Proc. Natl. Acad. Sci. USA
Vol. 77, No. 12, pp. 7054–7058, December 1980
Biochemistry

Cloning and complete nucleotide sequence of human 5′-α-globin gene

(λgtWES phage/DNA sequence analysis)

STEPHEN A. LIEBHABER, MICHEL J. GOOSSENS, AND YUET WAI KAN

Howard Hughes Medical Institute Laboratory and Department of Medicine, University of California, San Francisco, California 94143

Communicated by Choh Hao Li, September 16, 1980

ABSTRACT We have cloned one of the two human α-globin genes and report its complete nucleotide sequence. The gene is 832 base pairs (bp) long from the 5′-cap site to the 3′-polyadenylylation site. The amino acid coding sequences are separated into three segments (exons) by two short (117 and 140 bp) intervening sequences. Highly conserved regions are identified in the 5′-flanking region, intron–exon junctions, and 3′ noncoding regions that may have functional significance.

The genes coding for the human globin polypeptide chains are organized into two tightly linked clusters. The α-gene cluster, which contains the two coexpressed adult α-globin genes, two embryonic α-globin-like genes (ζ-globin), and one or more nonfunctional α-globin-like gene remnants (pseudogenes), is located on chromosome 16 (1, 2). The β-globin gene cluster, which contains the adult β-globin gene, the minimally expressed δ-globin gene, two coexpressed fetal γ-globin genes, the embryonic ϵ-globin genes, and β-globin related pseudogenes, is found on chromosome 11 (3–5). The genes within each family are expressed according to a strictly observed ontologic schedule, and the quantitative expression of genes from each of these two families is strictly balanced and coordinated (6). Such intrachromosomal and interchromosomal control of gene expression implies an elaborate set of regulatory mechanisms. Some insight into these control systems may be afforded by the study of dysfunctional globin genes. A detailed structural knowledge of the normal globin genes predicate such studies. Comparison of the structure of normal genes may also yield possible structure–function relationships if conserved sequences in strategic areas can be identified. We describe the cloning of a normal human α-globin gene and present its primary structure.

MATERIALS AND METHODS

DNA Isolation and Fractionation. Liver tissue was obtained from a second-trimester Caucasian fetus aborted for psychosocial indications. The fetus showed no obvious anatomic abnormalities, and there was no abnormal genetic lineage. Highly polymerized DNA was prepared as described (7). The DNA was digested with *Eco*RI (New England BioLabs) according to the manufacturer's recommendations and fractionated according to size by discontinuous horizontal agarose gel electrophoresis (8). An aliquot of each fraction was assayed for globin sequences by analytic agarose gel electrophoresis followed by filter hybridization (9) to a globin cDNA probe labeled to 10^8 cpm/μg by nick translation (10).

Recombinant Phage Construction. The arms of the EK2 vector λgtWES·λB (11) were separated from the central B fragment by RPC-5 chromatography (12). The arms were ligated to DNA from the 3- to 4.5-kilobase (kb) fraction of the *Eco*RI digest and the resultant recominant DNA was packaged in phage coat proteins as described by Blattner *et al.* (13). The packaging and subsequent cloning were performed in accordance with the National Institutes of Health guidelines for recombinant DNA research.

Plaque Screening. *In situ* hybridization of phage plaques transferred to nitrocellulose filters was performed as described by Benton and Davis (14); hybridization to the globin cDNA probe was performed on duplicate filters. Positive plaques were purified by low-density plating, and their DNA was subsequently purified from lysates of *Escherichia coli* DP50supF (15). DNA from each positive clone was digested with *Eco*RI, run on a 1.0% agarose gel, and analyzed by Southern hybridization using nick-translated cDNA probes specific for α-, β-, or γ-globin [JW101, -102, and -151 plasmids (16), respectively, generously supplied by B. Forget].

Mapping. The DNA insert containing the α-globin gene was digested with *Eco*RI, and the insert was subsequently separated from the phage arms on a 0.7% agarose gel and recovered by electroelution. The isolated DNA was digested with various restriction endonucleases under conditions recommended by the manufacturers (New England BioLabs, Beverly, MA, and Bethesda Research Laboratories, Rockville, MD). Bands were visualized under UV light after staining with ethidium bromide, and the α-globin-containing fragments were identified by Southern hybridization using specific α-globin probes. The orientation of the gene was determined by using probes specific for the 5′ and 3′ ends of the α-globin structural region (17).

Sequence Determination. The nucleotide sequence of the α-globin gene and its flanking areas was approached as diagrammed in Fig. 1. Appropriately digested fragments were labeled at the 5′ end with polynucleotide kinase (BRL) and [γ-^{32}P]ATP (Amersham, 3000 Ci/mmol; 1 Ci = 3.7 × 10^{10} becquerels) after dephosphorylation of the DNA with bacterial alkaline phosphatase (Bethesda Research Laboratories) or at the 3′ end by *E. coli* DNA polymerase I [Klenow fragment (Bethesda Research Laboratories)] and the appropriate [α-^{32}P]dNTP (3000 Ci/mmol). These labeled fragments were isolated on 5% acrylamide gels (20) and then either strand separated (21) or digested again with another endonuclease followed by isolation of end-labeled pieces on acrylamide gels. Partial chemical modification and cleavage of bases was done as described by Maxam and Gilbert (21) with one major reaction modification in the adenosine reaction (22). The thin gel system of Sanger and Coulson (23) was used with a urea concentration of 8 M and acrylamide concentration of 6.6–15%, depending upon the desired area of optimal reading. All areas were subjected to sequence determination at least three times and, where indicated (areas of ambiguous reading or reading at variance from the literature), on both strands as well.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U. S. C. §1734 solely to indicate this fact.

Abbreviations: kb, kilobase(s); IVS, intervening sequence(s).

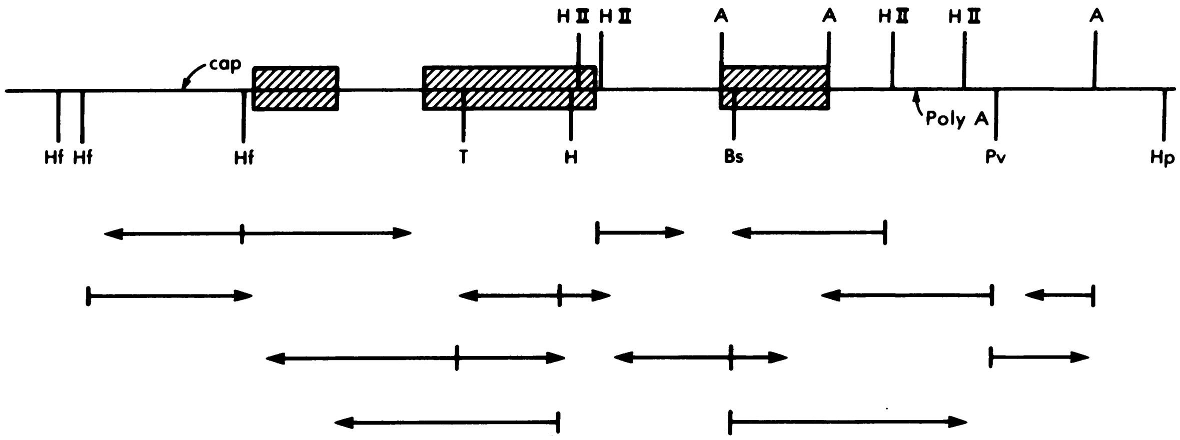


FIG. 1. Sequencing strategy for the human α-globin gene. The restriction map represents a 1.35-kb segment of DNA containing the 5′ α-globin gene. The transcribed region is defined by the 5′ mRNA cap site (cap) and the 3′ mRNA polyadenylylation site (Poly A) as previously determined (18, 19). The coding areas are shown as hatched boxes. Only restriction sites actually used in preparing fragments for sequencing are shown. Fragments were labeled with ^{32}P at sites indicated by the short vertical lines, and the distance of sequence determination on each fragment is indicated by the horizontal arrows.

RESULTS AND DISCUSSION

Cloning the α-Globin Gene. Both α-globin genes are normally located together on a 23-kb *Eco*RI fragment (17, 24). However, during an attempt to clone β- and γ-globin gene fragments from a purified 3- to 4.5-kb fraction of *Eco*RI digested normal human DNA recombined into λgtWES·λB, we detected a signal that was reproducibly positive with a total reticulocyte cDNA probe but negative with specific β- and γ-globin probes. This plaque gave a strongly positive signal when hybridized with a pure α-globin cDNA probe (JW101). DNA from the positive clone was isolated from the λgtWES arms and analyzed by a combination of endonuclease digestions and Southern blotting. Orientation of the α-globin gene was established by using probes specific for the 5′ and 3′ ends of the gene on Southern blots of the restriction digests (17). Although the DNA fragment containing the α-globin gene was cloned from a 3- to 4.5-kb fraction, the DNA insert isolated was 7.4 kb. It is likely that, during the cloning procedure, two previously unassociated DNA fragments, one of which contained the α-globin gene, ligated *in vitro*. This is consistent with the results achieved by aligning the restriction sites flanking the gene in the insert with those flanking the natural α-globin genes (Fig. 2). Such an alignment of restriction sites also identified the cloned gene as the more 5′ of the two adjacent α-globin genes.

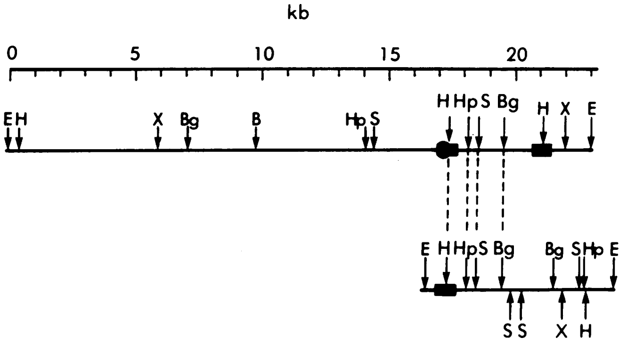


FIG. 2. Identification of the cloned α-globin gene as the 5′ α-globin gene. The restriction map of the two adjacent α-globin genes as found on the 23-kb *Eco*RI restriction fragment of normal genomic DNA (17, 24) is compared with the map of the 7.6-kb gene fragment cloned in the present experiment. The identification of the cloned α-globin gene as the more 5′ can be made by lining up the unique restriction sites flanking the genes (dashed vertical lines). The 3′ half of the cloned fragment is apparently unrelated to the α-gene complex (see text).

It is uncertain how a small (approximately 3 kb) DNA fragment containing the 5′ α-globin gene was generated, but a Southern analysis of the original *Eco*RI-digested DNA prior to cloning identified a faint band hybridizing to α-globin probe

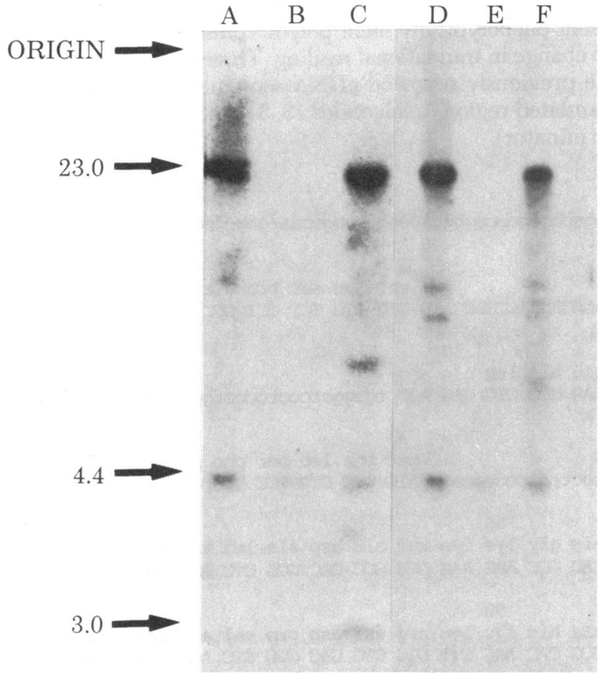


FIG. 3. Autoradiographs of *Eco*RI-digested human DNA hybridized with an α-globin cDNA probe (JW101). The numbers at the left are fragment length in kb as determined by double-stranded DNA size markers. Tracks: A and D, DNA from a normal blood donor; B and E, DNA from α-thalassemic hydrops abortus; C and F, DNA used for the cloning experiments (isolated from the liver of a normal abortus). The described α-globin gene was cloned from the digest analyzed in track C. This pattern is compared to that in track A which contains DNA digested with a different (commercial) lot of *Eco*RI and in tracks E and F containing DNA digested in parallel with a third (noncommercial) lot of *Eco*RI. The major band at 23 kb contains both α-globin genes (17, 24). The identity of the consistent band at 4.4 kb, seen in all samples including hydrops, is undetermined. The inconsistent minor banding varied with enzyme lots and was not uniquely associated with the particular DNA used. The α-globin gene cloned here probably originates from the 3.0-kb band seen in track C.

in the 3-kb size range (Fig. 3). Two subsequent Southern analyses of the same DNA digested with two other preparations of *Eco*RI yielded either no extra α-globin gene-containing bands (not shown) or faint bands with a different distribution from that seen in the original digest. We conclude that the α-globin-containing fragment that we cloned was generated by a contaminating minor endonuclease activity and was isolated fortuitously after ligation to an unassociated DNA fragment. Because the most likely contaminating activity in *Eco*RI is *Eco*RI*, the newly generated *Eco*RI site at the 5′ end of the insert probably can be attributed to this enzyme and not an aberrant *Eco*RI site in the native DNA.

Nucleotide Sequence of the α-Globin Gene. Fig. 4 shows the nucleotide sequence of the 5′ human α-globin gene along with 98 bp of 5′ flanking sequences and 152 bp of 3′ flanking sequences. The gene is 832 bp long from the 5′ cap site to the 3′ polyadenylylation site [both as defined by previous mRNA sequencing (18, 19)]. The amino acid coding sequences are separated into three segments (exons) by two intervening sequences (introns).

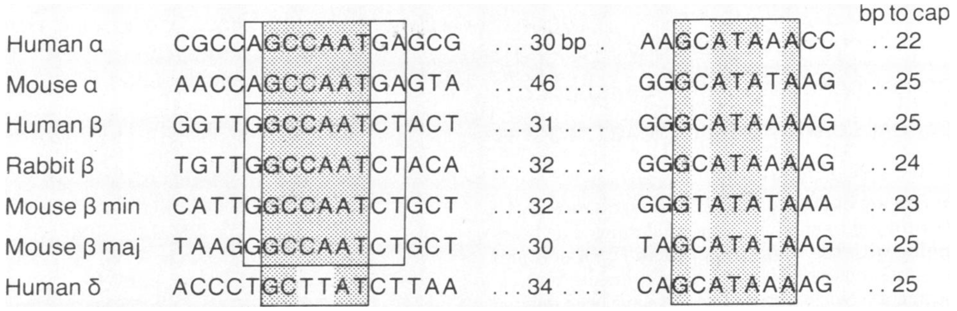
The nucleotide sequence of the three exons encodes the normal α-globin amino acid sequence (25). There are five base differences in the structural (amino acid coding) region from that previously derived from the sequence of α-globin mRNA (26, 27). These are at codons 13, 17, 54, 60, and 123. Our sequence of AAG at codon 60 predicts the appropriate amino acid (lysine). Assuming these differences are not due to the infidelity of reverse transcriptase (28) or errors in the cDNA sequence determination, the remaining four base differences may represent phenotypically silent polymorphisms because they cause no change in translational reading. Three base differences from the previously reported cDNA sequence occur in the 3′ untranslated region (nucleotides 18, 31, and 33 bp after the UAA terminator).

5′ Flanking Area. A 98-bp sequence was determined proximal to the mRNA cap site. Within this area, we found two segments homologous in sequence to other globin genes (Fig. 5). Located 24 bp 5′ to the cap site is a 7-base stretch, G-C-A-T-A-A-A, whose position and composition are similar to those in other globin and nonglobin eukaryotic genes. By analogy to the bacterial Pribnow box (T-A-T-Pu-A-T-G-T) (34, 35) this area has been postulated to be a possible polymerase promotor site. The center (T) of this conserved sequence (A-T-A box) is located 29–30 bp from the cap site in the previously studied globins. This distance, three turns of a double helix, has a postulated steric role in the interaction of polymerase with the transcription initiation site (36). However, because the distance from the center of the A-T-A box to the origin of transcription is 3 bp less (27 bp) than 30 bp in human α-globin and 2 more (32 bp) in chicken ovomucoid, this relationship may not be stringent. There is a 12-bp stretch (C-A-A-T box) 33 bp proximal to this A-T-A box that is identical to a 12-bp stretch 51 bp proximal to the mouse α-globin A-T-A box (Fig. 5). The internal six bases of this sequence match with a stretch 37 bp proximal to the human β-globin, rabbit β-globin, and mouse β-globin A-T-A box.

It should be noted that the β genes have further homology among themselves extending on either side of the 6-bp stretch. Four of the central six bases are also present at the same position in human δ-globin, with some divergence from the α and β in the surrounding bases. Thus, this area of six bases 5′ to the A-T-A box is highly conserved among all the globin genes with further α-specific or β-specific homology surrounding it. The prokaryotic Pribnow box is preceded by a similar homology area centered −35 bp from the AUG (37). Although the function of this region is not fully known, it contains certain sequences essential for efficient initiations in prokaryotes (38). Whether any

```
aggccgcgccccgggctccgcgccagccaatgagcgccgcccggccgggcgtgcccccgcgccccaagcataaaccctggcgcgctcgcggcccggcACTCTTCTGGTCCCCACAGACTC

                  1                                   10                                      20
                  val leu ser pro ala asp lys thr asn val lys ala ala trp gly lys val gly ala his ala gly glu tyr gly ala
AGAGAGAACCCACCATG GTG CTG TCT CCT GCC GAC AAG ACC AAC GTC AAG GCC GCC TGG GGT AAG GTC GGC GCG CAC GCT GGC GAG TAT GGT GCG

            30
glu ala leu glu arg
GAG GCC CTG GAG AGG tgaggctccctcccctgctccgacccgggctcctcgcccgcccggacccacaggccaccctcaaccgtcctggccccggacccaaaccccacccctcactc

                 32                              40                                      50
                 met phe leu ser phe pro thr thr lys thr tyr phe pro his phe asp leu ser his gly ser ala gln val lys gly
tgcttctccccgcagg ATG TTC CTG TCC TTC CCC ACC ACC AAG ACC TAC TTC CCG CAC TTC GAC CTG AGC CAC GGC TCT GCC CAA GTT AAG GGC

        60                                      70                                      80
his gly lys lys val ala asp ala leu thr asn ala val ala his val asp asp met pro asn ala leu ser ala leu ser asp leu his
CAC GGC AAG AAG GTG GCC GAC GCG CTG ACC AAC GCC GTG GCG CAC GTG GAC GAC ATG CCC AAC GCG CTG TCC GCC CTG AGC GAC CTG CAC

        90                                  99
ala his lys leu arg val asp pro val asn phe lys
GCG CAC AAG CTT CGG GTG GAC CCG GTC AAC TTC AAG gtgagcggcgggccgggagcgatctgggtcgaggggcgagatggcgccttcctctcagggcagaggatcacgc

                                                                    100                                     110
                                                                    leu leu ser his cys leu leu val thr leu ala ala his
gggttgcgggaggtgtagcgcaggcggcggcgcggcttgggccgcactgaccctcttctctgcacag CTC CTA AGC CAC TGC CTG CTG GTG ACC CTG GCC GCC CAC

                            120                                     130                                     140
leu pro ala glu phe thr pro ala val his ala ser leu asp lys phe leu ala ser val ser thr val leu thr ser lys tyr arg OC
CTC CCC GCC GAG TTC ACC CCT GCG GTG CAC GCT TCC CTG GAC AAG TTC CTG GCT TCT GTG AGC ACC GTG CTG ACC TCC AAA TAC CGT TAA

GCTGGAGCCTCGGTAGCCGTTCCTCCTGCCCGCTGGGCCTCCCAACGGGCCCTCCTCCCCTCCTTGCACCGGCCCTTCCTGGTCTTTGAATAAAGTCTGAGTGGGCGGCagcctgtgtgtg

cctgggttctctctgtcccggaatgtgccaacaatggaggtgtttacctgtctcagaccaaggacctctctgcagctgcatggggctggggagggagaactgcagggagtatgggaggga

agctgaggtgggcctgctcaagagaaggtgctgaaccatcccctgtcctgagaggtgccagcctgcaggcagtggc
```

FIG. 4. Nucleotide sequence of the human α-globin gene. The nucleotide sequence of the coding strand is displayed 5′ to 3′. Bases found in the mature mRNA are shown in upper case; nontranscribed flanking sequences and the two intervening sequences are shown in lower case. Numbering refers to the encoded amino acids indicated above their respective codons.



			bp to cap
Human α	CGCCAGCCAATGAGCG	. . . 30 bp .	AAGCATAAACC . . 22
Mouse α	AACCAGCCAATGAGTA	. . . 46	GGGCATATAAG . . 25
Human β	GGTTGGCCAATCTACT	. . . 31	GGGCATAAAAG . . 25
Rabbit β	TGTTGGCCAATCTACA	. . . 32	GGGCATAAAAG . . 24
Mouse β min	CATTGGCCAATCTGCT	. . . 32	GGGTATATAAA . . 23
Mouse β maj	TAAGGGCCAATCTGCT	. . . 30	TAGCATATAAG . . 25
Human δ	ACCCTGCTTATCTTAA	. . . 34	CAGCATAAAAG . . 25

FIG. 5. Comparison of the 5′ flanking region of the genes coding for human α-globin, mouse α-globin (29), human β-globin (30), rabbit β-globin (31), mouse β-minor and β-major globin (32), and human δ-globin (33). Homologous sequences are stippled, and additional base homologies shared within the α-globin or β-globin subsets are included within the open boxes.

functional constraints have limited sequence divergence of this area in eukaryotes is yet to be determined.

Intervening Sequences (IVS). Several interesting observations can be made from the comparisons shown in Table 1. The base composition of the intron–exon junction is remarkably conserved among the globin genes. There is an identical pattern of splicing site ambiguity; A-G-G can be located on either side of the first IVS and A-G can be on either side of the second IVS. Furthermore, it is possible to arrange the splice point in each of these genes as indicated so that they conform to the general rule that IVS begin with G-T and end with A-G. In this case the second IVS is located between codons, and the first IVS interrupts a codon which is subsequently spliced together.

The site of intron insertion within the coding sequence is invariate *vis-a-vis* the secondary helical structure of the protein (helix positions B12 and G6). This may relate in some way to the hypothesis (39) that introns code for functionally distinct regions in a polypeptide (e.g., hemoglobin) (40) and that such a coding structure serves some positive evolutionary function.

A computer search of both IVS for possible secondary structure failed to yield dyads >4 bp or significant stem loop configurations. These negative findings are consistent with a similarly negative search for secondary structure in three of the ovalbumin IVS (41) and with the finding that large segments of the IVS in simian virus 40 genes can be deleted without affecting the accuracy of splicing (42). Recent evidence that the larger IVS in β-globin may be excised in steps has led to broadened requirements for site-specific splicing signals (43). Benoist *et al.* (41) proposed that such sites must contain the tetranucleotide A-G-G-T in order to correspond to the finding of G-T at the donor site and A-G at receptor sites of intron–exon junctions. A search for such a sequence revealed one centrally located in the larger (second) IVS (positions 83–86) of the α-globin gene. Because this sequence occurs elsewhere in the structural area (codons 16–17) as it does in ovalbumin, the sequence cannot alone signal a splice.

An internal 20-of-23-base match (87% homology) exists in IVS 1 between bases 41–63 and 75–98. By segmenting the entire first IVS into 34-bp stretches, a larger repeating unit can be constructed with a 23/34 (68%) match for two central segments (Fig. 6). Two separate sequences partially homologous to this IVS repeat unit were found in mouse α-globin IVS 1 with 63% homology (data not shown). Similar homologies were not found in IVS 2 of mouse and human α-globin, or in IVS 1 and 2 of human β-globin. A random match of this length is quite unlikely, and its absence from the second IVS of both human and mouse α-genes, as well as the first IVS of the human β-gene, makes it specific to this IVS 1. Whether the match has functional significance or is a residue of an earlier reduplication that formed the IVS is open to question. If it does indicate a series of reduplication events that formed this intron from an earlier small unit, it would be evidence that introns grow rather than shrink during evolution.

3′-Untranslated Area. The 3′ flanking area contained the highly conserved hexanucleotide sequence A-A-T-A-A-A found in eukaryotic messages between the terminator codon and the polyadenylylation site (44). Again, we failed to locate a significant secondary structural configuration in this area that might serve as a transcription termination or polyadenylylation signal.

G+C Content of the α-Globin Gene. The high G+C content (65%) of the structural α-globin sequence has been noted and commented on by others (27, 45). In the present sequence data we find an equally high content in the IVS and an even higher percentage (83%) in the 5′ flanking area, which is greater than 2 times the average value of the total human DNA of 39.5% (46). Such a high G+C content has a significant effect on the

Table 1. Splicing junctions of globin genes of known sequence

	IVS 1					IVS 2				
Human α	AGG	TGA	113	GCAGG	ATG	AAG	GTGAGC	141	ACAG	CTC
	30				31	99				100
Human β	AGG	TTG	130	TTAGG	CTG	AAG	GTGAGT	903	ACAG	CTC
	30				31	104				105
Human δ	AAG	TTG	128	TCAGG	GTG	AAG	GTGAGT	878	GCAG	CTC
	30				31	104				105
Mouse α	AGG	TGA	121	CCAGG	ATG	AAG	GTATGC	135	GCAG	CTC
	31				32	99				100
Mouse β	AGG	TTG	116	TTAGG	CTG	AGG	GTGAGT	650	ACAG	CTC
	30				31	104				105
Rabbit β	AGG	TTG	126	TCAGG	CTG	AGG	GTG	573	ACAG	CTC
	30				31	104				105

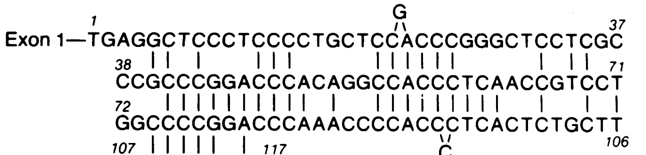

1
G
37
Exon 1—TGAGGCTCCCTCCCCTGCTCCACCCGGGCTCCTCGC
38
71
CCGCCCGGACCCACAGGCCACCCTCAACCGTCCT
72
GGCCCCGGACCCAAACCCCACCCTCACTCTGCTT
107
117
106
C
CTCCCCGCAGG—Exon 2

FIG. 6. Tandem repeats in IVS 1. The primary sequence of the globin gene IVS 1 is displayed to maximize base matching. The bases are numbered beginning with the first nucleotide in IVS 1 as 1 and the last base in this segment as 117. Vertical lines connect identical bases in sequential segments. Nucleotides 22 and 94 have been deleted as shown to maximize homology.

DNA duplex melting temperature, but whether this relates to any selective advantage in the α-globin gene is entirely speculative. The stability of hairpin loops in the 5′ flanking region is probably related to this high G+C content.

CONCLUSIONS

The primary structure of the human 5′ α-globin gene and its flanking areas has been completely determined. Analysis of this gene and comparison with other sequenced globin genes points out several highly conserved characteristics. Two homology regions (C-A-A-T and A-T-A boxes) occur in the 5′ flanking region. The similarity in position of these two regions with sequences of probable promotor activity in prokaryotes suggests a possible function. By determination of sequence in this area in naturally occurring mutants with a low level of expression (i.e., nondeletion α-thalassemia syndromes) and by inducing controlled mutations at strategic points in a normal gene and studying consequent effects upon transcription, this function could be tested. As with all of the globin genes whose sequences have been determined, the human α-globin gene has two introns that interrupt the coding sequence at a highly conserved position. It must be assumed that this conserved structure of the globin genes is maintained by as yet unrecognized evolutionary constraints, possibly relating to effective splicing activity. A search of the two introns revealed no secondary structure in the coding strand that might direct splicing activity, although conservation of a small number of nucleotides was again observed at intron–exon junctions.

Evidence of a large tandem repeat within the first IVS was found. Although this finding cannot be generalized to the other globins, and hence functional and evolutionary inferences are difficult, this large tandem repeat may indicate that the intron is enlarging by duplication.

1. Lauer, J., Shen, C. J. & Maniatis, T. (1980) *Cell* **20**, 119–130.
2. Deisseroth, A., Nienhuis, A., Turner, P., Velez, R., Anderson, W. F., Ruddle, F., Lawrence, J., Creagan, R. & Kucherlapati, R. (1977) *Cell* **12**, 205–218.
3. Lawn, R. M., Fritsch, E. F., Parker, R. C., Blake, G. & Maniatis, T. (1978) *Cell* **15**, 1157–1174.
4. Fritsch, E. F., Lawn, R. M. & Maniatis, T. (1978) *Nature (London)* **279**, 598–603.
5. Deisseroth, A., Nienhuis, A., Lawrence, J., Giles, R., Turner, P. & Ruddle, F. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 1456–1460.
6. Bunn, H. F., Forget, B. G. & Ranney, H. M. (1977) *Human Hemoglobins* (Saunders, Philadelphia), pp. 101–140.
7. Gross-Bellard, M., Oudet, P. & Chambon, P. (1973) *Eur. J. Biochem.* **36**, 32–38.
8. Tilghman, S. M., Tiemeier, D. C., Polsky, F., Edgell, M. H., Seidman, J. G., Leder, A., Enquist, L. W., Norman, B. & Leder, P. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 4406–4410.
9. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
10. Maniatis, T., Jeffrey, A. & Kleid, D. G. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 1184–1188.
11. Tiemeier, D., Enquist, L. & Leder, P. (1976) *Nature (London)* **263**, 526–527.
12. Hardies, S. C. & Wells, R. D. (1976) *Proc. Natl. Acad. Sci. USA* **73**, 3117–3121.
13. Blattner, F. R., Blechl, A. E., Denniston-Thompson, K., Faber, H. E., Richards, J. E., Slightom, J. L., Tucker, P. W. & Smithies, O. (1978) *Science* **202**, 1279–1284.
14. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
15. Maniatis, T., Hardison, R. C., Lacy, E., Lauer, J., O'Connell, C., Quon, D., Sim, G. K. & Efstratiadis, A. (1978) *Cell* **15**, 687–701.
16. Wilson, J. T., Wilson, L. B., deRiel, J. K., Villa-Komaroff, L., Efstratiadis, A., Forget, B. G. & Weissman, S. M. (1978) *Nucleic Acids Res.* **5**, 563–581.
17. Embury, S. H., Lebo, R. V., Dozy, A. M. & Kan, Y. W. (1978) *J. Clin. Invest.* **63**, 1307–1310.
18. Baralle, F. E. (1977) *Cell* **12**, 1085–1095.
19. Proudfoot, N. J. & Longley, J. I. (1976) *Cell* **9**, 733–746.
20. Peacock, A. C. & Dingman, C. W. (1967) *Biochemistry* **6**, 1818–1827.
21. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
22. Cooke, N. E., Coit, D., Weiner, R. I., Baxter, J. D. & Martial, J. A. (1980) *J. Biol. Chem.* **255**, 6502–6510.
23. Sanger, F. & Coulson, A. R. (1978) *FEBS Lett.* **87**, 107–110.
24. Orkin, S. H. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 5950–5954.
25. Dayhoff, M. O. (1972) *Atlas of Proteon Sequence and Structure, 1969* (National Biomedical Research Foundation, Washington, DC), Vol. 5, p. D56.
26. Wilson, J. T., deRiel, J. K., Forget, B. G., Marotta, C. A. & Weissman, S. M. (1977) *Nucleic Acids Res.* **4**, 2353–2368.
27. Wilson, J. T., Wilson, L. B., Reddy, V. B., Cavallesco, C., Ghosh, P. K., deRiel, J. K., Forget, B. G. & Weissman, S. M. (1980) *J. Biol. Chem.* **255**, 2807–2815.
28. Gopinathan, K. P., Weymouth, L. A., Kunkel, T. A. & Loeb, L. A. (1979) *Nature (London)* **278**, 857–859.
29. Nishioka, Y. & Leder, P. (1979) *Cell* **18**, 875–882.
30. Lawn, R. M., Efstratiadis, A., O'Connell, C. & Maniatis, T. (1980) *Cell* **21**, 647–653.
31. Van Ooyen, A., van den Berg, J., Mantei, N. & Weissmann, C. (1979) *Science* **206**, 337–344.
32. Konkel, D. A., Tilghman, S. M. & Leder, P. (1978) *Cell* **15**, 1125–1132.
33. Spritz, R. A., deRiel, J. K., Forget, B. & Weissman, S. (1980) *Cell* **21**, 639–647.
34. Schaller, H., Gray, C. & Herrmann, K. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 737–741.
35. Pribnow, D. (1975) *Proc. Natl. Acad. Sci. USA* **72**, 784–788.
36. Konkel, D. A., Maizel, J. V., Jr. & Leder, P. (1979) *Cell* **18**, 865–873.
37. Takanami, M., Sugimoto, K., Sugisaki, H. & Okamoto, T. (1976) *Nature (London)* **260**, 297–302.
38. Rosenberg, M. & Court, D. (1979) *Annu. Rev. Genet.* **13**, 319–353.
39. Gilbert, W. (1978) *Nature (London)* **271**, 501.
40. Craik, C. S., Buchman, S. R. & Beychok, S. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 1384–1388.
41. Benoist, C., O'Hare, K., Breathnach, R. & Chambon, P. (1980) *Nucleic Acids Res.* **8**, 127–142.
42. Thimmappaya, B. & Shenk, T. (1979) *J. Virol.* **30**, 668–673.
43. Kinniburgh, A. J. & Ross, J. (1979) *Cell* **17**, 915–921.
44. Proudfoot, N. J. & Brownlee, G. G. (1976) *Nature (London)* **263**, 211–214.
45. Forget, B. G., Wilson, J. T., Wilson, L. B., Cavallesco, C., Reddy, V. B., deRiel, J. K., Biro, A. P., Ghosh, P. K. & Weissman, S. M. (1979) *Cellular and Molecular Regulation of Hemoglobin Switching* (Grune & Stratton, New York), pp. 569–591.
46. Sober, H. A. (1968) *Handbook of Biochemistry* (The Chemical Rubber Co., Cleveland, OH), pp. 1–11.