes were washed twices in each, 1×SSC at 65°C for 1 hour, and 0.2×SSC at 55°C for 30 minutes (medium stringency washes). High stringency washing included supplementary steps in 0.1×SSC at 65°C during 20 minutes.

DNA sequencing

Was carried out according to the original protocol of Maxam and Gilbert (24).

Extraction of nuclear proteins from chicken erythroblasts

Nuclei of bone marrow cells were obtained from anemic adult chicken, according to methods detailed elsewhere (19). Briefly, bone marrow cells were washed by 4 cycles of centrifugation (1500 rpm) and resuspended in 5 volumes of KSB solution (10 mM KCl, 3mM $MgCl_2$, 1 mM $MnCl_2$, 10 mM Tris HCl (pH 6.9), 0.25 M sucrose). Floating material and the upper layer of white cells were fully discarded. Thereafter, all isolation steps were done at 4°C; all solutions contained 0.5 mM PMSF and 7 mM mercaptoethanol. Cells were lysed in KSB solution containing 0.4 % of the detergent Cemusol NP-6, and disrupted by several strokes of a hydraulically driven Dounce homogenizer. Nuclei were pelleted and washed 3 times in 10 volumes of KSB but at a pH of 7.4. The nuclear pellet was resuspended and extracted by dropwise additon of 5 to 7 volumes of the extraction buffer (0.35 M NaCl, 10 mM Trs-HCl (pH 7.5), 7 mM mercaptoethanol, 0.5 mM PMSF). After incubation at 4°C for 45 minutes with frequent stirring on a Vortex mixer, the supernatant was recovered after centrifugation at 10.000 rpm for 20 minutes. After addition of glycerol to 20% (v/v), the extract was stored at −70°C. Titration of the protein concentration in the extract was done by a Biorad assay according to the suppliers instruction manual.

The electrophoretic retardation assay was performed essentially as described by Strauss and Varshavsky (25). Incubation mixtures containing fixed quantities of the DNA probes (1−2 ng, 10.000 cpm), nuclear protein (2−3 μg) and increasing quantities of sonicated *E. Coli* competitor DNA (0.25, 0.5, 1 and 2 μg) were added to a final volume of 10 μl of a solution containing 10 mM Tris-HCl (pH 7.6), 100 mM NaCl, 3−4% glycerol, 0.5 mM PMSF and 7 mM mercaptoethanol. After incubation at 25°C

during 30 minutes, mixtures were loaded on a 4% polyacrylamide gel in low ionic strength buffer, prepared and run as described (25). After electrophoresis, gels were dried and autoradiographed.

Footprint assay

Proteins (5–10 μg) and *E. Coli* DNA (5 μg) were mixed in a volume of 45 μl. After 10 minutes of pre-incubation, the DNA probe labeled at the 5′-end was added and incubation continued for 30 minutes. The solution (complemented to 5 mM $MgCl_2$ final concentration) was digested for 1 minute with 200 μg of DNAse I at 25°C; digestion was arrested by adding EDTA to 5 mM final concentration. The solution was loaded and fractionated in a preparative gel at low ionic strength. Bands of differential electrophoretic mobility were detected by autoradiography and DNA was eluted from gel slices. Equivalent amounts of Cerenkov radioactivity of the isolated DNA fractions were resolved by denaturing 8% polyacrylamide gel electrophoresis and autoradiography. G+A specific sequencing reactions (24) were carried out and resolved in parallel to the DNA footprint fractions.

RESULTS

Localization of the repetitive sequence in the 3.2 Kbp Bam HI fragment

The position within the α globin gene domain of the DNA fragments analyzed as well as the restriction map of the pCAG21 insertion established in preliminary experiments, are shown in Fig.1.

To find the position of the repetitive element within the pCAG21 insertion, the DNA of plasmid pCAG21 (1–2 μg) was digested with the appropriate restriction enzymes, and the resulting fragments were separated on a 1.5% agarose gel, transferred to 'Gene Screen' membrane (NEN) and hybridized with nick-translated genomic chicken DNA. After repetitive washing in medium stringency conditions, the blot was dried and exposed for autoradiography.

The conditions used permit detection of hybridization signals only for restriction fragments containing repetitive DNA sequences. The hybridization pattern obtained and the position of the corresponding restriction fragments are shown in Fig. 2. The minimal DNA segment giving a hybridization signal with the total DNA probe was found to be flanked on its 5′-side by the first HpaII restriction site; a very strong signal was observed on the 230 bp HpaII restriction fragment. On the 3′-side, DNA fragments giving hybridization signals are flanked by the first Sau 3A restriction site. The hybridization signal obtained with the 2.5 Kbp HpaII fragment was quite low; this indicated that most of the repetitive element is contained in the 230 bp long HpaII restriction fragment. Since most if not all of the rapidly denaturing DNA is in the 2.5 Kbp HpaII fragment (Moreau et al., unpubl. obs.), this result indicates that the AT-rich region and the repetitive sequences are two different, although closely linked genomic elements.

Characterization of the repetitive sequence

The genomic segment about 1 Kbp long containing the repetitive and AT-rich DNA was sequenced according to the strategy shown in Fig. 3A; the DNA sequence is presented (Fig. 3B) in the direction of globin gene transcription. Particular features noticed will be discussed in the 'Discussion' section.

The computer analysis of this DNA sequence revealed the presence of a region (in positions 36–347) sharing extensive homology with the consensus of the CR1 repetitive element (14). This made it possible to conclude that the repetitive DNA element found in the hybridization experiments within the pCAG21 insertion belongs to the family of CR1 repeats. According to the polarity assigned in correlation to the related mammalian Alu sequences, the identified CR1 repeat is oriented towards the nearby structural gene (namely the α^A gene). The same type of orientation (toward the nearby structural gene) is typical of all the other CR1 elements that have been characterized up to now.

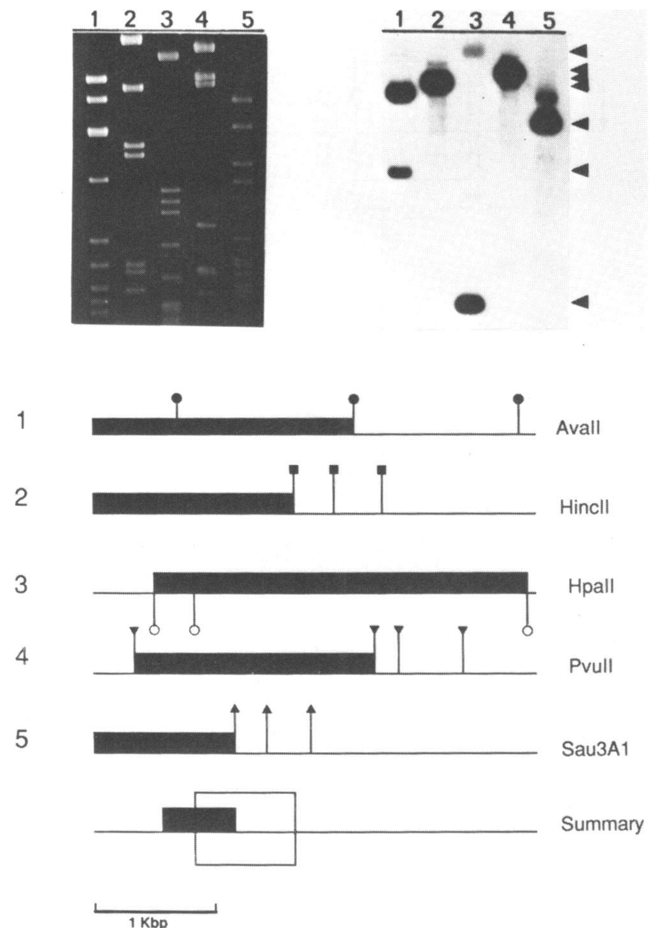

Figure 2–Localization of the repetitive sequence contained in the plasmid pCAG21– DNA of plasmid pCAG21 (1–2 μg) was digested with the restriction enzymes indicated and fractionated (upper left panel). Restriction fragments were transfered to a Gene Screen membrane and hybridized with nick translated chicken genomic DNA. After washing in conditions of medium stringency (see 'Methods'), the membrane was dried and autoradiographed. The right panel shows the autoradiography of the membrane after hybridization. Slots 1 to 5 show respectively: Ava II, Hinc II, Hpa II, Pvu II and Sau 3AI digestion of plasmid pCAG21 DNA in combination with the enzyme Bam HI. The arrows point to the main hybridizing band. The bottom maps represent the restriction fragments (black boxes) hybridizing with the genomic probe. The summary map gives the position of the minimal region which contains the repetitive sequence, and indicates the position of the A+T segment (open box). Symbols for restriction enzymes are as for Fig. 1.

Localization of CR1 and AT-rich sequences elsewhere in the chicken and duck α and β globin gene domains

The localization of the CR1 repetitive sequence at the 3′ end of the α globin gene domain, precisely at the boundary of the undermethylated and DNAse I sensitive region, made it possible

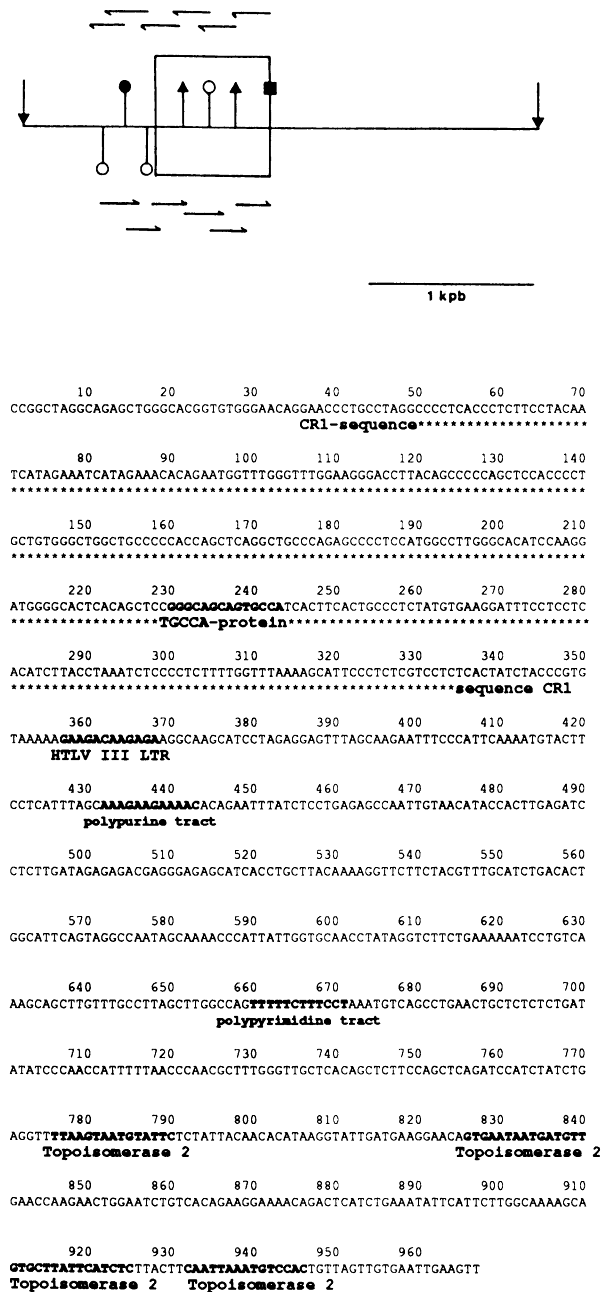

Figure 3 – Sequencing strategy and nucleotide sequence of the region containing the repetitive element and the AT-rich region – (A) Sequencing strategy: symbols for restriction enzymes are as in Figure 1. The open box indicates the position of the A+T rich sequence. (B) The complete nucleotide sequence of the region contained between the first Hpa II restriction site and the first Hinc II restriction site. Start and end of the CR1 sequence are indicated, as well as the topoisomerase II sites (nucleotides 776–790, 826–840, 911–925 and 932–946). Notice the polypurine (354–364, 432–443) and polypyrimidine (660–668) stretches (see text). Multiple regions possessing homology with other sequence elements (see text) are outlined.

to suggest that CR1 elements in general participate in the organization of the boundaries of avian genomic structural and functional domains. To test the idea it seemed reasonable to analyze the distribution of CR1 elements within the large cloned areas of the chicken and duck genomes. In the present investigation the distribution of CR1 elements within the domains of α and β globin genes was studied; the positions of all repetitive

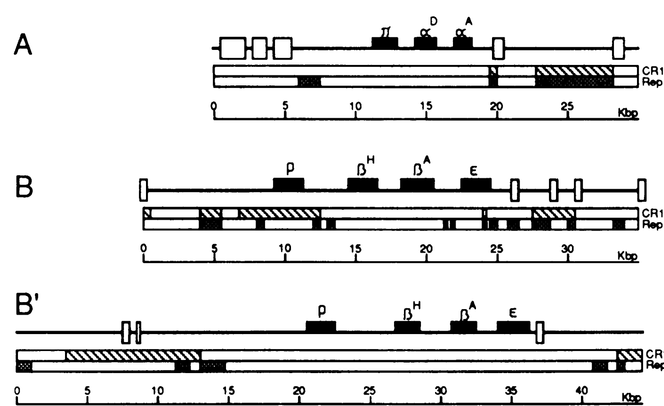

Figure 4 – Maps of the localization of CR1 sequences in the α and β globin gene domains ofchicken and duck (data not shown) – (A) The chicken recombinant phages pCαG6 and pCαG2 (19, 22, 26, 28) DNA was digested respectively with Hind III and Bam HI restriction enzymes. Restriction fragments were separated on a 1% agarose gel, transfered to a 'Gene Screen' membrane and hybridized with the nick-translated, 230 bp long Hpa II fragment, which contains a large part of the CR1 sequence (see Figure 2). The membrane was washed in medium stringency conditions, dried and autoradiographed. The panels show the maps indicating the positions of genes (black boxes), the A+T rich regions (white boxes), and repetitive sequences (grey boxes). (A) The positions of the fragments hybridizing with the CR1 probe are outlined by hatched boxes. No hybridization was detected with any restriction fragment contained in the pCαG6 clone (26, 28). – (B) The CR1 probe was hybridized with recombinant lambda phages containing the chicken β globin gene domain digested with various restriction enzymes (Bam HI, Eco RI, Hind III) and transferred to a 'Gene Screen' membrane. The map represents the chicken β globin gene region with the positions of genes (black boxes) (22), A+T rich regions (white boxes) (13) and repetitive sequences (grey boxes) (27). Restriction fragments of each clone hybridizing to the CR1 probe are indicated by hatched boxes. (B') The same probe as in (A) and (B) was hybridized with recombinant cosmid and plasmid DNA containing the duck β globin gene domain (29). Presentation as for (B).

DNA sequences in these gene domains have been previously reported (26, 27).

To find the positions of the CR1 elements, the 230 bp HpaII fragment of the pCAG21 insertion (i.e. the fragment containing the largest part of the CR1 repetitive sequence localized at the 3′ boundary of the α globin gene domain) was used as a hybridization probe ('CR1 GlA1' probe; for 'CR1 globin α 1'). The results (not shown) of the hybridization experiments are summarized in Fig. 4. In the chicken α globin gene domain, in addition to the 3.2 Kbp long Bam HI segment, only the repetitive sequence located downstream of the globin gene domain (26) in the 7 Kbp long Bam HI fragment of the λCαG2 clone gives an appreciable hybridization signal with the CR1 probe. The repetitive sequences contained in the 5′-part of the domain (the λCαG6 clone; Fig. 4A) did not hybridize with the probe containing the CR1 GlA1 sequence. CR1 is hence present at a unique site within the available part of the α globin gene domain; this is in contrast with the chicken and duck β globin domains discussed below, in which CR1 elements were found on both sides of the domain.

It is interesting that the CR1 probe did not hybridize to any fragment of the α globin domain of the duck (data not shown) cloned in a cosmid, spanning 15 Kbp of DNA upstream and 12 Kbp downstream of the three α type globin genes (29). In contrast, when the non-repetitive (AT-rich) segment of the pCAG21 insertion was labelled and used in hybridization experiments, strong signals were obtained for fragments localized

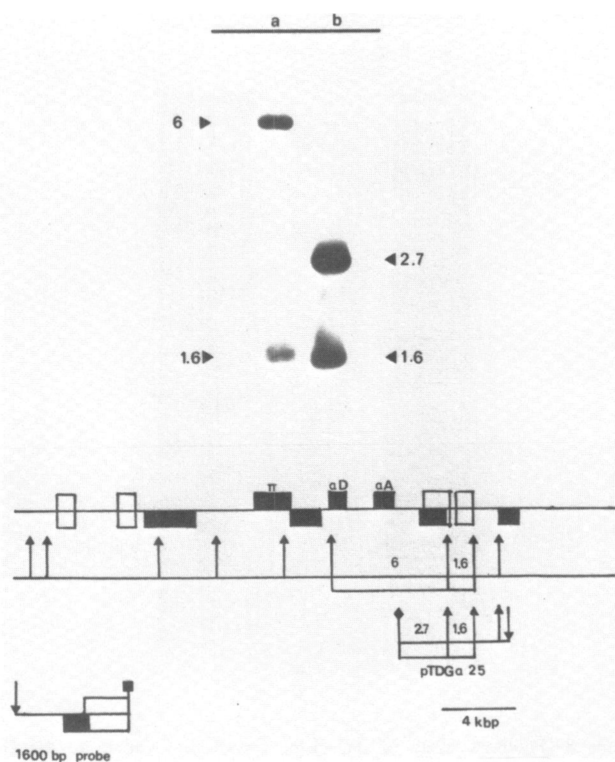

Figure 5—Hybridization between the AT-rich regions and flanking sequences mapping downstream of the chicken and duck α globin domains—DNA from the duck cosmid (29) cDGA (slot a) and plasmid (29) pTDG (slot b) were digested with Eco RI and with a combination of Eco RI (↑) Bam HI (↓) and Xho (↑) enzymes respectively. Resulting restriction fragments were separated on a 0.9% agarose gel, transfered to a 'Gene Screen' membrane and hybridized to a probe (see bottom left hand side of the Figure) consisting of a chicken 1600 bp long Bam HI- Hinc II restriction fragment containing both, the repetitive sequence and the AT-rich region we have sequenced. After several washes in high stringency conditions (see 'Methods') the membrane was dried and autoradiographed. The upper map shows the positions of the duck α type globin genes (upper black boxes, the AT-rich segments (white boxes), and the repetitive sequences (lower black boxes). Bands in which a high intensity of hybridization was observed are outlined in the bottom maps with hatched boxes.

in equivalent positions of the duck α globin domain (Fig. 5). The hybridization signals observed in this experiment were resistant to washing in high stringency conditions, suggesting that the DNA sequence of the AT-rich segment and flanking regions might have been conserved in evolution and thus might be of a certain functional importance.

The distribution of repetitive sequences in the chicken β globin gene domain was previously shown to be very complex, in contrast to the duck (27, 29). The hybridization pattern obtained using the CR1 GlA1 sequence as a probe shows that at least 5 CR1 elements are present in this genomic domain; three of them in the 5′ moiety of the domain (Fig. 4B). In the duck 2 CR1 elements frame the domain (Fig. 4C).

Interestingly, one of these CR1 elements is placed at the border of the domain of sensitivity to DNAse I in the chicken (9 and close to the AT-rich region (13), i.e. in a configuration analogous to that found at the 3′-end of the α globin domain. Thus, in the chicken α and β globin gene regions of the chicken and duck, clearly identified boundaries of the domains of sensitivity to DNAse I seem characterised by the presence of an AT-rich region closely linked to a CR1 sequence.

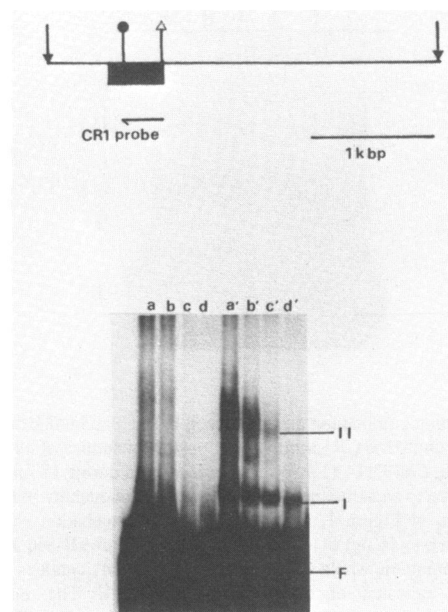

Figure 6—Nuclear factors binding specifically to the CR1 sequence—Nuclear proteins from adult chicken bone marrow cells were extracted as described in 'Methods'. Mixtures containing 2–3 μg of proteins, 1–2 ng of end-labeled DNA fragments and increasing quantities of sonicated *E. Coli* DNA (0.25, 0.5, 1 and 2 μg for a-a′, b-b′, c-c′ and d-d′, respectively) were incubated and electrophoretically fractionated on a polyacrylamide gel in low ionic strength buffer. Shifts in the electrophoretic mobility of the labelled fragment were detected by autoradiography of the dried gels. The specific CR1 probe (see map of Fig. 1) was in slots (a′) to (d′), and the pBR322 control in slots (a) to (d). The migration position of the free fragment (F) and of the shifted band I and II are indicated to the right of the gel.

Some nuclear proteins associate preferentially to the CR1 sequence flanking the 3′-end of the α globin gene domain

The electrophoretic behaviour of a DNA fragment derived from the CR1 sequence of the α globin gene domain was analysed after exposure to the proteins present in a nuclear extract of chicken erythroid cells.

The 196 bp Dra I-Aca II restriction fragment containing the 5′-end (representing the 3′-segment relative to globin gene transcription of the CR1 sequence (Fig. 1)) was end-labelled and used as a probe. Nuclear proteins extracted from bone marrow cells were added to this labelled DNA. The mixture was fractionated by electrophoresis on a polyacrylamide gel at low ionic strength. The 207 bp long Sau 3AI restriction fragment of plasmid pBR 322 was used as a control.

Autoradiography of the dried gel indicated that the nuclear extract of the bone marrow cells contains factors inducing two mobility shifts of the fragment derived from the CR1 sequence, whereas that derived from the pBR 322 DNA, used as a control, did not change its position (Fig. 6; see bands I and II in slots a′ to d′). Other bands, visible in the same gel, but of fairly low intensity, were also produced by the non-specific DNA probe, indicating that they might involve factors associating to DNA in a non-specific manner. After prolonged autoradiographic exposure of the dried gels, no supplementary bands were observed; in this case a high background signal developed throughout the migration distance of both fragments (not shown).

The specificity of the DNA-protein interaction was tested by the ability of a cold restriction fragment containing the CR1 sequence to compete successfully with the labelled probe for the

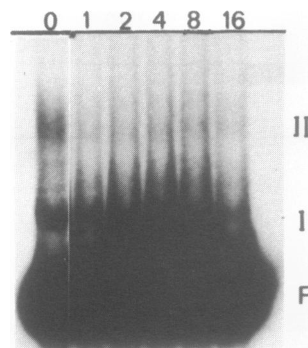

Figure 7—Binding competition assay between labelled and unlabelled restriction fragments containing the CR1 sequence—Constant quantities of nuclear proteins (1–2 μg) and *E. Coli* DNA (1 μg) were pre-incubated during 15 minutes at room temperature; then to each mixture was added a constant quantity of the same CR1 probe as shown in Figure 1. Competition was by incubation with increasing quantities (from 0 to 16 ng) of an unlabelled 1000 bp Bam HI-Sau 3AI restriction fragment containing the whole CR1 sequence (nanogram quantities of competitor fragment are indicated at the top of each slot). The mixtures were electrophoretically fractionated on a gel in low ionic strength buffer. The dried gel was autoradiographed without an intensifying screen. The position of migration of free DNA (F) and shifted DNA (I and II) is indicated to the right of the gel.

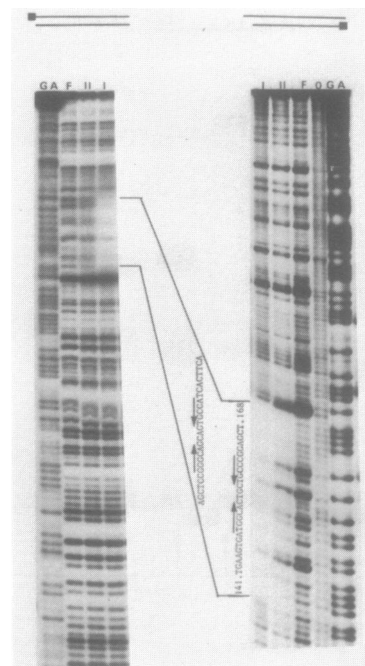

Figure 8—Footprint assay of the factor specifically binding to the CR1 sequence—The CR1 probe shown in Figure 1 was 5′-end labelled at the Ava II site (left), or at the Dra I site (right), and incubated with nuclear proteins, treated with DNAse I and fractionated in a gel at low ionic strength as described in 'Methods';. Bands of different electrophoretic mobility were visualized by autoradiography, and the DNA eluted from the corresponding gel slices. The DNA fractions thus obtained (F, I and II indicated at the top of the gels) were resolved in a denaturing 8% polyacrylamide gel in parallel with the fragments resulting from a G+A sequence reaction of the same labeled DNA segment (GA; in the right panel, slot O contained the probe incubated in presence of the nuclear proteins, but not digested by DNAse I). The nucleotide sequences of the binding site in each DNA strand are shown. The arrows on the sequence indicate the location of inverted repeats.

same factor. The same quantity of nuclear extract and a constant amount of *E. Coli* DNA (1 μg) were incubated during 15 minutes. Then a mixture containing a constant concentration of the labelled probe (1–2 ng), and increasing quantities of a cold restriction fragment containing the CR1 sequence (the 1000 bp Bam HI-Sau 3A1 restriction fragment) was added. The results presented in Fig. 7 show that the intensity of band I gradually decreased with the increase in concentration of the cold fragment present in the mixtures. This indicated that the shift of band I was produced by a nuclear factor present in the extract, interacting with a specific fragment of the CR1 sequence.

The two shifted bands (band I and II) reacted differently when the non-specific DNA was added as competitor. At a low concentration of sonicated *E. Coli* DNA (0.25 μg), both bands were present in almost the same amount. However, when the concentration of the competitor DNA was raised to 2 μg, band I still remained, but band II could no longer be detected (Fig. 6, slots a′ and d′). Furthermore, when the nuclear extract was preincubated for 15 minutes in presence of non-specific competitor DNA prior to addition of the radioactive probe and the homologeous competitor, band II was fully reduced by as little as 1 μg of homologeous DNA, but band I remained, gradually responding to the increasing amounts of the competitor (Fig. 7, slots 1–16). These observations suggest that the factors interacting with the DNA probe in band I and in band II are not identical, and furthermore, that the DNA-protein complexes revealed in band II do not represent a sequence-specific interaction of DNA with protein factors.

Identification of the binding site for the factor(s) producing bands I and II

In order to identify the nucleotide sequence of the site interacting with the factor(s) present in the nuclear extract, footprint analysis was performed taking advantage of the different electrophoretic mobilities of the DNA fragments bound (bands I and II), and unbound (band F) by factors present in the nuclear extract.

Preparative amounts of reaction mixture containing the nuclear extract and one of the end-labelled fragments were treated with DNAse I and fractionated for footprinting as described in 'Methods'. DNA fractions of different mobility were recovered from the gel slices and fractionated in a 8% agarose gel, in parallel with the same fragment hydrolysed by the G+A sequencing reaction.

Results presented in Fig. 8 show the DNAse I digestion pattern of the fragments contained in the DNA fractions eluted from the gel slices. In the band I fraction, a DNA segment of about 28 bp was resistant to DNAse I, when compared to the banding pattern of the equivalent DNA sequence in fraction F. Protection was observed for both DNA strands. The band II DNA showed a slight degree of protection in the same segment, represented by a diminished intensity of the bands. This may be due, however, to the contamination of the material with that present in band I. The possibility that band II originates as the result of a weak interaction of the same DNA region with a degraded form of the protein(s), revealed in band I, can also not be excluded.

It was not possible to determine the exact borders of the protected region in the CR1 sequence with the footprints we obtained. Nevertheless, one may conclude that it spans, at least, from positions 224 to 251 of the sequence (Fig. 3 and 8). This protected sequence contains in its center a symmetric region consisting of a 4 bp inverted repeat separated by a central, 4 bp long, non-symmetric sequence (see arrows in Fig. 8).

The central region of the protected area is clearly similar to

TABLE I. AT-RICH ELEMENTS WITHIN THE pCAG21 INSERTION

Nb	From	To	Size	AT %	Sequence 5'.............3'
1)	774	798	25	80.0	GTTTTAAGTAATGTATTCTCTATTA
2)	927	943	17	82.3	TTACTTCAATTAAATGT
3)	883	898	16	81.2	ATCTGAAATATTCATT
4)	66	80	15	80.0	TACAATCATAGAAAT
5)	305	317	13	84.6	TTTTGGTTTAAAA
6)	468	479	12	83.3	AATTGTAACATA
7)	613	624	12	83.3	TTCTGAAAAAAT
8)	709	720	12	83.3	AGTTTTTTCTTT
9)	659	669	11	81.8	AGTTTTTCTTT
10)	350	360	11	81.8	TGTAAAAAGAA
11)	445	455	11	81.8	ACAGAATTTAT
12)	406	415	10	80.0	CATTCAAAAT
13)	426	435	10	80.0	ATTTAGCAAA
14)	286	295	10	80.0	TTACCTAAAT
15)	805	814	10	80.0	ATAAGGTATT
16)	826	835	10	80.0	AGTGAATAAT
17)	696	705	10	80.0	TCTGATATAT
18)	916	925	10	80.0	TTATTCATCT
19)	577	586	10	80.0	AATAGCAAAA
20)	959	968	10	80.0	AATTGAAGTTI

the binding site for other nuclear factors detected in chicken cells, in particular to the binding site for the so-called TGGCA-protein (21). The major difference from the classical TGGCA-protein binding site is the length of the non-symmetrical central region which is one nucleotide longer in our case.

Another binding site sharing sequence homology with the protected DNA sequence is present in a hypersensitive DNA site, located upstream of the chicken β^A globin gene. This region was found to be capable of interacting *in vivo* (30) and *in vitro* (20, 31) with a protein factor only present in the nuclear extract derived from chicken erythroid cells (20), although it appears also to be related to the TGGCA binding protein site (31).

DISCUSSION

Boundaries of the chicken and duck α and β globin gene domains

One of the most interesting aspects raised by our results concerns the characteristics of the DNA sequences flanking the chicken globin gene domains sensitive to DNAse I. CR1 sequences were first mapped at boundaries of genomic domains (15), suggesting that this kind of repetitive DNA could contain signals involved in the definition of chromosomic domains in the chicken genome.

The results reported here tend to indicate that the distribution of the CR1 sequences in the chicken α and β globin gene domains is not random in respect to the boundaries of the domains of activity, as revealed by DNAse I sensitivity. Furthermore, in both genomic domains, clearly established boundaries of the transcriptionally active domain (17, 27, 32) are characterized by the presence of a CR1 sequence closely linked to an AT-rich segment suggesting functional cooperativity. Such structural links suggest that these two elements are involved in the establishment, and possibly in the control of variations of chromatin structure of these chromosomic domains in relation to activation. Several lines of evidence suggest that the increased DNAse sensitivity characteristic of active chromosomal domains is to some extent due to the torsional stress in the DNA regulated by topoisomerase II (27). Matrix attachment regions containing topoisomerase II sites (see below) are ideally placed to regulate the activity of specific chromosomal domains by altering their topological state. The juxtaposition of a CR1 element and an AT-rich segment including topoisomerase II sites in the matrix attachment segment under investigation is particularly significant in this respect.

The AT-rich segment placed at the 3'-end of the chicken α globin gene domain

In the nucleotide sequence reported here, the AT-rich segment was found immediately adjacent to the end of the CR1 element (position 351 onwards in Fig. 3B). This DNA segment corresponds to that identified as an ATRL ('AT-rich linker') by partial DNA denaturation (13; and Moreau, unpubl. res.); the sequence given is the first published for a genuine ATRL (12, 13). Interestingly, the overall content in Adenine and Tymidine is about 60% in the sequence spanning from nucleotide 351 to 967 in Fig. 3B). However, the A and T nucleotides are not uniformly distributed throughout this region. Blocks of very high AT-content (> 80%) are found scattered over a distance of about 600 bp (Table I). The proximity of such blocks of very high AT-content placed in an AT-rich environment is certainly responsible for the rapid denaturation of this region in partially denaturing conditions.

The border of the CR1 sequence and the AT-rich segment is marked by the presence of an A-rich polypurine tract (positions 352–364 in Fig. 3B). This tract overlaps a region showing 100% homology with a segment of the HTLV III LTR (AAGAAGACAAGAGA), found in positions 355–369. Most interestingly, the same HTLV III LTR sequence is found also in the 5'-side matrix attachment site of the globin domain under investigation (19). At the 3'-broder, an extensively homologeous sequence is found 80 bp downstream (AAAGAAGA; positions 432–443), in a direct repeat arrangement. Further downstream, at positions 659–671, a T-rich polypyrimidine tract is found which exhibits a somewhat imperfect dyad symmetry with the above mentioned A-rich polypurine tracts. These three regions share between 70 and 80% homology with the A and T rich stretches abundantly found at the scaffold attachment sites of certain *Drosophila* genomic domains ('A and T boxes'; 33).

We have also found in this AT-rich region 4 sites which exhibit a high although variable match with the *Drosophila* topoisomerase II consensus sequence (GTNA/ TAYATTNATNNG/A; 34),

ranging from a 13/15 to a 12/15 bp match (outlined in Fig. 3B). Interestingly, all four sites are clustered within a region of 180 bp only. No other regions exhibiting such sequence homology were found either upstream of these positions or in the CR1 sequence and its 5′-flanking region.

Identification of these topoisomerase II recognition sites might be interesting in view of the existence of the permanent site of DNA attachment to the nuclear matrix within the analysed DNA fragment (18, 19). It was demonstrated that DNA sequences capable of specific *in vitro* interactions with nuclear matrix proteins share extensive homology with topoisomerase II recognition (cleavage) sites on DNA (35), and that topoisomerase II directly participates in these interactions (36).

Comparison of the AT-rich non-repetitive part of the sequence shown in Fig. 3B with sequences present in computerized Gene Banks made it possible to identify two distinct regions possessing multiple homologies with different non-coding consensus sequences (not shown). The functional significance of many of these signals is as yet to be understood. Particularly interesting is, however, a homology with two DNA binding sequences described as a positive autoregulation signal of homeotic genes, the *fuji tarazu* gene (*ftz*) of *Drosophila* (see 37). The *ftz*Δgene expression involves two different cis-acting DNA elements; at the molecular level, the *ftz* product recognizes the sequence TCAATTAAAT located within its enhancer and autoregulates its own expression through this motif. At the 3′-border of the chicken α globin gene domain, the sequence TCAATAAATA was identified in the AT-rich segment at positions 931 to 940, overlapping the fourth *Drosophila* Topoisomerase II consensus. This sequence was also found in other homeotic genes in *Drosophila* , like the *engrailed (en*) gene, *even-skipped* (*eve*), *zerknüllt* (*zen*) and *paired* (*prd*) gene (37). Another DNA binding sequence was found at position 583 to 592 of our sequence, matching 100% the consensus ANNNNCATTA (37), which was suggested to represent a putative control region of *ultrabithorax*, *Antennapedia*, and also a sequence binding the product of the *fuji tarazu* gene, the *ftz* protein (37). Finally a sequence about 80% homologeous to the replication origin of Polyoma virus JC was identified in positions (351−371 and 892−906) of the sequence shown in Fig. 3B.

The CR1-binding protein

The most interesting finding concerning the CR1 element located at the 3′-boundary of the chicken α globin gene domain consists probably in the identification of a specific protein binding site within this element. One major question raised by this result is whether or not other CR1 elements, and in particular those flanking the β globin domains in chicken and duck, as well as the ovalbumin gene domains of sensitivity to DNAse I, can interact with the nuclear factor binding to the CR1 sequence of the α globin gene domain. Two lines of indirect evidence suggest that this possibility can not be taken for granted. First, even if the binding site characterized here was located in one of the two conserved regions of the CR1 sequences, detailed analysis of its equivalent position in other previously published CR1 element shows that the element concerned in this work presents a deletion of one nucleotide localised in the central non-palindromic spacer of the binding site, which in other CR1 elements, is normaly 5 bp long. Second, results presented by Sanzo et al. (38) demonstrated that in chicken oviduct cells, a nuclear protein preferentially binds to a restriction fragment derived from the 3′ moiety of several CR1 sequences and not to other segments. These considerations suggest that the DNA-protein interaction described in the present work does not necessarily concern other characterized CR1 elements.

A further question concerns both the nature and the significance of the DNA-protein interaction observed. Our results suggest that the factor binding to the CR1 element is related to other nuclear proteins detected in chicken cells. The similarity between the binding sites (shown in Fig. 3B) suggests the possibility that the factor interacting with the CR1 element may be related to either the TGGCA binding protein or the factor binding to the hypersensitive site of the β globin gene. However, when considering the major differences between the binding sites, it seems unlikely, that these factors are identical, at least as far as the TGGCA protein (or the equivalent nuclear factor I (NF-1 in mammals) is concerned. Indeed, the characterization of a considerable number of binding sites (21, 39−43) makes it possible to attribute a certain importance to the length of the central, non-symmetrical spacer which in some cases differs from the CR1 site observed here.

On the other hand, recent results (31) suggest that the factor binding the sequence upstream of the β^A and β^H globin genes is closely related, but not identical (see also 20) to the TGGCA binding protein. These observations indicate that chicken cells contain DNA binding proteins which share some common structural identities, as revealed by homologies in the DNA sequences they bind. Furthermore, when these observations are tentatively correlated with our results, it is possible to speculate that these proteins are part of a group of DNA-binding proteins involved in the regulation of the changing pattern of chromatin structure, according to the transcriptional activity of chromosome domains.

Finally, recent observations indicate that in a few genomic domains enhancer sequences are closely linked or overlap with matrix attachment regions (33, 35). This seems also to be the case for the globin domain investigated, in which enhancers were identified within 1 Kbp of the matrix attachment points on both sides of the domain in chicken and duck (44−46). Possibly, proteins interacting with DNA regions in the proximity of matrix attachment sequences could modify some of their structural properties. This could involve, for instance, the regulation of the exposure of preferential cleavage sites for topoisomerase II present in matrix attachment regions, modifying the topological state of the DNA in the domain and, hence, its transcriptional state. The binding of proteins to CR1 sequences present at the boundaries of chromosome domains could thus be one of the factors involved in the regulation of the functional properties of matrix attachment sites.

ACKNOWLEDGMENTS

The authors are grateful to François STRAUSS for valuable discussion on the technical approach and the results of this work. The help of Chantal CUISINIER and Richard SCHWARTZMANN and Antonia KROPFINGER in the preparation of the manuscript are duly acknowledged. Gabriel FARACHE was supported by a fellowship from the Fundacion Gran Mariscal de Ayacucho (Venezuela). Félix RECILLAS TARGA is recipient of a fellowship from the Consejo Nacional de Ciencia y Tecnologia (CONACYT), Mexico. S.V. Razin had a short term EMBO fellowship. This work was supported by the French CNRS, the Fondation pour la Recherche Médicale, the INSERM, the Ministère de la Recherche et de l'Enseignement

Supérieur, the Association pour la Recherche sur le Cancer (ARC) and the Ligue Nationale Française contre le Cancer.

REFERENCES

1. Hackstein, J.H.P., Leoncini, O., Beck, H., Peelen, G. and Hennig, W. (1982) Genetics **101**, 257–277
2. De Loes, F., Dijkhof, R;, Grend, C.J. and Hennig, W. (1984) EMBO J. **3**, 2845–2849
3. Gautz, J., Bencze, G., Gyurkovics, H., Ashburner, M., Ish-Horowicz, O. and Holden, J.J. (1982) Genetics **93**, 917–934
4. Bossy, B., Hall, L.M.C. and Spierer, P. (1984) EMBO J. **3**, 2537–2541
5. Cook, P.R. and Brazell, I.A. (1980) Nucl. Acids Res. **8**, 2895–2907
6. Adolph, K.W., Cheng, S.M., Paulson, J.R. and Laemmli, U.K. (1977) Proc. Natl. Acad. Sci. USA **74**, 4937–4941
7. Weisbrod, S. (1982) Nature **297**, 289–295
8. Lawson, G.M., Knoll, B.J., March, C.J., Woo, S.L.C., Tsai, M-J. and O'Malley, B.W. (1982) J. Biol. Chem. **257**, 1501–1507
9. Stalder, J., Larsen, A., Engel, J.D., Dolan, M., Groudine, M. and Weintraub, H. (1980) Cell **20**, 451–460
10. Alevy, M.C., Tsai, M.J. and O'Malley, B.W. (1984) Biochemistry **23**, 2309–2314
11. Jantzen, K., Fritton, H.P. and Igo-Kimenes, T. (1986) Nucl. Acids Res. **14**, 6085–610
12. Moreau, J., Matyash-Smirniaguina, L. and Scherrer, K. (1981) Proc. Natl. Acad. Sci. USA **78**, 1341–1345
13. Moreau, J., Matyash-Smirniaguina, L., Kezlarova-Lepesant, J., Lepesant, J-A. and Scherrer, K. (1982) Nature **295**, 260–262
14. Stumph, W.E., Kristo, A. and Groudine, M. (1981) Nucl. Acids Res. *9*, 5383–5397
15. Stumph, W.E., Baez, M., Beattie, W.G., Tsai, M.J. and O'Malley, B. (1983) Biochemistry **22**, 306–315
16. Weintraub, H., Larsen, A. and Groudine, M. (1981) Cell **24**, 333–344
17. Broders, F. and Scherrer, K. (1987) Mol. Gen. Genetics **209**,210–220

18a. Razin, S., Rzeszowska-Wolny, J., Moreau, J. and Scherrer, K. (1985a) Molekularnaja Biologia **19**, 456–466

18b. Razin, S., Rzeszowska-Wolny, J., Moreau, J. and Scherrer, K. (1985b) Mol. Biol. **19**, 376–384

19. Farache, G., Razin, S., Rzeszowska-Wolny, J., Moreau, J., Recillas Targa, F. and Scherrer, K. (1990) Mol. Cell. Biol.; in press
20. Emerson, B.M., Lewis, C.D. and Felsenfeld, G. (1985) Cell **41**, 21–30
21. Borgmeyer, V., Nowock, J. and Sippel, A.E. (1984) Nucl. Acids Res. 10, 4295–4311
22. Dodgson, J.B., Strommer, J. and Engel, J.D. (1979) Cell **17**, 879–887
23. Maniatis, T., Fritsch, E.F. an,d Sambrook, J. (1982) in 'Molecular Cloning. A Laboratory Manual.' Cold Spring Harbor Laboratory Ed.
24. Maxam, A.M. and Gilbert, W. (1980) Methods in Enzymology **65**, 499–560
25. Strauss, F. and Varshavsky, A. (1984) Cell **37**, 889–901
26. Broders, F., Razin, S., Farache, G. and Scherrer, K. (1986) Mol. Biol. Rep. **11**, 177–187
27. Villeponteau, B., Landes, G.M., Pankratz, M.J. and Martinson, H. (1982) J. Biol. Chem. **257**, 11015–11023
28. Engel, J.D. and Dodgson, J.B. (1980) Proc. Natl. Acad. Sci. USA **77**, 2596–2600
29. Kretsovali, A., Marcaud, L., Moreau, J. and Scherrer, K. (1986) Mol. Gen. Genetics **203**, 193–201
30. Jackson, P.D. and Felsenfeld, G. (1985) Proc. Natl. Acad. Sci. **82**, 2296–2300
31. Plumb, M.A., Lobanenkov, V.V., Nicolas, R.H., Wright, C.A., Zavou, S. and Goodwin, G.H. (1986) Nucl. Acids Res. **14**, 7675–7693
32. Broders, F., Zahraoui, A. and Scherrer, K. (1990) Proc. Natl. Acad. Sci. USA; in press
33. Gasser, S.M. and Laemmli, U.K., (1986) Cell **46**, 521–530
34. Sander, M. and Hsieh, T. (1985) Nucl. Acids Res. **13**, 1057–1071
35. Cockerill, P.N. and Garrard, W.T. (1986) Cell **44**, 273–282
36. Vassetzky, Y.S., Razin, S.V. and Georgiev, G.P. (1989) Biochem. and Biophys. Res. Comm. **159**, 1263–1268
37. Serfling, E. (1989) Trends in Gen. **55**, 964–971
38. Sanzo, M., Stevens, B., Tsai, M.J. and O'Malley, B., (1984) Biochemistry **23**, 6491–6498
39. Nowock, J., Borgmeyer, V., Puschel, A.W., Rupp, R.A.W. and Sippel, A.E. (1985) Nucl. Acids Res. **13**, 2045–2061
40. Gronojstajski, R.P., Adhya, S., Nagata, K., Guggenheimer, R.A. and Hurwitz, J. (1985) Mol. Cell Biol. **5**, 964–971
41. Henninghausen, L., Siebenlist, U., Danner, D., Leder, P., Rawlins, D., Rosenfeld, P. and Kelly, T. Jr (1985) Nature **314**, 289–292
42. Shaul, Y., Ben-Levy, R. and De-Medina, T. (1986) EMBO J. *5*, 1967–1971
43. Gronajstaski, R.M. (1986) Nucl. Acids Res. **14**, 9117–9148
44. Razin, S.V., Vassetzky, Y.S., Kvartskhava, A.I., Grinenko, N.F. and Georgiev, G.P. (1989) J. Mol. Biol., in press
45. Kretsovali, A., Huesca, M. and Marcaud, L. (1988) C.R. Acad. Sci. **307**, 563–568
46. Knezetik, J. and Felsenfeld, G. (1989) Mol. Cell. Biol. **9**, 893–901