Volume 11 Number 22 1983 Nucleic Acids Research

Intergenic DNA sequences flanking the pseudo alpha globin genes of human and chimpanzee

Ikuhisa Sawada[1], Marc P.Beal[1], Che-Kun James Shen[2], Barbara Chapman[3], Allan C.Wilson[3] and Carl Schmid[1]

Departments of [1]Chemistry and [2]Genetics, University of California, Davis, CA 95616, and [3]Department of Biochemistry, University of California, Berkeley, CA 94720, USA

Received 5 August 1983; Revised and Accepted 20 October 1983

Abstract

We have determined the sequence of 2400 base pairs upstream from the human pseudo alpha globin ($\psi\alpha$) gene, and for comparison, 1100 base pairs of DNA within and upstream from the chimpanzee $\psi\alpha$ gene. The region upstream from the promoter of the $\psi\alpha$ gene shows no significant homology to the intergenic regions of the adult $\alpha 2$ and $\alpha 1$ globin genes.

The chimpanzee gene has a coding defect in common with the human $\psi\alpha$ gene, showing that the product of this gene, if any, was inactivated before the divergence of human and chimpanzee. However the chimpanzee gene contains a normal ATG initiation codon in contrast to the human gene which has GTG as the initiation codon.

The $\psi\alpha$ genes of both human and chimpanzee are flanked by the same Alu family member. The structure and position of this repeat have not been altered since the divergence of human and chimpanzee, and it is at least as well conserved as its immediate flanking sequence. Comparing human and chimpanzee, the 300 bp Alu repeat has accumulated only two base substitutions and one length mutation; the adjacent 300 bp flanking region has accumulated five base substitutions and twelve length mutations.

INTRODUCTION

The evolution of repetitive sequences has been extensively studied by methods which primarily detect changes in the average consensus sequence of the entire family rather than changes in the sequences of particular members of the family. Examples of such methods include determinations of the melting temperature of interspecies DNA heteroduplexes (10,11) or interspecies sequence comparisons of randomly selected members of repetitive sequence families (8,12,13,14). The results of these studies show that the base sequences of families of interspersed repeats, including especially the Alu family, are well conserved in evolution. The average base sequence of a family might be conserved by a high degree of sequence conservation of each individual member (11). In this case particular Alu repeats in human and chimpanzee would have nearly identical sequences. Alternatively the base sequence

of a family could be conserved by a rapid replacement or correction of individual member sequences (11). Following such a correction in one species we could expect major interspecies sequence differences in any particular Alu family member. The excellent conservation of the α globin gene cluster in primate evolution provides a unique opportunity to study this question. In particular the restriction maps of regions containing the genes ψα, α2 and α1 globin genes in human and chimpanzee are nearly identical (9). We isolated a chimpanzee genomic clone containing a part of the ψα globin gene and its flanking Alu family member to study the evolution of a particular Alu family repeat which is located on the 5' side of the ψα gene.

The information obtained from this study of the ψα gene is also relevant to the evolution of the α-like globin gene cluster. Five genes encoding α-like globins are clustered in a 30 kb region of the human genome (1). The cluster includes a pair of embryonic genes ζ, and ψζ, a nonfunctional adult gene ψα and a pair of adult genes, α2 and α1 (1,2,3, 4,5) which are ordered as shown in Fig. 1. The five genes and much of their intergenic regions have been sequenced (1-8). For brevity we refer to the 5' flanking sequence of a gene and its known control elements as an intergene. In particular the ψα interegene refers to the sequences 5' to the ψα gene and 3' to the ψζ gene (Fig. 1). The α2 and α1 genes have almost identical base sequences (4,5). The close homology of the coding regions has been ascribed to a recent sequence correction event which also extends into the α2 and α1 intergenes (1,4,5 9).

The ψα gene is ancestrally related to the adult α globin gene (2). However available sequences (ca. 100 bp) and restriction mapping shows that the ψα intergene is nonhomologous to the α2 and α1 intergenes (1,2). It is curious that closely related and clustered genes would have totally nonhomologous intergenes. There is evidence for gene duplication by unequal crossing over which requires a concomitant duplication of intergenes (28). In this regard the 5' flanking sequences of the ψα, α2 and α1 genes have similar lengths (1, Fig. 1). Perhaps the ψα and α2/α1 intergenes are ancestrally related and share vestiges of homology which were not detected by comparing restriction maps. To detect any possible residual homology between the ψα and α2/α1 intergenes we have determined the sequence of 2.4 kb of DNA lying 5' to the human ψα gene. With these additional sequence data the base sequence of the α globin gene cluster

has been determined with a few minor gaps from the Bam site lying 2.4 kb 5' to the ψα gene, through the ψα gene and the α2 and α1 genes and their intergenic regions (Fig. 1).

The comparison of the human and chimpanzee ψα sequences might also provide insight into the evolution of pseudogene. Proudfoot and Maniatis (3) have clearly shown that the ψα gene was inactivated significantly further back in time than the divergence of human and chimpanzee. Until recently, all the information about pseudogene evolution came from comparing a given pseudogene with its functional counterpart in the same species. By comparing homologous pseudogenes in closely related species of known divergence time, one can examine the rate and pattern of sequence divergence in a way that is relatively free from the problem of multiple hits, as pointed out by Martin et al. (31).

METHODS

The isolation of human ψα globin gene as a lambda genomic clone and the restriction map of this region has been reported by Lauer et al. (1). The chimpanzee clone was isolated from a library prepared by λ Charon 30 by J. Slightom using chimpanzee DNA provided by E. Zimmer (9). The chimpanzee library was screened by use of avilable human ψζ gene fragments (1). The clone derived by this procedure was verified by comparison to be restriction map of genomic DNA from the donor animal. The donor animal is polymorphic for an Eco R1 site which is located approximately 700 at 5' to the ψα gene. The clone selected in this study does not contain the Eco R1 site. Base sequence determinations were performed by the M13 modification of Sanger et al.'s dideoxy method (15,16). For M13 dideoxy sequencing, restriction fragments 5' to the ψα gene was subcloned into appropriate restriction sites in M13 strains mp 8, 9, 10 and 11 (17). The sequencing protocol is described in New England Biolabs M13 cloning/sequencing system handbook.

Blot hybridization studies of restriction digests of the alpha globin regions were performed according to the modification of Souther's procedure (18,19). Appropriate M13 subclones, which were constructed as described above, were used as radiolabelled hybridization probes.

RESULTS

1) Determination of DNA Sequences

The sequences of the human and chimpanzee DNAs were determined by dideoxy sequencing of the M13 subclones which are depicted in Fig. 1. All

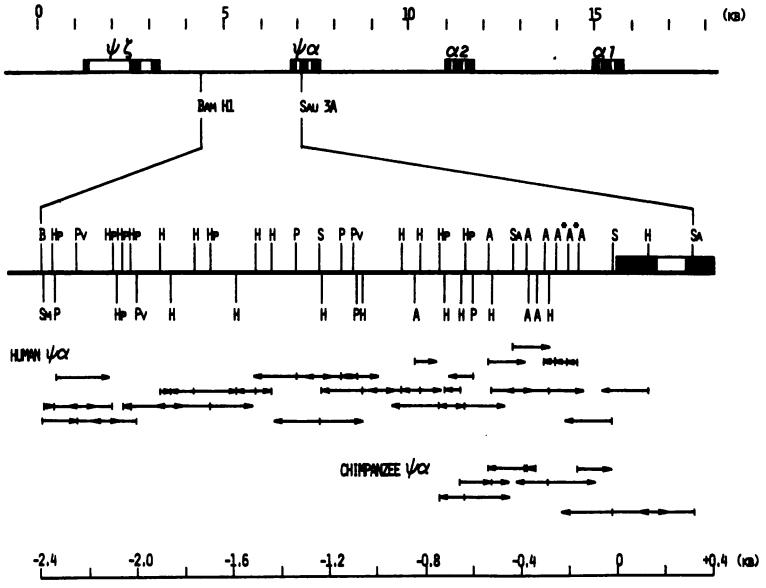

Figure 1. The human alpha like globin genes, which are arranged in the order 5' ζ, ψζ, α2 and α1, are clustered within a 30 kb region (1). The positions of major restriction sites 5' and within the chimpanzee ψα gene agrees very closely with those depicted above for the human ψα gene (9). As depicted above we determined the DNA base sequences of regions lying 5' to the human and chimpanzee ψα genes as well as the 5' coding region of the chimpanzee gene. The small arrows indicate M13 subclones and the extent to which any subclone was read to construct the final sequence. The base sequence through the human ψα gene has been previously determined to approximately position -70 (3) and overlaps the present determination. The restriction sites are B, Bam H1; Sm, SmaI; P, Pst I; Hp, HpaII; Pv, PvuII; H, HaeIII; S, Sst I; A, AluI, and Sa, Sau 3A. The asterisk designates a site which is present in human and absent in chimpanzee.

regions were sequenced several times and most regions are confirmed by complementary sequencing. Normally the sequence determination results from readings that are 150 nucleotides or less from the M13 cloning site, Fig. 1. In the region of comparison between human and chimpanzee it was possible to scrutinize each purported difference so that we are especially confident of our sequence in this region, Fig. 1. In spite of this care, the readings at certain peculiar regions are unavoidable ambiguous, Fig. 2. Examples of such ambiguities include long runs of a particular base which result in compressions on the sequencing gel, Fig. 2.

It should be recognized that a few of the differences reported for the human and chimpanzee sequences are supported by cleavage of predicted

restriction sites. In particular a number of human but not chimpanzee DNA clones were generated by cleavage with the enzyme Alu I (cleaves AGCT) at sites -195 and -216 (Figs. 1,2).

2) The ψα Intergene is Unique in the α Globin Gene Cluster

With one exception, computer searches do not reveal any significant homology between the ψα (Fig. 2) and the α2/α1 intergenes (5,7). The one exception is a 300 nt long Alu family member (positions -710 to -370, Fig. 2) which flanks the chimpanzee and human ψα genes. Like other members of the family it terminates in an A rich 3' end, which is depicted in this case as a T rich 5' end (positions -650 to -700, Fig. 2). The entire repetitive unit is flanked by recognizable short direct repeats (positions -690 vs -370) which result from the duplication of the genomic entry site upon insertion of the Alu family member (8). This Alu family member is not related in any special way to the other five Alu family members in the α globin gene cluster which have been previously sequenced (6,7).

Blot hybridization was employed as a direct test of this lack of homology between the ψα and α2/α1 intergenes. Restriction digests of a clone containing the entire α2 intergenic region as well as the ψα intergenic region depicted in Fig. 2 were probed with M13 subclones from the ψα intergenic region. The M13 subclones were from positions -20 to -250, positions -756 to -1259 and positions -1128 to -2055, which exclude the ubiquitous Alu family member. One of these probes maps up to the ψα promoter region which is precisely the region where homology between the ψα and α2/α1 genes is disrupted (3). In all three cases the M13 subclones probed only the ψα globin intergene and not the α2 intergene (data not shown). By the criteria of blot hybridization the 5' flanking sequence of the ψα gene studied here (Fig. 2) is unrelated to the 5' flanking region of the α2 gene. This conclusion merely confirms the accuracy of the computer assisted base sequence comparison described above. We also find that the ψα intergenic region from -20 to -250 does not hybridize to a clone containing the adjacent ψζ intergenic region (1). We therefore conclude that most of the 5' ψα intergenic region is probably unique within the α globin gene cluster.

3) Sequence Similarities With and Between the ψα and α2/α1 Intergenes

The ψα and α2/α1 genes clearly result from an ancestral duplication (3). As just one illustration of the close relationship between these gene duplicates, even their intervening sequences are partially homologous. It is surprising that the closely related genes have entirely different

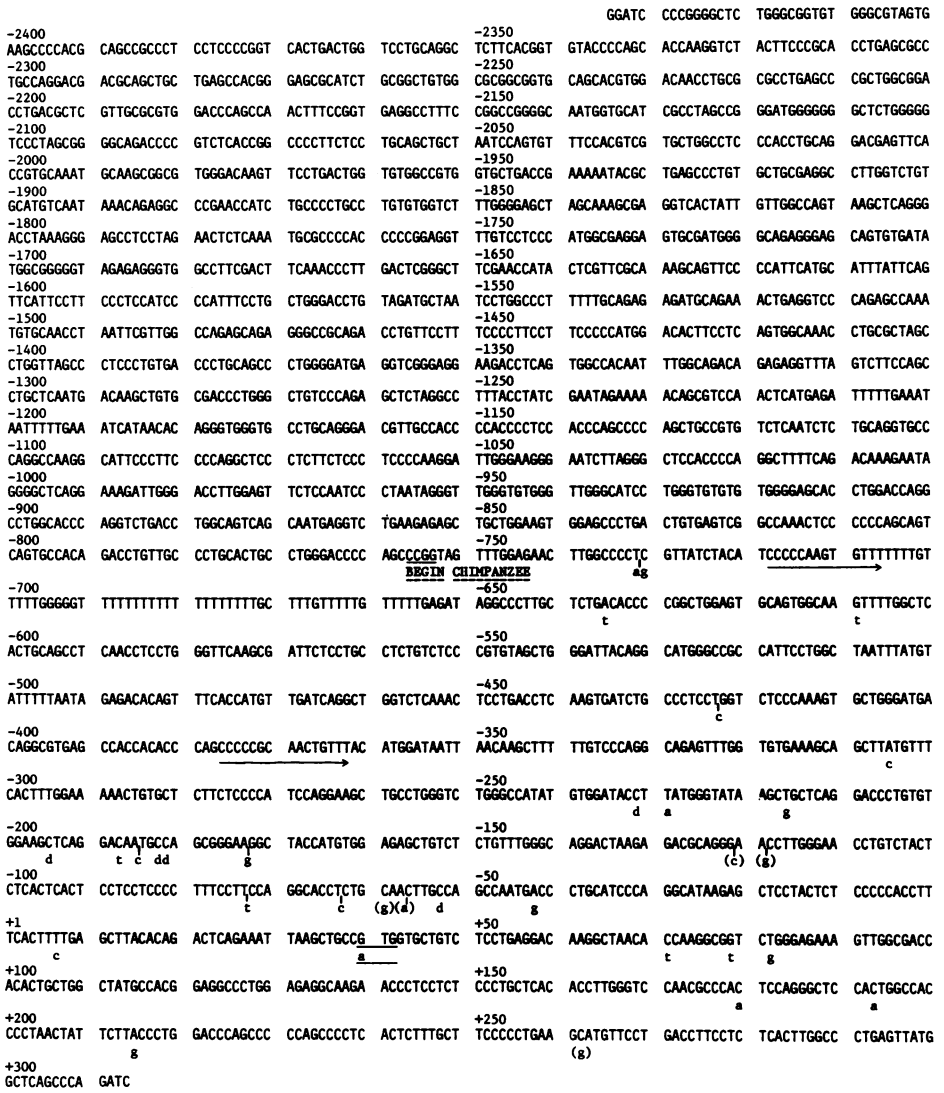

Figure 2. The DNA base sequence determined for the region 5' to the human ψα gene as well as the previously determined coding regions (3). The cap site is numbered as position +1 (3). Also shown are approximately 300 nucleotides from the 5' end of the ψα gene (3). The present results confirm the sequence determination of Proudfoot and Maniatis (3) except for two nucleotides near the extreme 5' end of their sequence. Starting at position -757 and reading to position +314 we also report the chimpanzee DNA sequence. In most positions chimpanzee and human are identical; differences are reported as a lower case letter for chimpanzee. The letter "d" designates are deletion in chimpanzee relative to human; insertion are indicated by a

fiducial mark. Parenthesis are used to indicate positions where a difference between human and chimpanzee is uncertain. An Alu family member between positions -700 and -320 is flanked by short direct repeats, which are designated by an underlined arrow. The chimpanzee and human initiation codons, position +39, are also indicated by an underline. Of particular note the human and chimpanzee sequences are identical at position +279 which contains a twenty nucleotide deletion relative to the functional human α2 gene (3).

flanking sequences. Base sequence comparisons presented below suggest that intergenic regions may rapidly diverge by duplications of existing regions, amplifying runs of simple sequences and perhaps by a high mutation rate associated with runs of simple sequences.

Scattered throughout the 5' flanking region are a number of short repetitive elements or runs of either a particular base or a short simple sequences (Fig. 2). The presence of internal repeats suggests that part of the intergenic region may result from a duplication of existing sequences. For example an imperfect 9 base pair repeat is present at positions -2206 and -1921 within the ψα intergene (Fig. 3). An imperfect twelve base pair repeat lies the same distance upstream from each of these 9 base pair repeats (Fig. 3). Further upstream from the 12 bp repeat is a perfect eight nucleotide repeat which is contained in a 40 nucleotide region having 65% overall homology (Fig. 3). We conclude that either these two 150 bp units were formed by duplicating an ancestral sequence or as a minimum that much intergenic DNA may result from duplicating oligonucleotides.

Polypyrimidine runs (as well as complementary polypurines) occurs at a number of potentially significant sites within the intergenic region (Fig. 2). To better understand the probable implication of these polypyrimidine runs it is helpful to consider a 230 bp region of nonhomology which is present in the α1 intergenic region and is absent in the α2 intergenic region, Fig. 4. Because homology between the α2 and α1 intergenes resumes on either side of this 230 bp sequence, it can be regarded as an insertion into the α1 intergene or as a deletion from the α2 intergene (7). The 5' end of this sequence (position -1100) is essentially a ninefold tandem repetition of the trinucleotide CCN where N is usually the base T. The resulting 28 nucleotides contains 26 pyrimidines. Three tandem copies of CCN are located at position -1050 and a tandem run of five less perfect copies of recognizable at -1000 (Fig. 4). We suggest that this nonhomology region results in part by tandemly expanding simple

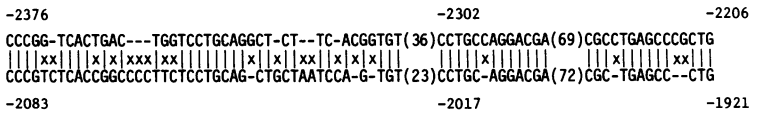

Figure 3. Comparison of oligonucleotide homologes which are colinearly repeated within the region flanking the human ψα gene.

sequences. As further evidence for this hypothesis approximately 22 bases at position -950 are an imperfect tandem array of alternating TG residues. This shows that tandem expansions are not restricted to polypyrimidines. Alternating pyrimidine purine homopolymers have attracted considerable attention because of the ease with which they convert to the left handed Z helix. Poly TG has received particular attention as it is a ubiquitous component of the eukaryotic genome which occurs at a frequency of about once per gene (or intergene) (20,21). The alternating TG tract at position -950 is quite possibly an incipient run of poly TG. Additional repeats can be recognized at other sites within this nonhomology region. The formation of α2/α1 intergenic regions by expansion of simple sequences is not limited to this particular 230 bp region but can be recognized elsewhere in the α2/α1 intergenic regions (5,7).

The last vestige of homology between the ψα and α2/α1 genes disappears upstream from their promoters at approximately position -80 (3, Fig. 5).

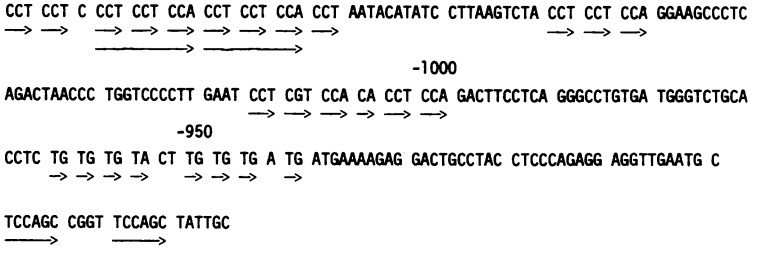

Figure 4. A region of non homology which is present in the α1 intergene but absent from the α2 intergene (7) can be partially represented as a tandem expansion of simple sequences. Polypyrimidine runs such as those indicated above can be recognized at a number of regions within the α2 and α1 intergenes (7) as well as the ψα intergene (Figure 2). Less perfect variants of these expansions, such as those at position -1000, are recognizable in that they contain sequence elements in common with other more certain expansions, e.g. CCT and CCA which are present at around position -1100. As illustrated by the run of alternating TG, sequence tandem expansions are not restricted to polypyrimidines.

Figure 5. A comparison of the ψα and α2/α1 promoter regions (3). Differences between the human and chimpanzee sequences in the ψα promoter region are depicted by the higher letter for human and the lower letter for chimpanzee. D is used to designate a deletion. Direct repeats are indicated by arrows.

This position corresponds to a "boundary" between the gene and intergene for both the ψα and α2/α1 genes (Fig. 5). It is perhaps significant that this boundary region is compounded by a number of short tandem repeats. For example six imperfect copies of CCT are present in the chimpanzee ψα intergene boundary and five copies are present in this region in the human sequence, Fig. 5. The CCT run is preceded by two copies of ACTC. The boundary for the α2 intergene is also occupied by a tandem expansion of a simple sequence CCG (Fig. 5). Multiple runs of CCG in various forms are found throughout the 500 nucleotide region which immediately flanks the 5' end of the α2 and α1 genes (5). We conclude that in part α-like globin intergenic regions result from the tandem amplification of short sequences; the amplification of different sequences in the ψα and α2/α1 intergenes may in part be responsible for the nonhomology. As reviewed in the Discussion there are several examples in which the 5' flanking regions of closely related genes diverge by the tandem amplification of short sequences, such as CCT.

4) The Corresponding Chimpanzee Gene is Also a Pseudogene

Proudfoot and Maniatis (3) have identified a number of defects that render the human ψα globin gene nonfunctional. The human and chimpanzee ψα gene share an identical 20 nt deletion at position 279 relative to the active α2 gene (Fig. 2; 6). In the human ψα gene this twenty nucleotide frameshift deletion results in three downstream termination condons (3). As the chimpanzee clone does not extend into this downstream region we cannot ascertain the presence of these three termination codons in chimpanzee. However even if these terminators are absent, the twenty nucleotide deletion and ensuing frameshift is strong evidence that any polypeptide resulting from this gene is unlikely to be a functional α-

globin. The homologous chimpanzee ψα gene is therefore certainly a pseudogene in agreement with Proudfoot and Maniatis's (3) estimate that this gene was inactivated about 45 million years ago; well before the divergence of human and chimpanzee.

Unlike the human ψα gene which has GTG at the position (position 40) of the initiation codon, the chimpanzee ψα gene contains the normal initiation codon ATG (Fig. 2; 6). Consequently the single nucleotide mutation in the initiation codon in the human ψα was probably not responsible for silencing the ψα gene (3).

Many apes including individual chimpanzees are known to express a third α globin gene product (27). The polypeptide encoded by the open reading frame of the chimpanzee ψα gene would not account for this gene product.

5) Divergence Between Human and Chimpanzee

The divergence between coding regions of the human and chimpanzee ψα genes is due entirely to 10 base substitutions (Table 1, Fig. 6), the percentage divergence being 2.7%. By contrast, in the Alu repeat, there is less point mutational divergence (0.7%). The intergenic region exclusive of the Alu repeat is intermediate in this respect, exhibiting 1.6% divergence (Table 1). Striking differences exist among the three regions in the number of additions and deletions, as shown in Table 1, (Fig. 6). The high incidence in the intergenic region exclusive of the Alu repeat

TABLE 1. Types of mutational differences between human and chimpanzee in the vicinity of the pseudo α globin gene.

Region	Size bp	Base substitutions		Length mutations
		Transitions	Transversions	
Alu sequence (-720 to -364)	357	0	2	1
Intergene (-363 to -51)	312	1	4	12
Gene (-50 to +314)	364	5	5	0
Total (-757 to +314)	1071	6	11	15

exceeds the value for typical noncoding regions of nuclear DNA (33). The absence of length mutations in the coding regions of the ψα gene is also notable and may imply that this region is not free of functional constraints. Combining both length and base changes there is a total of 2.6% divergence between the human and chimpanzee sequences. This value may be compared to 1.5% divergence for the whole α region as estimated from restriction mapping (9) and 0.7% to 2.5% divergence for total single copy DNA as estimated from DNA heteroduplex melting studies (10,25,26,32).

DISCUSSION

i) The Evolution of Alu Family Members

As reviewed in the Introduction the results of previous studies show that the consensus Alu family sequence is well conserved in evolution as a result of either the conservation of its individual members or by the rapid replacement or correction of its members. The presence of the same Alu family member at the same genomic position in human and chimpanzee argues that this member of the family has not been corrected since the divergence of human and chimpanzee (see Introduction). Correction this Alu family member in either lineage would have resulted in the substitution of a recognizably distinct member of the family. Another unlikely possibility is that both were corrected to the same Alu sequence. The tremendous sequence diversity of the Alu family rules out such a single master Alu sequence. As reviewed in the Introduction,

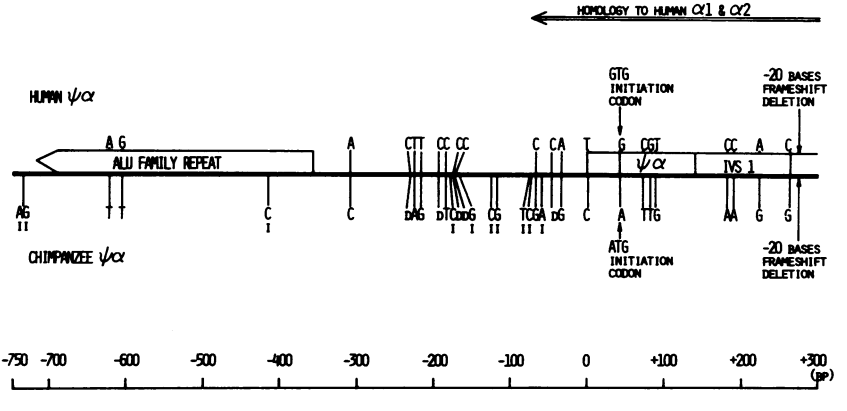

Figure 6. Differences between the human and chimpanzee sequences reported in Figure 2 are depicted schematically as either point mutations or deletions "D" and insertions "I" of the indicated base.

if the sequence of repetitive DNA families is not conserved by a mechanism which operates on the family as a whole then each individual member must be well conserved.

The Alu family member 5' to the ψα gene also appears to be conserved relative to adjacent single copy sequences, in agreement with results from DNA heteroduplex melting studies (10,11). The sample size (three mutations in one 300 bp Alu repeat and thirty-two mutations in the 1071 bp of flanking DNA) is too small to be statistically significant. However the qualitative difference between the mutational divergence of this Alu repeat and its immediate flanking region (Fig. 6) leads us to propose that the base sequences of other Alu family members will be highly conserved since the divergence of human and chimpanzee. If indeed this proposal is proven to be correct, it implies that the sequences of Alu repeats are subject to selection and may serve a biological function.

ii) Intergenic DNA

There is good evidence for gene duplication by unequal crossing over (28). Unequal crossing over would also duplicate intergenes as is the case for α1 and α2 (1). The 2 kb ψα intergenic region studied here is completely unrelated to the α2/α1 intergenes and probably the ψζ intergene. In contrast the α2 and α1 intergenes include regions of identical sequence (1,5,7). One possibility is that the 5' flanking region of the ψα gene was not part of the original duplication unit that gave rise to the α and ψα genes. This would require that the ancestral duplication resulted from several rounds of unequal crossing over rather than one. A second possibility is that the ψα and α2/α1 intergenic regions were part of ancestral duplication unit, but that intergenic regions are subject to rapid mutational changes which erased the ancestral homology of the ψα and α2/α1 intergenes.

Our sequence studies show that intergenic regions are subject to rapid change by several different mechanisms. First, many regions flanking the α2/α1 and ψα genes appear to result from the tandem amplification of simple sequences, such as polypyrimidines. Consequently the resulting ψα and α2/α1 intergenes contain regions of sequence similarity , e.g. runs of CCT, as opposed to regions of strict sequence homology. Second, we have evidence for the duplication of either oligonucleotides or extended regions within the ψα intergene. Third, there is the suggestion that the rate for length mutations may be higher

in intergenes than in genic sequences.

There is good evidence that the tandem amplification of simple sequences observed within the ψα intergene is a general phenomenon within noncoding regions. The intervening sequences within the ψζ and ζ globin genes are largely composed of simple sequence runs (2). Amplification of these simple sequences has the effect of expanding the intervening sequences (2). The tandem amplification of simple sequence DNA e.g. alternating poly CT, has also been identified as a source of length polyporphism within the ribosomal genes of subspecies of mouse and slime mould (22,23). This is good evidence for the rapid (intraspecies) amplification of simple sequences within flanking DNA. The 230 bp non-homology region in the α1 intergenic region is another example of the rapid expansion of simple sequence DNA (Fig. 4); in this case it is likely that the expansion occurred since the last correction of this particular region in the α2 and α1 intergenes (7). A 500 nucleotide rat insulin intergenic region also includes a tandemly amplified sequence (27). Interestingly this region exhibits a high degree of length polymorphism consistent with our suggestion that such regions may rapidly diverge.

Our finding that the closely related human ψα and α2/α1 genes are flanked by totally unrelated intergenes agrees with previous findings on two other clustered multigene families: the chicken ovalbumin and the goat α globin gene families (24,30). Sequences flanking the genes of the ovalbumin cluster are unrelated even though these genes are thought to have arisen by tandem duplication. Of particular interest is a 300-nucleotide expansion of the sequence CCTT that occurs in the 5' flanking region of the Y gene but is absent in the regions flanking both the X and ovalbumin genes (24). Homology between the three ovalbumin-like genes extends approximately 80 nucleotides upstream from their cap sites to include the recognizable promoter elements. The 5' end of the promoter region can then be recognized as the boundary between gene and intergene in exact analogy to our interpretation of the breakdown in homology between the ψα and α2/α1 genes (Fig. 5). There is also a distinct boundary between homologous and nonhomologous sequences in the two functional α globin genes in goat (30). Again in this case homology between the duplicate genes extends approximately 85 nucleotides upstream from the cap site to include known promoter elements (31). After this boundary sequence the excellent homology between the two goat α-globin

genes (1% divergence) breaks down into completely nonhomologous intergenic sequences. This abrupt change from almost exactly homologous to completely unrelated sequences is evidence for the existence of a sequence conversion unit which includes the essential elements of the gene (30). According to this view, the intergenic regions which are part of the ancestral duplication unit but lie outside the conversion unit eventually diverge into unrelated sequences (30). The boundary between intergene and gene at position -94 in the goat I α gene is marked by the sequence: ...CCTCCACCTCCT.... This agrees with our finding that an inexact tandem aplification of the sequence CCT marks the boundary between the human $\psi\alpha$ gene and intergene at position -78, Fig. 5. In the case of the goat α globins, as in the case of the human $\psi\alpha$ and $\alpha 2/\alpha 1$, intergenic nonhomology in part results from tandemly amplified simple sequences.

These conclusions conform with the widely held view, that much intergenic DNA is filler sequence. This does not imply that all intergenic DNA is irrelevant to the structure of a gene. Intergenic regions may contain functional elements which are not recognizable by a base sequence determination. Conceivably the function of intergenic DNA is satisfied by other structural features such as base composition, pyrimidine runs, or sequence length. In this context it is noteworthy that the $\psi\alpha$, $\alpha 2$, and $\alpha 1$ intergenes all have approximately the same length (1).

Acknowledgements

This work was supported in part by USPHS grants GM 21346 (C.S.), AM 29800 (C.-K.J.S.) and an NSF grant DEB81-12412 (A.C.W.).

*To whom correspondence should be communicated

REFERENCES

1. Lauer, J., Shen, C.-K. J. and Maniatis, T. (1980) Cell 20, 119-130.
2. Proudfoot, N.J., Gill, A. and Maniatis, T. (1982) Cell 31, 553-563.
3. Proudfoot, N.J. and Maniatis, T. (1980) Cell 21, 537-544.
4. Liebhaber, S., Goosens, J., Poon, R. and Kan, Y.W. (1980) Proc. Nat. Acad. Sci. USA 77, 7054.
5. Michelson, A.M. and Orkin, S.H. (1983) J. Biol. Chem. in press.
6. Shen, C.-K. J. and Maniatis, T. (1981) J. Mol. Appl. Genet. 1, 343-360.
7. Hess, J.F., Fox, G.M., Schmid, C.W. and Shen, C.-K. J. (1983) Proc. Nat. Acad. Sci. USA 80, 5970-5974.
8. Schmid, C.W. and Jelinek, W.R. (1982) Science 216, 1065-1070.

9. Zimmer, E.A., Martin, S.L., Beverly, S.M., Kan, Y.W. and Wilson, A.C. (1980) Proc. Nat. Sci. USA 77, 2158-2162.
10. Deininger, P.L. and Schmid, C.W. (1976) Science 194, 846-848.
11. Deininger, P.L. and Schmid, C.W. (1979) J. Mol. Biol. 127, 437-460.
12. Grimaldi, G., McCutchan, T. and Singer, M. (1982) Proc. Nat. Acad. Sci. USA 79, 1497-1500.
13. Dhruva, B.R., Shenk, T. and Subramanian, K.N. (1980) Proc. Nat. Acad. Sci. USA 77, 4514.
14. Fox, G.M. (1982) Ph.D. Thesis, University of California, Davis.
15. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) Proc. Nat. Acad. Sci. USA, 74, 5463-5467.
16. Messing, J., Crea, R. and Seeburg, P.H. (1981) Nucleic Acids Res. 9, 309-323.
17. Messing, J. and Viera, J. (1982) Gene 19, 269-276.
18. Southern, E. (1975) J. Mol. Biol. 98, 503-517.
19. Smither, G.E. and Summers, M.D. (1980) Anal. Biochem. 109, 123-129.
20. Hamada, H., Petrino, M.G. and Kakunaga, Y. (1982) Proc. Nat. Acad. Sci. USA 59, 6465-6469.
21. Miesfeld, R., Krystal, M. and Arnheim, N. (1980) Nucleic Acids Res. 9, 5931-5937.
22. Kominami, R., Yrano, Y., Mishima, Y. and Muramatsu, M. (1983) J. Mol. Biol. 165, 209-228.
23. Emery, H.S. and Weiner, A.M. (1981) Cell, 26, 411-419.
24. Heilig, R., Muraskowsky, R. and Mandel, J.-L. (1982) J. Mol. Biol. 156, 1-19.
25. Kohne, D.E., Chiscon, J.A. and Hoyer, B.H. (1972) J. Hum. Evol. 1, 627.
26. Hoyer, B.H., Van de Velde, N.W., Goodman, M. and Roberts, R.B. (1972) J. Hum. Evol. 1, 645.
27. Bell, G.I., Selby, M.J. and Rutter, W.J. (1982) Nature 7, 31-35.
28. Jeffreys, A.J. and Harris, S. (1982) Nature 296, 9-10.
29. Boyer, S.H., Noyes, A.N., Boyer, M.L. and Marr, K. (1973) J. Biol. Chem. 248, 992-1003.
30. Schon, E.A., Wernke, S.M. and Lingrel, J.B. (1982) J. Biol. Chem. 257, 6825-6835.
31. Martin, S.L., Vincent, K.A. and Wilson, A.C. (1983) J. Mol. Biol. 164, 513-528.
32. Zimmer, E. (1980) Ph.D. Thesis, University of California, Berkeley.
33. Cann, R.L. and Wilson, A.C. (1983) Genetics 104, 699-711.