Acta Biochimica Polonica

Vol. 46 No. 3/1999

777–784

QUARTERLY

Characterization of DNA sequences localized 11 to 17.5 kb upstream of the chicken embryonic π-globin gene°

Monika Pietrowska[1], Marek Rusin[2], Piotr Widłak[1], Sergey V. Razin[3] and Joanna Rzeszowska-Wolny[1]✉

[1]*Department of Experimental and Clinical Radiobiology and* [2]*Department of Tumor Biology, Center of Oncology, Gliwice, Poland*
[3]*Institute of Gene Biology, Russian Academy of Science, Moscow, Russian Federation*

Received: 07 April, 1999; accepted: 05 July, 1999

Key words: alfa globin domain, regulatory sequences, nuclear matrix, matrix attachment region, inverted repeats

We have analyzed the DNA fragment localized about 11 to 17.5 kb upstream of the chicken α-globin gene domain (the fragment was designed as α-0). The nucleotide sequence of its 3.3 kb-long 5′ part was established and interactions with nuclear matrix proteins were studied. The DNA region localized about 16 kb upstream of the embryonic π-globin gene showed high affinity to nuclear matrices *in vitro*. Two palindromes and a cluster of inverted repeats were co-localized in the same region. The whole 6.6 kb α-0 fragment decreased the activity of linked CAT reporter gene when transfected into chicken erythroblastoid cells.

The chicken α-globin gene domain is one of the best known eukaryotic genomic domains. It consists of three genes: the gene coding for embryonic protein and two π genes expressed in adult birds: α_D and α_A, in the 5′ to 3′ direction. The domain contains several previously established structural and functional elements: DNase I hypersensitive sites at its both ends, nuclear matrix attachment regions, topoisomerase II recognition sites, nuclease S1 hypersensitive sites, enhancers, silencers, a replication origin and individual promoters with transcription-controlling-elements (Farache *et al.*,1990; Razin *et al.*, 1991; De Moura

°The work was supported by the State Committee for Scientific Research (KBN, Polish-Russian Scientific Cooperation Project and grant 8T11E03315)

✉Corresponding author: Joanna Rzeszowska-Wolny, Phone: (48 32) 278 9677; Fax: (48 32) 231 3512; E-mail: jwolny@onkol.instonko.gliwice.pl.

Abbreviations: bp, base pair; kb, thousand of base pairs; CAT, chloramphenicol acetyl transferase, LCR, locus control region; BSA, bovine serum albumin; MAR, matrix attachment region; SAR, scaffold associated region.

Gallo *et al.*, 1992; Recillas Targa & De Moura Gallo,1993; Razin *et al.*, 1994; Recillas Targa *et al.*, 1994).

The positions of particular globin genes in the domain correspond to the order of their expression during development. In the human β-globin gene domain the developmentally coordinated expression of five different β-globin genes is regulated by the locus control region (LCR) located over 11.5 kb upstream of the first of the β-globin genes (Grosveld *et al.*, 1987; Wijgerde *et al.*, 1995). Chicken α-globin genes are also developmentally regulated; however, until now none of the LCR type sequences have been detected and the mechanism of switching of different α-globin genes remains unclear.

In this paper we describe DNA sequences localized from about 11 to 17.5 kb upstream of the embryonic π-gene. The DNA fragment (designated as α-0) was isolated from chicken genomic library and subcloned in the pUC19 plasmid. This region was initially characterized by the presence of DNase I hypersensitive sites (Ioudinkova *et al.*, 1997). Recent work by Razin *et al.* (1999) and this paper show that the region possesses properties of a matrix attachment region.

MATERIALS AND METHODS

DNA sequencing. The 6.6 kb *Bam*HI–*Bam*HI fragment of $\lambda F\alpha G0$ clone from chicken DNA library was cloned into pUC19 as described earlier (Ioudinkova *et al.*, 1997). A 3.3 kb-long fragment of α-0 was sequenced using an automatic ABI 377 DNA sequencer (Perkin Elmer). The ABI PRISM™ Dye Terminator Cycle Sequencing Ready Reaction Kit with Ampli Taq Polymerase from Perkin Elmer was used. The first sequencing reaction was performed with standard primer, then next primers were designed on the basis of the determined DNA sequence (the primer walking approach). DNA sequence was analyzed using the PC/GENE program.

Preparation of nuclear matrices. Nuclear matrix proteins used in the studies of DNA–protein interactions were obtained from blood cells, liver, kidney and heart of chicken as described earlier (Farache *et al.*, 1990), except that stabilization of nuclei with copper was omitted. Nuclei were purified from the frozen homogenized tissues by centrifugation in 0.25 M sucrose. Purified nuclei were incubated with DNase I and extracted with buffers containing increasing amounts of salt. The pellet obtained after the extraction of chromatin with 2 M NaCl was used as nuclear matrix in further studies.

DNA molecular probes. Four subfragments (S1, S2, S3 and S4) partially covering the α-0 sequence were obtained by PCR amplification. DNA fragments were 5′ dephosphorylated with alkaline phosphatase and end-labeled with [γ-^{32}P]ATP using T4 polynucleotide kinase, then purified by polyacrylamide gel electrophoresis (Sambrook *et al.*, 1989).

Complex formation between DNA and nuclear matrices. The binding of DNA to nuclear matrix proteins *in vitro* was studied according to the method described by Cockerill & Garrard (1986). About 25 μg of matrix proteins were suspended in 0.2 ml of binding buffer comprising 50 mM NaCl, 2 mM EDTA, 0.25 M sucrose, 10 mM Tris/HCl (pH 7.4), 50 μg BSA, 25 ng of each ^{32}P-end-labeled DNA probe and sonicated *E. coli* DNA (0.4 μg or 2 μg). After 2-hour incubation at 25°C matrices were recovered by centrifugation. DNA from both matrix-bound (pellet) and unbound (supernatant) fractions was purified, then resolved by electrophoresis on agarose gels and autoradiographed.

Southwestern blot analysis. Nuclear matrix proteins (100 μg) were fractionated on a 10% polyacrylamide/SDS gel and electrotransferred onto PVDF membrane (Hybond-P; Amersham) in 25 mM Tris, 190 mM glycine and 20% methanol. Filter-bound proteins were renatured by incubation in a hybridization oven for 5 h at 20°C with 25 mM Tris/HCl (pH 7.6), 50 mM NaCl, 1 mM EDTA,

1 mM dithiothreitol and 2.5% BSA. After washing with binding buffer (same composition as above except for 0.25% BSA added) filters were incubated for 5 h at 20°C in the binding buffer supplemented with 500 ng of ^{32}P-end-labeled DNA probe and 100 μg of *E. coli* DNA (final volume 10 ml). Filters were then washed and autoradiographed.

CAT assay. The Promega assay system was used according to the manufacturer's protocol. The minimal promoter of the α^D gene was cloned into pCAT3 basic vector (Promega) to obtain the α^D pCAT3 vector. The 6.6 kb α-0 fragment (in either direct or reverse orientation) was cloned downstream from the CAT gene to obtain α^D pCAT3α-0 constructs. Equimolar amounts of each construct were transfected into cultured chicken avian erythroblastosis virus (AEV)-transformed erythroblasts (line HD3) with Lipofectin™ (Gibco-BRL). Cells were lysed and the activity of the CAT gene was assayed 60 h after transfection. Thin-layer chromatography was used for the separation of butyrylated and non-butyrylated (substrate) forms of chloramphenicol labeled with ^{14}C. The spots were cut out from the plate and their radioactivity was measured in a liquid scintillation counter.

RESULTS

The α-0 DNA fragment analyzed in this paper is localized from about 11 to 17.5 kb upstream of the 5′ end of the well characterized domain of developmentally regulated chicken α-globin genes (Fig.1). This localization makes it an interesting object for studies as the sequences which regulate specific developmental expression of α-globins are still not known and they could be potentially placed in this region. We started establishing the nucleotide sequence of this 6.6 kb fragment from both ends using the "primer walking" method.

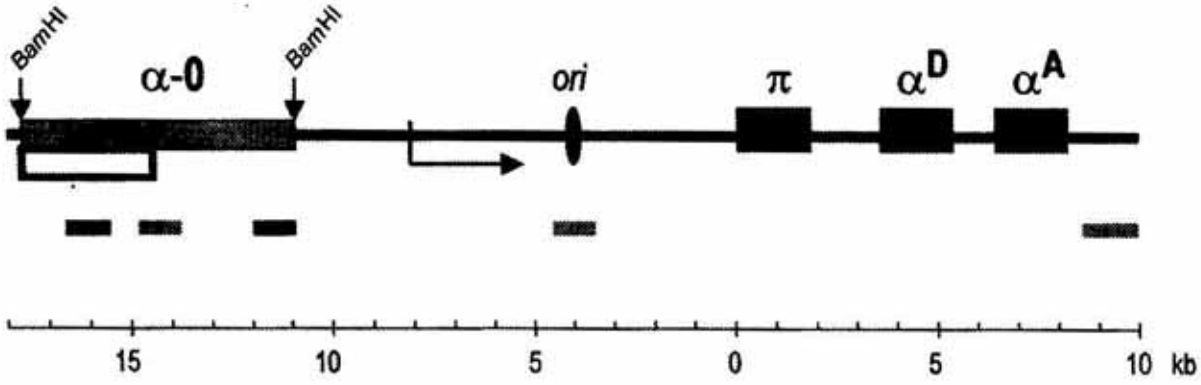

Figure 1. Schematic diagram of the chicken α-globin gene domain.

Black boxes represent α-globin genes, grey box represents 6.6 kb α-0 DNA (*Bam*HI–*Bam*HI fragment), empty box represents DNA fragment sequenced in this work, black oval represents replication origin (*ori*), arrow represents transcription initiation site. Bars below diagram represent MAR sequences (those described in this work are black ones). Scale (in kb) starts at the 5′ end of the π-globin gene.

In this paper we present the sequence obtained for the 5′ part of the fragment (about 14.5 to 17.5 kb upstream of the π gene) as its 3′ part (about 11 to 15 kb) was studied in parallel in the Jacques Monod Institute and its sequence has already been published (Razin *et al.*, 1999). The sequence of the 3.3 kb 5′ part of the α-0 DNA is available from the EMBL Nucleotide Sequence Database (accession no. Y18681), and partially overlaps the sequence of the 3′ part of the α-0 DNA (Gene Bank, accession No. AF067138).

The nucleotide sequence of the α-0 DNA was analyzed using PC-Gene. We looked for the

presence of nucleotide motifs known to be characteristic for matrix attachment regions (MARs). Such regions are usually AT-rich and contain the following motifs: AAATATTT, ATTA, ATTTA, ATTTTA, and other motifs in which repeated regions composed of A and T bases alternate with other sequences of 8 to 12 nucleotides (reviewed in Boulikas, 1993). Figure 2A shows the summarized frequency of such subsequences along the whole α-0 DNA. Both the 5′ end and the middle part of the fragment (about 14.5 kb upstream of the π gene) are enriched in AT-containing motives characteristic for MARs. The presence of inverted repeats was also analyzed along the α-0 DNA (Fig. 2B). Such motifs are potentially important for loop formation in transcripts and

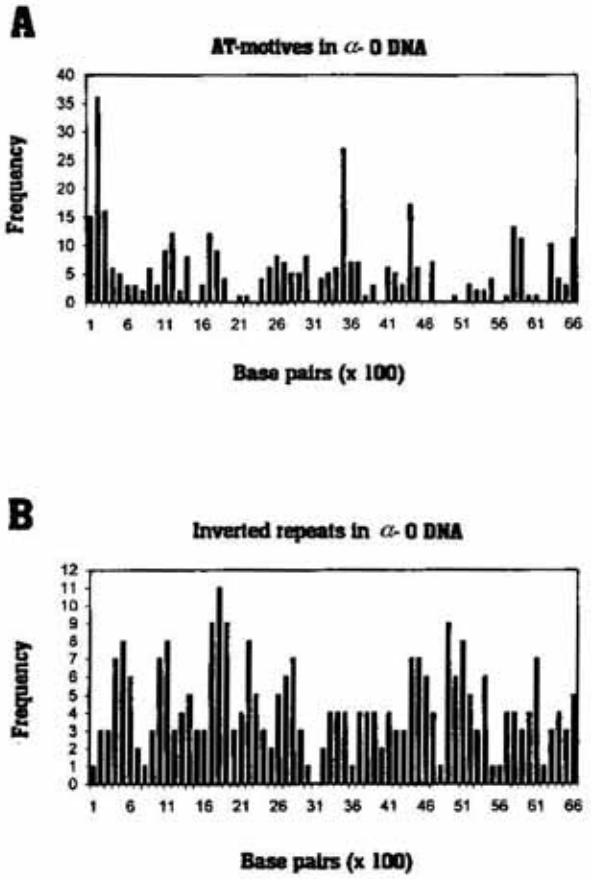

Figure 2. Distribution of AT-rich motives specific for MAR sequences (panel A) and inverted repeated sequences (panel B) along the 6.6 kb α-0 DNA fragment. Numbering starts from the 5′ end of the fragment.

are frequently localized in regulatory sequences (and also in some MAR sequences) (reviewed in Boulikas, 1993). The α-0 region contains such sequences in all its parts, however, a particularly rich cluster of inverted repeats is localized 1.7–1.9 kb from the 5′ end of α-0 DNA (about 15.5 kb upstream of the π gene). Two 12 bp long palindromes (5′-ATGTTATAACAT-3′ and 5′-TCAGGGCCCTGA-3′) are also localized in the same region. The α-0 DNA contains several sequences which are potentially recognized by different transcription factors. Among these protein factors are ADR-1, ANTP, AP-1, c/EBP, F3, F5, GATA-1, HSTF, LBP-1, NF-E1, OTF-2B, TEF-2, TRF; however, most of them are localized in the 3′ part of the fragment.

In order to study the interactions with nuclear matrix proteins the α-0 DNA was divided into four smaller subfragmetns. Three subfragments covered the 5′ part of the α-0 DNA from about 15 to 17.5 kb upstream of the π gene: 1 to 1375 (**S1**, positions in bp from the 5′ end), 1142 to 1870 (**S2**), 1617 to 2716 (**S3**), and the fourth covered 1 kb from the 3′ end of the α-0 DNA (**S4**, see Fig. 3A). Nuclear matrix proteins obtained from different chicken tissues (liver, kidney, heart and blood cells) were complexed *in vitro* with radioactive DNA probes, then the matrix-bound material was separated from the unbound one and analyzed electrophoretically (Fig. 3B). We found that only nuclear matrices prepared from liver efficiently bound the analyzed DNA probes while matrices prepared from other tissues studied did not form such complexes even at low concentration of competing *E. coli* DNA. In the presence of low amounts of non-radioactive competitor (2 μg/ml of *E. coli* DNA) all four probes were found in both matrix-bound and unbound fractions when liver preparations were tested. However, a 5-fold increase of the competitor DNA amount enabled us to show that only the S2 and S4 probes resisted the competition and were bound to matrices. The high affinity of the S2 and S4 subfragments to nuclear matrix proteins was confirmed by

Southwestern blot analysis, in which DNA probes bind to electrophoretically resolved proteins. Figure 3C shows the results of an experiment in which all four subfragments of the α-0 DNA were incubated with nuclear matrix proteins from liver and blood cells. Only

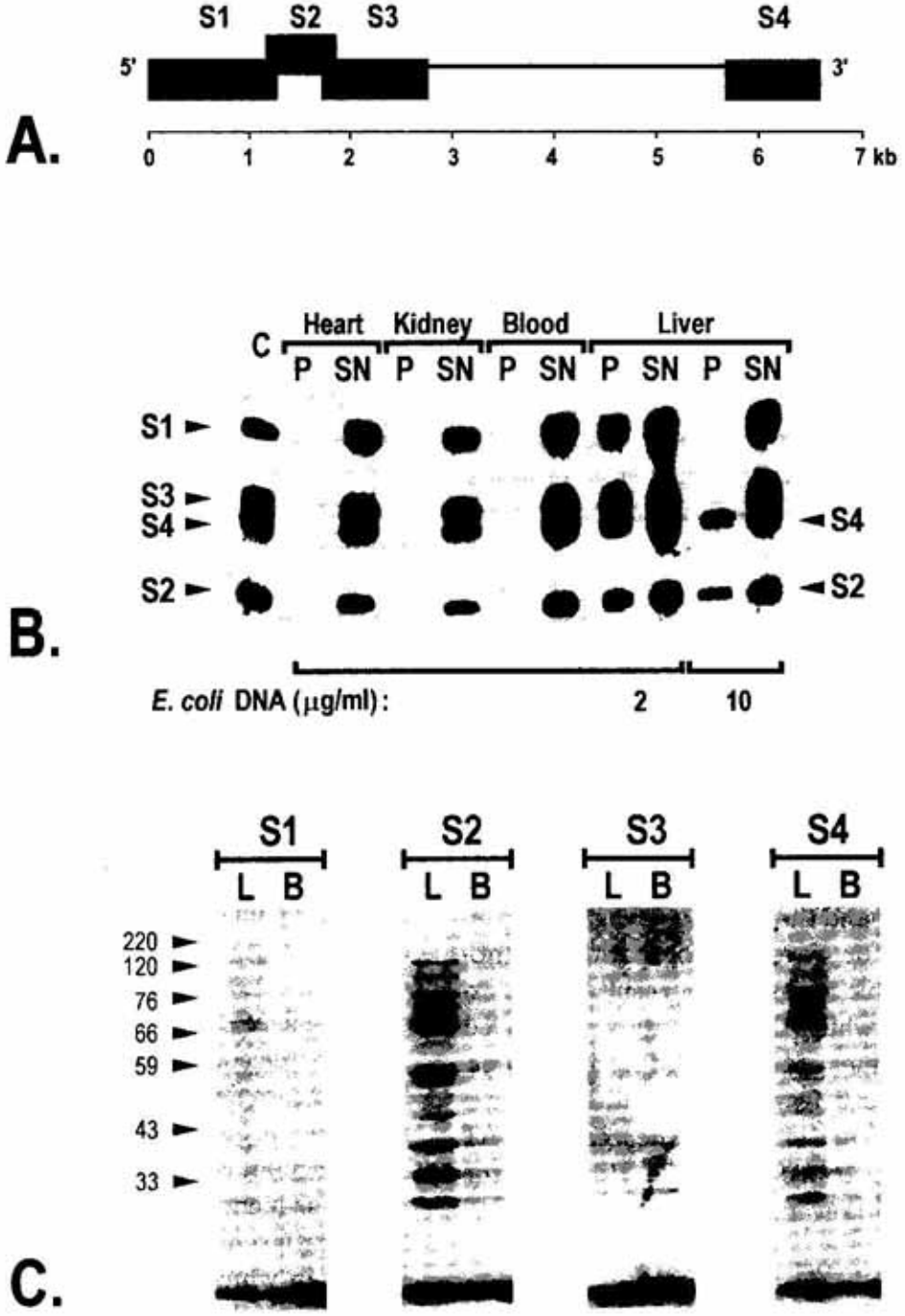

Figure 3. Binding of α-0 subfragments to nuclear matrix proteins.

Panel A. Localization of subfragments S1, S2, S3 and S4 along α-0 DNA. Panel B. The analysis of *in vitro* complex formation between α-0 subfragments and nuclear matrices from chicken tissues. Complexes were formed in the presence of 2 and 10 μg/ml of *E. coli* DNA. **P** (pellet) represents matrix-bound fraction, **SN** (supernatant) represents unbound fraction, **C** (control) represents mixture of non-fractionated probes. Panel C. Southwestern blot analysis of proteins binding the α-0 DNA subfragments. Nuclear matrix proteins were isolated from liver (**L**) or blood cells (**B**). The positions of molecular size markers are denoted along the figure (in kDa).

Table 1. The influence of the α-0 DNA fragment on the expression of the CAT gene under the control of the minimal promoter of the α^D-globin gene.

The levels of CAT activity were normalized to that of the ^{D}pCAT3 vector. Values are mean for 3 independent transfection experiments.

Vector	CAT activity
α^DpCAT3	100
α^DpCAT3α-0 direct orientation	44
α^DpCAT3α-0 reverse orientation	79
pCAT3	2

S2 and S4 subfragments form strong complexes with the liver proteins, while proteins isolated from blood cells form very weak complexes with all probes.

To check whether DNA sequences localized in the α-0 fragment could influence transcription of the globin genes, the effects of this DNA fragment on the CAT gene expression was studied. Chicken erythroblastoid cells HD3 were transfected with vectors containing CAT gen alone (pCAT3), CAT gene under control of the minimal promoter of the α^D gene (α_D pCAT3), or CAT gene under control of the α^D promoter and the 6.6 kb α-0 fragment in either direct or reverse orientation (α pCAT3α-0). Table 1 shows the results of an experiment in which CAT activity was assayed in different lines of transfected cells. The presence of the α-0 fragment (in either orientation) resulted in a significant decrease of CAT expression in transfected cells. However, the effect of α-0 in reverse orientation was lower.

DISCUSSION

In this work we analyzed a DNA fragment spanning the close vicinity of the well-characterized domain of chicken α-globin genes. The chicken α-globin gene domain contains three functional genes π, α^D and α^A. The similar human domain contains genes ζ, α2 and α1. In humans, the ζ-globin gene is expressed exclusively in embryonic primitive erythroblasts. The developmental silencing of the ζ-globin gene is directed by both transcriptional and posttranscriptional mechanisms (Liebhaber & Russel, 1998). The chicken α-globin gene domain has similar structure and also is developmentally regulated, yet the mechanisms of this regulation are not known. The α-0 DNA fragment which was characterized here is a new sequence spreading in the 5′ direction from the known 5′ end of the α-globin gene domain. At this moment it is difficult to conclude whether this sequence still belongs to the α-globin gene domain or rather represents a boundary region separating two different domains. An interesting observation shows that some non-coding sequences from the 5′ end of the domain are present in transcripts of opposite direction (Sjakste *et al.*, 1998). This suggests that a separate transcriptional unit is localized upstream of the α-globin genes. If this is true, sequences analyzed in this paper may contain some regulatory sequences for both units. On the other hand, regulatory sequences controlling the transcription of human globin genes were found far upstream of the genes (e.g. the locus controlling region of human globin genes cluster) and the existence of such regions in the α-0 fragment or even further in the 5′ direction is still possible.

In the nucleotide sequence of the α-0 DNA fragment we detected many sequences recognized by protein factors which potentially have regulatory functions. Among them we found sequence motifs observed in MARs from different genes. In previous experiments, in which nuclear matrices were purified from chicken erythroblastoid cells, a DNA fragment localized about 14.5 kb upstream of the π gene was identified to posses properties of a MAR (Razin *et al.*, 1999). This fragment contains a cluster of AT-rich motives, which are characteristic for many known MAR/SAR sequences (Boulikas, 1993). Here we show that two additional fragments, localized about

16 kb and 11 kb upstream of the π gene, also possess properties of a MAR (when complexed with matrices from liver cells). Neither of them is particularly enriched in AT-rich motives. However, a fragment localized 16 kb upstream of the π gene (in the S2 subsequence) contains a cluster of inverted repeats and palindromes. These sequences are characteristic for a particular class of a MARs divergent from the AT-rich one (Boulikas & Kong, 1993). We tested the affinity of DNA to the matrix protein under relatively mild conditions (5–10 higher concentration of competitor DNA is frequently used), however, the matrices were purified without the copper-stabilization step (which is used to increase their DNA-binding capacity). The presence of MARs may suggest that the α-0 DNA possibly contains a boundary region separating two different domains.

Tracts of inverted repeated sequences were found inside the α-0 DNA fragment. It is difficult to speculate what function could such inverted repeats play in the regulation of α-globin genes' expression. One could imagine that such sequences, as well as palindromes also found in this region, might induce the appearance of RNA loops and participate in its splicing. It was shown that the chicken α-globin gene domain is transcribed as a full domain transcript (De Moura Gallo *et al.*, 1992; Broders & Scherrer, 1987; Broders *et al.*, 1990) which suggests posttranscriptional mechanisms for the developmentally regulated expression of these genes. On the other hand, we do not know yet whether any part of the α-0 fragment is transcribed. The CAT-transfection experiments clearly show that sequences localized within the α-0 DNA fragment interfere with transcriptional activity, however, any speculation on the role of its inverted repeats in this effect is premature at the moment.

Authors are greatly indebted to J. Rogoliński for help in database analyses and A. Sochanik for help in the preparation of this manuscript.

REFERENCES

Boulikas, T. (1993) Nature of DNA sequences at the attachment regions of genes to the nuclear matrix. *J. Cell. Biochem.* **52**, 14–22.

Boulikas, T. & Kong, C.F. (1993) Multitude of inverted repeats characterizes a class of anchorage sites of chromatin loops to the nuclear matrix. *J. Cell. Biochem.* **53**, 1–12.

Broders, F. & Scherrer, K. (1987) Transcription of the α-globin gene domain in normal and AEV-transformed chicken erythroblasts: Mapping of giant globin-specific RNA including embryonic and adult gene. *Mol. Gen. Genet.* **209**, 210–220.

Broders, F., Zahraoni, A. & Scherrer, K. (1990) The chicken α-globin gene domain is transcribed into a 17-kilobase polycistronic RNA. *Proc. Natl. Acad. Sci. U.S.A.* **87**, 503–507.

Cockerill, P.N. & Garrard, W.T. (1986) Chromosomal loop anchorage of the kappa immunoglobulin gene occurs next to the enhancer in a region containing topoisomerase II sites. *Cell* **44**, 273–282.

De Moura Gallo, C.V., Vassetzky, Y.S., Huesca, M. & Scherrer, K. (1992) A transcription-dependent Dnase I-hypersensitive site in a far upstream segment of the chicken α-globin gene domain coincides with a matrix attachment region. *Biochem. Biophys. Res. Commun.* **184**, 1181–1189.

Farache, G., Razin, S.V., Rzeszowska-Wolny, J., Moreau, J., Recillas Targa, F. & Scherrer, K. (1990) Mapping of the structural and transcription-related matrix attachment sites in the α-globin gene domain of avian erythroblasts and erythrocytes. *Mol. Cell. Biol.* **10**, 5349–5458.

Grosveld, F., Blom van Assendelft, M., Greaves, D.R. & Kollias, G. (1987) Position-independent high-level expression of the human β-globin gene in transgenic mice. *Cell* **51**, 975–985.

Ioudinkova, E., Lagarkova, M.A. & Razin, S.V. (1997) Molecular cloning and characteristics of the 5'-terminal end of alpha-globin genes of chickens including potential regulator ele-

ment of domain level. *Dokl. Akad. Nauk.* **356**, 407–408.

Liebehaber, S.A. & Russel, J.E. (1998) Expression and developmental control of the human alpha globin gene cluster. *Ann. N.Y. Acad. Sci.* **850**, 54–63.

Razin, S.V., Vassetzky, Y.S., Jr., Kvartskhava, A.J., Grinenko, N.F. & Georgiev, G.P. (1991) Transcriptional enhancer in the vicinity of a replication origin within the 5′ region of the chicken α-globin domain. *J. Mol. Biol.* **217**, 595–598.

Razin, S.V., De Moura Gallo, C.V. & Scherrer, K. (1994) Characterization of the chromatin structure in the upstream area of the chicken α-globin gene domain. *Mol. Gen. Genet.* **242**, 649–652.

Razin, S.V., Shen, K., Ioudinkova, E. & Scherrer, K. (1999) Functional analysis of DNA sequences located within a cluster of Dnase I hypersensitive sites colocalizing with a MAR element at the upstream border of the chicken α-globin gene domain. *J. Cell. Biochem.***74**, 38–49.

Recillas Targa, F. & De Moura Gallo, C.V. (1993) Analysis of the distribution of protein binding DNA motives in the vicinity of the 3′-side chicken α-globin enhancer. *Biochem. Biophys. Res. Commun.* **190**, 1163–1172.

Recillas Targa, F., Razin, S.V., De Moura Gallo, C.V. & Scherrer, K. (1994) Excision close to matrix attachment regions of the entrie chicken α-globin gene domain by nuclease S1 and characterization of framing structures. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 4422–4426.

Sambrook, J., Fritsch, E.F. & Maniatis, T. (1989) *Molecular Cloning, A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

Sjakste, N., Jarovaia, O., Linares-Cruz, G., De Moura Gallo, C.V. & Scherrer, K. (1998) Transcription of the chicken alpha-globin domain; in *The Wilhelm Bernhard Workshop on the Cell Nucleus*, Quebec, Canada, pp. 58–59.

Wijgerde, M., Grosveld, F. & Fraser, P. (1995) Transcription complex stability and chromatin dynamics *in vivo*. *Nature* **377**, 209–213.