Volume 8 Number 20 1980 Nucleic Acids Research

Identification by nucleotide sequence analysis of a goat pseudoglobin gene

Michael L.Cleary, Joel R.Haynes, Eric A.Schon and Jerry B.Lingrel

University of Cincinnati College of Medicine, Department of Biological Chemistry, 231 Bethesda Avenue, Cincinnati, OH 45267, USA

Received 29 July 1980

ABSTRACT

We have recently prepared a recombinant library of goat genomic DNA and have isolated clones containing the known goat globin genes. These include the α, γ, β^C and β^A genes. In addition to these, another β-like sequence has been observed. In this communication we report the complete nucleotide sequence of this gene, excluding a portion of the large intervening sequence. Several features suggest that this is a non-functional or pseudoglobin gene. The alterations include a frameshift mutation, substitution of the heme-binding histidines, a mutated termination codon, a change in the GT/AG excision sequence of the 5' end of the first intervening sequence, an AT rich sequence in the 3' untranslated region, and a mutated Hogness-Goldberg box. We conclude that this gene cannot function in the synthesis of globin.

INTRODUCTION

We have recently cloned the globin genes of the goat (1) and have been studying their structural organization in an attempt to determine how their expression is regulated during development. Overlapping recombinants carrying the genes for the major globins have been identified (2), including the γ gene which is expressed during fetal development, the β^C or preadult gene which is switched on just before birth and is expressed in the preadult, and the β^A or adult globin gene. Two other globin-like genes, β^X and β^Y, have been identified and these are linked to the β^C globin gene. In order to understand the nature of these genes, the nucleotide sequence of one of these, the β^X gene, has been determined and it is concluded that this gene does not code for a functional globin protein.

METHODS

Propagation of recombinant bacteriophage and isolation of phage DNA was as reported by Blattner et al (3) and Maniatis et al (4). Agarose gel electrophoresis was performed as decribed previously (1). Polyacrylamide gel electrophoresis in Tris-borate-EDTA buffer was performed as described by Maniatis et al (5). Bam H1 cut Charon 4A and Hae III cut pBR322 DNA were used as molecular weight size

standards for agarose and polyacrylamide gel electrophoresis. The isolation of restriction fragments from agarose and polyacrylamide gels, end labeling of DNA restriction fragments for DNA sequencing, and DNA strand separation have been described elsewhere (2,6,7). Sequencing of end labeled DNA was performed as described by Maxam and Gilbert (8) and DNA sequencing gels were prepared as described by Sanger et al (9).

[α-^{32}P] Nucleoside triphosphates were obtained from Amersham or New England Nuclear. Restriction enzymes were obtained from Boehringer-Mannheim, Bethesda Research Labs, Inc. or New England Biolabs and all reaction conditions were as recommended by the suppliers. DNA polymerase I (Klenow fragment) was obtained from Boehringer-Mannheim.

Sequencing strategy

The sequencing strategy is presented in Figure 1. Briefly, in order to obtain sequence data of the first coding block and IVS-1, the 368 bp Hae III fragment was end-labeled with DNA polyermase I (Klenow fragment) by 3' terminal exchange using [α-^{32}P] dGTP. This labeled fragment was then strand separated and the complementary DNAs were sequenced. In addition, the two Pvu II sites were used to generate free ends which were labeled by the 3' terminal exchange method after blockage of terminal Eco R1 ends with dATP and TTP. The two large end-labeled Pvu II-Eco R1 fragments and the separated strands of the smaller 156 bp Pvu II fragment were then used for sequencing. The nucleotide sequence of coding block 2, the heme binding region, was generated by 3' terminal labeling at the Sal I site and then sequencing in both the 5' and 3' directions. This sequence overlaps that generated from the 368 bp Hae III fragment and also indicates that a small 37 bp Hae III fragment joins the 368 bp and 790 bp Hae III fragments. Sequence representative of coding block 3 was generated as described earlier (2) by labeling at the Eco R1 site joining the 2800 and 1300 bp Eco R1 fragments and sequencing in both the 5' and 3' directions.

RESULTS

The restriction maps of three recombinants derived from a partial Eco R1 library of goat DNA are shown in Fig. 1. Two of these (clones 10 and 13) contain the β^C or preadult globin gene of the goat. The amino acids coded for by the nucleotide sequence of the β^C gene correspond to those of the published amino acid sequence and thus establish the identity of this gene (10). An additional sequence designated β^X, which hybridizes to globin cDNA, is also found in clone 10 and in an overlapping clone (clone 6). Also, an additional hybridizing sequence, β^Y,

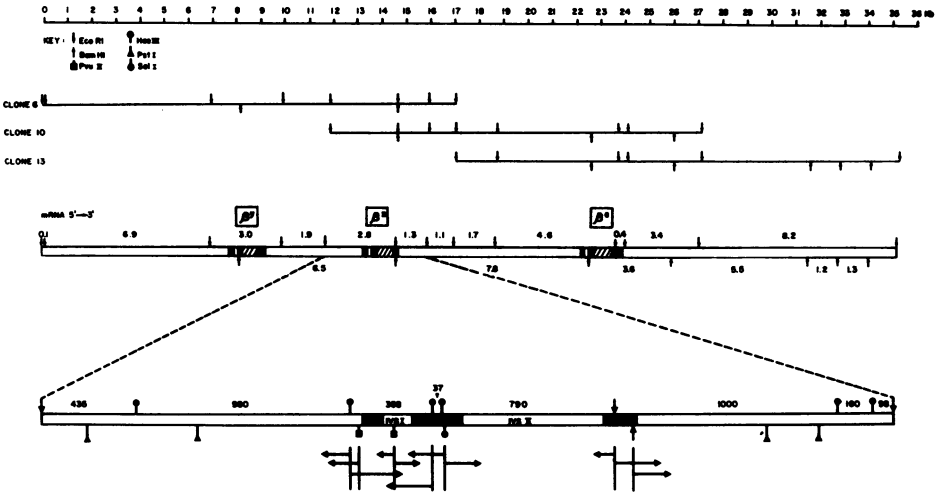
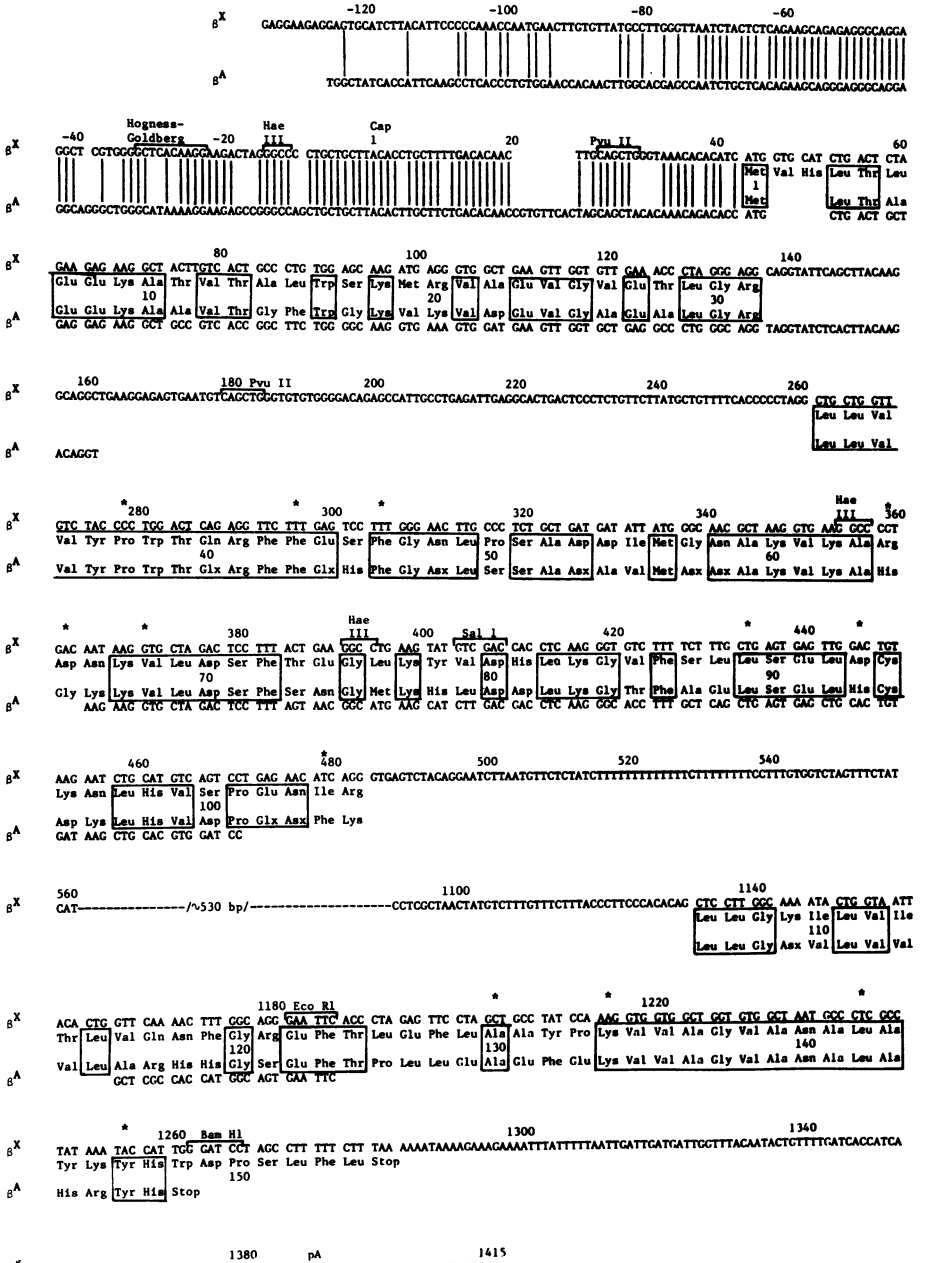


Figure 1. Restriction endonuclease map and nucleotide sequencing strategy for the β^X pseudoglobin gene. Three overlapping recombinants derived from a partial Eco R1 library of goat DNA are depicted along with map positions of Eco R1 and Bam H1 restriction enzyme cutting sites. The area of the goat genome represented by these clones is shown below them with the globin coding regions denoted by solid blocks. The 2800 and 1300 bp Eco R1 fragments which together contain the β^X gene have been enlarged to demonstrate restriction enzyme cutting sites pertinent to the present work. The nucleotide sequencing strategy is shown beneath the β^X gene enlargement.

is located 6 kb upstream from the β^X gene, a total of 15 kb upstream from the β^C gene. The recombinants containing these genes cover a total of 35.5 kb of the goat genome.

The β^X gene spans two Eco R1 restriction fragments 2800 and 1300 bp in length, as determined by agarose gel electrophoresis. Southern transfer of an Eco R1 digest of clone 10 DNA and hybridization to a β-globin cDNA probe indicated that both the 2800 and 1300 bp fragments contain globin coding sequences. Preliminary nucleotide sequence data from the Eco R1 site between these two fragments indicated that this region coded for a globin-like amino acid sequence yet did not code for any known goat globin (2). We therefore obtained additional sequence data in order to elucidate the specific nature of this β^X gene.

For this purpose, the 2800 and 1300 bp fragments were more extensively mapped using various restriction endonucleases; restriction sites relevant to the present work are presented in Fig. 1. These two Eco R1 fragments were excised from recombinant clone 6 and separated by electrophoresis in order to obtain DNA



specific for this region. In addition, the 5' end of this gene (i.e. the left hand Eco R1-Sal I fragment of the 2800 bp Eco R1 fragment) was subcloned into pBR322 to facilitate the preparation of fragments required for sequencing. A Hae III digestion of the 2800 bp Eco R1 fragment followed by electrophoresis and Southern transfer indicated that the 368 bp Hae III fragment hybridized strongly and the 790 bp Hae III fragment hybridized weakly to a β-globin cDNA probe. Therefore, sequence data were obtained for the entire 368 bp Hae III fragment, the 3' end of the adjacent 980 bp Hae III fragment and the 5' end of the 790 bp Hae III fragment in order to supplement the sequence data already obtained in the vicinity of the Eco R1 site joining the 2800 and 1300 Eco R1 fragments.

The nucleotide sequence of the β^X gene is depicted in Fig. 2. Where it has been determined, the nucleotide sequence for the β^A globin gene is included for comparison. An ATG codon for the chain-initiating methionine residue is found 18 nucleotides downstream from the 5' Pvu II site at nucleotides 44-46. Commencing at this point, amino acid assignments can be deduced from the nucleotide data. Within or immediately following codon 11, a single base insertion has shifted the remaining translational reading frame such that two in-phase stop codons are created 11 and 15 codons downstream. Therefore, in the event that an mRNA transcript is synthesized from the β^X gene, translation of the mRNA would result in a non-functional globin polypeptide due to the frameshift and the resultant premature chain termination. If this insertion were not present, a polypeptide sequence would be encoded as represented in Fig. 2. The amino acid sequence deduced for β^A globin is also shown and shared identical amino acids are represented with boxes in this figure. IVS-1 begins at nucleotide number 138 since, beyond this point, the codons no longer code for globin. This results in a coding block 1 for β^X which is 31 amino acids long; the same region in β^A is only 29 amino acids long due to a 6 bp deletion with respect to β^X immediately following the translation initiation methionine residue. Haynes et al (7) have shown that this same region of the β^C gene contains a 9 bp deletion with respect to the β^A gene resulting in a 3 amino acid deletion in the β^C globin when compared to β^A. Therefore, the β^A gene has a 6 bp deletion and the β^C, a 15 bp deletion when compared to the analogous region of the β^X gene. The translated sequence of the

Figure 2. Nucleotide sequence of the β^X pseudoglobin gene and comparison with the goat β^A globin gene. The nucleotide sequence of β^X is displayed 5' to 3' on the top line; the deduced β amino acid sequence is displayed on the second line. The third and fourth lines display the β^A amino acid and nucleotide sequences respectively. Areas of shared nucleotide sequence homology in the 5' flanking region are indicated with vertical lines; shared identical amino acids are enclosed within boxes. Asterisks indicate highly conserved amino acids in vertebrate globins.

initial coding block is 62% homologous to the amino acid sequence of β^A and 54% homologous to that of β^C.

The amino acid sequence of the heme binding region is flanked at either end by a recognizable intervening sequence and is uninterrupted by chain terminating codons. Although the nucleotide sequence of the β^A gene is only known for the 3' half of this region (2), several interesting observations can be made by comparison of the published amino acid sequence of β^A (10) with that deduced for β^X. There are 74 amino acids in the region between IVS-1 and IVS-2 in both β^X and β^A with 22 differences between them resulting in only a 70% amino acid homology (there is also a 70% homology to the same region of β^C). These differences are interspersed among blocks of shared identical sequence. The amino acid sequences of both polypeptide domains immediately following the 3' splice junction of IVS-1 is flanked by a large stretch of identical sequence beginning with Leu-Leu-Val and continuing for 13 amino acids. Nine amino acids in coding block 2 have been shown to be highly conserved for most vertebrate globins (10); these are represented with asterisks in Fig. 2. All nine are conserved in β^A but in β^X four have been altered, including the two absolutely invariant histidine residues intimately associated with heme binding at positions 64 and 93. It is possible to hypothesize single base missense mutations of these codon triplets causing the alteration of histidine at position 64 to an arginine and at position 93 to aspartate. Likewise, single base missense mutations can account for the replacement of glycine at position 65 by aspartate and the replacement of phenylalanine at position 104 by isoleucine.

The nucleotide sequence generated after fill-in labeling of the Eco R1 site between the 2800 and 1300 bp Eco R1 fragments represents coding block 3 or the carboxy-terminal portion of the globin polypeptide. Again there is marked similarity between these sequences and those for β^A (Fig. 2). A comparison of the amino acids coded for by this region with the corresponding region of β^A shows about 55% homology. There are four highly conserved amino acids in this region and all four are conserved in β^A and β^X. There is a recognizable intervening sequence 48 bp on the 5' side of the Eco R1 site since the nucleotide sequences at this point fail to code for globin-like amino acid residues. This position is identical to the intervening sequence seen in the β^A and β^C genes and globin genes from other species. Analysis of the sequence data generated from the Eco R1 site leading in the 3' direction revealed the absence of a termination codon in the position predicted from the amino acid sequence of known goat β globins (2). Instead, a tryptophan codon is present (the result of a single base missense mutation) and the nearest in-phase termination codon is located 21 bp, or 7 codons downstream.

Thus, if the single base insertion mutation at nucleotide 77 is neglected, a globin-like polypeptide is encoded which does not correspond to any of the known goat globins. This polypeptide is only about 65% homologous to either β^A or β^C goat globin when the entire amino acid sequences are compared. The genomic coding sequence for this polypeptide is interrupted by two intervening sequences precisely at the same points as those found in rabbit and mouse β globins. The first IVS interrupts the coding sequence at amino acids 31 and 32 and is 125 bp in length. Virtually all introns described to date have a G-T dinucleotide at the 5' end and an A-G dinucleotide at the 3' end of the excised segment (11). The β^X IVS-1 does not conform to this rule since it has a G-C instead of a G-T at its 5' border (nucleotide numbers 137 and 138). Another notable exception to this rule is the mouse immunoglobin J3 segment which is not found in any published myeloma amino acid sequence, presumably due to an altered RNA splice point in its germ-line gene (12). β^X is also interrupted by a second intron approximately 650 bp in length (as judged by restriction fragment sizing) between amino acids 105 and 106. The splice points for this larger IVS conform to the GT/AG excision rule.

Nucleotide sequence data of the 5' untranslated region were generated from the left hand Pvu II-Eco R1 fragment after blockage of the Eco R1 end with dATP and TTP and labeling at the Pvu II site. To resolve ambiguous nucleotide assignments, this region was also sequenced from the 3' Hae III end of the 980 bp Hae III fragment after end labeling and digestion with Pst I or from the 5' Hae III end of the 368 bp Hae III fragment after strand separation to provide fragments labeled at one end. The sequence data obtained are shown in Fig. 2 and for comparison, the 5' adjacent regions of the adult β^A gene as determined by Haynes et al (7) have been included; areas of shared identical sequences are represented with vertical lines. There are numerous regions of homology between the 5' non-coding regions of these two genes. Generally, as one progresses further in a 5' direction away from the coding region, the homology between these two genes decreases. Of particular interest is the 10 nucleotide segment at nucleotides -32 to -23 designated the Hogness-Goldberg box (13) which occurs in the same relative position in many eukaryotic genes including mouse β^{maj} (14), β^{min} (15) and α globin (16), ovalbumin (17), ovomucoid (18), silk fibroin (19), adenovirus major late mRNA (20), preproinsulin (21), and the rabbit β globin gene (22,23), as well as the goat β and γ globin genes (7). This area is very similar to the procaryotic Pribnow box (24), and may function in RNA polymerase binding (25). Within these 10 nucleotides, β^X differs from β^A at three locations where single base alterations have occurred.

A presumed transcription initiation position is designated Cap in Fig. 2. The

location of the transcription initiation site is based on the finding by Weaver and Weissmann (26) that the 5' end of the 15S globin mRNA precursor maps at the A designated by nucleotide number 1. Mature globin mRNA contains the same 5' sequence and since no larger precursor to globin mRNA has been found in erythroid cells (27), it is likely that the 15S RNA is the initial transcription product and supports the assignment of ACA as the transcription initiation site. In addition, Luse and Roeder (28) have found that transcription initiation occurs at the Cap site when the mouse β^{maj} globin gene is transcribed by an in vitro RNA polymerase II system. Haynes et al (7) have demonstrated that this area of the goat β^{A}, β^{C}, and γ globin genes is so similar to those of the mouse that it too can be presumed to contain the transcription initiation point for the β and γ globin genes of the goat. In the β^{X} gene, the region extending from 9 nucleotides in the 5' direction to 17 nucleotides in the 3' direction from the Cap site is highly conserved when compared to the corresponding region of the β^{A} gene. This area of homology is terminated in β^{X} by a 9 bp deletion located at the immediate 5' side of the Pvu II site in the region between Cap and the ATG translation initiation codon at the 5' end of the coding sequence.

Nucleotide sequence data of the 3' untranslated region were also obtained by cutting the 1300 bp Eco R1 fragment with Bam H1, labeling at the Bam H1 site, then cleaving with Pst 1 to produce a suitable fragment for sequencing. This was done in an attempt to identify possible transcription termination and poly(A) addition signals. In the mouse β^{maj} and β^{min} genes, and human and rabbit β genes the hexanucleotide AATAAA is located 18 base pairs before the poly(A) addition site (29). It is even found in Xenopus β globin mRNA 15 bp from the poly(A) addition site (30), and also in the human, rabbit and Xenopus α globin and chicken ovalbumin and mouse immunoglobin light chain mRNAs in close proximity to their 3' poly(A) ends (29,31). The chicken and rat preproinsulin genes also contain this sequence in their 3' untranslated regions (32,33). In the mouse β^{maj} and β^{min} genes this hexanucleotide is located approximately 106 nucleotides downstream from the translation termination codon. The actual poly(A) addition site in mouse β^{maj} and rabbit and human β globin genes is represented by the triplet TGC (15). The mouse β^{min} gene appears to have lost the internal G and has only a TC pair at the poly(A) addition site. The goat β^{X} gene has a similar but presumably diverged hexanucleotide (AATTAA) 105 nucleotides downstream from the altered stop codon. Eighteen nucleotides downstream from this sequence is a TC couplet possibly representative of a poly(A) addition site. Homologous sequence from the β^{A} gene is not yet available for comparison. This region in β^{X} is also unusually AT-rich in comparison to the known 3' untranslated regions of mouse (15) and

rabbit (22) globin genes.

DISCUSSION

The general structure of β globin genes consists of three distinct amino acid coding blocks interrupted by a small 5' intervening sequence and a larger 3' intervening sequence. Our nucleotide sequence data indicate that β^X conforms to this general structure but has diverged to such an extent that it could not function in the synthesis of globin. This conclusion is based on the data summarized in Fig. 3. Within the β globin coding region several significant mutations occur: (1) the early frameshift mutation with resultant in-phase nonsense codons, (2) the missense obliteration of both invariant histidines, and (3) an alteration in the normal translation termination codon. In addition, there are other suggestions that this sequence represents a pseudoglobin gene; sequences coding for two significant RNA processing sites have been altered in β^X. These sites are (1) the ubiquitous hexanucleotide in the 3' untranslated region of the mRNA which is thought to be associated with poly(A) addition and (2) the universal GT/AG excision sequence at the RNA splicing points for IVS-1.

Other alterations make it unlikely that this gene is even transcribed. The Hogness-Goldberg box (13), a nucleotide sequence found at the 5' end of other eukaryotic mRNA encoding genes and thought to be involved in the initiation of transcription, is markedly altered in the β^X gene. Both an A and T are replaced by C's, eliminating the AT rich character of the sequence. If this region is required for transcription, then these nucleotide differences may affect the expression of this gene. While it has not yet been determined if an RNA is synthesized from this

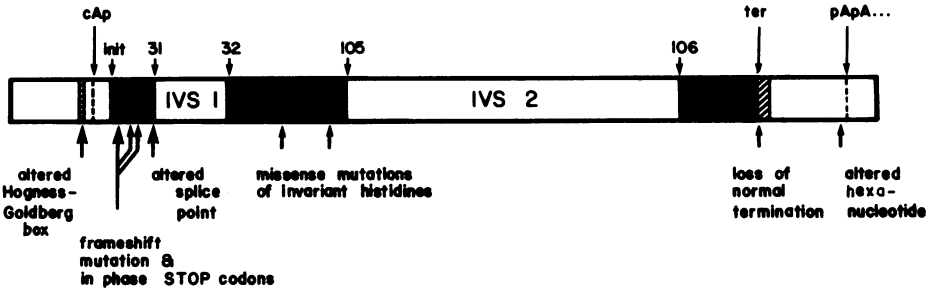


Figure 3. Diagramatic representation of the β^X pseudoglobin gene. Alterations discussed in the text are indicated at appropriate points in the gene. Globin coding regions are indicated as solid black areas. Black diagonal lines at the 3' end indicate additional amino acids past the normal termination codon.

gene in vivo, it is not transcribed in a cell-free RNA polymerase II system (34).

While it is possible that the alterations observed in β^X are due to cloning artifacts, this is considered unlikely because of the extensive and diverse nature of the changes seen. In addition, unexpected globin-like sequences have been found in other globin gene clusters. A β-like sequence designated β2 is located 6 kb upstream from the rabbit adult β globin gene (35,36). This gene has limited sequence homology to adult and embryonic β globin probes and a mature mRNA homologous to this sequence has not been detected in adult or embryonic erythropoietic tissue. Likewise, Fritsch et al (37) have described two genes within the human β globin linkage map (ψβ2 located 5' to the ε globin gene and ψβ1 located between the γ^A and ε globin genes) which have diverged considerably from ε, γ and β globin sequences. Similar unaccountable β-like sequences have been described within the mouse β globin linkage group (38).

Nishioka et al (39) and Vanin et al (40) have described a mouse α globin-like gene which has cleanly lost both globin intervening sequences. This pseudoglobin gene has a distorted translational sequence with early in-phase termination codons, and does not seem to be active in the production of stable mRNA in either adult or embryonic erythrocytes. The human α-like globin gene cluster contains an α-like sequence which does not encode any known globin polypeptide (41). Nucleotide sequence data suggest that it also does not encode a functional polypeptide.

The origin of pseudogenes is unknown but some of them may arise following gene duplication events. Since only one member of the resultant pair is required to function, a deleterious mutation would not result in the loss of that sequence from the gene pool. For β^X it is difficult to estimate when the postulated gene duplication event occurred. Based simply on amino acid sequence changes between the β^X and β^A genes (51 amino acid differences) the duplication occurred approximately 250 million years ago (10). However, this is undoubtedly an overestimate since the selective forces acting on an expressed sequence may no longer act on β^X to maintain a functional sequence. In this case one would expect the divergence to be more rapid. By comparison, β^A and β^C diverged approximately 70 million years ago and γ and the primordial β diverged about 120 million years ago.

The mutational process acting upon the β^X gene and responsible for most of the alterations depicted in Fig. 3 appears to be one of point mutations and not of insertions or deletions. This is surprising in light of the finding by Konkel et al (15) that non-coding sequences, such as flanking and IVS-2 sequences, diverged rapidly by an insertion/deletion mechanism; therefore one might expect a similar process to be acting upon a defective β globin gene such as β^X. We also find that the

"wobble" position of β^X has not diverged more rapidly than other codon positions, but this is also true for comparisons of β^A/β^C (unpublished observations) and β^{maj}/β^{min} (15). In spite of the defective nature of this gene, it remains a possibility that the β^X sequences have been preserved to serve some other role, possibly regulatory in nature.

ACKNOWLEDGEMENTS

We are very grateful for the excellent technical assistance of Mr. Donald VanLeeuwen and Ms. Gail Sears. We are also grateful to Rose Alden for expert assistance in the preparation of this manuscript.

This work was supported by Grant PCM 76-80222 from the National Science Foundation and Grants AM 20119 and GM 10999 from the United States Public Health Service.

REFERENCES

1. Robbins, J., Rosteck, P., Jr., Haynes, J.R., Freyer, G., Cleary, M.L., Kalter, H.D., Smith, K. and Lingrel, J.B. (1979) J. Biol. Chem. **254**, 6187-6195.
2. Haynes, J.R., Rosteck, P., Jr., Schon, E.A., Gallagher, P.M., Burks, D.J., Smith, K. and Lingrel, J.B. (1980) J. Biol. Chem. **255**, 6355-6367.
3. Blattner, F.R., Williams, B.G., Blechl, A.E., Thompson, K.D., Faber, H.R., Furlong, L.A., Grunwald, D.J., Kiefer, D.O., Moore, D.D., Schumm, J.W., Sheldon, E.L., Smithies, D. (1977) Science **196**, 161-169.
4. Maniatis, T., Hardison, R.C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G.K., Efstradiadis, A. (1978) Cell **15**, 687-701.
5. Maniatis, T., Jeffrey, A., and Van de Sande, J. (1975) Biochemistry **14**, 3783-3794.
6. Donelson, J.E. and Wu, R. (1972) J. Biol. Chem. **247**, 4654-4660.
7. Haynes, J.R., Rosteck, P., and Lingrel, J.B. (1980) Proc. Natl. Acad. Sci. USA (in press).
8. Maxam, A.M. and Gilbert, W. (1979) Methods in Enzymology (in press).
9. Sanger, F. and Coulson, A.R. (1978) FEBS Lett. **87**, 107-110.
10. Dayhoff, M.O. and Eck, R.V. (1968) in Atlas of Protein Sequence and Structure, pp 154-158, The National Biomedical Research Foundation, Silver Spring, MD.
11. Breathnach, R., Benoist, C., O'Hare, K., Gannon, F. and Chambon, P. (1978) Proc. Natl. Acad. Sci. USA **75**, 4853-4857.
12. Max, E.E., Seidman, J.G., and Leder, P. (1979) Proc. Natl. Acad. Sci. USA **76**, 3450-3454.
13. Goldberg, M. (1979) Ph.D. thesis, Stanford University, Stanford, California.
14. Konkel, D.A., Tilghman, S.M. and Leder, P. (1978) Cell **15**, 1125-1132.
15. Konkel, D.A., Maizel, J.V. and Leder, P. (1979) Cell **18**, 865-873.
16. Nishioka, Y. and Leder, P. (1979) Cell **18**, 875-882.
17. Gannon, F., O'Hare, K., Perrin, F., LePennec, J.P., Benoist, C., Cochet, M., Breathnach, R., Royal, A., Garapin, A., Cami, B. and Chambon, P. (1979) Nature **278**, 428-434.
18. Lai, E.C., Stein, J.P., Catterall, J.F., Woo, S.L.C., Mace, M., Means, A.R. and O'Malley, B.W. (1979) Cell **18**, 829-842.
19. Tsujimoto, Y. and Suzuki, Y. (1979) Cell **18**, 591-600.
20. Ziff, E.B. and Evans, R.M. (1978) Cell **15**, 1463-1475.

21. Lomedico, P., Rosenthal, N., Efstratiadis, A., Gilbert, W., Kolodner, R. and Tizard, R. (1979) Cell **18**, 545-558.
22. van Ooyen, A., van den Berg, J., Mantei, N. and Weissmann, C. (1979) Science **206**, 337-344.
23. Hardison, R.C., Butler, E.T., III, Lacy, E., Maniatis, T., Rosenthal, N. and Efstratiadis, A. (1979) Cell **18**, 1285-1297.
24. Pribnow, D.(1975) Proc. Natl. Acad. Sci. USA **72**, 784-788.
25. Grosschedl, R. and Birnstiel, M.L. (1980) Proc. Natl. Acad. Sci. USA **77**, 1432-1436.
26. Weaver, R.F. and Weissmann, C. (1979) Nucl. Acids Res. **7**, 1175-1193.
27. Haynes, J.R., Kalb, V.F., Rosteck, P., Jr., and Lingrel, J.B. (1978) FEBS Lett. **91**, 173-177.
28. Luse, D.S. and Roeder, R.G. (1980) Cell(in press).
29. Proudfoot, N.J. and Brownlee, G.G. (1976) Nature **263**, 211-214.
30. Richardson, C., Capello, J., Cochran, M.D., Armentrout, R.W. and Brown, R.D. (1980) Develop. Biol.(in press).
31. Proudfoot, N.J. and Longley, J.I. (1976) Cell **9**, 733-746.
32. Lomedico, P.T., Rosenthal, N., Efstratiadis, A., Gilbert, W., Kolodner, R. and Tizard, R. (1979) Cell **18**, 545.
33. Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. and Dodgson, J. (1989) Cell **20**, 555-566.
34. Luse, D.S. and Roeder, R.G.(personal communication).
35. Hardison, R.C., Butler, E.T., III, Lacy, E., Maniatis, T., Rosenthal, N. and Efstratiadis, A. (1979) Cell **18**, 1285-1297.
36. Lacy, E., Hardison, R.C., Quon, D. and Maniatis, T. (1979) Cell **18**, 1273-1283.
37. Fritsch, E.F., Lawn, R.M. and Maniatis, T. (1980) Cell **19**, 959-972.
38. Jahn, C.L., Hutchinson, C.A. III, Phillips, S.J., Weaver, S., Haigwood, N.L., Voliva, C.F. and Edgell, M.H. (1980) Cell(in press).
39. Nishioka, Y., Leder, A. and Leder, P. (1980) Proc. Natl. Acad. Sci. USA **77**, 2806-2809.
40. Vanin, E.F., Goldberg, G.I., Tucker, P.W., and Smithies, O.(1980) Nature(in press).
41. Lauer, J., Shen, C.J. and Maniatis, T. (1980) Cell **20**, 119-130.