Vol. 182, No. 1, 1992 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
January 15, 1992 Pages 78-85

Matrix Attachment Sites in the Murine α-globin Gene

Zoya Avramova* and Elena Paneva

Institute of Molecular Biology, Bulgarian Academy of Sciences,
Sofia 1113, Bulgaria

Received November 15, 1991

DNA sequences with a high affinity for nuclear matrix proteins have been identified and localized in the mouse α-globin gene. These matrix association regions (MARs) are adjacent, covering the first intron and part of the 5'-coding sequence. The binding sites are in close proximity to DNase I hypersensitive sites and other important signal sequences. The proteins of the nuclear lamina do not bind the α-globin gene MARs in the *in vitro* binding assay. The finding of MARs in the mouse α-globin gene creates an apparent paradox, since works from other authors and our results presented here indicate that this gene is not bound to the nuclear matrix *in vivo*. This contradiction is difficult to explain at present but different possibilities are accounted for in the text.

According to a current model eukaryotic DNA is folded into topologically constrained loops formed through the attachment of chromatin to the nuclear skeletal structures. It is believed that the chromatin structural units comprise also independently regulated domains playing an important role in the control of gene activity [reviewed in 1-3].

The proteinaceous residue, obtained after high salt- or LIS- extraction of nuclease digested nuclei is termed "matrix" or "scaffold", respectively [4-7]. Independently of some differences in the composition and the appearance of the residual structures obtained by the two methods [8], the important fact is that both types recognize and interact with the same class of DNA sequences. These sequences, called "matrix association regions" (MARs) or "scaffold association regions" (SARs) represent the sites where chromosomal DNA binds the nuclear skeleton [reviewed in 1, 7, 9]. These sites were found located at the borders of different genes, close to 5'- or 3'- end *cis*-regulatory sequences or DNase I hypersensitive sites [6, 10-20].

*To whom correspondence should be addressed at present address: Department of Biologica Sciences, Purdue University, Lilly Hall of Life Sciences, West Lafayette, IN 47907. Fax (317) 494-0876.

0006-291X/92 $1.50

Analysis of the 90 kbp region around the human β-globin gene complex revealed the presence of at least eight SARs, while none was identified in the very similar α-globin gene system [21]. No attachment to the nuclear matrix has been registered for the mammalian α-globin gene in either its active, or inactive state [21-23].

Based on these results we have chosen the murine α-globin gene as a negative control for the *in vitro* binding assays carried out in our laboratory. Unexpectedly, a MAR with a very high affinity for the nuclear matrix has been found. Here we describe the detection and precise location of the mouse α-globin gene MAR.

MATERIALS AND METHODS

Nuclei from normal mouse liver, or from Ehrlich ascites tumor (EAT) cells were isolated by conventional procedures. Nuclear matrices consisting of nuclear lamina plus an internal network of matrix proteins were prepared by high-salt extraction of DNase I digested nuclei as described [4, 10]. Empty nuclear shells consisting of nuclear lamina only were isolated as in [24]. The morphology and the integrity of the isolated nuclear skeletal structures were checked under the electron microscope (not shown). Equal amounts of matrices were present in all binding experiments. The *in vitro* binding assay was performed exactly as described in [10].

The plasmid pML2αG, containing a 2.1 kbp fragment from adult mouse α-globin gene was a gift by Dr. Rougeon, Inst. Pasteur, Paris. The fragments used as binding substrates, or hybridization probes are shown schematically in Fig. 1, and discussed in detail in the text. When used in the binding experiments, the fragments obtained after cutting of the plasmid with the appropriate restriction nuclease were 5'-end labeled with T4-polynucleotide kinase and reacted with isolated nuclear matrices or nuclear shells in the presence of increasing concentrations of unlabeled competitor DNA [10]. For the precise location of the MAR sequences in the gene domain, shorter fragments from the region (shown schematically in Fig.1 B) were isolated from low-melting agarose and recloned in pUC 8.

The possible *in vivo* attachment of α-globin gene sequences to the nuclear matrix was assessed after limit digestion of isolated nuclei with Sac I, followed by isolation of the matrices by either high-salt [4, 10], or LIS [6] - extractions. DNA associated with the skeletal structure was recovered from the residual insoluble pellet, while chromatin DNA not bound to the skeleton was recovered from the soluble fraction. After deproteinization equal amounts of DNA from the pellet, from the supernatant or total unfractionated DNA were checked for the presence of α-globin gene sequences by slot-hybridization assay, using the whole pML2αG plasmid as a probe. A cloned fragment containing the κ-immunoglobulin gene MAR (a kind gift from Dr. W.T. Garrard, University of Dallas, Texas) was used as a control for matrix associated DNA.

RESULTS AND DISCUSSION

Interaction of pML2αG plasmid with the nuclear matrix

As shown in Fig.1, the plasmid pML2αG contains a fragment of the mouse gene including the transcribed region plus about 1 kbp upstream and 0.7 kbp downstream

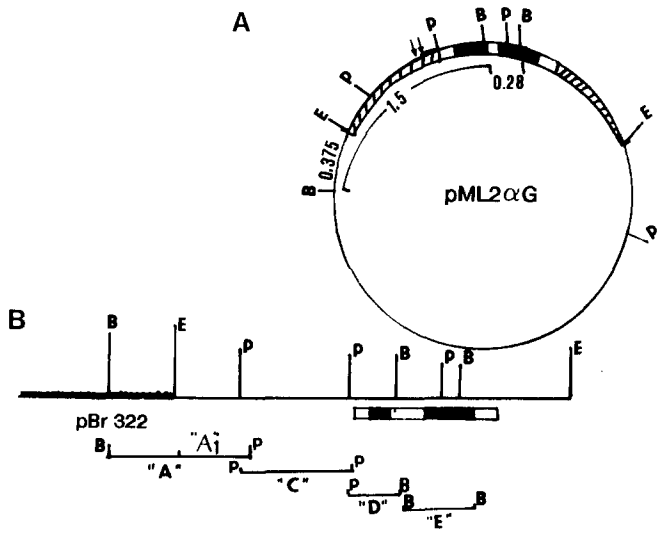

Figure 1. (A) - A schematic representation of the whole plasmid pML2αG. The region between the two Eco RI sites is the inserted sequence of the mouse α-globin gene. E, P and B mark the restriction sites for Eco RI, PstI and Bam HI, respectively. The arrows mark the position of the DNase I hypersensitive and the transcription initiation sites. The open boxes mark the introns, while the black boxes - the exons; (B) - the inserted mouse α-globin gene sequence (the map is according to 28).The fragments used as probes are shown below.

flanking regions. Digestion with Bam HI yielded three fragments covering an intragenic region (280 bp), the immediate 5'-flanking 1.5 kbp region (including 375 bp from the vector) and the 3'-flanking gene segment, as well as the rest of the vector (see Fig.1A and Fig. 2 a). 5'-end labeled fragments were reacted with isolated liver nuclear matrices in an *in vitro* model binding assay. As shown in Fig. 2 b-c, two fragments corresponding to the 1.5 kbp located in the 5'-end of the region and a 280 bp piece from the intragenic region have been specifically retained and precipitated by the matrix in the presence of increasing concentrations of competitor (up to 300 μg/ml) E.coli DNA. The two fragments bound to the matrix with a notable preference over the bacterial DNA and over the largest labeled fragment containing the vector and a 610 bp α-globin segment located 3'-from the third Bam H1 site. Note that the smallest Bam HI-Bam HI fragment bound to the matrix with a greater affinity than the 5'-adjacent sequence (Fig.2 b-c). The result indicates that there are matrix binding regions in the α-globin gene domain, located in the 1.4 kbp fragment spanning the 5'-region between the Eco RI and the third Bam HI sites.

Precise location of the attachment regions in the α-globin gene

In an attempt to define the matrix binding sites in more detail, individual fragments spanning the 5'-α globin genomic region, recloned in pUC 8 were used. The corresponding plasmid, containing the intragenic 280 bp sequence between the

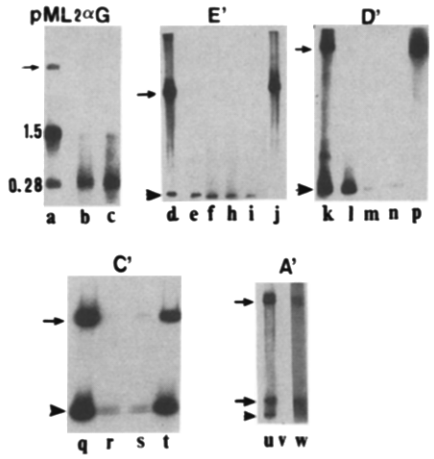

Figure 2. *In vitro* binding of 5'-end labeled fragments obtained after cutting of the plasmids containing different inserted mouse α-globin gene sequences with the corresponding restriction nucleases. **pML2αG-** the plasmid containing the whole globin gene sequence after cutting with Bam H1. **(a)** - fragments bound by the nuclear matrix in the absence of competitor E. coli DNA; **(b)** - fragments precipitated by the nuclear matrix in the presence of 200 μg/ml competitor DNA; **(c)** - binding in the presence of 300 μg/ml E.coli DNA. The largest fragment not found in the matrix bound fraction is the vector plus the 3'-end globin gene sequences flanking the Bam H1 site (Fig.1 A). Everywhere arrows mark the position of the vector; **E`-** the plasmid containing the cloned fragment "E" from Fig. 1B. **(d)-** the input fragments; Binding to the matrix in the presence of 200, 250, 300 and 400 μg/ml of E.coli DNA **(lanes e-i)**, respectively; **(j)** - the supernatant corresponding to the pellet fraction shown in (e); **(D`)** - the plasmid containing the cloned fragment "D" (Fig.1 B); **(k)** - the input fragments; **(l-n)** - the fragments bound to the matrix in the presence of 50, 100 and 200 μg/ml of E.coli DNA, respectively; **(p)** - the supernatant corresponding to the pellet fraction shown in (l); **(C`)** - the cloned plasmid "C"; **(q)** - the input fragments; **(r-s)** - the matrix bound sequences in the presence of 25 and 50 μg/ml of E.coli DNA, respectively; **(t)-** the supernatant corresponding to the pellet in (r); **(A`)** - binding of the cloned fragment "A". The plasmid has been cut additionally with Eco RI, yielding the 300bp fragment "A1" belonging to the mouse sequence and two other fragments belonging to the vector (marked with arrows) - **(u)**; No fragment was found associated with the matrix fraction in the presence of 25 μg/ml of E. coli DNA **(v)**, all the label recovered in the supernatant **(w)**. The sequences of mouse origin are shown with arrowheads.

Bam HI sites (fragment "E") was tested for its capacity to bind to the nuclear matrix. The results shown in Fig.2 E` confirmed the conclusion that in fragment "E" there was a MAR, binding to the matrix in the presence of even 350 μg/ml of E.coli DNA (lane i), while the fragment containing the vector was found in the supernatant exclusively (lane j).

Since it was found that the whole region upstream of this fragment was capable of an interaction with the matrix as well (Fig.2 b-c) we searched for the attachment of sequences adjacent to "E". As seen in Fig. 2 D`, the small fragment (180 bp) between PstI and Bam HI sites (fragment "D") bound specifically to the matrix too (lane l), but in contrast to fragment "E", lower concentrations of the competitor abolished the binding (lanes m-n). The specificity of this interaction is

supported by the complete absence of fragment "D" from the soluble fraction and the specific accumulation of the vector at low concentrations of competitor DNA(lane p). In all experiments the labeled vector acted as an internal control for the specificity of the binding.

Fragment "C" located still further upstream displayed an only insignificant capacity to bind the matrix, the interacton being almost completely abolished by 5(µg/ml E.coli (Fig.2 C`, r-s); lane "t" shows that compeitor DNA concentrations as low as 25 µg/ml prevent the binding of fragment "C" to the matrix, both fragments being recovered exclusively from the soluble fraction.

The specificity of the binding of the tested fragments was confirmed also by the lack of attachment of the region located about 600 bp upstream from the transcribed region (fragment "A1", Fig.2, A , u-w).

The results presented thus far, therefore, suggest that in the α-globin gene there are sequences located between the second PstI and the third Bam HI sites capable of a specific interaction with the nuclear matrix. These are either two separate MARs (in fragments "E" and "D") differing in their affinity for the matrix proteins or they constitute one matrix association region with multiple attachment points differing by their degree of involvement in the interaction [11].

When isolated lamina [24] were reacted with labeled fragments obtained after Bam HI or PstI digestions, no binding of any MAR-containing fragment occurred ever at competitor concentrations as low as 5 µg/ml E.coli DNA (not shown). This result is in agreement with previous reports demonstrating that the lamins are not involved in the attachment or the specific recognition of SARs [14, 31] and with a very recent observation that in fact chromosomal DNA is not anchored on the nuclear lamina as previously considered [32].

In general MARs cohabitate with elements of importance for gene regulation and there is growing evidence for the role of MARs in the control of gene expressior [7, 9, 25-27]. In our case, fragment "D" contains the transcription initiation site anc is adjacent to DNase I hypersensitive sites [28, 29]. The most strongly binding fragment - "E", covers the first intron as well as part of the second exon. Although often sequences closely related to the cleavage consensus of topoisomerase II are enriched in MARs [reviewed in 1, 11] in the MAR-containing region of the α-globin gene, there is no enrichment in topo II sites. However, matrix association regions lacking topoisomerase recognition sites have been reported for other genes and evidently a homology to the topo II consensus sequence is not a necessary condition for a MAR function [11, 30, 31].

The α-globin gene is not associated with the nuclear matrix *in vivo*

The results obtained to this end demonstrate only the capacity of certain sequences to form specific precipitable complexes with isolated nuclear matrices.

The results describing the existence of a MAR in the α-globin gene, however, are in contrast to several previous studies showing that the mammalian α-globin gene is not attached to the nuclear matrix *in vivo*. In these cases isolated nuclear matrices [21], or nuclear scaffolds [21, 23], still containing the neighboring bound

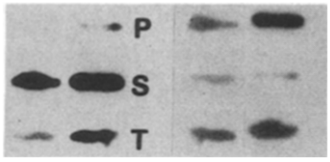

Figure 3. Nuclear matrices prepared after 2 M NaCl-extraction of Sac1 digested mouse liver nuclei. 5 μg and 10 μg of matrix-associated (P), unbound soluble (S) and total unfractionated (T) DNAs, respectively, were hybridized to nick-translated whole pML2αG- **(a)**; the same preparations of (P), (S) and (T) DNAs hybridized to a labeled plasmid containing the κ-immunoglobulin gene MAR - **(b).**

DNA, have been tested for a possible enrichment in the α-globin gene sequences. Independently of whether the analysis has been based on restriction fragment- or dot-blot hybridizations, the results invariably indicated that *in vivo* this gene is not attached to the matrix in mammalian nuclei.

It was necessary, therefore, to test whether the *in vivo* approach will give similar results in our hands. Isolated mouse liver nuclei were extensively digested with Sac1 and the fragment of about 3.1 kbp, containing the α-globin gene sequence used as a probe [28]. After digestion, the nuclei were extracted with an equal volume of 4 M NaCl and equal amounts of DNA from the solubilized and from the matrix-bound fractions were hybridized to nick-translated pML2αG in a slot-blot. It is shown that practically no α-globin gene sequences remained in the pellet with the matrix fraction (Fig.3, a-P), while all of them were recovered from the soluble fraction (Fig.3, a-S). In contrast, the same preparation of matrix-bound and soluble genomic DNA showed an enrichment of the P-fraction in κ-immunoglobulin MAR sequences (Fig. 3, b-P).

Therefore, in our *in vivo* analysis we obtained results in full agreement with previous conclusions that the α-globin gene does not seem to be attached to the nuclear matrix independently of whether the skeletal structures have been obtained by the high-salt or LIS- procedures [21-23]. Moreover, the α-globin gene is not matrix-attached either when inactive or when actively transcribed [21, 22]. The existence of MAR sequences shown here, therefore, presents an apparent paradox: on the one hand, there is a sequence with a high affinity for the matrix proteins and on the other, this sequence does not interact with the proteins *in vivo.* Two questions immediately arise: 1. What prevents the sequences located between the second PstI and the third Bam HI sites from interaction with the proteins of the matrix?; 2. Why does such a sequence exist and could it have some biological significance? It is not possible to answer these questions at present but several possibilities may be considered: 1).- post-synthetic modifications - i.e. methylation in the MAR-containing sequence of the α-globin gene prevents it from interaction with the matrix proteins or,

2) - a soluble protein specifically interacts and covers the MAR, hindering its interaction with the matrix.

We consider the second possibility more probable, since the α-globin gene has not been found associated with the nuclear matrix even when actively transcribed [22, 23], i.e. when the level of methylation is low. Besides, the observed enrichment of the matrix fraction in α-globin sequences when high-salt extractions preceded the nuclease digestion [22] could be interpreted as an evidence for the existence of a factor which is extracted from the nuclei by the salt treatment. The MAR-sequence of the α-globin gene would be left unprotected which would favour its interaction and precipitation with the matrix.

This supposition requires also that at least two different proteins - one soluble and one insoluble (component of the nuclear matrix) should be able to bind the MAR. This seems quite possible since it has been shown that there are multiple and possibly overlapping binding sites for the matrix proteins on the MARs [29]. We have obtained evidence for the existence of an extractable factor reacting specifically with the α-globin gene MAR (unpublished results) and its study is currently underway.

Although MARs (SARs) are usually found at the borders of different genes, considered to mark the boundaries of a chromatin domain, there have been several reports localizing the matrix attachment sites within the transcribed sequences [10, 21, 34, 35]. Of particular interest for us is the MAR found in the β-gene intron which might correspond to the MAR detected in the intron of the α-gene described here. It is thought that mammalian α-and β -globin genes have evolved from a common ancestral sequence and the general structure of the genes, the position and the size of the exons and introns is very similar [28]. However, both MARs seem not to be utilized *in vivo,* but their proximity to cis-regulatory sequences suggests a possible involvement in regulation [9, 25, 30]. As the mammalian α-globin gene does not attach to the matrix either when transcribed or while silent, the intragenic MAR might reflect a regulatory mechanism for this gene lost in evolutior or during ontogenesis. The finding of matrix binding regions in the chicken and duck α-globin gene loci and the attachment of these genes *in vivo* to the nuclear matrix supports such a hypothesis [35, 36].

REFERENCES

1. Gross, D.S., and Garrard, W.T. (1987). *Trends in Biochem. Sci.* **12**, 293-297.
2. Bodnar, J.W. (1988). *J. Theor. Biol.* **132**, 479-507.
3. Goldman, M.A. (1988). *BioEssays* **9**, 50-55.
4. Berezney, R., and Coffey, D. (1974), *Biochem. Biophys. Res. Commun.* **60**, 1410-1417.
5. Kaufmann, S.H., Coffey, D.C., and Shaper, J.H. (1981). *Exp. Cell Res.* **132**, 105-123.
6. Mirkovitch, J., Mirault, M.-E., and Laemmli, U.K. (1984). *Cell* **39**, 223-232.
7. Gasser, S.M. (1988). In: Architecture of Eukaryotic Genes, G. Kahl, ed. (VCH Verlagsgeselschaft, Wienheim); pp. 461-471.
8. Smith, H.C., Ochs, R.L., Lin, D., and Chinault, A.C. (1987). *Mol. Cell. Biochem.* **77**, 49-61.

9. Blasquez, V.C., Xu, M., Moses, S.C., and Garrard, W.T. (1989). *J. Biol. Chem.* **264**, 21183-21189.
10. Cockerill, P.N., and Garrard, W.T. (1986). *Cell* **44**, 273-282.
11. Gasser, S.M., and Laemmli, U.K. (1986). *Cell* **46**, 521-530.
12. Gasser, S.M., and Laemmli, U.K. (1986). *EMBO J.* **5**, 511-517.
13. Cockerill, P.N., M.-H. Yuen, and W.T. Garrard (1987). *J. Biol. Chem.* **262**, 5394-539.
14. Phi-Van, L., and Strätling, W.H. (1988). *EMBO J.* **7**, 655-664.
15. Bode, J, and Maass, K. (1988). *Biochemistry* **27**, 4706-4711.
16. Dijkwell, P.A., and Hamlin, J.L. (1988). *Mol. Cell. Biol.* **8**, 5398-5409.
17. Amati, B.B., and Gasser, S.M. (1988). *Cell* **54**, 967-978.
18. Hofmann, J.F.-X., Laroche, T., Brand, A.H., and Gasser, S.M. (1989). *Cell* **57**, 725-737.
19. Cockerill, P.N. (1990). *Nucleic Acids Res.* **18**, 2643-2648.
20. Zenk, D.W., Ginder, G.D., and Brotherton, T.W. (1990). *Biochemistry* **29**, 5221-5226.
21. Jarman, A.P. and Higgs, D.R. (1988). *EMBO J.* **7**, 3337-3344.
22. Kirov, N., Djondjurov, L., and Tsanev, R. (1984). *J. Mol. Biol.* **180**, 601-614.
23. Greenstein, R.J. (1988). DNA **7**, 601-607.
24. Krachmarov, C., Tasheva, B., Markov, D., Hancock, R., and Dessev, G. (1986). *J. Cell. Biochem.* **30**, 351-359.
25. Xu, M., Hammer, R.E., Blasquez, V.C., Jones, S., and Garrard, W.T. (1989). *J. Biol. Chem.* **264**, 21190-21195.
26. Stief, A., Winter, D.M., Strätling, W.H., and Sippel, A.E. (1989). *Nature* **341**, 343-345.
27. Bonifer, C., Vidal, H., Grosveld, D., and Sippel, A.E. (1990). *EMBO J.* **9**, 2843-2849.
28. Nishioka, Y., and Leder, P. (1979). *Cell.* **18**, 875-882.
29. Sheffery, M., Marks, P.D., and Rifkind, R.A. (1984). *J. Mol. Biol.* **172**, 417-436.
30. Ito, T., and Sakaki, Y. (1987). *Biochem. Biophys. Res. Commun.* **149**, 449-454.
31. Sperry, A.O., Blasquez, V.C. and Garrard, W.T. (1989). *Proc. Natl. Acad. Sci. USA* **86**, 5497-5501.
32. Izaurralde, E., Mirkovitch, J., and Laemmli, U.K. (1988). *J. Mol. Biol.* **200**, 111-125.
33. Paddy, M.R., Belmont, A.S., Saumwebel, H., Hamid, D., and Sedat, J. (1990). *Cell* **62**, 89-106.
34. Käs, E., and Chasin, L.A. (1987). *J. Mol. Biol.* **198**, 677-692.
35. Farache,G., Razin, S.V., Rzeszowska-Wolny, J., Moreau, J., Recillas-Targa,F. and Scherrer, K. (1990) *Mol. Cell. Biol.* **10,** 5349-5358.
36. Kretsovali, A., Marcaud, L., Moreau, J. and Scherrer, K. (1986) *Mol. Gen. Genet.* **203,** 193-201.