Highly variable regions of DNA flank the human α globin genes

D.R.Higgs, S.E.Y.Goodbourn, J.S.Wainscoat, J.B.Clegg and D.J.Weatherall

M.R.C. Molecular Haematology Unit, Nuffield Department of Clinical Medicine, John Radcliffe Hospital, Oxford OX3 9DU, UK

Received 23 July 1981

ABSTRACT

A series of restriction fragment length polymorphisms which are due to DNA rearrangements have been identified within two highly variable regions flanking the human α globin genes. The existence of such highly polymorphic areas provides a large number of individual genetic markers for the α globin gene cluster on chromosome 16. If, as seems likely, such regions occur frequently throughout the human genome they should be of considerable value in the antenatal diagnosis of genetic disease.

INTRODUCTION

The genes that code for the human globin polypeptide chains are arranged in two separate multigene clusters on chromosomes 11 and 16 (1). The complete nucleotide sequences of most of the globin genes are now known and analysis of cloned DNA, together with Southern blotting analysis of genomic DNA, has enabled extensive restriction endonuclease maps of the two complexes to be constructed (2).

Variations from the normal map due to deletions, insertions, inversions and point mutations may be seen in DNA prepared from individuals with different forms of thalassemia (1). Restriction maps differing from normal may also be seen in DNA obtained from non-thalassemic individuals. They are thought to be due to alterations in DNA sequence which have destroyed or created a restriction endonuclease cleavage site (2,3,4). These polymorphisms, which have no apparent effect on the expression of the globin genes, are usually detected as size variations in specific restriction fragments.

During a recent study of a number of individuals with α

thalassemia, many of them appeared to have typical examples of restriction fragment polymorphisms. However, further investigation revealed that these polymorphisms, unlike those described in the β -like gene cluster, seem to result from DNA rearrangements, within at least two segments of the α globin gene complex, rather than from simple base pair alterations. Furthermore, they are not confined to individuals with α thalassemia and can therefore be considered to be part of the 'normal' variation within the human genome.

MATERIALS AND METHODS

2.1 Restriction endonuclease analysis

DNA was prepared from the buffy coat of 10-20 ml of heparinised blood by phenol-chloroform extraction (5). 10 μ g-15 μ g of DNA was digested with 20-30 units of the restriction endonuclease Bam HI (New England Biolabs) or its isoschizomer Bst I (prepared by a modification of the method in ref. 6), Bgl II, Hpa I or Sac I (New England Biolabs) in the appropriate buffer at 37°C for 6-24 h. The restricted DNA was then electrophoresed in 0.8% agarose gel for 16 h, 40 h or 72 h at 1.0-2.0 volts/cm (with recirculation of buffer) and transferred to nitrocellulose filters. These were then hybridised with a 32 P-labelled probe as previously described (7), except that 5% dextran sulphate was incorporated into the hybridisation mixture and the filters were incubated for 24 h.

2.2 Preparation of probes

Full-length α -specific probe was obtained by nick translation of plasmid JW101. A probe complementary to part of the first intron and the whole of the middle exon of the embryonic ζ 1-globin gene (exon II ζ probe), was prepared by nick translation of a fragment isolated from the plasmid pBR ζ as previously described (7).

A third probe comprising the 5' end of the ζ 1 gene and 350 base pairs of flanking region (5' ζ probe) was prepared by digesting to completion 20 μ g of the plasmid pBR ζ with Bgl II in a total volume of 50 μ l. 100 picomoles of dGTP, 10 μ Ci (α - 32 P) dATP (2000-3000 Ci/mmol) and 2 units of E.Coli DNA pol.1 large fragment (New England Biolabs) were added to this reaction

mixture, which was incubated at 20°C for 5 minutes to end-label the DNA. The reaction was stopped by phenol extraction and the DNA precipitated with ethanol. The pellet was washed in 70% ethanol and then resuspended in 40 μ l of Sac I buffer. Following digestion with Sac I the reaction products were fractionated on a 0.36 mm x 20 cm x 40 cm 6% polyacrylamide gel and identified by autoradiography. The smaller of the two bands was cut from the gel and eluted. This fragment was nick translated and used as a probe for the 5' regions of the ζ globin genes.

RESULTS

3.1 Subjects Investigated

The initial aim of investigating this group of individuals was to characterise the molecular defects which give rise to α thalassemia; these include both deletion and non-deletion abnormalities (1). During the course of the study, we detected several variations within the α gene complex which appeared to be unrelated to defects in globin synthesis, since they were present in both normal and thalassemic individuals and they were identified in areas outside of the DNA region known to be involved in many of the thalassemia defects.

The subjects described here include 68 individuals who are phenotypically normal, 39 individuals heterozygous for α thalassemia (α^0 thalassemia or α^+ thalassemia) and 28 individuals who are homozygotes or compound heterozygotes for α^0 and α^+ thalassemia. In addition, we examined 9 individuals with various acquired types of α thalassemia.

3.2 Evidence for non-polymorphic regions of DNA in the α gene complex

It is apparent from our studies that at least some regions of the α globin gene complex are not commonly polymorphic. When normal DNA is restricted with the enzyme Bgl II two α -specific fragments can be identified, 12.6 kb and 7.4 kb in length (fragments A and B in Fig. 1). Fragment A contains the 5' ($\alpha 2$) gene, a pseudo α gene ($\psi\alpha 1$) and the 3' portion of the embryonic $\zeta 1$ gene. Fragment B contains the 3' ($\alpha 1$) globin gene. We examined 115 haplotypes in which the α genes were intact and found no apparent variation in either of these Bgl II fragments.

Figure 1

The α globin gene complex in which the genes are arranged in the order: 5'- $\zeta 2$ - $\zeta 1$ - $\psi\alpha 1$ - $\alpha 2$ - $\alpha 1$ -3' (8). Restriction sites are indicated: Ba, Bam HI; B, Bgl II; H, Hpa I; S, Sac I. as described in ref. 8 except that the position of the Bgl II sites relative to the coding sequences of the ζ -globin genes are different from those previously described and are based on preliminary sequencing data (N.J. Proudfoot, personal communication) and mapping data from this laboratory. Segments, A, B and C correspond to the constant regions of DNA described in the text. Area X represents the variable, overlapping area of Hpa I and Sac I restriction fragments outside of the constant area B. Area Y corresponds to the variable DNA segment overlapping Bgl II and Bam HI restriction fragments.

The resolution obtained by Southern blotting under the conditions used here would not detect small variations of the order of 0.5-1 kb in these segments of DNA.

Restriction of normal DNA with Bam HI produces two ζ -specific fragments (7,8). One of these (fragment C in Figure 1), which contains the $\zeta 2$ embryonic globin gene, appears to be constant in length (5.9 kb). We studied 132 haplotypes in which the α gene loci appeared to be intact and found that in all cases the length of fragment C was 5.9 kb. Thus, in normal people the sizes of the restriction fragments A, B and C seem to be constant.

3.3 Restriction fragment length polymorphism at the 3' end of the α gene complex

In contrast to the apparent lack of heterogeneity within segments A, B and C, we identified considerable variation in DNA sequence arrangements outside those segments. Hpa I cuts normal DNA to yield two α -specific fragments, 4.3 and 14 kb in length, which contain the $\alpha 2$ and $\alpha 1$ genes respectively. In DNA from the 44 individuals that we studied using this enzyme we could detect no variation in the 4.3 kb fragment, which falls within segment A (Figures 1 and 2). In 58 haplotypes the Hpa I fragment containing the $\alpha 1$ gene was not distinguishable from the normal 14 kb fragment. However, in 22 haplotypes this fragment could be resolved from the more common 14 kb type. (The other

Figure 2

Autoradiograph demonstrating the variability of Hpa I α -specific restriction fragments in various individuals (A-G). This gel was run for 72 h. at 1.0 volt/cm. In one case (A) only a single restriction fragment was observed when the sample was originally investigated using a 24 h. electrophoresis time. Although the 14.0 kb and 14.5 kb fragments cannot be readily distinguished on this reproduction two distinct bands are clearly visible on the original autoradiograph. Figures indicate sizes of fragments in kilobases.

8 were associated with various α thalassemia deletions which alter the size of the Hpa I fragment). These restriction fragment polymorphisms were not all the same length but their size ranged from 0.5 kb to approximately 3.0 kb greater than the common 14 kb fragment. To determine how many different polymorphic fragment lengths were present within this range some DNA samples in which such polymorphisms had been identified were electrophoresed for 72 h. in agarose gel prior to Southern blotting. This procedure revealed at least 5 different sized fragments and also showed that one individual (A, Fig. 2) who was previously considered to be homozygous, did in fact have two different sized Hpa I α 1-specific fragments.

The finding that several different Hpa I α -specific fragments exist makes it unlikely that restriction site mutations are responsible for the polymorphism since several independent changes would be necessary to produce the observed patterns. We were able to confirm that these polymorphisms are due to DNA rearrangements by using a second restriction enzyme Sac I. This enzyme cuts normal DNA within 200-300 base pairs of the Hpa I sites around the α globin genes to give similar sized fragments to Hpa I (8). If the variation in Hpa I fragments were due to a series of restriction site polymorphisms, the size of the corresponding Sac I fragment would stay constant despite the Hpa I fragment variation. However, if these polymorphisms were due to DNA rearrangement within the area X (Figure 1), between the two pairs of Hpa I and Sac I sites, then variation in one of these fragments would be mimicked by variation in the other.

We digested DNA from 23 individuals separately with Hpa I and Sac I. In 29 haplotypes we could detect no variation from the common 14 kb fragment in either digest. In 17 haplotypes both Hpa I and Sac I fragments were 0.5-3.0 kb larger than the normal 14 kb fragment (Figure 2). Therefore, it seems that these polymorphisms are probably due to a sequence rearrangement (insertion or deletion) within the region X (Figure 1) rather than a mutation at a specific restriction site. In two families we have shown that these polymorphisms follow a Mendelian pattern of inheritance through two generations.

3.4 Restriction fragment length polymorphisms between the two embryonic ζ -globin genes

Bgl II digestion of normal DNA gives rise to two ζ -specific fragments (8) one of which is the same as the 12.6 kb α -specific fragment A (Figure 1) discussed in 3.3. The size of the other fragment, which spans the intergenic DNA between the two embryonic ζ globin genes, varies between 10.5 kb and 12.0 kb. We observed the common 11.3 kb fragment in 33 haplotypes. In 5 haplotypes this fragment was larger (~ 0.5 kb) than normal, whereas in 6 haplotypes it was approximately 1 kb smaller. As before, the resolution of this method does not allow clear distinction of small differences between the fragments so these groups may include smaller size variations (Figure 3). In two haplotypes we observed a much smaller fragment of 6.5 kb.

Again, the existence of several different specific restriction fragments suggested that it was unlikely that single site mutations were responsible for the observed variation. Hence, we used a similar approach to that described above for the α -specific fragments to establish the nature of these polymorphic ζ -specific fragments using Bam HI. Two Bam HI ζ -specific fragments are seen in normal individuals (8). The 5.9 kb fragment C has already been discussed (3.2). The second fragment spans a similar length of DNA to that in the variable Bgl II ζ -specific fragment. It varies between 11.3 and 10.0 kb, with a 10.8 kb fragment being the most common. If these polymorphic fragments were due to DNA rearrangement within the area marked Y (Figure 1) a similar variation would be present in both Bgl II and Bam HI fragments from any particular haplotype.

Therefore, we digested DNA from 19 individuals separately with Bgl II and Bam HI (Figure 3). In 30 haplotypes there was no variation in either digest from the pattern of restriction fragments most commonly seen (e.g. lanes H and K in Figure 3). In 4 (e.g. I and J in Figure 3), both Bgl II and Bam HI fragments were ~ 1 kb shorter than the most commonly seen fragments and in 4 (e.g. L in Figure 3), both fragments were ~ 0.5 kb longer than normal. The variation in Y could be detected by identification of the polymorphic Bgl II fragment by either the exon II ζ probe which is complementary to the $\zeta 2$ gene sequence in this fragment, or the 5'

Figure 3

Autoradiographs of Bam HI and Bgl II ζ -specific fragments showing coincidental variation in the size of the ζ -specific fragments in five individuals. The non-variable Bgl II (12.6 kb) fragment and Bam HI (5.9 kb) fragment correspond to segments A and C respectively in Figure 1. Figures indicate the sizes of fragments in kilobases.

ζ probe, which detects the $\zeta 1$ gene in the same fragment. These findings argue in favour of DNA rearrangement as the cause of these restriction length polymorphisms rather than restriction site alterations. We have shown in two families that these polymorphisms follow a Mendelian pattern of inheritance through two generations.

DISCUSSION

Many polymorphisms of structural genes have been detected in humans by screening normal individuals for electrophoretic variants of many different proteins (9). As a result, it follows that a corresponding number of DNA sequence variants within coding sequences must exist and it has been argued that there must also be comparable variability in non-coding sequences. In fact, several restriction site polymorphisms in the non- α globin complex have been located in the intervening sequences of the fetal (G_γ and A_γ) globin genes and the minor adult (δ) gene (4). Similar restriction site variants have also been seen on the 3' side of the major adult (β) globin gene (3,10). It has been estimated that for a haploid genome of 3×10^9 base pairs at least 3×10^7 variants of this type will be present in the human population (4).

DNA polymorphisms of the α globin complex have been noted previously but not characterised in any detail (8,11). The underlying mechanism for the variation in α -specific restriction fragments reported here is clearly different from that in the well characterised polymorphisms of the non- α locus which are thought to result from restriction site mutations. Our data show that variation in the fragments of DNA to the right of the $\alpha 1$ gene and between the two ζ globin genes is due to DNA rearrangements rather than restriction site mutations. Although we were unable to detect variations in the remainder of the α globin gene complex we cannot exclude the possibility that small variations do occur in these regions.

It seems likely that DNA rearrangements may be a relatively common mechanism by which restriction fragment length polymorphisms arise. Recently, Wyman and White (12) reported a highly variable, but as yet unidentified, locus with at least eight

alleles present. The polymorphism in this case also appeared to result from DNA rearrangements which were inherited in a Mendelian fashion, as in the case of the α -loci polymorphisms described here. Of the eight alleles described by Wyman and White the size range of the Eco RI restriction fragments was 14 kb to 29 kb and they stressed the difficulty in resolving fragments within these size ranges, so that there may in fact have been more than eight alleles present. In the cases we have studied polymorphisms at the 3' end of the α -loci and in the intergenic region between the two ζ genes were due to variations in DNA fragment length of 0.5-3 kb. We experienced similar difficulties in resolving fragments in the higher molecular weight (>10 kb) ranges, especially where they differ by as little as 0.5 kb.

Despite the limitation of the present technique we were able to demonstrate at least three different fragment sizes for each DNA segment. Furthermore, when gels were run for 72h. even those fragments could be resolved into several different sizes (Fig 2). Therefore, it seems likely that these regions of DNA will prove to be highly polymorphic. Eventually it will be possible to assess the full extent of the variability within these regions of DNA as unique probes become available which are sufficiently close to them to enable them to be identified and resolved in small restriction fragments differing by as little as 100-200 base pairs.

Wyman and White (12) suggested that the molecular basis for this type of polymorphism might be due to the presence of an active DNA element similar to those identified in yeast and *Drosophila* (13,14). However, other mechanisms may be responsible for this phenomenon. It is possible that inverted repeats exist within this DNA which form loop structures which are susceptible to rearrangement. Alternatively, these segments may consist of a series of direct repeats which could misalign and cross over to give rise to both larger and smaller DNA segments. Both of these explanations avoid the question of how such segments of DNA may have arisen but, whatever the reason, it seems likely that this phenomenon occurs widely throughout the genome.

Despite the fact that some restriction endonuclease site polymorphisms exist in linkage disequilibrium with certain abnormalities of the β globin gene cluster, there is no evidence

to suggest that they are in any way responsible for the abnormal expression of these loci (3,4,10). We have previously observed that two relatively common α^0 thalassemia haplotypes are consistently associated with abnormal Bgl II ζ -specific fragments (15). It is possible that these are polymorphic fragments of the type described here, which are associated with, but not part of, the deletion event which is responsible for the α^0 thalassemia. This suggests that the DNA rearrangements which give rise to these polymorphisms are at least as old as the deletion events which resulted in the α thalassemia mutations. Such findings may give useful insight into the frequency with which such rearrangements occur in the human genome.

Although, we have found no consistent association between any single one of these polymorphisms and abnormal expression of the α gene complex, we cannot rule out the possibility that, in certain cases, these rearrangements do affect the expression of the α gene locus. It is interesting that in all cases of acquired α thalassemia associated with myeloproliferative disorders (1) that we have studied we have observed these size polymorphisms at the 3' end of the complex. Similarly, polymorphisms at this site or between the ζ -genes were also found in three out of three cases of a congenital type of α thalassemia associated with mental retardation (1) (in a fourth case the entire α gene complex is deleted).

The finding of these rearrangement polymorphisms opens up further possibilities for examination of the structure and function of the α globin gene complex. They should be particularly useful for carrying out detailed genetic analyses within families, and may ultimately prove of considerable value in the antenatal diagnosis of genetic disease. It seems likely that any general principles learnt from the α globin gene model will apply to other regions of the human genome.

ACKNOWLEDGEMENTS

Plasmid JW101 was a gift from Dr. B. Forget and pBR ζ was a gift from Drs. T. Maniatis and J. Lauer. We are grateful to Bev Aldridge and Janette Lamb for excellent technical assistance.

REFERENCES

1. Weatherall, D.J. and Clegg, J.B. (1981) *The Thalassemia Syndromes*, 3rd. Ed. Blackwell Scientific Publications, Oxford.
2. Maniatis, T., Fritsch, E.F., Lauer, J. and Lawn, R.M. (1980) *Ann. Rev. Genet.*, 14, 145-178.
3. Kan, Y.W. and Dozy, A.M. (1978) *Proc. Natl. Acad. Sci. USA.*, 11, 5631-5635.
4. Jeffreys, A.J. (1979) *Cell*, 18, 1-10.
5. Gross-Bellard, M., Oudet, P. and Chambon, P. (1973) *Eur. J. Biochem.* 36, 32-38.
6. Clarke, C.M. and Hartley, B.S. (1979) *Biochem. J.* 177, 49-62.
7. Pressley, L., Higgs, D.R., Clegg, J.B., and Weatherall, D.J. (1980) *Proc. Natl. Acad. Sci. USA.*, 77, 3586-3589.
8. Lauer, J., Shen, C-K.J. and Maniatis, T. (1980) *Cell*, 20, 119-130.
9. Harris, H. (1974) *J. Clin. Path.* 27, (Suppl) 8, 32-37.
10. Kan, Y.W., Lee, K.Y., Furbetta, M., Anguis, A. and Cao, A. (1980) *N. Engl. J. Med.*, 302, 185-188.
11. Embury, S.H., Lebo, R.V., Dozy, A.M. and Kan, Y.W. (1979) *J. Clin. Invest.* 63, 1307-1310.
12. Wyman, A.R. and White, R. (1980) *Proc. Natl. Acad. Sci. USA.*, 77, 6754-6758.
13. Cameron, J.R., Loh, E.Y. and Davis, R.W. (1979) *Cell*, 16, 739-751.
14. Potter, S.S., Brorein, W.J., Dunsmuir, P. and Rubin, G.M. (1979) *Cell*, 17, 415-427.
15. Higgs, D.R., Pressley, L., Aldridge, B., Clegg, J.B., Weatherall, D.J., Cao, A., Hadjiminis, M.G., Kattamis, C., Metaxatou-Mavromati, A., Rachmilewitz, E.A., and Sophocleous, T. (1981) *Proc. Natl. Acad. Sci. USA.* In Press.