# Sheep α-Globin Gene Sequences: Implications for Their Concerted Evolution and for the Down-Regulation of the 3' Genes

Maria S. Ristaldi,[1]* Stefania Casula,[1] Andrea Rando,[2] Rita Vestri[3]

[1] Istituto di Ricerca sulle Talassemie e Anemie Mediterranee, CNR, via Ospedale 107, Cagliari, Sardinia, Italy
[2] Istituto di Produzione Animale, Università della Basilicata, Pontenza, Italy
[3] Istituto di Fisiologia Umana, Università Cattolica del Sacro Cuore, Roma, Italy

**Abstract.** In sheep as in man and most other mammals, there are two α-globin genes ($^I\alpha$ and $^{II}\alpha$), which are expressed at different levels, the upstream gene being the most efficient. In α-globin gene triplication and quadruplication, this trend is confirmed, i.e., the α-chain output of the downstream genes progressively decreases. In this study, we have determined the complete sequence of the cDNAs and of both the introns in a triple-α haplotype in which each gene could be recognized for the presence of distinct alleles. The sequence analysis reveals that the bodies of the three α-globin genes are essentially identical (99.9% homology) and moreover indicates that the down-regulation of additional α-globin genes in sheep is not the effect of sequence variation from the Cap to the Poly(A) addition sites. This striking similarity among α-genes is higher than that seen in other mammals and is probably sustained by particularly efficient mechanisms of gene conversion and cross-over fixation.

**Key words:** DNA sequencing — Sheep α-globin genes — Gene conversion — Cross-over fixation — Selection — Gene expression

## Introduction

The duplication of the α-globin genes predates the mammalian radiation and the ancestor cluster of the α-genes is thought to be ζ-α-α-θ- (Hardison et al. 1987). Through gain and loss of genes, each species has diverged from the common ancestral order, leading to the contemporary arrangements. In most mammals studied so far, there are two α-globin genes which are highly homologous, the intraspecies similarity of their sequence being higher than the interspecies one. This type of coevolution is called "concerted evolution" (Zimmer et al. 1980; Liebhaber et al. 1981). The molecular mechanisms that have been proposed to explain the maintenance of this striking similarity are cross-over fixation and gene conversion (Michelson and Orkin 1983). Nevertheless, the two genes in humans and other mammals are expressed at different levels, the 5' gene being the most efficient (Liebhaber and Kan 1981; Liebhaber et al. 1986).

In a previous paper we have demonstrated, at the protein level, that in sheep the upstream α-gene ($^I\alpha$) is expressed almost twice as much as the downstream one ($^{II}\alpha$) (Vestri et al. 1983), and also that in ααα and αααα haplotypes (not unusual in these mammals) the downstream genes progressively reduce their expression (Vestri et al. 1991). This expression gradient appears to be regulated transcriptionally, since the steady-state ratio among the distinct mRNAs parallels that among the α-globin chains (Vestri et al. 1994).

In this study we have completely sequenced the cDNAs and both intervening sequences (IVSs) of the three α-globin genes of a sheep homozygous for the ααα haplotype, where each gene could be recognized by the presence of allelic variants in homozygosity.

The sequencing data show an almost complete homology among the three α-genes, which presupposes a very

---

*Present address: Erasmus University, Department of Cell Biology and Genetics, P.O. Box 1738, 3000 DR Rotterdam, The Netherlands
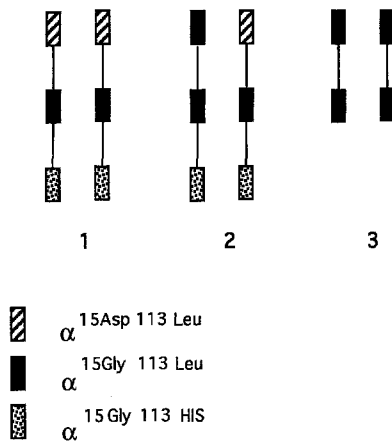Correspondence to: Dr. M.S. Ristaldi

Fig. 1. Schematic representation of the sheep genotypes used in the present study.



Fig. 2. Schematic representation of the position of the oligonucleotides used inside the sheep α-globin gene. Their sequences are as follows: (1) 5'ATGGTGCTGTCTGCCGCC 3', (2) 5'ATGTTCCT-GAGCTTCCCC 3', (3) 5'CGTGTGGACCCGGTCAACT 3', (4) 5'TACCCTGGCCTGGAGCGCC 3', (5) 5'CGGCGGCAGACAG-CACCAT 3', (6) 5'GCCGTGGCCCTTGACCTGGGC 3', (7) 5'CAG-CACGGTGCTCACGTTGGC 3', (8) 5'TTAACGGTATTTGGAGGT-CAG 3'.

## Results

In sheep, the upstream α-globin gene, $^{I}\alpha$, encodes an α-chain-carrying leucine at position 113 ($\alpha^{113Leu}$) whereas the downstream gene, $^{II}\alpha$, produces an α-chain having either leucine or histidine ($\alpha^{113His}$) at that position, the latter being a very frequent allele (Vestri et al. 1983). Moreover, at the $^{I}\alpha$ locus, a rather rare allele occurs which carries a codon for aspartic acid instead of glycine at position 15 ($\alpha^{15Asp}$, also called $\alpha^{D}$) (Huisman et al. 1968).

From a group of selected and inbred sheep, we obtained a sheep homozygous for the triple α-gene arrangement (ααα/ααα) and for both the $\alpha^{15Asp}$ and $\alpha^{113His}$ alleles (genotype n 1 in Fig. 1). From the reticulocytes of this animal we prepared total RNA, and after RT-PCR (using a $(dT)_{18}$ oligo and oligo n 1 in Fig. 2) and M13 cloning, we isolated and sequenced a total of 12 clones (clones 1–12 in Fig. 3). Each clone contained an α-gene cDNA sequence from the ATG to the polyadenylation site. We identified each of the three possible genes in each clone on the basis of the codon present at positions 15 and 113. Five of them (clones 1–5 in Fig. 3) were fully sequenced, while the remaining seven were sequenced only partially (clones 6–12 in Fig. 3).

In order to sequence the 5' untranslated region, we carried out a modified RACE (rapid amplification of cDNA ends) (Frohman 1990) and M13 cloning. The resulting recombinants were screened for the presence of α-globin gene sequences using a $^{32}$P-labeled oligonucleotide specific for the sheep sequence (oligonucleotide n 2 in Fig. 2). We sequenced six positive clones spanning from the Cap site to the Poly(A) addition site, four of them totally and two partially (clones 13–18 in Fig. 2).

By using this approach we were able to unambiguously sequence the Cap site of each cDNA, since the stretch of poly(A) added in vitro was clearly visible. On the other hand, the polyadenylation site could not be located in 11 sequenced clones because of the rearrangements that occurred in this cDNA region during the clon-

efficient mechanism of homogenization and implies that nucleotide variations, in the sequenced region, are not the cause of the observed down-regulation.

We also demonstrate that the break point of the crossing-over which originated the hybrid ($^{II}\alpha/^{I}\alpha$) middle gene in the ααα arrangement is located between position -196 from the Cap site of the $^{II}\alpha$-gene and codon 137 of the $^{I}\alpha$-globin gene.

## Materials and Methods

Sheep were Altamurana, Sarda × Frisona, and an undefined local breed (n 1, 2, and 3, respectively, in Fig. 1).

Each genotype (Fig. 1) was determined by Southern blotting as in Rando et al. (1986) and, at protein level, by isoelectric focusing (IEF) of the hemolysate carried out as in Vestri et al. (1991). Direct detection of the relative position of each gene in the ααα haplotype and sequencing of the IVSs has been carried out on EcoRI-digested and size-fractionated DNA, as previously described (Vestri et al. 1994).

For cDNA synthesis, total RNA was extracted from reticulocytes of anemic sheep by guanidine-HCl lysis and separation through a CsCl gradient (Sambrook et al. 1989). cDNA synthesis was carried out using 5 μg of total RNA and 1 μg of a $(dT)_{18}$ oligo, as reverse primer, in the presence of 20 U of murine myeloblastosis virus reverse transcriptase (BRL), following the manufacturer's instructions.

Polymerase chain reaction (PCR) was performed using standard techniques and a Perkin Elmer apparatus, model 9600. The oligonucleotide for the RT-PCR reaction (n 1 in Fig. 2) was devised on the basis of the goat's α-globin gene coding sequence (Schon et al. 1982), since sheep and goat α-chains are almost identical (Wilson et al. 1968).

A modification in the rapid amplification of cDNA ends (RACE) (Frohman 1990) was carried out using a $(dT)_{18}$ oligo in the reverse transcription as well as in the PCR reactions. The PCR product was ligated to a HincII-digested M13mp18 vector. Ligation, phosphorylation, dephosphorylation, transformation, single-strand DNA extraction, and sequencing were carried out by standard techniques (Sambrook et al. 1989). Screening for the positive clones was performed either by color development or by hybridization with specific $^{32}$P end-labeled oligonucleotides and tetramethylammonium chloride (TMAC) washing (Wood et al. 1985).

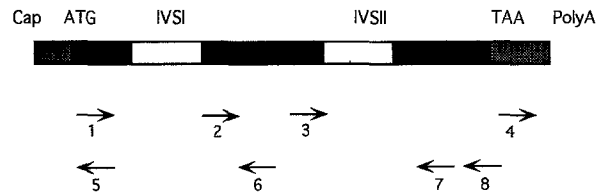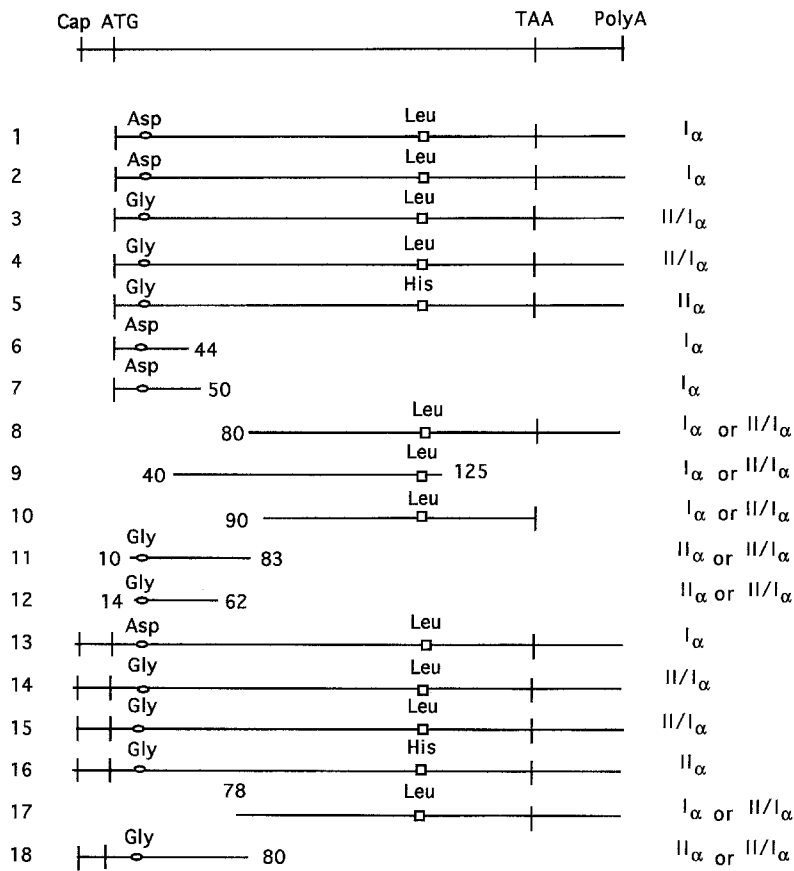The sequences and positions of all oligonucleotides used are shown in Fig. 2.

**Fig. 3.** Schematic representation of the regions sequenced in the 18 clones containing sheep α-globin cDNA. On the *right column* the α-globin locus corresponding to each clone is marked. —○— denotes codon 15; —□— denotes codon 113.

ing in the M13 vector. To overcome this problem, we sequenced and located the cDNA 3′ end by direct sequencing the product of an asymmetric PCR. This was carried out on cDNA prepared from total RNA and using a $(dT)_{18}$ oligo and primer $n$ 3 of Fig. 2. By the same method, using cDNA 5′ Poly(A) tailed in vitro, a $(dT)_{18}$ oligo, and primer $n$ 6 of Fig. 2, we confirmed the position of the Cap site. The entire nucleotide sequence turned out to be practically identical for the three α-globin cDNAs, the only difference, besides those due to the known variants, being a codon synonymous at position 137, which is ACT in $^{I}α$ and ACC in $^{II}α$ gene.

In normal αα arrangement, a *Kpn*I site is present in the 5′ region of the $^{I}α$ globin gene (position –196 based on goat's α-gene sequence [Schon et al. 1982]). The absence of this site, as shown by the restriction map (Rando et al. 1986), and the presence of the ACT codon at position 137 in the middle gene of the ααα arrangement, allows us to deduce that the break point of the crossing-over that originated that gene must be located between these two positions.

To more precisely locate the break point, we sequenced the intervening regions, which, being under a less-powerful selective pressure, were expected to be less conserved than the coding ones. First we physically sep-

arated the three individual α-globin genes of a sheep homozygous for the ααα arrangement (genotype $n$ 2 Fig. 1). In fact, as the restriction map shows (Rando et al. 1986), the *Eco*RI digestion of the DNA from a triple-α sheep produces three fragments of 8.7, 3.6, and 1.5 kb, containing the 5′, the middle, and the 3′ gene, respectively. We therefore *Eco*RI digested 50 μg of total genomic DNA from this sheep and after electrophoresis on agarose gel we excised and eluted the 8.7-, 3.6-, and 1.5-kb fractions (Vestri et al. 1994). We then amplified these fractions in vitro, each containing the individual gene, using specific primers which bridged the two introns, oligonucleotides $n$ 1 and 6 for IVSI and $n$ 3 and 7 for IVSII (their position and sequence are shown in Fig. 2), and sequenced the product of an asymmetrical PCR reaction. Surprisingly, the sequence analysis revealed a complete identity even among the IVSsI of the three genes, as well as among the IVSsII.

In order to exclude the possibility that this identity was a particularity of the specific sheep under investigation, we used the same approach to sequence the IVSs of a totally unrelated sheep (genotype $n$ 3 in Fig. 1) with a normal α-globin gene arrangement (αα/αα). Even in this sheep, the IVSs were identical to those of the triple-α sheep, confirming the previous data.

## Discussion

In this study we have completely sequenced the cDNAs of the three α-globin genes of a sheep homozygous for the ααα arrangement and, in addition, we have determined the nucleotide sequences of both intervening sequences of each of the three genes ($^{I}$α, $^{II}$α, and hybrid $^{II/I}$α genes). This analysis revealed some noteworthy features. Apart from the expected nucleotide differences of the $α^{15Asp}$ and $α^{113His}$ variants, the three α-genes present a striking homology, which extends to the IVSs and to the untranslated regions: one difference out of 768 nucleotides sequenced (99.9% homology). Because of the high homology among the three α-genes, we could not locate the break point of the crossing-over which gave origin to the $^{II/I}$α gene, but we could establish that it occurred in a 854-bp region between position −196 of the $^{II}$α gene and codon 137 of the $^{I}$α gene. Also in humans the two α-globin genes are highly homologous, but this similarity decreases in the IVSII and 3′ untranslated region, where the divergence is 16% (Michelson and Orkin 1983). A deletion of 7 bp in the IVSII of the human α2-gene (or an insertion in the α1-gene) is thought to be important in explaining this diversity. In fact, during a mispairing between α1- and α2-genes, the IVSII gap would hamper the branch migration, which is a necessary mechanism for the model of gene conversion, creating two contiguous but independent conversion units. The downstream unit is less extended than the upstream one (283 bp vs 1436) and would therefore undergo gene conversion less frequently, thus explaining the difference in homology between the two blocks (Michelson and Orkin 1983). In sheep, it can be hypothesized that the absence of gaps in one α-gene, relative to the other, allowed the gene conversion to operate on the entire sequence, hence opposing divergence. Despite the high homology between sheep and goat α-genes (Schon et al. 1982), the two goat α-genes do not reach the identity found in sheep. In this regard it is noteworthy that a 7-bp gap is present in the IVSI of the goat $^{II}$α gene relative to the $^{I}$α, which subdivides the small (about 900 bp) conversion unit into two smaller subunits, likely reducing the conversion rate.

Contrary to gene conversion, the unequal crossing-over determines an expansion-contraction of the gene copy number, the −α and ααα arrangements being considered necessary intermediates in the process leading to the fixation of two α-genes. The ααα arrangement is very frequent in sheep (gene frequency of at least 0.1) and the αααα arrangement is not rare, but surprisingly, no −α haplotypes have been found, despite extensive population studies (more than 500 sheep examined, Rando et al. manuscript in preparation). The high frequency of the triple-α chromosome indicates that the unequal crossing-over contributed to maintain the homology. As to the unusually high frequency observed, a genetic drift effect is excluded by the observation that the

triple-α is found in sheep of different origin. On the other hand, the finding that additional α-genes in sheep do not lead to overproduction of α-globin chains, the output from ααα and αααα being similar to that of αα chromosomes (Vestri et al. 1991), suggests that these arrangements are neutral with regard to the selective pressure. Therefore we think that a high frequency of crossing-over, facilitated by an increased gene copy number and operating for many generations, coupled with absence of selection, may explain the wide spreading of the ααα chromosome. As to the absence of the single-α chromosome, the possibility that this genotype is under negative selective pressure must be considered. This hypothesis seems supported by the positive selection of the −α haplotype in malaric areas in humans (Flint et al. 1986; Modiano et al. 1991). The high frequency detected in these regions implies that α-thal red cells are to some extent different from normal ones. This datum, together with the observation that in nonmalaric areas the −α haplotype is uncommon, strongly suggests that this haplotype confers some kind of selective disadvantage in nonmalaric areas. If this is the case also in sheep, it is conceivable that lacking the positive selection by malaria, the −α haplotype has been lost in the course of many generations.

Another implication of the identity among the three α-globin genes is that the previously observed down-regulation of the downstream α-genes in the ααα (and αααα) haplotypes is not due to structural variations in the sequence from the Cap to the Poly(A) site. Even though the promoter region has still to be sequenced, this region most likely is not responsible for the different expression of the genes. In fact, the genes in second and third positions in the triple-α arrangement have very likely the same 5′ region as deduced from the restriction map. Also in humans the 5′ α-globin gene is expressed two to three times more than the 3′ one (Liebhaber et al. 1986); nevertheless, the promoter region of the two genes is practically identical up to −900 bp from the Cap site (Michelson and Orkin 1983). It is also noteworthy to observe that the αααα haplotype originates, in all probability, from an unequal crossing-over between αα and ααα haplotypes that does not involve the 5′ region of the gene in position 3′. As a result, the gene in third position in the ααα haplotype has simply moved from the third to the fourth position in the αααα haplotype, while the itergenic region, and thus the promoter, remains intact. Nevertheless the two genes are expressed at different levels, the expression of the 3′ gene being 6% in the ααα and 1% in the αααα haplotypes (Vestri et al. 1991, 1994). In conclusion, the decreasing expression level from the 5′ to the 3′ end in multiple α-gene arrangements seems to be a general phenomenon correlated to the relative position of each gene in the haplotype. This cline of expression could reflect a transcriptional interference (Proudfoot 1986). The occurrence of a 5′ locus control region (LCR) similar to that reported in man (Higgs et al.

1990) may also be hypothesized. This region has an enhancerlike effect in vitro (Pondel et al. 1992), and a preferential interaction with proximal compared to distal promoters has been shown for some enhancers (Kadesh and Berg 1986). It is thus possible that a similar phenomenon may occur also in vivo, explaining the observed gradient.

# References

Flint J, Hill AVS, Dowden DK, Oppenheimer SJ, Sill PR, Serjeantson SW, Bana-Koiri J, Bhatia K, Alpers MJ, Boyce AJ, Wetherall DJ, Clegg JB (1986) High frequencies of a thalassemia are the result of natural selection by malaria. Nature 321:744–750

Frohman MA (1990) RACE: rapid amplification of cDNA ends. In: Innis MA, Gelfand DH, Sninsky JJ, White JJ (eds) PCR Protocols. pp 28–38

Hardison R, Cheng JF, Demers GW, Margot JB (1987) Rabbit α like and β like globin gene clusters: comparisons among mammalian globin gene clusters. In: Developmental control of globin gene expression. Alan R Liss, New York, pp 91–105

Higgs DR, Wood WG, Jarman AP, Sharpe J, Lida J, Pretorius IM, Ayyud H (1990). A major positive regulatory region located far upstream of the human α-globin gene locus. Genes Dev 4:1588–1601

Huisman THJ, Dozy AM, Wilson JB, Efremov GD, Vaskov B (1968) Sheep hemoglobin D, an α-chain variant with one apparent amino acid substitution ($\alpha^{15Gly-Asp}$). Biochim Biophys Acta 160:467–469

Kadesh T, Berg P (1986) Effect of the position of the simian virus 40 enhancer on expression of multiple transcription units in a single plasmid. Mol Cell Biol 6:2593–2601

Liebhaber SA, Gossens MJ, Kan YW (1981) Homology and concerted evolution of the α1 and α2 loci of the human α-globin. Nature 290:26–29

Liebhaber SA, Kan YW (1981) Differentiation of mRNA transcripts originating from the α1 and α2-genes in normals and α-thalassemics. J Clin Invest 68:439–448

Liebhaber SA, Cash FE, Ballas SK (1986) Human alfa globin gene expression. J Biol Chem 261:15327–15333

Michelson AM, Orkin SH (1983) Boundaries of gene conversion within the duplicated human α-globin genes. J Biol Chem 258:15245–15253

Modiano G, Morpurgo G, Terrenato L, Novelletto A, Di Rienzo A, Colombo B, Purpura M, Mariani M, Santachiara-Benerecetti S, Brega A, Dixit KA, Shrestha SL, Lania A, Wanachiwanawin W, Luzzato L (1991) Protection against malaria morbidity: near-fixation of the α thalassemia gene in a Nepalese population. Am J Hum Genet 48:390–397

Pondel MD, George M, Proudfoot NJ (1992) The LCR-like α-globin positive regulatory element functions as an enhancer in transiently transfected cells during erythroid differentiation. Nucleic Acid Res 20:237–243

Proudfoot NJ (1986) Transcriptional interference and termination between duplicated α-globin gene constructs suggests a novel mechanism of gene regulation. Nature 322:562–565

Rando A, Ramunno L, Masina P (1986) Variation of the number of α globin loci in sheep. Mol Biol Evol 3:168–176

Sambrook J, Fritsh EF, Maniatis T (1989) Molecular cloning—a laboratory manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY

Schon EA, Wernke SM, Lingrel JB (1982) Gene conversion of two functional goat α-globin genes preserves only minimal flanking sequences. J Biol Chem 257:6825–6835

Vestri R, Giordano PC, Bernini LF (1983) Different quantitative Expression of the hemoglobin α-chain genes in sheep. Biochem Genet 21:1089–1099

Vestri R, Pieragostini E, Yang F, Di Gregorio P, Rando A, Masina P (1991) Expression of triplicated and quadruplicated α globin genes in sheep. Br J Haematol 77:110–116

Vestri R, Pieragostini E, Ristaldi MS (1994) Expression popgradient in sheep αα and ααα globin gene haplotypes: mRNA levels. Blood 83:2317–2322

Wilson JB, Brandt G, Hisman THJ (1968) The structure of sheep hemoglobins. III. Structural studies of the α chain of hemoglobin A. J Biol Chem 243:3687–3692

Wood WI, Gitschier J, Lasky LA, Lawn RM (1985) Base composition independent hybridization in tetramethylammonium chloride: a method for oligonucleotide screening of high complex gene libraries. Proc Natl Acad Sci USA 82:1585–1588

Zimmer EA, Martin SL, Beverly SM, Kan YW, Wilson AC (1980) Rapid duplication and loss of genes coding for the α chains of the hemoglobin. Proc Natl Acad Sci USA 77:2158–2162