Structural and evolutionary analysis of the two chimpanzee α -globin mRNAs

Stephen A. Liebhaber¹ and Kathi A. Begley²

Howard Hughes Medical Institute Laboratory, and Division of Molecular Hematology, Department of Medicine, University of California, San Francisco, CA 94143, USA

Received 15 August 1983; Revised and Accepted 8 November 1983

ABSTRACT

Two distinct α -globin mRNAs were detected in chimpanzee reticulocyte mRNA using a primer extension assay. DNA copies of these two mRNAs were cloned in the bacterial plasmid pBR322, and their sequence was determined. The two α -globin mRNAs have obvious structural homology to the two human α -globin mRNAs, $\alpha 1$ and $\alpha 2$. Comparison of the two chimpanzee α -globin mRNAs to each other and to their corresponding human counterparts revealed evidence of a recent gene conversion in the human α -globin complex and a marked heterogeneity in the rate of structural divergence within the α -globin gene.

INTRODUCTION

The structural evolution of the α -globin gene appears to be mediated by an interplay of three molecular mechanisms: gene duplication, point mutation, and gene correction. Duplicated α -globin loci have been structurally mapped in chickens (1), mice (2), goats (3), apes (4) and man (5-7). Most humans have two α -globin genes ($\alpha 1$ and $\alpha 2$) per haploid genome (5-7), although individuals with one (6-8) and three genes (9-11) have been identified in certain populations. Similarly, the haploid number of α -globin genes in apes is usually two, although evidence for subpopulations of chimpanzee with one and three has been presented (4,12). The structure of the α -globin gene has diverged between species by accumulation of point mutations while the structures of the adjacent copies within each individual species remain highly homologous. This is most clearly demonstrated in studies of the primate α -globin genes (4,12). This maintenance of homology among the duplicated α -globin genes within a species implies parallel or concerted evolution of these linked genes. This pattern of genetic interaction seems to be shared by a number of linked (3,13,14), as well as unlinked (15), gene families. Although gene conversion, or non-reciprocal exchange of DNA sequences, has been demonstrated experimentally between duplicated genes in

yeast (16), the underlying mechanisms remain undefined. Each of the three molecular mechanisms described above therefore appears to play an important role in the evolution of the α -globin gene family.

We report the molecular cloning and sequence analysis of the α 1- and α 2-globin mRNAs of the chimpanzee and compare these data to the corresponding α -globin gene sequences in man (17-19). Such a structural comparison of genes which have diverged from a common precursor over a relatively short and well-defined evolutionary period (5 million years) can facilitate detection of recent and frequent events in molecular evolution. Our data suggests that a gene conversion event has occurred between the two human α -globin genes 5' to the UAA termination codon at a time subsequent to the divergence of human from chimpanzee lineage. We also find, independent of gene conversion events, a substantially lower rate of point mutation in this region than in the 3' noncoding sequences in both the human and chimpanzee α -globin genes. The net result of these two effects is a marked heterogeneity in the rate of structural divergence within the α -globin gene.

MATERIALS AND METHODS

RNA

Chimpanzee (Pan troglodytes) RNA extracted from the blood of a phenylhydrazine treated individual was supplied by the Yerkes Regional Primate Center, Emory University, Atlanta, Georgia 30322. The RNA was prepared by selective lysis of erythrocytes (20) and acid precipitation of polysomes, followed by phenol extraction, according to published methods (21). This total polysomal RNA (kindly supplied by E. Zimmer) from chimpanzee 2 in Ref. 12 was used for subsequent procedures without further purification.

cDNA Synthesis

For first strand α -globin cDNA synthesis, 100 ug of total polysomal RNA was mixed with 16 ng of a 5'-³²P-labeled 31 base α -globin primer (10⁷CPM/ug). This mixture was then incubated with 600 U/ml avian myeloblastosis virus (A.M.V.) reverse transcriptase at 55°C for 45 minutes in conditions previously described (22). This primer is complementary to the terminal 9 bases of the 3' nontranslated region of human α -globin mRNA and the contiguous poly A tail. The reaction was terminated by extraction with phenol:chloroform:isoamyl alcohol (50:49:1) saturated with Tris HCl pH 7.8. The aqueous phase was ethanol precipitated, dissolved in 0.6 M NaOH and incubated at 37°C for 1 hour, neutralized with an equimolar quantity of acetic acid and concentrated from the exclusion peak of a Sephadex G-75 column by ethanol precipitation.

Two percent of this sample was assayed on a 5% polyacrylamide, 8 M urea gel to confirm full-length α -globin cDNA synthesis, and a second aliquot was digested with 24 units HaeIII (New England Biolabs) and run on the same gel to detect the presence of distinct α -globin cDNA species.

Second strand synthesis was self-primed by the hairpin loop of the first strand in a 70 μ l reaction volume containing 50 mM NaCl, 6 mM Tris HCl pH 7.4, 6 mM 2-mercaptoethanol, 2 mM $MgCl_2$, 300 μ M dATP, dTTP, dGTP, 90 μ M dCTP, 30 nM α - ^{32}P -dCTP (Amersham 2-3000 Ci/mM) in the presence of 6 units DNA polymerase I (*E. coli*, large fragment, Bethesda Research Labs). The reaction mix was heated to 65° for 5 minutes, then the DNA polymerase was added and incubated for 1 hour at 37°C. This reaction was terminated by phenol-chloroform extraction and the aqueous phase was applied to a G-75 Sephadex column pre-equilibrated with 1 mM Tris HCl pH 7.4. The exclusion peak was pooled, phenol and chloroform extracted, ethanol precipitated and the dissolved pellet digested in 50 μ l with 50 units of S_1 nuclease (Boehringer-Mannheim) at 37°C for 60 minutes in the buffer recommended by the distributor. The reaction was terminated by phenol extraction and ethanol precipitation then 3' poly dCMP tails were added in a 70 μ l reaction containing 20 units terminal deoxynucleoside transferase (TdT) (Ratliff, Los Alamos, New Mexico) in the recommended $CoCl_2$ buffer. Tailed DNA from this reaction was annealed to an equimolar concentration of Pst I digested and (dGMP)₈ tailed pBR322 (kindly supplied by Claude Benicourt) at a final plasmid concentration of 0.5 ng/ μ l in 1 M Tris (pH 7.5), 100 mM EDTA, 2.5 M NaCl by heating to 70°C for one minute and then cooling slowly overnight to a final temperature of 4°C.

Transformation

Competent (23) *E. coli* HB101 (24) were transformed with the annealing mix and plated on LB agar containing 15 μ g/ml of tetracycline. Colonies were replica plated and 44 Tet^R Amp^S colonies were selected for further evaluation.

Colony Screening and Characterization

Replica-plated colonies were screened in situ (25) using ^{32}P -end labeled probes specific for the 5' terminus and the 3' terminus of the α -globin cDNA. The 5'-terminus specific probe was the Ava I restriction fragment of the human α 1-globin gene subclone pRB α 1 (7) (generously contributed by T. Maniatis) which encompasses the Exon 1 coding region. The 3' specific probe was the 250 bp HaeIII x Ava I fragment from the same clone which encompasses the terminal 73 bp of Exon 3 of the α -globin gene. Plasmid DNA was isolated from 1 ml clarified lysates of all colonies positive with either of these two

probes, digested with Pst I then electrophoresed on 5% polyacrylamide gels to size the cDNA inserts. Clones hybridizing to the 3'-terminus probe were additionally digested with Rsa I (New England Biolabs) which specifically cleaves the 3' nontranslated region of the $\alpha 1$ (human) cDNA (18,19).

DNA Sequencing

Restriction sites with 5' protruding ends were dephosphorylated with bacterial alkaline phosphatase (BRL) and ^{32}P -end labeled in a 25 μl reaction containing 5% glycerol, 50 mM Tris pH 9.4, 10 mM MgCl_2 , 5 mM DTT and 200 uCi of [γ - ^{32}P]-ATP (Amersham 3000 Ci/mM) in the presence of 5 units of T4 polynucleotide kinase (BRL) at 37°C for 30 minutes. Alternatively, 3' recessed ends were labeled with *E. coli* DNA polymerase, Klenow fragment (BRL) and the appropriate [α - ^{32}P] dNTP (3000 Ci/mM) in the recommended buffer. 3' protruding ends were labeled with ^{32}P -cordycepin using TdT and buffers provided by New England Nuclear in kit form. The chemical degradation method of Maxam and Gilbert (26) was used for all DNA sequence determinations.

RESULTS AND DISCUSSION

Two distinct α -globin mRNAs are present in the chimpanzee reticulocytes.

Chimpanzee erythrocytes contain a single α -globin protein (27). Southern blot analysis of chimpanzee DNA reveals chromosomes with two α -globin loci (4) arranged identically to the $\alpha 1$ and $\alpha 2$ loci in human, as well as chromosomes with three α -globin genes. In humans, both α -globin genes, $\alpha 1$ and $\alpha 2$, are expressed (22,27,28). To determine if both these α -globin genes are also transcribed and processed to stable mRNA in the chimpanzee, we assayed reticulocytes for $\alpha 1$ - and $\alpha 2$ -globin mRNA by a previously described primer extension technique (22). A 31 base DNA fragment complementary to the 3' terminus of human α -globin mRNA was labeled at its 5' terminus and used to prime the synthesis of α -globin cDNA from chimpanzee reticulocyte RNA. Fig. 1, Lane 2, demonstrates that under the conditions described, full-length α -globin cDNA of approximately 560 bases is synthesized as the major product. It has been previously demonstrated in the case of the human α -globin mRNAs that the differences in the HaeIII restriction map can be used to differentiate $\alpha 1$ - from $\alpha 2$ -cDNA (22). HaeIII cleavage sites in the 3' noncoding region result in different size terminal ^{32}P -labeled fragments from $\alpha 1$ - and $\alpha 2$ -globin human cDNA. When ^{32}P primer extended chimpanzee α -globin cDNA is digested to completion with HaeIII and the reaction products are separated on a denaturing gel, two bands (besides the unincorporated 31 base primer) are present (Fig. 1, Lane 3). The smaller

Fig. 1: α -globin cDNA mapping of chimpanzee and human reticulocyte mRNA. mRNA was isolated from the peripheral reticulocytes of a phenylhydrazine treated chimpanzee and from peripheral reticulocytes of a human subject with a normal α -globin genotype ($\alpha^{\alpha}/\alpha^{\alpha}$). α -globin cDNA was synthesized from this mRNA by extension of a 32 P-end labeled 31 base pair primer complementary to the poly A tail and the last 9 bases of α -globin mRNA (22). The resultant end labeled cDNA was digested with HaeIII (BRL, Gaithersburg, Maryland). The labeled products were separated on an 8% polyacrylamide, 8 M urea gel and quantitated by soft laser densitometer (LKB). Lane 1, end-labeled digested HindIII PM2 molecular weight markers; Lane 2, 32 P-end labeled primer extended chimpanzee α -globin cDNA; Lane 3, HaeIII digested 32 P-end labeled primer extended chimpanzee α -globin cDNA; Lane 4, HaeIII digested 32 P-end labeled primer extended human α -globin cDNA. The bands specific for $\alpha 1$ - and $\alpha 2$ -globin mRNA, the full length α -globin cDNA, and the unincorporated primer are indicated.

band co-migrates with the 60 bp fragment specific for the human $\alpha 2$ -globin mRNA while the upper band of 83 bases is 12 bases shorter than the 95 base specific fragment specific for human $\alpha 1$ -globin mRNA. An C to G change in the human $\alpha 1$ mRNA at position +52 would create a HaeIII site at this position and result in the observed 83 base 3' terminal fragment (see below). This data implies that two α -globin mRNA species are present in chimpanzee reticulocyte mRNA. We tentatively designated them as $\alpha 1$ and $\alpha 2$ because of the similarity in the HaeIII restriction patterns to human (Fig. 1, Lane 4).

Since the α -globin product of both α -globin gene loci, $\alpha 1$ and $\alpha 2$, are identical in both humans and chimpanzee (27), relative levels of expression from the two loci cannot be distinguished under normal circumstances. It is possible, however, to distinguish expression of the two α -globin loci at the level of mRNA. In human reticulocytes and bone marrow, there is a three-fold excess of $\alpha 2$ -globin mRNA as measured by primer extension ($\alpha 1/(\alpha 1 + \alpha 2) = 26\%$) (22). A densitometric analysis of Figure 1, Lane 3, using the same technique reveals a four-fold excess of $\alpha 2$ -globin mRNA ($\alpha 1/(\alpha 1 + \alpha 2) = 18\%$) which is significantly below human values (22). This difference may result from technical factors, a gene dosage effect in this particular chimpanzee individual, or a real difference in relative expression of the two α -globin loci in chimpanzee and man. Since the primary sequence of the terminal 9 bases of chimpanzee α -globin mRNA has not been directly determined, it is possible that a minor mismatch between the primer and the $\alpha 1$ mRNA may result in less than proportionate $\alpha 1$ -globin cDNA synthesis. A previously noted pattern of sequence conservation in this region of the human α -globin gene (19) which appears to be paralleled in the chimpanzee (see below) and the highly stringent conditions of our reverse transcriptase reaction make this possibility less likely. The α -globin genomic map of the presently studied chimpanzee is $\alpha\alpha\alpha/\alpha\alpha$ (Ref. 12 and B. Chapman, personal communication). A human with a $\alpha\alpha\alpha/\alpha\alpha$ genotype has previously been determined to have four-fold excess of $\alpha 2$ -globin mRNA ($\alpha 1/(\alpha 1 + \alpha 2) = 18\%$). Although these results would imply similarity in α -globin expression ($\alpha 1$ relative to $\alpha 2$) between man and chimpanzee, direct comparison cannot be firmly made for a number of reasons. Whereas the extra α -globin gene on the human triple α chromosome is $\alpha 2$ (5-7,9), and limited restriction enzyme mapping analysis shows an identical pattern between the chimpanzee (12) and human (9) triple α chromosome, the identity of the third α -globin gene on the chimpanzee triple α chromosome as $\alpha 2$ is not yet directly established. In addition, expression of the α -globin genes on the two α -globin gene chromosome of this chimpanzee may not be representative of normal due to an unusual deletion in the α -globin gene complex beginning 5' to the $\alpha 2$ -globin gene (B. Chapman, personal communication). Therefore, conclusions regarding the relative expression of the two α -globin genes in human versus chimpanzee must await further structural analysis of the chimpanzee α -globin complex and assay of $\alpha 1:\alpha 2$ -globin mRNA levels amongst chimpanzees with a variety of genotypes.

Cloning the two chimpanzee α -globin cDNAs.

α -globin cDNA was synthesized by priming reverse transcription of reticulocyte RNA with a 31 base α -globin specific primer. This cDNA primes the synthesis of two α -globin species: $\alpha 1$ and $\alpha 2$ (Fig. 1, Lane 3). After second strand cDNA synthesis and S_1 trimming, the cDNA was 3' tailed with dCTP and annealed into Pst I digested and dGMP tailed pBR322. Clones containing α -globin were identified by in situ hybridization using ^{32}P -end labeled probes specific for the 5' end and 3' ends of the α -globin cDNA. Plasmid minipreps of all positive colonies were prepared and digested with Pst I to determine recombinant insert sizes. The α -globin cDNA inserts containing 3' sequences were also digested with Rsa I which can differentiate human $\alpha 1$ - from $\alpha 2$ -globin cDNA by recognizing a unique site in the 3' nontranslated region of $\alpha 1$.

Nine clones hybridizing to the described probes and possessing or lacking the 3' Rsa I site were selected for sequence analysis. pCh α 18 was felt to be a full-length $\alpha 2$ cDNA by Pst I and Rsa I restriction analysis. pCh α 2, 16, 25, 34, and 38 were all partial length and contained 5' sequences, while pCh α 23, which contained 3' sequences and an Rsa I site, was preliminarily assigned as a 3' $\alpha 1$ -globin cDNA fragment. All the above clones were sequenced to determine their identity ($\alpha 1$ vs. $\alpha 2$) and their borders. pCh α 18, pCh α 16 and pCh α 23 were sequenced in their entirety (Fig. 2a,b).

Sequence analysis of the two chimpanzee α -globin cDNAs.

The entire sequence of three α cDNA clones, pCh α 18, pCh α 16, pCh α 23, was determined according to the scheme in Fig. 2 and is displayed in Fig. 3. pCh α 18 is designated $\alpha 2$ by its homology to the human $\alpha 2$ sequence. Likewise, pCh α 23 is designated $\alpha 1$ by its strong sequence homologies to human $\alpha 1$. The 5' α -globin cDNA fragment, pCh α 16, which overlaps pCh α 23 by 16 bp is assigned as $\alpha 1$ on the basis of two base differences from pCh α 18 ($\alpha 2$). This last assignment assumes that only two α -globin mRNA species are present based upon detection of two α -globin mRNA species by primer extension and the presence of two species of α -globin genes in humans ($\alpha 1$ and $\alpha 2$) (9-11) with the same genotype ($\alpha\alpha\alpha/\alpha\alpha$) as the chimpanzee being studied. The 5' termini of both pCh α 18 and pCh α 23 cDNAs is 24 nucleotides 5' to the AUG translation initiation codon. The 5' NT region of chimpanzee cDNA from the cap addition site to the beginning of the coding region has previously been determined (12). This sequence overlaps and agrees with our sequence. Since this previously reported 5' noncoding sequence of chimpanzee α -globin was

Fig. 2: Strategy for sequencing the chimpanzee $\alpha 1$ cDNA clones (pCh α 16, pCh α 23) and the $\alpha 2$ -globin cDNA pCh α 18. The cross hatched regions represent pBR322 sequences bordering the inserted cDNA. The position of the restriction sites used for end labelling and the position of the translation initiation (AUG) and termination (UAA) codons are indicated in the figure. The solid circles indicate the ends of the fragments that were labeled; the horizontal arrows show the extent and direction of sequencing. All regions were either sequenced on both strands or multiple times on the same strand.

determined from unfractionated chimpanzee reticulocyte mRNA and was unambiguous, we assume that $\alpha 1$ and $\alpha 2$ are identical to each other and to human from -24 to the cap. Based upon these assumptions, this sequence has been included in Fig. 3 in brackets. The 5' nontranslated region of chimpanzee α -globin is therefore identical to human α -globin except for a substitution of A for G at position -14. The primary sequence of the 3' terminal 9 bases of chimpanzee $\alpha 1$ - and $\alpha 2$ -globin mRNAs have also not been directly determined

	-40	-30	-20	-10	1	10
Ho 1			6			
Cho 1	[ACTCTTCTGGTCCCCAC]AGACTCAGAAAGAACCCACCATG GTG CTG TCT CCT GCC GAC AAG ACC AAC GTC AAG GCC GCC TGG GGT AAG					
Cho 2	[ACTCTTCTGGTCCCCAC]AGACTCAGAAAGAACCCACCATG GTG CTG TCT CCT GCC GAC AAG ACC AAC GTC AAG GCC GCC TGG GGT AAG					
Ho 2	6					
Ho 1	20		30		40	
Cho 1	GTC GGC GCG CAC GCT GGC GAG TAT GGT GCG GAG GCC CTG GAG AGG ATG TTC CTG TCC TTC CCC ACC ACC AAG ACC TAC					
Cho 2	GTC GGC GCG CAC GCT GGC GAG TAT GGT GCG GAG GCC CTG GAG AGG ATG TTC CTG TCC TTC CCC ACC ACC AAG ACC TAC					
Ho 2						
Ho 1	6		50		60	
Cho 1	TTC CCC CAC TTC GAC CTG AGC CAC GGC TCT GCC CAG GTT AAG GGT CAC GGC AAG AAG GTG GCC GAC GCG CTG ACC AAC					
Cho 2	TTC CCG CAC TTC GAC CTG AGC CAC GGC TCT GCC CAG GTT AAG GGC CAC GGC AAG AAG GTG GCC GAC GCG CTG ACC AAC					
Ho 2						
Ho 1	70		80		90	
Cho 1	GCC GTG GCG CAC GTG GAC ATG CCC AAC GCG CTG TCC GCC CTG AGT GAC CTG CAC GCG CAC AAG CTT CGG GTG GAC					
Cho 2	GCC GTG GCG CAC GTG GAC ATG CCC AAC GCG CTG TCC GCC CTG AGC GAC CTG CAC GCG CAC AAG CTT CGG GTG GAC					
Ho 2						
Ho 1	100		110		120	
Cho 1	CCG GTC AAC TTC AAG CTC CTA AGC CAC TGC CTG CTG GTG ACC CTG GCC GCC CAC CTC CCC GCC GAG TTC ACC CCT GCG GTG					
Cho 2	CCG GTC AAC TTC AAG CTC CTA AGC CAC TGC CTG CTG GTG ACC CTG GCC GCC CAC CTC CCC GCC GAG TTC ACC CCT GCG GTG					
Ho 2						
Ho 1	130		140 141		+10 +20	
Cho 1	CAC GCC TCC CTG GAC AAG TTC CTG GCT TCT GTG AGC ACC GTG CTG ACC TCC AAA TAC CGT TAAGCTGGAGCCTCGGTGGCCATGCTTCT					
Cho 2	CAC GCC TCC CTG GAC AAG TTC CTG GCT TCT GTG AGC ACC GTG CTG ACC TCC AAA TAC CGT TAAGCTGGGAGCTCGGTGGCCGTTCTCTCC					
Ho 2	A C A					
Ho 1	+30	+40	+50	+60	+70	+80 +90 +100 +110
Cho 1	TGCCCCCTGGGCTCTCGCCAGGCCCTCTCTCTCTGACACTGTACCCCCCTGGTCTTTGAATAAAGCTGA[GTGGGCGGC]					
Cho 2	TGCCCGCTGGGCTCTCCCAACGGGCCCTCTCTCCCTCTTGACCGGCCCTTCC--TGGTCTTTGAATAAAGCTGA[GTGGGCGGC]					
Ho 2						

Fig. 3: Nucleotide sequence of the chimpanzee $\alpha 1$ - and $\alpha 2$ -globin cDNA. The nucleotide sequence of the coding (mRNA) strand is displayed 5' to 3'. The two base gap in chimp $\alpha 2$ -globin mRNA is positioned to align with the identical human $\alpha 2$ mRNA. Bases not directly determined in this report (see text) are in brackets. An asterisk marks differences between the chimpanzee $\alpha 1$ and $\alpha 2$ sequences. At positions of divergence from the respective human residue is noted. Deletions are indicated by a hyphen. The numbers refer to the translated codons. The 5' nontranslated region is numbered from -40 to -1 and the 3' nontranslated region from +1 to +114. The human sequence is taken from references 17-19.

Table 1

a)	-14	44	57	84	+9	+11	+18	+22	+24	+26	+29	+35	+36	+45
Ho 1	G	CCG	GGC	AGC	A	C	G	A	G	T	T	C	T	C
Cho 1	A	CCC	GGT	AGT	A	C	G	A	G	T	T	C	T	T
Cho 2	A	CCG	GGC	AGC	G	A	G	G	T	C	C	G	C	C
Ho 2	G	CCG	GGC	AGC	A	C	A	G	T	C	C	G	C	C
Class	1	2	2	2	2	2	2	3	3	3	3	3	3	2
	+47	+48	+50	+52	+61	+65	+67	+74	+76	+77	+79	+80	+83	+84
Ho 1	C	C	A	C	C	T	C	C	T	A	C	C	-	G
Cho 1	G	C	A	G	T	T	C	T	T	A	C	C	C	C
Cho 2	A	A	G	G	C	C	T	G	C	C	T	T	-	-
Ho 2	A	A	G	G	C	C	T	G	C	C	T	T	-	-
Class	2	3	3	2	2	3	3	2	3	3	3	3	2	2
	3(2)							3(2)						3(2)
b)	5' NT				Coding				3' NT					
Class	1	1				-				-				
Class	2	-				3				10				
Class	3	-				-				17				

Table 1: Classification of sequence differences amongst the human and chimpanzee $\alpha 1$ - and $\alpha 2$ -globin mRNAs. The positions containing divergent bases are grouped into their respective regions. Each base divergence is categorized. Class 1 changes are consistent with a gene conversion event within the α -globin cluster postdating human/chimpanzee divergence. Class 2 changes are consistent with nonconcerted changes postdating human/chimpanzee divergence. Class 3 changes are consistent with nonconcerted changes prior to human/chimpanzee divergence. The numbers above the bases refer to their position in the mRNA sequence (see Fig. 3). 5' NT (5'-nontranslated region), 3' NT (3'-nontranslated region). At positions +47, +74 and +84, the nonconcerted point mutations cannot be definitely assigned to one of the two species and in Fig. 4, one is therefore assigned to each. Furthermore, since it is possible that two Class 2 mutations could have occurred at these positions rather than a Class 2 and a Class 3, this is indicated in parentheses. It is also assumed that the +83, +84 deletion in $\alpha 2$ is a single event in Fig. 4.

in this report. However, based upon the ability of both chimpanzee $\alpha 1$ and $\alpha 2$ mRNAs to recognize the human α -globin cDNA primer at high stringency, we have assumed they are identical to each other and to human and have included them as such in Fig. 3 in brackets.

Recent evolutionary history of the higher primate α -globin genes.

A comparison of the sequences of the two chimpanzee α -globin cDNAs and the corresponding human mRNAs are displayed in Fig. 3. In Table 1, the sequence differences among the four α -globin RNAs are summarized and categorized. A single difference exists in the 5' noncoding region at position -14 (14 bases 5' to the translation initiation codon) at which both chimpanzee species are A and both human are G. This parallel divergence at -14 may have occurred in one of two ways; two independent mutational events at identical positions in both α genes in one of the species after divergence of humans from chimpanzees or, alternatively, a single mutation in one of the genes after species divergence followed by a transfer of that mutation to the adjacent α -globin gene by a gene conversion event. The second explanation is consistent with previous reports that this region of the two α -globin genes normalizes its primary sequence structure at a high frequency (4,19). Did the mutation occur in human or chimpanzee α -globin gene? The sequence of the 5' nontranslated region of the Cebus monkey (*Cebus albifrons*) α -globin cDNA has previously been determined (12). The Cebus species shares a common ancestry with the chimpanzee and human. The -14 position in the Cebus is A. This suggests that an A to G mutation occurred in one of the two human α -globin genes possibly followed by its transfer to the adjacent gene by a gene conversion event.

The coding region of the chimpanzee $\alpha 1$ - and $\alpha 2$ -globin mRNAs differ by three base differences. Each of these differences are at the third position of a codon and do not change in encoded amino acid. Such mutations are known as nonreplacement or silent substitutions. The presence of these three divergent positions argue against, but do not rule out, $\alpha 1$ - to $\alpha 2$ -globin gene conversion in the chimpanzee subsequent to human-chimpanzee divergence.

The evolutionary history of the α -globin gene 3' to the UAA terminator codon is quite different in both the tempo and characteristics of fixed mutational events. Differences between the $\alpha 1$ - and $\alpha 2$ -globin sequences in this region can be divided into those which predate and those which postdate human-chimpanzee divergence. In comparing this 114 base region of the two human and two chimpanzee mRNAs, there are 10 positions at which one of the mRNAs differ from the other three. These ten mutations most probably occur-

Figure 4: Recent evolutionary history of the human and chimpanzee α -globin genes. The divergence of human and chimpanzee species 5 million years ago is indicated diagrammatically. The numbers on each fork of the diagram indicate the number of separate mutational events observed in the human and chimpanzee lineage by comparison of α 1- and α 2-globin sequence data (see text and Table 1). The circular arrow on the top line represents a hypothesized gene conversion event. 5' nontranslated, 5' NT; 3' nontranslated 3' NT; million years, Myr.

red after human-chimpanzee separation. This group of mutations is categorized as class 2 in Table 1. In the same comparison of the four sequences in Table 1, there are 17 positions at which the human and chimpanzee α 1 mRNAs are either identical to each other, but different from the human and chimpanzee α 2 or the human and chimpanzee α 2-globin mRNAs are identical to each other, but different from both α 1-globin mRNAs. This pattern implies that these sequence differences between α 1 and α 2 occurred prior to the human-chimpanzee divergence and is grouped in Table 1 as a class 3 change.

A hypothetical scheme of the evolutionary history of the α -globin genes in chimpanzee and man subsequent to the species separation can be constructed by grouping the divergent bases in the manner described above. This scheme is displayed in Fig. 4. Using a pairwise comparison of the two sets of α -globin mRNAs, human and chimpanzee, the combined number of mutations 5' to the UAA translation termination codon divided by the combined target area (4/892) results in a divergence rate of 0.45% in 5 million years (the period since the human-chimpanzee separation)(29,30). The unit evolutionary period (U.E.P.; the number of years for 1% divergence (31)) for this region of the mRNA is therefore approximately 11 million years (Myr). It should be noted that the estimated rate of sequence divergence 5' to the UAA terminator may

be somewhat underestimated as a number of mutations in this region could have been erased by the hypothesized gene conversion. However, for this to substantially affect the rate estimates, a large number of mutations would have had to accumulate asymmetrically on the gene ($\alpha 1$ or $\alpha 2$) subject to correction. A comparison of the 3' noncoding region reveals ten mutations subsequent to divergence of humans from chimpanzee (class 2 mutation; Table 1) in a target area of 228 bases (114 x 2) resulting in a U.E.P. of approximately 1.1 Myr. Therefore, with the stated assumptions and reservations, the region of the α -globin mRNA 5' to the UAA terminator undergoes sequence divergence at a rate one-tenth that 3' to this boundary.

Analysis of the human α -globin gene structure suggests that the 3' noncoding region of the human α -globin gene last underwent a gene conversion event well before human-chimpanzee separation (19). The timing of this gene conversion event can be estimated from the sequence comparisons presented above. There are 17 positions in the 3' nontranslated region at which the identical $\alpha 1$ versus $\alpha 2$ differences are found in both humans and chimpanzee (class 3 mutation; Table 1). Assuming these differences existed prior to human divergence from chimpanzee, the extent of sequence divergence 5 Myr. ago was 15% in this region (17 differences in a target area of 114 bases). If the rate of divergence directly calculated for this region over the past 5 Myr. (U.E.P. = 11 Myr.) is applied, then the 3' nontranslated sequences of the two α -globin genes were identical (i.e.--last converted to identity) approximately 21 Myr. ago (1.1 Myr. x 15% = 16 Myr. plus 5 Myr. since human-chimpanzee divergence).

The above calculations, based upon analysis of the chimpanzee and human $\alpha 1$ - and $\alpha 2$ -globin mRNAs, indicate that the evolutionary history of the 3'noncoding region differs from the remainder of the α -globin gene in two respects: its rate of sequence divergence is approximately ten-fold greater and its most recent gene conversion was considerably more remote in time. The net result of these two differences is a substantially higher rate of structural drift in the 3' noncoding region of the gene. Comparisons of globin gene structure within the β -globin cluster over a broad evolutionary range (bird to mammal) suggests that the 3' nontranslated region has in general a higher rate of sequence drift than regions more 5' within these genes (32). While such lack of conserved structure suggests a general lack of function in this 3' region, detailed comparisons of evolutionary drift in primate β -globin genes demonstrate that the 5' and 3' noncoding regions of these genes have diverged at a very slow rate (33). The opposite findings of

these two cited studies demonstrates that even in the same gene system (β -globin), results of evolutionary comparisons may differ depending upon the time span and species studied.

Some insight into the question of why silent site divergence rates vary so greatly within a single gene might be gained by further investigations of the α -globin mRNA structure-function relationships, as well as by sequence determinations in more remote primates to determine if and when this rate of sequence divergence changed. In addition, in order to further interpret the distinctly rapid divergence rate of the 3' noncoding region, it will be necessary to define the function(s) of this region and its structural basis.

ACKNOWLEDGEMENTS

We thank Y.W. Kan for generous support of this project, E. Zimmer for the gift of chimpanzee RNA and B. Chapman for sharing her unpublished data.

This investigation was partially supported by National Institutes of Health Grant AM-16666. S.A. Liebbhaber is a fellow of the John A. and George L. Hartford Foundation and a recipient of a Basil O'Connor Starter Grant from the National Foundation - March of Dimes.

¹Present address: Department of Human Genetics, University of Pennsylvania School of Medicine, Philadelphia, PA 19104, USA

²Present address: Codon Inc., Brisbane, California, USA

REFERENCES

1. Engle J.O. and Dodgson J.B. (1980) Proc Natl Acad Sci USA 77, 2596-2600.
2. Nishioka Y. and Leder P. (1979) Cell 18, 875-882.
3. Schon E.A., Wernke S.M., and Lingrel J.B. (1982) JBC 257, 6825-6835.
4. Zimmer E.A., Martin S.L., Beverley S.M., Kan Y.W., and Wilson A.C. (1980) Proc Natl Acad Sci USA 77, 2158-2162.
5. Orkin S. (1978) Proc Natl Acad Sci USA 75, 5950-5954.
6. Embury S.H., Lebo R.V., Dozy A.M., and Kan Y.W. (1979) J Clin Invest 63, 1307-1310.
7. Lauer J., Shen C-K.J., and Maniatis T. (1980) Cell 20, 119-130.
8. Embury S.H., Miller J.A., Dozy A.M., Kan Y.W., Chan V., and Todd D. (1980) J Clin Invest 66, 1319-1325.
9. Goossens M., Dozy A.M., Embury S.H., Zachariades Z., Hadjiminias M., Stamatoyannopoulos G., and Kan Y.W. (1980) Proc Natl Acad Sci USA 77, 518-521.
10. Higgs D.R., Old J.M., Pressley L., Clegg J.B., and Weatherall D.J. (1980) Nature 284, 632-635.
11. Lie-Injo L.E., Herrera A.R., and Kan Y.W. (1981) Nucleic Acids Res 9, 3707-3717.

12. Zimmer E.A., Ph.D. Thesis, Berkeley, California, 1981.
13. Slightom J.L., Blechl A.E., and Smithes O. (1980) *Cell* 21, 627-638.
14. Weaver S., Comer M.E., Jahn C.L., Hutchinson C.L. III, and Edgell M.H. (1981) *Cell* 24, 403-411.
15. Brown A.J.L. and Ish-Horowicz D. (1981) *Nature* 290, 677-682.
16. Klein H.L. and Petes T.D. (1981) *Nature* 289, 144-148.
17. Liebhaver S.A., Goossens M., and Kan Y.W. (1980) *Proc Natl Acad Sci USA* 77, 7054-7058.
18. Michelson A.M. and Orkin S.H. (1980) *Cell* 22, 371-377.
19. Liebhaver S.A., Goossens M., and Kan Y.W. (1981) *Nature* 290, 26-29.
20. Kan Y.W., Dozy A.M., Varmus H.E., Taylor J.M., Holland J.P., Lie-Injo L.E., Ganeson J., and Todd D. (1975) *Nature (Lond.)* 255, 255-256.
21. Temple G.F., Chang J.C., and Kan Y.W. (1977) *Proc Natl Acad Sci USA* 74, 3047-3051, 1977.
22. Liebhaver S.A. and Kan Y.W. (1981) *J Clin Invest* 68, 439-446.
23. Dagert M. and Ehrlick D.S. (1979) *Gene* 6, 23-28.
24. Boyer H.W. and Roulland-Dussoix D.J. (1969) *J Mol Biol* 41, 459-472.
25. Grunstein M. and Hogness D.S. (1975) *Proc Natl Acad Sci USA* 72, 3961.
26. Maxam A.M. and Gilbert W. (1977) *Proc Natl Acad Sci USA* 74, 560-564.
27. Dayhoff M.O. *Atlas of Protein Sequence and Structure* (National Biomedical Research Foundation, Washington, D.C.), Vol 5 and Suppls 1 and 2, 1975.
28. Orkin S.H. and Goff S.C. (1981) *Cell* 24, 345-352.
29. Brown W.M., George M., and Wilson A.C. (1979) *Proc Natl Acad Sci USA* 76, 1967-1971.
30. Martin S.A., Zimmer E.A., Kan Y.W., and Wilson A.C. (1980) *Proc Natl Acad Sci USA* 77I, 3563-3566.
31. Wilson A.C., Carlson S.S., and White T.J. (1977) *Ann Rev Biochem* 46, 573.
32. Efstratiadis A., Posakony J.W., Maniatis T., Lawn R.M., O'Connell C., Spritz R.A., DeRiel J.K., Forget B.G., Weissman S.M., Slightom J.L., Blechl A.E., Smithes O., Baralle F.E., Shoulders C.C., and Proudfoot N.J. (1980) *Cell* 21, 653-668.
33. Martin S., Zimmer E.A., Davidson W., Wilson A., and Kan Y.W. (1981) *Cell* 25, 737-743.