Volume 13 Number 6 1985 Nucleic Acids Research

Localization and characterization of members of a family of repetitive sequences in the goat β globin locus

Sally E.Spence, Regina M.Young, Karen J.Garner and Jerry B.Lingrel*

University of Cincinnati College of Medicine, Department of Microbiology and Molecular Genetics, 231 Bethesda Avenue, Cincinnati, OH 45267-0524, USA

Received 9 November 1984; Revised and Accepted 4 February 1985

ABSTRACT

We have characterized a family of repetitive DNA elements in the β-globin locus of the goat. These sequences are structurally analogous to the Alu families of repeats of other mammals. Repetitive elements are located both in the intervening sequences and in the intergenic regions of the goat β-globin locus. Nucleotide sequence analysis of five repetitive elements located within the large intervening sequence of the β-like globin genes, and four repeats located 5' to the major developmentally regulated β-globin genes has resulted in the definition of a consensus sequence for this family of repeats.

INTRODUCTION

Short interspersed repetitive DNA sequences (SINES) are a common feature of all known mammalian genomes (1-3). Although no definitive role has been assigned to these sequences, many functions have been postulated to explain their retention in the genome. Some of these suggestions include organizers of chromatin structure (4), signals for tissue-specific regulation of transcription (5-7), origins of DNA replication (8), interruptions in stretches of homologous DNA sequences which block gene conversion (9,10), and "selfish DNA" sequences (11,12).

While defining the goat β-like globin locus, Schon et al. observed the presence of inserted repetitive DNA sequences in the large intervening sequences of the goat β^F, β^C, and β^A globin genes (13). The β^F gene in the goat, formerly referred to as γ, is analogous in function to the γ gene of humans. It is now referred to as β^F due to its evolutionary history. Further characterization of the locus indicated that related sequences are also present in the large intervening sequences of the ϵ^I and ϵ^V globin genes (14,15). To date, β-like globin genes containing inserted repetitive sequences in their intervening sequences have been identified in goats, sheep and cows (10,13,16). These sequences share homology with a repetitive element reported in and around the bovine corticotropin-β-lipotropin precursor gene region (17). Together these sequences define a major class of artiodactyl short interspersed repetitive DNA sequences which are structurally related to the Alu family of sequences in humans and B1 and B2 families in rodents (18-22).

We have found these repetitive sequences particularly interesting in view of their

presence in a triplicated gene cluster. As the goat β-globin gene locus is composed of a recently triplicated four-gene set, we have the opportunity to examine repetitive sequences within long stretches of homologous DNA sequence (23). We are also able to compare inserted repetitive sequences which are the only blocks of non-homology within extremely homologous genes expressed differentially during development. For these reasons, we chose to further characterize repetitive sequences found within the goat β-like globin locus with respect to their position in the globin gene cluster and in the genome, and their structures and sequence composition.

MATERIALS AND METHODS

Mapping of Repetitive Sequences in the Goat β-Globin Locus

Overlapping Charon 4A or EMBL4 phage clones spanning the entire goat β-globin locus were isolated as previously described (23,24). Repetitive hybridization probes were subcloned into pBR322 as follows. The 110 base pair (bp) Pst I/Eco RI fragment from the β^F intervening sequence (IVS) was subcloned into Pst I/Eco RI digested pBR322 and excised by Bam HI/Pst I digestion to yield a 488 base pair fragment for nick translation (25). The β^A IVS repetitive probe was constructed by ligation of Bam HI linkers to a 347 bp partial Alu I fragment containing the entire β^A IVS repeat and subcloned into Bam HI digested pBR322. Southern blots of phage clones were hybridized at 68°C in 5X SSPE (1X SSPE - 0.18 M NaCl, 10 mM $NaPO_4$, pH 7.1, 1 mM EDTA), 2X Denhardt's and 100 μg/ml salmon sperm DNA (26). Filters were washed in 5X SSPE, 0.1% SDS at room temperature. Prior to hybridization to a second radioactive probe, filters were stripped in 5 mM Tris, pH 8.0, 0.2 mM EDTA, pH 8.0, 0.05% Na pyrophosphate, and 0.1 X Denhardt's at 68°C for 3 hours.

DNA Sequence Analysis

Overlapping fragments containing repeats were subcloned into M13mp 8, 9, 10 or 11 digested with various restriction endonucleases and treated with calf intestine phosphatase (27,28). Sequencing reactions were carried out by the dideoxy chain termination method of Sanger (29,30). DNA sequences were analyzed using the computer program of Roger Larsen and Joachim Messing, University of Minnesota.

RESULTS

Initial characterization of the goat β-like globin genes by Schon et al. demonstrated the presence of inserted repetitive elements in the large intervening sequences (IVS) of the β^F, β^C and β^A genes (13). The location of these inserts within the intervening sequences is shown in Figure 1. In the β^F gene the inserted sequence begins 164 nucleotides from the 5' boundary of the IVS, while the β^C and β^A insertions are located 542 nucleotides from the 5' boundary of the IVS. All three insertions are in the same

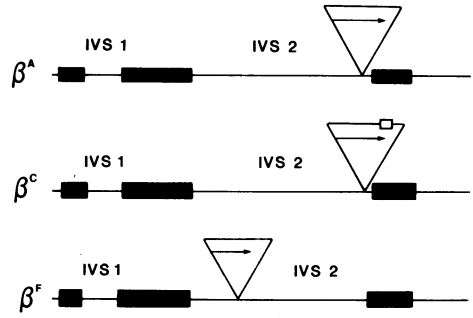


Figure 1: The location of inserted repetitive elements in the second intervening sequences (IVS) of the goat β^A, β^C and β^F globin genes is indicated by a triangle. β-globin coding sequences are shown in heavy black boxes. A deletion in the repetitive element in the β^C IVS relative to the β^A IVS repeat is indicated by an open box. Arrows show the orientation of repetitive elements with respect to transcription of the globin genes.

orientation with respect to transcription of the globin genes. Comparison of the sequence of the inserted repetitive elements (Fig. 2) shows two different sequence arrangements. In the β^F IVS, the inserted element consists of a head-to-tail dimer of sequence with the 5' most copy truncated by 32 base pairs. Perfect 13 base pair direct repeats surround the entire inserted sequence. The β^A globin gene IVS contains two copies of the same repetitive sequence present in the β^F IVS plus additional unrelated sequences located 5' to each copy of the β^F IVS sequence, all surrounded by a perfect 7 base pair direct repeat. Finally, the β^C IVS repeat shows a 60 base pair deletion with respect to the β^A IVS repeat. The imperfect 7 base pair direct repeats flanking the B^C IVS insertion are identified by comparison of non-repetitive portions of goat β and ψβ globin gene intervening sequences.

In order to identify additional repetitive sequences in the goat genome containing either the sequences common to the β^F, β^A and β^C repeats, or the sequences unique to the β^A and β^C repeats, or both, two different hybridization probes were constructed. A subclone of the 110 base pair Pst I/Eco RI fragment from the β^F IVS yields a probe which detects sequences homologous to the repetitive sequence shared by all three β-globin gene insertions (Fig. 2). A second probe detecting the sequences present only in the β^A and β^C IVS repetitive element would permit us to characterize additional repetitive sequences on the basis of their hybridization patterns. As there are no convenient sites in the β^A or β^C IVS to precisely excise the repetitive element, a 347 base pair partial Alu I fragment from the β^A IVS containing 38 base pairs of non-repetitive intervening sequence, sequence homologous to that in the β^F IVS probe, and additional repetitive sequence present only in the β^A and β^C IVS insertions was subcloned (Fig. 2).

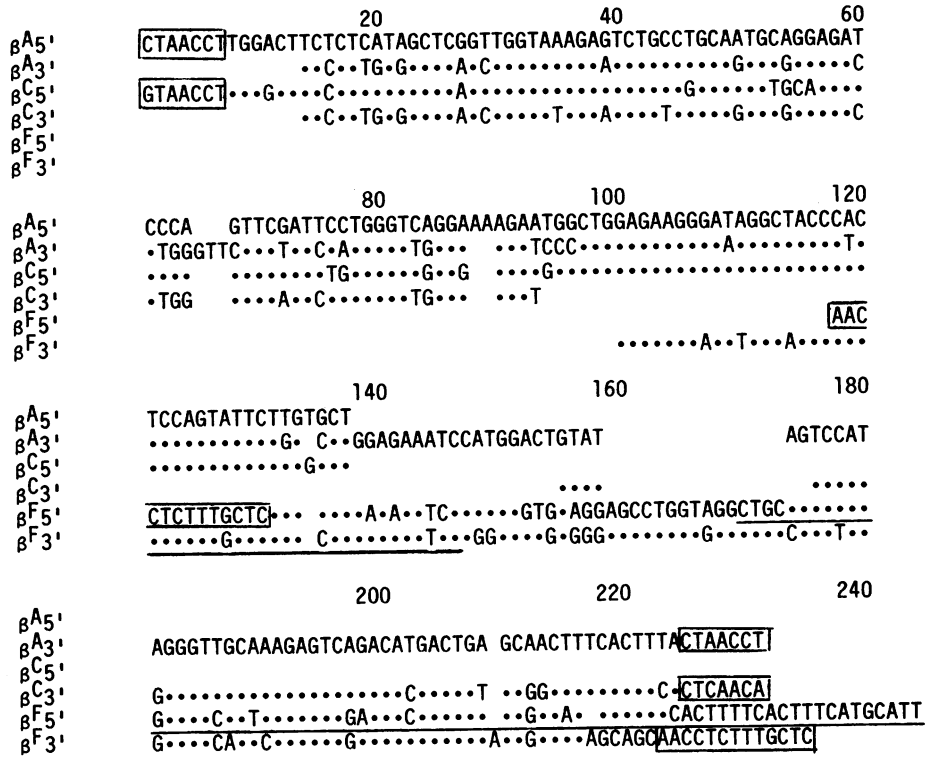


Figure 2: The complete nucleotide sequence of the inserted repetitive elements in the β^A, β^C and β^F IVS is shown. All sequences are aligned to demonstrate the dimeric nature of the repetitive elements, and to provide maximum homology. Terminal direct repeats are boxed; • indicates sequence identity with the uppermost sequence listed; gaps indicate deletions. The entire β^A IVS repetitive sequence and the underlined portions of the β^F IVS repetitive sequence were used as hybridization probes in Figure 3.

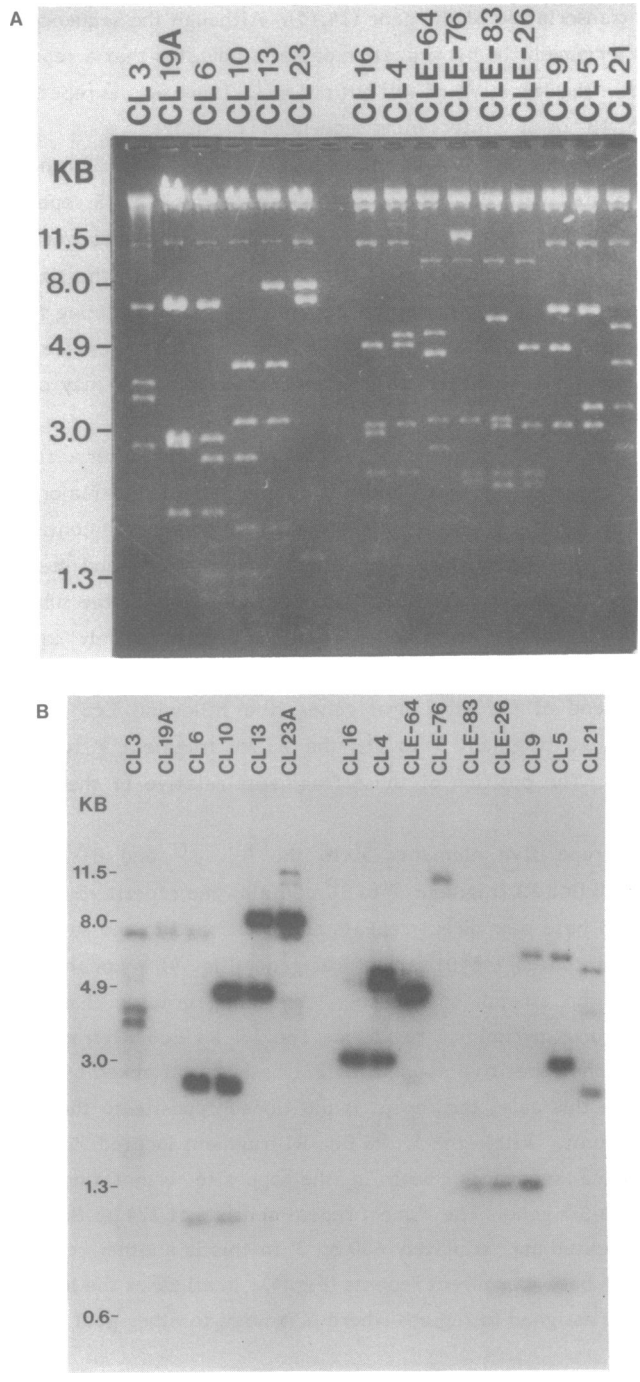
A bank of overlapping phage clones, shown in Figures 3A and 3D, representing the entire goat β-globin locus was used to map the distribution of repetitive elements in the globin gene cluster. Southern blots of these phage clones digested with Eco RI were hybridized sequentially to the β^A IVS and β^F IVS repetitive probes, resulting in the hybridization patterns seen in Figure 3B and 3C, respectively. Copies of the repetitive element are distributed throughout the locus with clustering of hybridizing Eco RI fragments at the 5' and 3' boundaries of the locus. Intensity of hybridization indicates that some Eco RI fragments may contain more than one copy of hybridizing sequence. Although the repetitive sequences in the second IVS of ϵ^I and ϵ^V are homologous to the β^A IVS and β^F IVS repetitive elements, they diverge sufficiently to prevent hybridization to the probes. Both ε IVS repetitive sequences are present in the opposite orientation

with respect to transcription of the gene (14,15). Although the sequences of ϵ^{III} and ϵ^{VI} have not been determined, the hybridization pattern indicates that a repetitive element is probably located in the large IVS of ϵ^{III} but not ϵ^{VI}. There are no repetitive sequences in the ϵ^{II}, ϵ^{IV}, $\psi\beta^{X}$, $\psi\beta^{Y}$ or $\psi\beta^{Z}$ intervening sequences (14,31,32).

To estimate the copy number of this repeat in the genome, 20 random phage clones were isolated, digested with Eco RI and hybridized to the β^{F} IVS repetitive probe (33). Hybridization patterns of these non-globin containing phage clones were similar to those seen in Figure 3 for globin gene clones. Calculations based upon the size of the genomic DNA insert in the 20 clones and the number and size of hybridizing Eco RI fragments result in an estimate of 300,000 copies of this repeat per goat genome. This number is most likely a conservative estimate as a given Eco RI fragment may contain more than one copy of the repetitive element.

We were particularly interested in the further characterization of repetitive elements located immediately 5' to and within the IVS of the major developmentally regulated β-globin genes β^{F}, β^{C} and β^{A}. Eco RI fragments containing 5' flanking sequence, the first and second coding sequences, and the first and second IVS of these three genes were subcloned into pBR325. Mapping showed that the subclones containing the β^{F}, β^{C} and β^{A} genes contained repetitive sequences only in the intervening sequences, not in the 5' flanking sequences. To analyze the repetitive elements located nearest to the 5' end of the structural genes, the following Eco RI fragments were subcloned: the 1.1 kb fragment 5' to β^{C}, the 5.1 kb fragment 5' to β^{A} and the 1.3 kb fragment 5' to β^{F}. The position of these fragments relative to the structural genes is shown in Figure 4.

Finally, the repetitive elements 5' to the β^{F}, β^{C} and β^{A} globin genes were sequenced. A 1.3 kb Eco RI fragment 5' to β^{F} contains one repetitive element of 254 base pairs flanked by 10 base pair direct repeats and positioned in the same orientation with respect to transcription as the β^{F} structural gene (Fig. 4). Located 50 bp 3' to the insertion is a 140 bp stretch of sequence containing greater than 90% pyrimidines, including 51 contiguous pyrimidine residues. The 1.1 kb Eco RI fragment 5' to the β^{C} gene also contains one repetitive element of 231 base pairs flanked by 14 base pair direct repeats (Fig. 4). In this case, the repeat is positioned opposite to the orientation of the β^{C} transcriptional unit. Within the 5.1 kb Eco RI fragment located 5' to β^{A} there are two copies of repetitive sequences, both in the opposite orientation with respect to transcription of the β^{A} gene. The 5' most repeat consists of 224 bp flanked by 6 base pair direct repeats. Located approximately 600 bp 3' to this is another repetitive element of 198 bp flanked by 6 base pair direct repeats (Fig. 4). In all cases the locations of terminal direct repeats were assigned to regions where homology to other goat repetitive elements ended.


A
CL 3
CL 19A
CL 6
CL 10
CL 13
CL 23
CL 16
CL 4
CLE-64
CLE-76
CLE-83
CLE-26
CL 9
CL 5
CL 21
KB
11.5
8.0
4.9
3.0
1.3
B
CL3
CL19A
CL6
CL10
CL13
CL23A
CL16
CL4
CLE-64
CLE-76
CLE-83
CLE-26
CL9
CL5
CL21
KB
11.5
8.0
4.9
3.0
1.3
0.6

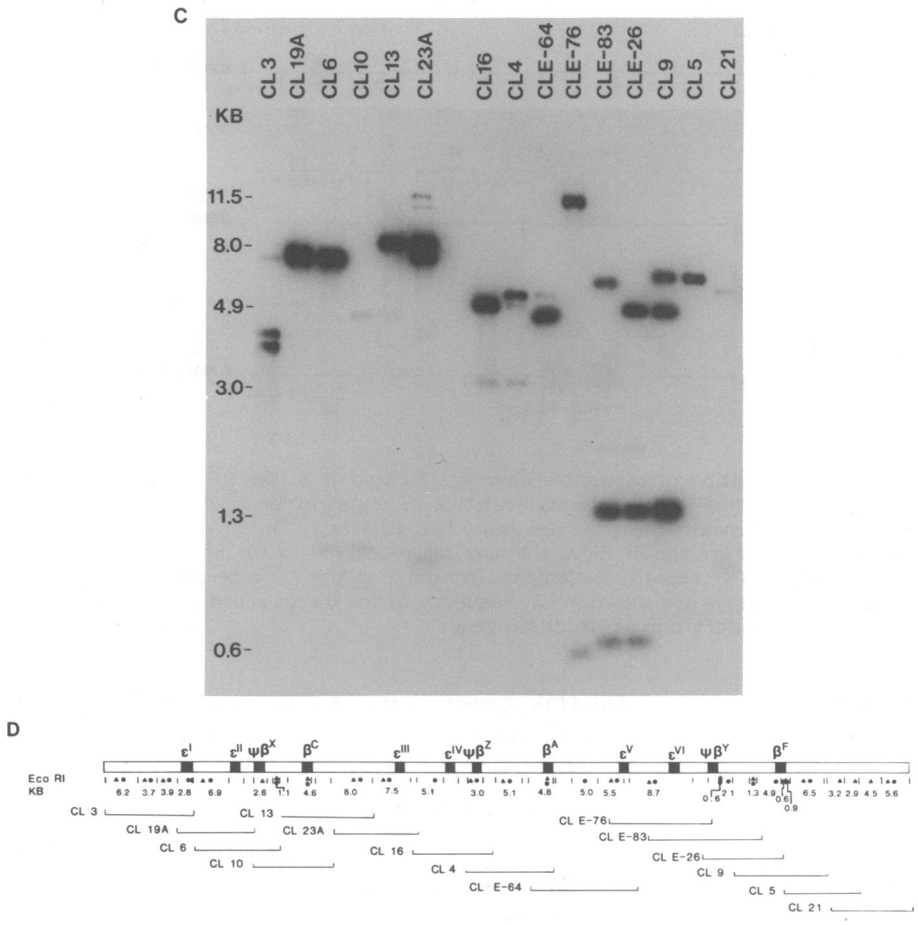


Figure 3: Ethidium bromide staining of a 0.8% agarose gel of Eco RI digested phage clones spanning the entire goat β-globin locus is shown in A (23). A Southern blot hybridized to the β^A IVS repetitive probe is shown in B. Subsequent hybridization to the β^F IVS repetitive probe is shown in C. A complete linkage map of the goat β-globin locus is shown in D. Location of the overlapping phage clones spanning the locus is indicated. The size in kilobases of Eco RI fragments hybridizing to either repetitive probe is shown; ▲, hybridization to the β^A IVS repetitive probe; ●, hybridization to the β^F IVS repetitive probe.

Complete sequence of the repetitive elements located 5' to the β^F, β^C and β^A globin genes, as well as the repetitive elements in the, β^F, β^A, β^C, ϵ^I and ϵ^V IVS, is shown in Figure 5. The top line of the figure defines a consensus sequence for this family of repeats. The consensus sequence beginning at the 5' end and ending at nucleotide 83 is derived from the collection of sequences presented in this paper, and previously identified

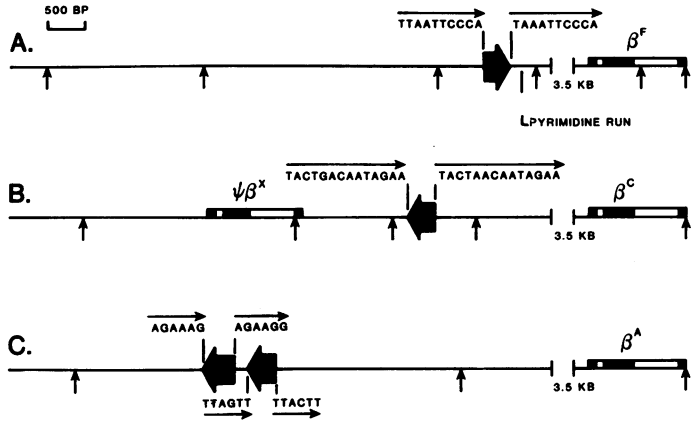


Figure 4: The position of repetitive elements located 5' to the β^F, β^C and β^A globin genes is shown by thick arrows oriented relative to transcription of the structural genes. Small arrows underneath the map represent Eco RI sites. Direct repeats flanking the repetitive elements are shown above the map for repeats 5' to the β^F gene (A) and 5' to the β^C gene (B). Two repetitive elements located 5' to the β^A gene and their respective flanking direct repeats are shown in C. Sequence of the direct repeats is written 5' to 3' with respect to transcription of structural genes.

as the C sequence by Schon et al. (13). Consensus sequence from nucleotide 84 to the 3' end is the sequence defined by Watanabe et al. as the bovine consensus sequence (BCS) (17). The BCS was derived upon characterization of repetitive sequences surrounding and within intervening sequences of the bovine corticotropin-β-lipotropin precursor gene. Sequences homologous to the consensus sequence presented in Figure 5 have been reported in the intervening sequences of bovine β and γ globin genes by Schimenti and Duncan (10). Within the consensus sequence shown in Figure 5, RNA polymerase III split promoter sequences defined by Fowlkes and Shenk (44) are located between nucleotides 10 and 27, and between nucleotides 53 and 64. Sequences showing homology to regions surrounding the origin of replication of papovaviruses previously identified by Watanabe et al. (17) are located between nucleotides 89 and 161 of the consensus sequence.

DISCUSSION

Sequence data presented here, in addition to sequences of repetitive elements from the bovine corticotropin-β-lipotropin precursor and the bovine β-like globin gene regions, define an artiodactyl family of repetitive sequences having many features in common with the Alu-like family of sequences. A major class of short, interspersed repetitive sequences in mammals, the Alu-like families have been characterized in humans (8,18), primates (34,35), mice (20-22,36), rats (37), Chinese hamsters (38) and rabbits (39).

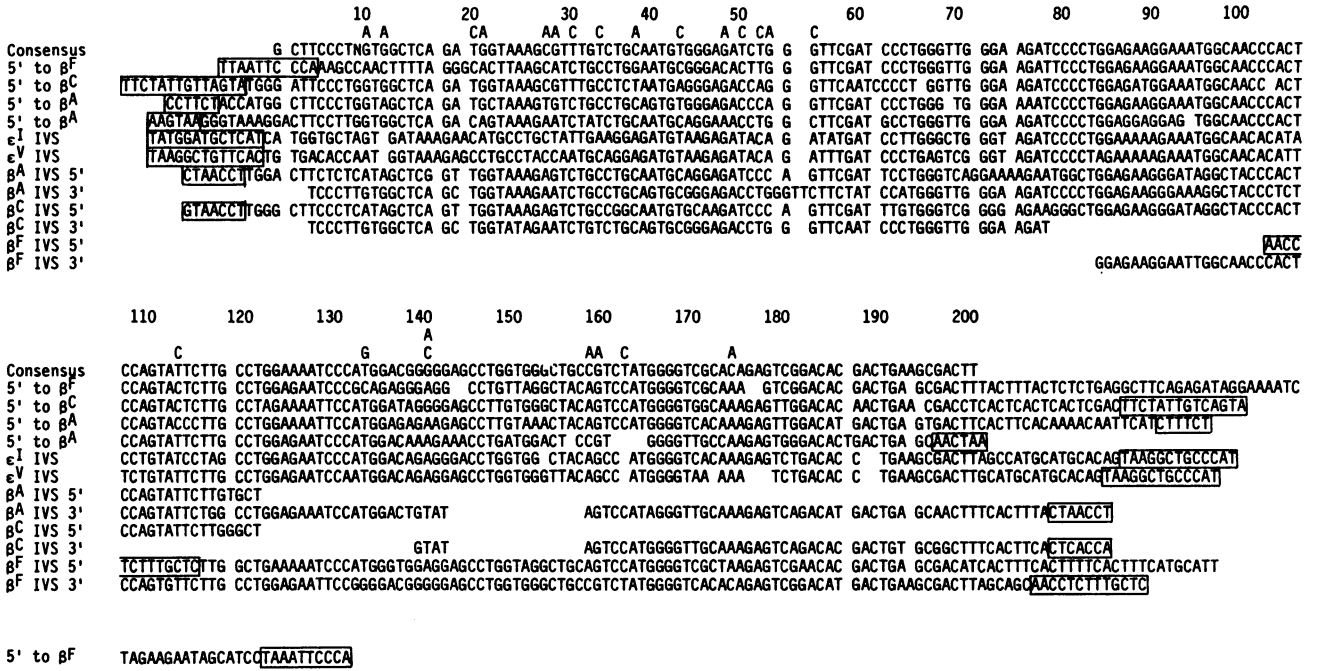


Figure 5: Complete nucleotide sequences of nine repetitive elements from the goat β globin locus, aligned for maximum homology are shown. All sequences are listed in a 5' to 3' orientation. The consensus sequence from nucleotide 1 to 83 was derived from data presented in this paper. Consensus sequence from nucleotide 84 to 200 was identified by Watanabe et al. (17) as a bovine consensus sequence. Direct terminal repeats of individual repetitive elements are boxed. The position of repeats located 5' to the β^F, β^C and β^A globin genes are shown in Figure 4; those within the intervening sequences of the β^F, β^C and β^A globin genes in Figure 1.

Although cross-hybridization between the repeats of these species is not observed under stringent conditions, they do have several features in common. Alu-like sequences are usually short, ranging from 100 to 300 base pairs in length. In the genome they are interspersed with single copy sequences and are present at approximately 300,000 copies. Human Alu sequences are present as a head-to-tail dimer of sequences approximately 135 base pairs in length, with an insertion of 30 base pairs in the 3' half of the dimer. Both halves of the dimer terminate in a stretch of adenine residues. Similar sequence arrangements have been observed in primates, in addition to a second type of sequence. These newly characterized Type II Alu repeats of the galago contain sequences analogous to human Alu sequences at their 3' end, with a non-analogous sequence approximately 100 base pairs in length at their 5' end (35). The other widely characterized sequences in this family, mouse B1 and B2, are observed as monomers of a consensus sequence averaging 129 and 180 base pairs in length, respectively.

Alu-like repetitive sequences from different species have several characteristic sequences in common. Most obvious among these is the presence of an A-rich tail. Several groups have postulated that this feature provides a mechanism for movement of repetitive sequences within the genome by reverse transcription of RNA molecules into DNA and re-insertion into chromosomal DNA (40-43). Consistent with this model is the observation that these repetitive sequence elements are usually surrounded by short direct repeats of various lengths and of non-homologous sequence. Alu-like elements contain the characteristic internal RNA polymerase III promoters defined by Fowlkes and Shenk (44), and it has been demonstrated that individual repetitive sequences can serve as RNA polymerase III templates in in vitro transcription assays (45-50). In addition, these sequences are represented in hnRNA, both as short unit length transcripts and in the intervening sequences and 3' untranslated regions of the structural gene transcripts (51-60). Finally, these repetitive elements share a short stretch of sequence homologous to the origin of replication of papovaviruses (8).

The goat repetitive elements described in this paper are identified as being related to the Alu class of repetitive sequences on the basis of several structural features. These short sequences, approximately 200 base pairs in length, are interspersed with single-copy sequences in the β-like globin locus, and are present at approximately 300,000 copies per genome. Seven copies of this repeat analyzed to date fit a consensus sequence with less than 80% divergence from the consensus, while the ε^I and ε^V IVS repeats diverge 75% from the consensus. Sequences homologous to the split RNA polymerase III promoter defined by Foulkes and Shenk can be identified in the C consensus sequence. Results of in vitro RNA polymerase III transcription assays (61,62) demonstrated that the ability of some individual goat repetitive sequences to act as templates can be correlated with the presence of promoter sequences (data not shown). The β^F intervening sequence repetitive

element, which lacks the C sequence, was unable to direct transcription in our assay. In addition, short sequences homologous to Alu family sequences from other species and to the papovavirus origins of replication have been identified by Watanabe et al. within the BCS.

However, artiodactyl SINES do have some features which differ greatly from other species' families of Alu-like sequences. Most notable is the lack of an A-rich tail in these sequences. If the prediction that the A stretch is required for reverse transcription and re-insertion of DNA copies of repetitive sequences into the genome is correct, then the artiodactyl sequences either do not move, or do so by a different mechanism. The differential placement of these sequences in the intervening sequences of the β^C, β^A and β^F globin genes in the goat argue in favor of their insertion after the duplication of an ancestral β globin gene to β and β^F, but before the duplication which gave rise to β^C and β^A. Also, the presence of direct repeats of non-homologous sequence of varying lengths at the boundaries of the repetitive elements suggsts that they are mobile in the genome. In a position analogous to that of the A-rich tail common in other Alu-like elements, a short repeated pyrimidine stretch can be seen in the goat sequences: insertions in the β^A and β^C IVS have CACTTT repetitive elements; 5' to β^A, $(ACTTC)_2$; 5' to β^C, $(CACT)_3$ and 5' to β^F, $(ACTTT)_2$. Similar short tailing sequences have been reported in the bovine β-like globin gene and corticotropin-β-lipotropin gene regions. Self-priming of reverse transcription of small RNAs lacking A-rich tails has been reported to occur in human U-3 RNA pseudogenes (63).

Another notable difference between the SINES of the goat and the Alu-like repeats of other species is the diversity in arrangement of related sequences. Inserted repetitive elements in the goat are composed of two basic sequence families; a family of sequences approximately 83 base pairs in length defined in Figure 5 as the C sequence, and the 117 base pair bovine consensus sequence defined by Watanabe et al. All artiodactyl SINES described to date contain sequences homologous to the BCS. The β^F IVS repeat is composed of a head-to-tail dimer of the BCS with a 32 base pair deletion in the 5' most copy and the sequence $(CACT_{[3-4]})_3$ separating the two copies. This sequence is similar to the tailing sequences found 3' to many repetitive elements discussed above. The inserted elements in the ϵ^I and ϵ^V IVS and located 5' to the β^C, β^A and β^F structural genes contain one complete copy of the C sequence linked to one copy of the BCS. A third type of repeat, present in the intervening sequences of the β^C and β^A globin genes, is a scrambled arrangement of the BCS and C sequences as seen in Figures 2 and 5. These elements contain head-to-tail dimers of the complete C sequence linked to partial copies of the BCS, the 5' most copy of the BCS containing nucleotides 1 through 30 of the consensus sequence, and the 3' copy containing the majority of the consensus sequence with a deletion of nucleotides 57 through 74. Relative to β^A, the β^C IVS insert contains a

deletion including the entire 5' end of the second BCS copy to nucleotide 74 of the consensus sequence. To date, no copies of the C sequence existing in the absence of the BCS have been identified. The observed variety of SINE sequence arrangements in the goat is very different from that observed in other species. Human and primate Alu sequences are similar to the dimeric arrangement of the BCS seen in the β^F intervening sequence, although the characteristic insertion of 30 nucleotides into the second half of the dimer in primates is not observed in the goat. However, the more common arrangement of the goat sequences with a copy of C 5' to a copy of the BCS is analogous to the recently reported Type II Alu sequences in the galago (35). It should be noted that we have not observed a consensus 5' end of the goat SINE family reported here.

Several explanations for the observed sequence arrangements in the goat may be postulated. First, it is important to note that consensus RNA polymerase III promoters have only been identified in the C sequences. Therefore, the repetitive sequences which contain a copy of the C sequence and a copy of the BCS fit current models designed to account for movement of Alu-like sequences within the genome. The scrambled arrangement of sequences observed in β^A and β^C intervening sequences may have resulted from insertion of two separate elements at the same time, or from insertion of a single element reverse transcribed from an RNA molecule containing two copies of the repetitive element. The presence of direct repeats flanking both of these elements argues against two independent, temporally separate insertion events. If the notion that one insertion event occurred before the duplication of an ancestral β globin gene to give rise to β^C and β^A is correct, the difference in sequence arrangement between the β^C and β^A repeats is most likely due to deletion. As the inserted element in the β^F intervening sequence lacks an RNA polymerase III promoter, yet is flanked by short direct repeats, it is possible that this particular sequence arrangement is due to incomplete reverse transcription of an RNA molecule that at one time contained the internal RNA polymerase III promoter. All repetitive elements characterized in the intergenic regions of the goat β globin cluster contain a single copy of the C sequence linked to a single copy of the BCS, flanked by short direct repeats. This sequence arrangement fits the accepted theory for the mechanism of retroposon insertion into the genome.

It has been noted by John Rogers (43) that the goat C sequence is 65% homologous to a eukaryotic tRNA consensus sequence as defined by Galli et al. (64) and Gauss and Sprinzl (65). Rogers postulates that the C sequence was derived from a tRNA gene and provides polymerase III promoter function for the BCS, which may contain a reverse-transcription priming site. He further states that the C-BCS composite repetitive element may have arisen following insertion of the BCS at the 3' end of a tRNA gene.

Of special interest to our laboratory are the repetitive elements located between the structural genes of the goat β-globin cluster. In contrast with all other characterized

β-globin gene loci, the goat locus is composed of a triplicated four-gene set, resulting in an arrangement of genes not linear with respect to developmental order of expression (23). We chose to examine repetitive elements located 5' to the major developmentally regulated genes β^F, β^C and β^A in an attempt to define the role of repeats in the evolution of goat β-globin locus, or in the regulation of differential gene expression. Alu-like repetitive elements have been characterized in intergenic regions of the human (66-69), mouse (70,71) and rabbit (39,72) β-globin gene loci. We have demonstrated that the goat β-like globin gene cluster, which spans 160 kilobases, contains at least 27 repetitive elements, including at least five and possibly six located in the intervening sequences of structural genes. Comparison of the location of repeats (Fig. 4) suggests that not all elements were present before the triplication of the locus. One proposed function for Alu-like sequences is to provide a short region of non-homology within a stretch of otherwise very homologous sequences in order to block gene conversion (9,10). In this way, subtle differences in duplicated genes which may account for differences in expression may be maintained. We have not detected any direct evidence of recent gene conversion within the goat β-globin structural genes. We do not yet have sufficient sequence information from intergenic regions of the goat β-globin locus to speculate on any relationship between repetitive elements and gene conversion in these areas.

An additional possibility has been suggested to explain the presence of Alu-like repeats in mammalian genomes. Alu-like sequences may be involved in the organization of chromatin structure. The DNase I sensitive domain extending 5' and 3' to the chicken ovalbumin gene has been shown by Stumph et al. to terminate in close proximity to chicken Alu-like repetitive sequences (4). To date, similar correlations have not been made between repeats and chromatin structure surrounding mammalian globin genes. Work in progress in our laboratory is designed to test the possibility that intergenic repetitive elements play a role in local organization of chromatin structure in the goat β-globin locus.

In summary, we have described here a family of short, interspersed repetitive elements located in the goat β-globin gene locus. These repeats are present in the intervening sequences of expressed globin genes as well as in intergenic regions. As the presence in the intervening sequences and sequence arrangement of these repetitive elements are the only significant differences between three very homologous but differentially expressed globin genes, it is possible that the repetitive elements play a role in the evolution or expression of the β^F, β^C and β^A globin genes. A similar function may be assigned to intergenic repetitive elements as there is extensive sequence homology between the triplicated gene sets extending into non-coding regions. We have characterized nine individual repetitive elements from the β-globin locus and have derived a consensus sequence for the 5' half of the most common sequence arrangement

observed in the goat. The 3' portion of this sequence is homologous to a previously characterized bovine family of repetitive sequences. Physical characteristics of this composite artiodactyl repetitive element indicate that it is related to the Alu family of sequences.

ACKNOWLEDGEMENTS

We would like to thank Ms. Rose Alden for help in preparation of the manuscript. This work was supported by National Research Service Award 5F32 HL 06510 (SES) and grants GM 10999 and HL 15996 from the National Institutes of Health (JBL).

*To whom correspondence and reprint requests should be addressed

REFERENCES

1. Singer, M.F. (1982) Int. Rev. Cytol. 76, 67-112.
2. Jelinek, W. and Schmid, C. (1982) Ann. Rev. Biochem. 51, 813-844.
3. Schmid, C. and Jelinek, W. (1982) Science 216, 1065-1070.
4. Stumph, W., Baez, M., Beattie, W., Tsai, M.J. and O'Malley, B. (1983) Biochemistry 22, 306-315.
5. Britten, R. and Davidson, E. (1969) Science 165, 349-357.
6. Davidson, E. and Britten, R. (1979) Science 204, 1052-1059.
7. Milner, R., Bloom, F., Lai, C., Lerner, R. and Sutcliffe, J.G. (1984) Proc. Natl. Acad. Sci. USA 81, 713-717.
8. Jelinek, W., Toomey, T., Leinwand, L., Duncan, C., Biro, P., Choudary, P., Weissman, S., Rubin, C., Houck, C., Deininger, P. and Schmid, C. (1980) Proc. Natl. Acad. Sci. USA 77, 1398-1402.
9. Hess, J., Fox, M., Schmid, C. and Shen, C-K.J. (1983) Proc. Natl. Acad. Sci. USA 80, 5970-5974.
10. Schimenti, J. and Duncan, C. (1984) Nucl. Acids Res. 12, 1641-1655.
11. Doolittle, W. and Sapienza, C. (1980) Nature 284, 601-603.
12. Orgel, L. and Crick, F.H.C. (1980) Nature 284, 604-607.
13. Schon, E.A., Cleary, M., Haynes, J. and Lingrel, J.B. (1981) Cell 27, 359-369.
14. Shapiro, S.G., Schon, E.A., Townes, T.M. and Lingrel, J.B. (1983) J. Mol. Biol. 169, 31-52.
15. Shapiro, S.G. and Lingrel, J.B. (1984) Mol. Cell. Biol., in press.
16. Kretschmer, P., Coon, H., Davis, A., Harrison, M. and Nienhuis, A. (1981) J. Biol. Chem. 256, 1975-1982.
17. Watanabe, Y., Tsukada, T., Notake, M., Nakanishi, S. and Numa, S. (1982) Nucl. Acids Res. 10, 1459-1469.
18. Houck, C., Rinehart, F. and Schmid, C. (1979) J. Mol. Biol. 132, 289-306.
19. Deininger, P., Jolly, D., Rubin, C., Friedmann, T. and Schmid, C. (1981) J. Mol. Biol. 151, 17-33.
20. Kramerov, D., Grigoryan, A., Ryskov, A. and Georgiev, G. (1979) Nucl. Acids Res. 6, 697-713.
21. Krayev, A., Kramerov, D., Skryabin, K., Ryskov, A., Bayev, A. and Georgiev, G. (1980) Nucl. Acids Res 8, 1201-1215.
22. Georgiev, G., Ilyin, Y., Chmeliauskaite, V., Rysokov, A., Kramerov, D., Skryabin, K., Krayev, A., Lukanidin, E. and Grigoryan, M. (1980) Cold Spring Harbor Symp. Quant. Biol. 45, 641-654.
23. Townes, T.M., Fitzgerald, M.C. and Lingrel, J.B. (1984) Proc. Natl. Acad. Sci. USA, in press.

24. Robbins, J., Rosteck, P., Haynes, J., Freyer, G., Cleary, M., Kalter, H., Smith, K. and Lingrel, J.B. (1979) J. Biol. Chem. 254, 6187-6195.
25. Rigby, P., Dieckmann, M., Rhodes, C. and Berg, P. (1977) J. Mol. Biol. 113, 237-251.
26. Southern, E.M. (1975) J. Mol. Biol. 98, 503-517.
27. Messing, J. and Vieira, J. (1982) Gene 19, 269-276.
28. Messing, J., Crea, R. and Seeburg, P. (1981) Nucl. Acids Res. 10, 6451-6463.
29. Sanger, F. and Coulson, A. (1975) J. Mol. Biol. 94, 441-448.
30. Sanger, F., Nicklen, S. and Coulson, A. (1977) Proc. Natl. Acad. Sci. USA 74, 5463-5467.
31. Cleary, M., Schon, E. and Lingrel, J.B. (1981) Cell 26, 181-190.
32. Townes, T.M., Shapiro, S.G., Wernke, S.M. and Lingrel, J.B. (1984) J. Biol. Chem. 259, 1896-1900.
33. Tashima, M., Calabretta, B., Torelli, G., Scofield, M., Maizel, A. and Saunders, G. (1981) Proc. Natl. Acad. Sci. USA 78, 1508-1512.
34. Grimaldi, G., Queen, C. and Singer, M. (1981) Nucl. Acids Res. 9, 5553-5568.
35. Daniels, G.R. and Deininger, P.L. (1983) Nucl. Acids Res. 11, 7595-7610.
36. Kalb, V.F., Glasser, S., King, D. and Lingrel, J.B. (1983) Nucl. Acids Res. 11, 2177-2184.
37. Page, G., Smith, S. and Goodman, H. (1981) Nucl. Acids Res. 9, 2087-2104.
38. Haynes, S., Toomey, T., Leinwand, L. and Jelinek, W. (1981) Mol. Cell. Biol. 1, 573-583.
39. Shen, C. and Maniatis, T. (1980) Cell 19, 379-391.
40. Jagadeeswaran, P., Forget, B. and Weissman, S. (1981) Cell 26, 141-142.
41. Grimaldi, G. and Singer, M. (1982) Proc. Natl. Acad. Sci. USA 79, 1497-1500.
42. Sharp, P. (1983) Nature 301, 471-472.
43. Rogers, J. (1984) Int. Rev. Cytol. 17, in press.
44. Foulkes, D. and Shenk, T. (1980) Cell 22, 405-413.
45. Duncan, C., Biro, P.A., Choudary, P., Elder, J., Wang, R., Forget, B., DeRiel, J. and Weissman, S. (1979) Proc. Natl. Acad. Sci. USA 76, 5095-5099.
46. Duncan, C., Jagadeeswaran, P., Wang, R. and Weissman, S. (1981) Gene 13, 185-196.
47. Elder, J., Pan, J., Duncan, C. and Weissman, S. (1981) Nucl. Acids Res. 9, 1171-1189.
48. Pan, J., Elder, J., Duncan, C. and Weissman, S. (1981) Nucl. Acids Res. 9, 1151-1170.
49. DiSegni, G., Carrara, G., Tocchini-Valentini, G.R., Shoulders, C.C. and Baralle, F. (1981) Nucl. Acids Res. 9, 6709-6722.
50. Shen, C-K.J. and Maniatis, T. (1982) J. Mol. Appl. Genet. 1, 343-360.
51. Federoff, N., Wellauer, P. and Wall, R. (1977) Cell 10, 597-610.
52. Bell, G., Pictet, R. and Reuter, W. (1980) Nucl. Acids Res. 8, 4091-4109.
53. Haynes, S. and Jelinek, W. (1981) Proc. Natl. Acad. Sci. USA 78, 6130-6134.
54. Dalla-Favera, R., Gelmann, E., Gallo, R., and Wang-Staal, F. (1981) Nature 292, 31-35.
55. Calabretta, B., Robberson, D., Maizel, A. and Saunders, G. (1981) Proc. Natl. Acad. Sci. USA 78, 6003-6007.
56. Balmain, A., Frew, L., Cole, G., Krumlauf, R., Ritchie, A. and Birnie, G. (1982) J. Mol. Biol. 160, 163-179.
57. Tsukada, T., Watanabe, Y., Nakai, Y., Imura, H., Nakanishi, S. and Numa, S. (1982) Nucl. Acids Res. 1471-1479.
58. Lee, M., Loomis, C. and Cowan, N. (1984) Nucl. Acids Res. 12, 5823-5836.
59. Robertson, H. and Dickson, E. (1984) Mol. Cell. Biol. 4, 310-316.
60. Allan, M. and Paul, J. (1984) Nucl. Acids Res. 12, 1193-1200.
61. Weil, P.A., Segall, J., Harris, B., Ng, S-Y, and Roeder, R. (1979) J. Biol. Chem. 254, 6163-6173.
62. Wu, G-J. (1978) Proc. Natl. Acad. Sci. USA 75, 2175-2179.
63. Bernstein, L., Mount, S., and Weiner, A. (1983) Cell 32,461-472.

64. Galli, G., Hofstetter, J. and Birnstiel, M. (1981) Nature 294, 626-631.
65. Gauss, D. and Sprinzl, M. (1983) Nucl. Acids Res. 11, 1-103
66. Fritsch, E., Shen, C., Lawn, R. and Maniatis, T. (1980) Cold Spring Harbor Symp. Quant. Biol. 45, 761-775.
67. Coggins, L., Grindlay, G., Vass, J., Slater, A., Montague, P., Stinson, M. and Paul, J. (1980) Nucl. Acids Res. 8, 3319-3333.
68. Kaufman, R., Ketschmer, P., Adams, J., Coon, H., Anderson, W. and Nienhius, A. (1980) Proc. Natl. Acad. Sci. USA 77, 4229-4233.
69. Malcolm, S., Barton, P. and Ferguson-Smith, M.A. (1981) Human Genetics 57, 388-393.
70. Haigwood, N., Jahn, C., Hutchinson, C. and Edgell, M. (1981) Nucl. Acids Res. 9, 1133-1150.
71. Coggins, L., Vass, K., Stinson, A., Lanyon, G. and Paul, J. (1982) Gene 17, 113-116.
72. Hoejmakers-van Dommela, H., Grosveld, G., de Boer, E., Flavell, R., Varley, J. and Jeffreys, A. (1980) J. Mol. Biol. 140, 531-547.