Am. J. Hum. Genet. 43:249–256, 1988

A New Hypervariable Marker for the Human α-Globin Gene Cluster

A. P. Jarman and D. R. Higgs

MRC Molecular Haematology Unit, Nuffield Department of Clinical Medicine, University of Oxford, John Radcliffe Hospital, Headington, Oxford

Summary

We have located a highly polymorphic region of DNA approximately 100 kb upstream of the human α-globin genes (the α-globin 5′ hypervariable region; 5′HVR). The element responsible is a minisatellite sequence comprising a variable copy number tandem repeat array of a G/C-rich 57-bp sequence. This increases the number of minisatellite elements in the vicinity of the α-globin genes to five, all of which share a region of sequence identity, thus raising questions concerning the distribution and origins of such tandem repeat sequences. The 5′HVR is highly polymorphic and, together with other hypervariable regions at this locus, provides a valuable genetic marker on the short arm of chromosome 16.

Introduction

The vertebrate genome contains many copies of a class of DNA elements, each member of which consists of tandem repeats of short (about 14–64 bp), usually G/C-rich sequence. These arrays of repeats are of limited size and are referred to as minisatellite sequences by analogy to the larger similarly structured segments of satellite DNA. In addition to being G/C rich, certain subsets of these arrays can be identified by virtue of short shared core sequences (Goodbourn et al. 1983; Jarman et al. 1986). This suggests that several discrete minisatellite "families" may exist, and indeed DNA probes constructed from individual minisatellites detect related minisatellites in both the same and other species (Jeffreys et al. 1985; Nakamura et al. 1987). In general, minisatellites appear to be scattered throughout the genome, examples being found on autosomes (Bell et al. 1982; Ullrich et al. 1982; Capon et al. 1983; Goodbourn et al. 1983; Jeffreys et al. 1985; Jarman et al. 1986; Knott et al. 1986), the X chromosome (Oberle et al. 1985), and the pseudoautosomal regions of sex chromosomes (Simmler et al. 1987). Several minisatellites have been found in close association with functional genes, although they show no consistent structural relationship to the expressed sequences.

The number of repeats in any particular minisatellite may be altered by unequal genetic exchange at mitosis and meiosis or DNA slippage during replication (Smith 1976; Goodbourn et al. 1984); the rate of such mutation may be as high as 1–15 × 10^{-4}/kb for some minisatellites (Jeffreys et al. 1985). Thus a proportion of minisatellites give rise to highly variable segments of the genome with multiple alleles detectable by Southern blot hybridization, segments referred to as hypervariable regions (HVRs). These loci are inherited in a stable, Mendelian fashion and, because of their high genetic information content, provide a valuable set of markers for genetic studies in higher-order species. Despite their undoubted usefulness, very little is known about the origins, evolution, or function of minisatellite sequences.

The human α-globin gene complex includes a number of HVRs in and around the structural genes (arranged 5′-ζ2-ψζ1-ψα2-ψα1-α2-α1-θ1-3′), providing an interesting segment of the genome in which to study the nature of minisatellite DNA. The interzeta (IZ), zeta-intron-1 (ζ), and downstream (3′) HVRs have been characterized (fig. 1*a*; Proudfoot et al. 1982; Goodbourn et al. 1983, 1984; Jarman et al. 1986) and consist of tandem repeats of different unit sequences which nevertheless share a core sequence. We describe here a new HVR associated with a minisatellite located approximately 100 kb upstream of the α-globin genes (and named the 5′ HVR). Like the 3′ HVR, this region is

Received January 12, 1988; revision received April 8, 1988.

Address for correspondence and reprints: Dr. A. P. Jarman, MRC Molecular Haematology Unit, Nuffield Department of Clinical Medicine, John Radcliffe Hospital, Headington, Oxford OX3 9DU, England.

0002-9297/88/4303-0004$2.00

highly polymorphic, and it promises to be a very useful genetic marker for α-globin and the short arm of chromosome 16 in general (Reeders et al. 1985; Higgs et al. 1986), especially when used in conjunction with the 3′ HVR. Its sequence conforms to the pattern of G/C-rich HVRs, and it shows a region of identity with the other α-globin HVRs. The presence of five apparently related HVRs clustered within 150 kb of DNA around the α-globin genes raises important questions about the evolution and distribution of minisatellite sequences and is of general relevance to the use of minisatellites as genetic markers.

Materials and Methods

DNA Clones

The isolation and characterization of the cosmid, cRN2103, of its subclone, pRN2103, and of the probe derived therefrom (RA0.6) have been described (Nicholls et al. 1987). The subclone, p5′HVR.14, was constructed as follows: pRN2103 was digested with *Dde*I, *Hinf*I, *Alu*I, and *Rsa*I, and the resulting 770-bp HVR-containing fragment was isolated by electroelution after fractionation on a 1.2% agarose gel (Maniatis et al. 1982). After "end repairing" with Klenow enzyme (Amersham), the DNA fragment was blunt-end ligated into the phosphatased *Hinc*II site of pSP64 (Melton et al. 1984). The insert could subsequently be released from the vector polylinker using *Eco*RI and *Hin*dIII and be used as a probe in Southern blot hybridizations. For orientation, a larger 4.3-kb *Bam*HI fragment was isolated from cRN2103 by digestion, agarose gel fractionation, and electroelution as before. This was ligated into the *Bam*HI site of the polylinker of pUC13 to give the plasmid, pAJ2103.

Growth of all plasmids was in a *rec*A strain of *E. coli* (strain 1046; see Maniatis et al. 1982) to minimize rearrangements of the HVR element. Under these conditions the 5′HVR element was both viable and stable, with no deletions occurring on subcloning or extended growth. The genomic source of the original cosmid library was unavailable, and so we cannot determine whether deletion has occurred during the construction of the library.

DNA Sequencing

Sequencing was carried out according to the base-specific chemical degradation method of Maxam and Gilbert (1980).

Genomic Blot Hybridizations

Genomic DNAs were isolated from peripheral blood as previously described (Old and Higgs 1982). Restriction digestion, gel electrophoresis, and Southern blot hybridization were carried out as described (Southern 1975). For most restriction analyses 5 μg of digested genomic DNA was fractionated on 0.8% agarose gels and transferred onto nylon filters (Hybond-N®; Amersham). For the *Dde*I analyses, 1.2% gels were used for higher resolution and 10-μg samples were used to counteract the poorer transfer of fragments below 300 bp. The probe, RA0.6, was isolated from pRN2103 as a 0.6-kb *Not*I/*Eco*RI fragment. The insert of p5′HVR.14 could be used as a direct probe for the HVR element; for use it was released from the vector by digestion with *Eco*RI and *Hin*dIII, fractionated on a 0.8% agarose gel, and electroeluted from the appropriate gel slice.

Hybridizations were performed at 42C in a buffer including 50% formamide, 3 × SSC, with posthybridization washing in 0.1 × SSC at 65 C (1 × SSC = 150 mM NaCl, 15 mM trisodium citrate). For low-stringency conditions, the formamide concentration in the hybridization buffer was reduced to 35% and the filters were washed in 2 × SSC at 65 C.

Results

Characterization

The cosmid, cRN2103, has been isolated and characterized as part of a chromosome walk upstream of the α-globin genes (Nicholls et al. 1987) and contains DNA from about 100 kb upstream of these genes (fig. 1*a*). From a subclone, pRN2103, a single-copy DNA fragment (RA0.6) has been localized. When used to probe genomic DNA samples digested with *Xba*I, RA0.6 detected a polymorphic segment of DNA (data not shown). Although the fragments observed in *Xba*I-digested DNA are large (about 10 kb) and hence poorly resolved, it was clear that RA0.6 detected a polymorphic locus with more than two alleles, i.e., an HVR. We named this region the 5′HVR (the HVR 5′ of the α-globin genes). Screening other digests with RA0.6 localized the polymorphism to a *Bam*HI-*Hpa*I fragment, corresponding to a 2-kb fragment in pRN2103.

Our strategy for isolating the element responsible for the variability was to digest pRN2103 with four frequently cutting enzymes (*Dde*I, *Hinf*I, *Alu*I, and *Rsa*I), upon which a 770-bp fragment remained intact along with numerous very small fragments. We reasoned that,

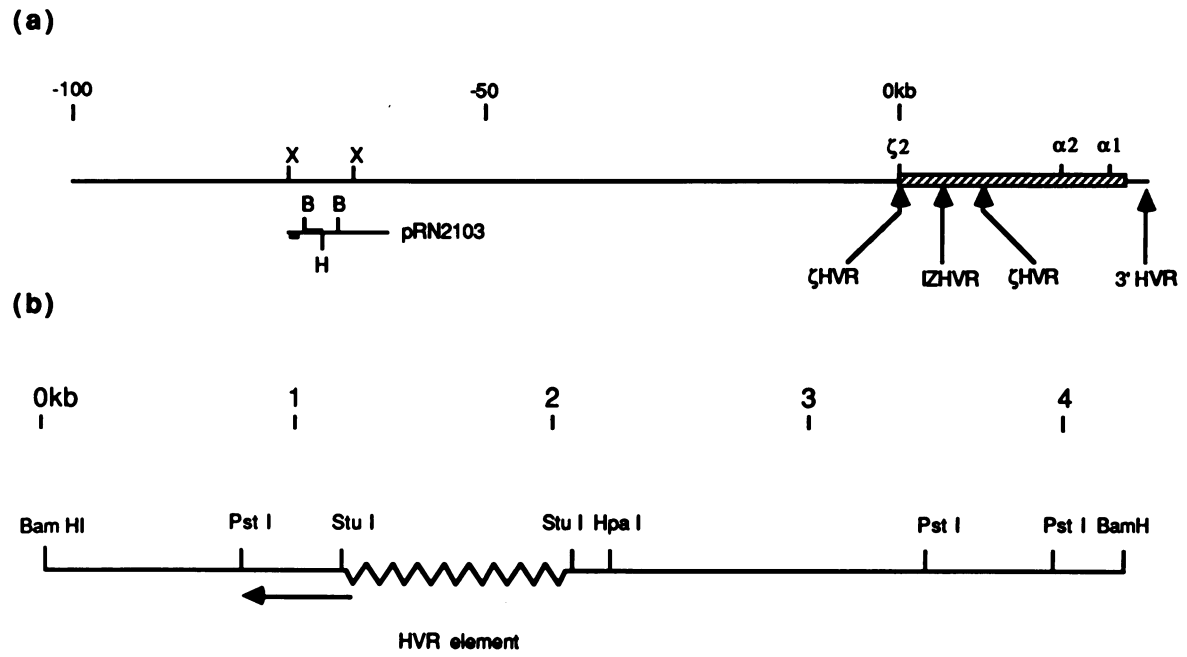

Figure 1 (*a*), Position of the variable region with respect to the α-globin gene complex. A 14-bp tandem repeat exists in the first intron of both ζ2 and ψζ1-globin genes (ζHVR; Proudfoot et al. 1982); a 36-bp repeat is found between the two ζ-globin genes (IZHVR; Goodbourn et al. 1983); 8 kb downstream of the α1-globin gene lies a 17-bp repeat array (3′ HVR; Jarman et al. 1986). The clone pRN2103 is derived from a cosmid, cRN2103, from a chromosomal walk (Nicholls et al. 1987). The probe RA0.6 (indicated by the box on pRN2103) initially detected the HVR in *Xba*I genomic digests (X = *Xba*I), which was subsequently localized to the 2-kb *Bam*HI/*Hpa*I fragment shown on pRN2103 (B = *Bam*HI; H = *Hpa*I). The 0-kb map coordinate represents the ζ-globin mRNA cap site. (*b*), Map of pAJ2103, the 4.3-kb *Bam*HI fragment subcloned from pRN2103 and used to orientate the 5′HVR. The HVR is located between the *Bam*HI/*Hpa*I sites as determined by the presence of multiple *Stu*I sites. The *Stu*I/*Pst*I fragment indicated was used to determine orientation (see text).

by analogy with other HVRs, this relatively large fragment should consist of an array of tandem repeats whose unit sequence lacked sites for these enzymes. Supporting this hypothesis, the 770-bp fragment still detected the polymorphism when isolated and used as a probe against *Rsa*I genomic digests and the intensity of the various genomic bands correlated with their length, an observation that we have found to associate with tandem repeat HVRs. The 770-bp fragment was therefore subcloned into pSP64 to give the plasmid, p5′HVR.14.

Sequencing

The insert in p5′HVR.14 was sequenced in both directions by specifically end-labeling sites in the vector polylinker flanking the insert and then sequencing inward. The sequence obtained (fig. 2) shows that the 770-bp fragment consists of tandem repeats of a 57-bp unit sequence, with very little flanking nonrepeat sequence, and we conclude that the tandem repeats are responsible for the hypervariability. Although the sequence data in each direction do not overlap, from the number of repeat units visible on the sequencing gels, and from the size of the allele sequenced, we can estimate that 13 copies of the repeat are present in p5′HVR.14. From the sequence of the vector/insert junctions it seems that the 770-bp fragment has *Dde*I ends (C/TNAG); thus this enzyme can be used alone to cut out the 5′HVR specifically.

The sequence shows that the 5′HVR is a member of the G/C-rich family of tandem repeats that includes most HVRs characterized to date; in common with these HVRs the sequence is highly asymmetric in that one strand is G rich (G = 47%). On closer inspection areas of high identity with other HVRs can be found, the best relationships being drawn with the ζ-globin HVR (fig. 3). Within this, a partial fit is seen with the previously defined core sequence [5′-GNGGGG(N)ACAG-3′] present in all other α-globin HVRs (Jarman et al. 1986).

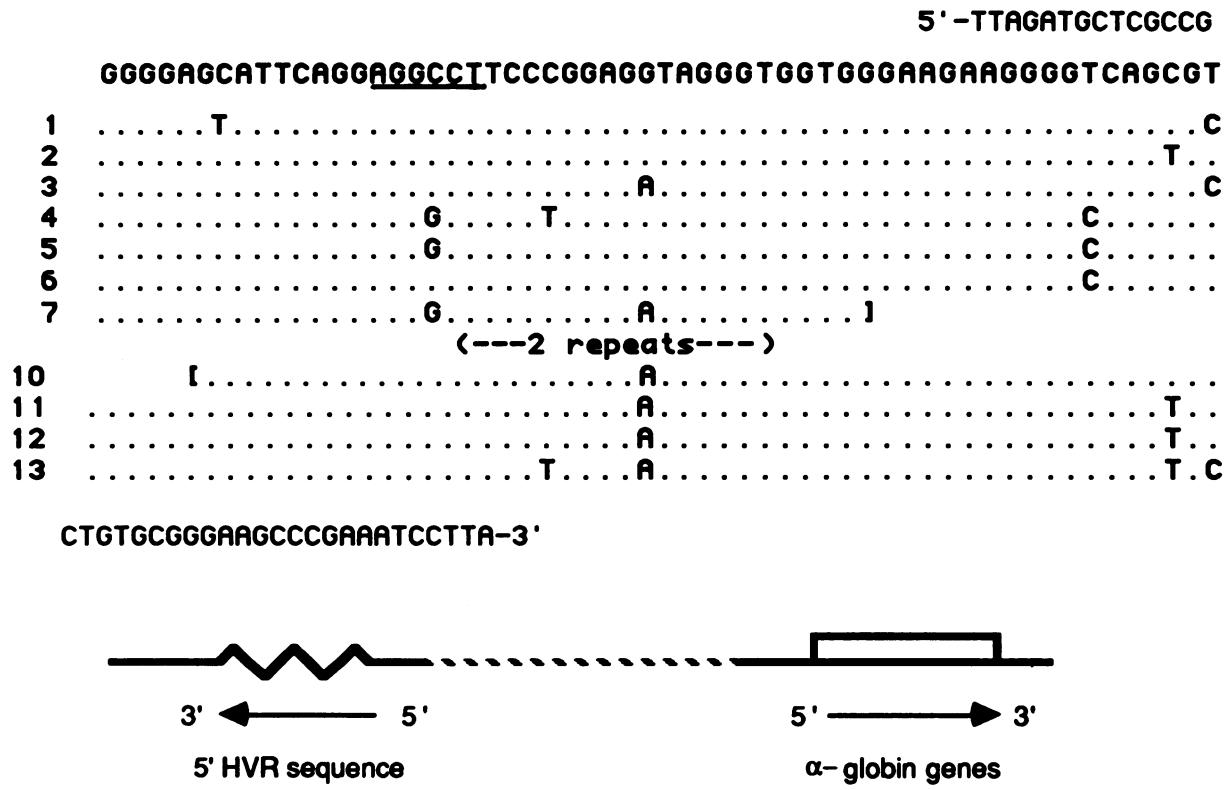

Figure 2 Sequence of the 5′ HVR. Deviations from the consensus repeat sequence are shown. The recurring *Stu*I site is underlined. The only flanking DNA sequence in the clone is shown above and below the array. As shown diagrammatically below the sequence, the G-rich strand represented here runs in the opposite direction to the α-globin gene sense strands.

Other alignments show weaker degrees of identity (with, for example, the minisatellite core of Jeffreys et al. [1985]), and the sequence element G----TGGG commonly found in G/C-rich HVRs is also present (Nakamura et al. 1987). However, in view of the G/C richness of these sequences the significance of such matches is unclear.

Sequence relationships between HVRs have successfully predicted the existence of minisatellite families detectable by probing at reduced hybridization stringency with a variety of HVR sequences including the myoglobin (Jeffreys et al. 1985), α-globin (Jarman et al. 1986), and insulin (Nakamura et al. 1987) repeats. We expected the 5′HVR to show the same characteristics, but when used at low stringency (see Methods) the 770-bp fragment continued to behave as a single-copy probe. Differences in repeat unit length may account for the variable behavior at low stringency of these otherwise related probes.

The sequence in figure 2 has been arbitrarily displayed as the G-rich strand, which illustrates the homologies with the other HVRs. It was of interest to know the orientation of this sequence with respect to the α-globin genes, as the other three HVRs in the cluster have a common orientation, with their G-rich strands in the same direction as the ζ- and α-gene coding sequences. The insert in p5′HVR.14 includes very little flanking

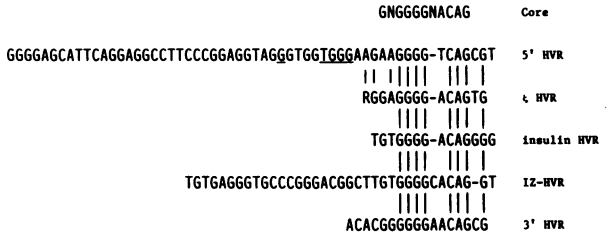

Figure 3 Sequence relationships between the HVRs. The 5′ HVR has a region of high identity with the ζHVR as detected by computer alignment using the Staden ALIGN program. Interestingly, this coincides with a partial identity with the core sequence previously identified in all α-globin complex HVRs and the insulin HVR (Jarman et al. 1986). The commonly occurring G----TGGG motif is underlined.

DNA (fig. 2). Therefore, to determine its orientation in the genome the following strategy was used: the 5′HVR was resubcloned from pRN2103 as a larger 4.3-kb *Bam*HI fragment to give pAJ2103 (fig. 1*b*). The sequence shows that most repeats in the array have a *Stu*I site. Mapping pAJ2103 with *Hpa*I and *Stu*I oriented the insert with respect to the chromosome, the array being 5′ of the *Hpa*I site. Mapping also showed a *Pst*I site close to the 5′ end of the multiple *Stu*I sites and, therefore, of the array. The 5′ *Stu*I/*Pst*I fragment was end-labeled at the *Stu*I site, isolated, and sequenced. The HVR sequence obtained was that of the 3′ end of the array, as represented in figure 2. Therefore, the 5′HVR is oriented such that the G-rich sequence is on the opposite strand to those of the other α-gene-complex HVRs.

Population Genetics

We have used the insert of p5′HVR.14 as a probe for the 5′HVR in population studies. Mostly, we have analyzed genomic *Rsa*I digests for a variety of populations (fig. 4*a*) (populations analyzed: Jamaican, Nigerian, Saudi Arabian, Southeast Asian, Sri Lankan, Icelandic, British, Filipino, and Mediterranean). In these digests, not only is the 5′HVR detected but also a common site RFLP close by, *Rsa*I− alleles being about 4 kb larger than corresponding *Rsa*I+ alleles. Therefore, when characterizing the range of alleles and heterozygosity displayed by the 5′HVR, we concentrated on chromosomes containing the *Rsa*I+ allele, as resolution is much higher.

The range of alleles found (on 304 *Rsa*I+ chromosomes) is from 1.7 to 4.6 kb. By comparison, digestion of pAJ2103 with *Rsa*I gives a 5′HVR-containing fragment of 2.1 kb. On the basis of the finding of 13 repeats within the sequenced allele, we estimate that *Rsa*I fragments contain 1.35 kb of nonrepeat sequence. From this the repeat-containing region varies from 0.35 to

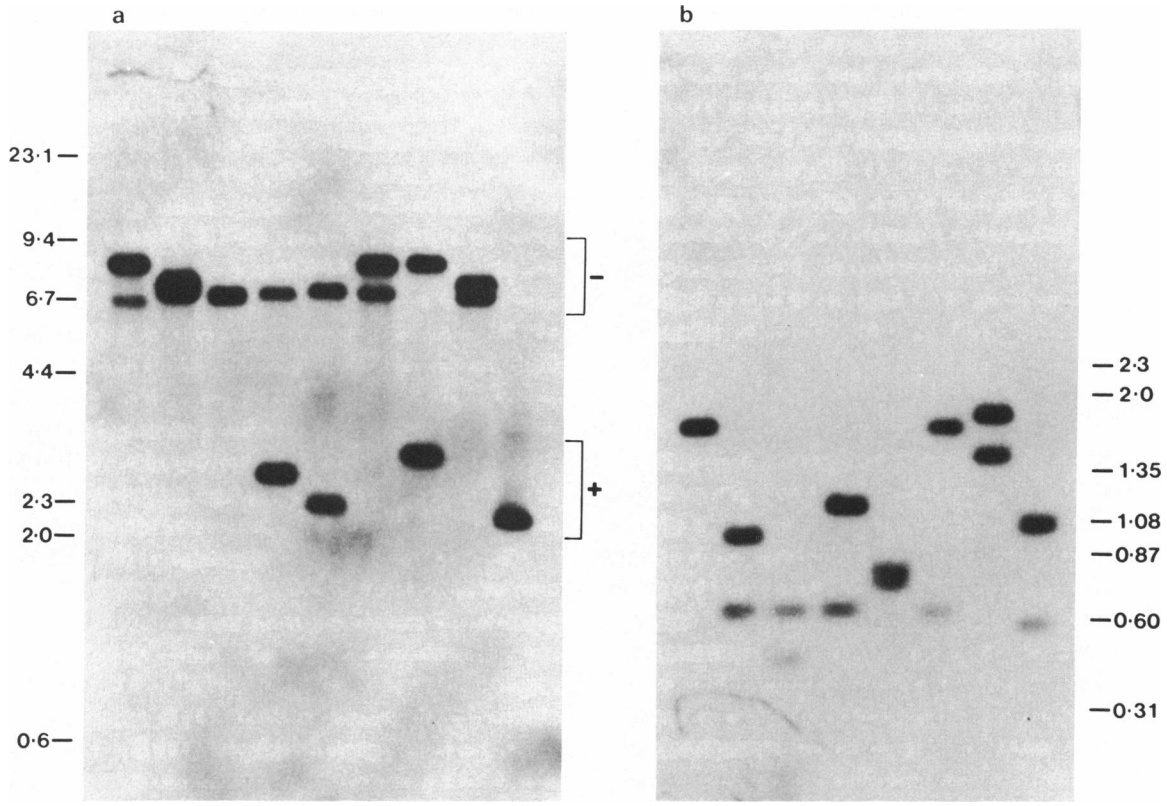

Figure 4 Detection of the 5′HVR in eight random Jamaican genomic DNA samples. *a*, Probing *Rsa*I digests with the 770-bp fragment allows the detection of a site polymorphism as well as the HVR. However, the additional 4 kb of DNA included in the *Rsa*I− alleles limits the resolution of the HVR-containing fragments. This is particularly noticeable in Jamaicans in whom the *Rsa*I− allele is common. The last track contains pAJ2103 digested with *Rsa*I for comparison. *b*, Probing the same samples digested with *Dde*I on a high-percentage agarose gel allows a much higher resolution of the HVR alleles. In this small sample, each individual has two distinguishable alleles (a very small 300-bp allele in the first track is only seen on a longer exposure; data not shown).

3.35 kb, corresponding to a range of approximately 5 to 55 repeats.

A large number of alleles are seen, dependent on gel resolution. Heterozygosity is lower than expected, however, because the distribution of allele sizes is biased, in all populations analyzed, toward smaller *Rsa*I alleles, with a mode of about 2.2 kb and *Rsa*I alleles larger than 3.5 kb being relatively uncommon. Even so, heterozygosities are about 70% for a number of populations. In addition, in a range of populations (British, Mediterranean, and Saudi Arabian) the *Rsa*I site polymorphism increases the heterozygosity to 75%–80%. In populations of African origin, however, heterozygosities are not improved, owing to the high frequency of the *Rsa*I− allele (Jamaicans + = .31; Nigerians + = .15), the resulting larger fragments decreasing the resolution of 5′HVR alleles.

The sequence shows that the p5′HVR.14 insert is actually a *Dde*I fragment, and therefore this enzyme cuts out the entire HVR with very little flanking sequence. Consequently, the best resolution of alleles should be obtained on genomic *Dde*I digests. *Dde*I analyses on a small group of individuals have shown that this is correct (fig. 4*b*): the range of alleles seen is 0.35 to 3.25 kb with a mode of approximately 0.8 kb; the cloned allele in p5′HVR.14 is 0.77 kb. Distinct small alleles can now be distinguished that differ by single repeat copies only, although the number of larger alleles may still be limited only by gel resolution. The limited population screening using *Dde*I suggests that very high heterozygosities, approaching 100%, may be observed.

Discussion

We have detected a new HVR in the human genome about 100 kb upstream of the α-globin gene cluster. Like others characterized, this HVR (the α-globin 5′HVR) consists of an array of tandem repeats of a short DNA sequence unit, in this case 57 bp long. This polymorphism thus arises from allelic variation in number of tandem repeats, the range of variation estimated as being between 5 and 55 repeats. The number of distinct alleles observed is large and increases with greater gel resolution, and heterozygosities are correspondingly high. It can thus be inferred that a high rate of new allele production occurs, in common with other HVRs, by the processes of unequal genetic exchange or DNA slippage during replication. The rate of allele production may be a reflection of unit length and copy number of repeats and in turn determines the degree of sequence homogenization through rapid spread of point mutations throughout the array (concerted evolution). Despite this high rate of genetic activity the 5′HVR was isolated from a standard cosmid library grown under recombination-proficient conditions and does not seem to undergo deletions on subcloning or extended growth. Thus it is both viable and stable in bacteria, unlike other highly variable HVRs such as the 3′HVR (Nicholls et al. 1985; Jarman et al. 1986). The lower tandem repeat copy number (13 vs. 227) and long unit sequence (57 vs. 17 bp) of the 5′HVR may be responsible for these differences.

Although the number of alleles is large, their frequency distribution is skewed such that small alleles (with fewer repeats) are much more common than other lengths. Similar allele frequency distributions have been observed with the 3′HVR (Higgs et al. 1986) and another highly variable repeat locus (Wong et al. 1986). This distribution may reflect the balance between unequal exchange, producing both large and small alleles, and intramolecular deletions that generate smaller alleles.

As a highly polymorphic marker the 5′HVR will be most useful in combination with the 3′HVR. The latter has been extensively used as a marker for the α-globin gene complex (Higgs et al. 1986) and was initially used to show the linkage of the complex to adult polycystic kidney disease (APKD; Reeders et al. 1985). For the analysis of APKD and as a marker for the short arm of chromosome 16 in general, the use of these two HVRs alone will be very powerful owing to their very high combined heterozygosities. This holds true because preliminary studies have shown that, although no recombination events have been observed between the 5′ and 3′ α-globin HVRs in 500 meioses (S. T. Reeders, personal communication), a high degree of linkage disequilibrium has not been observed (data not shown).

The 5′HVR sequence conforms to the general pattern of most HVRs: it is G/C rich with marked strand asymmetry, and the sequence motif G----TGGG is present in common with a wide variety of tandem repeat sequences (Nakamura et al. 1987). A short block of identity is also found with the ζHVR and, through this, to the other α-globin-complex HVRs, via a partial match to a previously identified core sequence (Goodbourn et al. 1983; Jarman et al. 1986). Although other repeat families such as the minisatellites have been described, it has been their method of isolation (cloning of related sequences by screening libraries at low stringency [Jeffreys et al. 1985; Nakamura et al. 1987]) that

has allowed the characterization of such families. The significance of the 5'HVR is twofold: first, five tandem repeat regions have now been described from 150 kb around and within the α-like globin genes; furthermore, although independently isolated and characterized, these all share sequence similarities. The reasons for these observations are at present unclear, owing to our lack of understanding of the behavior of HVRs, particularly with respect to their origins, function, and dispersal mechanisms. Thus it is not known whether the sequence relationships are due to functional convergence or to the dispersal and divergence of an ancestral tandem repeat sequence; the existence of blocks of homology suggests the former. It may be important that the α-globin genes are embedded within an extended G/C-rich domain or isochore which is characterized by the presence of HTF-islands which may have structural or functional significance (Fischel-Ghodsian et al. 1987). It has been suggested that such isochores may be invaded by mobile G/C-rich tandem repeat arrays, reinforcing their G/C richness and so lending isochores a self-perpetuating character (Zerial et al. 1986). Such dispersal of HVRs may be a by-product of a genetic function, such as their possible role in recombination (Jeffreys et al. 1985; Kobori et al. 1986); or else they may be spread by viral integration, as a number of viruses have tandem repeat arrays with known functions (Lupton and Levine 1985). Tandem repeat arrays have also been found in association with a dispersed member of a SINE family (Mermer et al. 1987).

If the clustering of HVRs is common, then this may have implications for HVR probes used for minisatellite detection, a situation where independent segregation of loci is an important feature. Although clustering has also been suggested for an A/T-rich group of repeat arrays on chromosome 19 (Das et al. 1987), it does not appear to be a problem so far with the minisatellite probes. It is also clear that, among the α-globin HVRs at least, clear sequence relationships drawn between independently isolated HVRs do not predict the possibility of their cross-hybridization at reduced stringencies. This is particularly highlighted by the inability of the 5'HVR to detect a clear minisatellite pattern in the low-stringency conditions used here.

Acknowledgments

We wish to thank R. D. Nicholls for helpful discussions and D. J. Weatherall for encouragement and support. We also thank H. Ayyub for excellent technical assistance and L. Roberts for preparing the manuscript.

References

Bell, G. I., M. J. Selby, and W. J. Rutter. 1982. The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. Nature **295**:31–35.

Capon, D. J., E. Y. Chen, A. D. Levinson, P. H. Seeburg, and D. V. Goeddel. 1983. Complete nucleotide sequence of the T24 human bladder carcinoma oncogene and its normal homologue. Nature **302**:33–37.

Das, H. K., C. L. Jackson, D. A. Miller, T. Leff, and J. L. Breslow. 1987. The human apolipoprotein C-II gene sequence contains a novel chromosome 19–specific minisatellite in its third intron. J. Biol. Chem. **262**:4787–4793.

Fischel-Ghodsian, N., R. D. Nicholls, and D. R. Higgs. 1987. Unusual features of CpG-rich (HTF) islands in the human α globin complex: association with non-functional pseudogenes and presence within the 3′ portion of the ζ gene. Nucleic Acids Res. **15**:9215–9225.

Goodbourn, S. E. Y., D. R. Higgs, J. B. Clegg, and D. J. Weatherall. 1983. Molecular basis of length polymorphism in the human ζ-globin gene complex. Proc. Natl. Acad. Sci. USA **80**:5022–5026.

———. 1984. Allelic variation and linkage properites of a highly polymorphic restriction fragment in humans. Mol. Biol. Med. **2**:223–238.

Higgs, D. R., J. S. Wainscoat, J. Flint, A. V. S. Hill, S. L. Thein, R. D. Nicholls, H. Teal, H. Ayyub, T. E. A. Peto, A. G. Falusi, A. P. Jarman, J. B. Clegg, and D. J. Weatherall. 1986. Analysis of the human α-globin gene cluster reveals a highly informative genetic locus. Proc. Natl. Acad. Sci. USA **83**:5165–5169.

Jarman, A. P., R. D. Nicholls, D. J. Weatherall, J. B. Clegg, and D. R. Higgs. 1986. Molecular characterization of a hypervariable region downstream of the human α-globin gene cluster. EMBO J. **5**:1857–1863.

Jeffreys, A. J., V. Wilson, and S. L. Thein. 1985. Hypervariable "minisatellite" regions in human DNA. Nature **314**:67–73.

Knott, T. J., S. C. Wallis, R. J. Pease, L. M. Powell, and J. Scott. 1986. A hypervariable region 3′ of the human apolipoprotein B gene. Nucleic Acids Res. **14**:9215–9216.

Kobori, J. A., E. Strauss, K. Minard, and L. Hood. 1986. Molecular analysis of the hotspot of recombination in the murine major histocompatibility complex. Science **234**:173–179.

Lupton, S., and A. J. Levine. 1985. Mapping genetic elements of Epstein-Barr virus that facilitate extrachromosomal persistence of Epstein-Barr virus–derived plasmids in human cells. Mol. Cell. Biol. **5**:2533–2542.

Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Maxam, A. M., and W. Gilbert. 1980. Sequencing end-labelled DNA with base-specific chemical cleavages. Methods Enzymol. **65**:499–560.

Melton, D. A., P. A. Krieg, M. R. Rebagliati, T. Maniatis, K. Zinn, and M. R. Green. 1984. Efficient *in vitro* synthesis of biologically active RNA and RNA hybridization probes from plasmids containing bacteriophage SP6 promoter. Nucleic Acids Res. **12**:7035–7057.

Mermer, B., M. Colb, and T. Krontiris. 1987. A family of short, interspersed repeats is associated with tandemly repetitive DNA in the human genome. Proc. Natl. Acad. Sci. USA **84**:3320–3324.

Nakamura, Y., M. Leppert, P. O'Connell, R. Wolff, T. Holm, M. Culver, C. Martin, E. Fujomoto, M. Hoff, E. Kumlin, and R. White. 1987. Variable number of tandem repeat (VNTR) markers for human gene mapping. Science **235**:1616–1622.

Nicholls, R. D., N. Fischel-Ghodsian, and D. R. Higgs. 1987. Recombination at the human α-globin gene cluster: sequence features and topological constraints. Cell **49**:369–378.

Nicholls, R. D., A. V. S. Hill, J. B. Clegg, and D. R. Higgs. 1985. Direct cloning of specific genomic DNA sequences in plasmid libraries following fragment enrichment. Nucleic Acids Res. **13**:7569–7578.

Oberle, I., D. Drayna, G. Camerino, R. White, and J. L. Mandel. 1985. The telomeric region of the human X chromosome long arm: presence of a highly polymorphic DNA marker and analysis of recombination frequency. Proc. Natl. Acad. Sci. USA **82**:2824–2828.

Old, J. M., and D. R. Higgs. 1982. Gene analysis. Pp. 74–102 *in* D. J. Weatherall, ed. Methods in haematology. Churchill Livingstone, London.

Proudfoot, N. J., A. Gil, and T. Maniatis. 1982. The structure of the human zeta-globin gene and a closely linked, nearly identical, pseudogene. Cell **31**:553–563.

Reeders, S. T., M. H. Breuning, K. E. Davies, R. D. Nicholls, A. P. Jarman, D. R. Higgs, P. L. Pearson, and D. J. Weatherall. 1985. A highly polymorphic DNA marker linked to adult polycystic kidney disease on chromosome 16. Nature **317**:542–544.

Simmler, M. C., C. Johnsson, C. Petit, F. Rouyer, G. Vergnaud, and J. Weissenbach. 1987. Two highly polymorphic minisatellites from the pseudoautosomal region of the human sex chromosomes. EMBO J. **6**:963–969.

Smith, G. P. 1976. Evolution of repeated DNA sequences by unequal crossing over. Science **191**:528–535.

Southern, E. M. 1975. Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. **98**:503–517.

Ullrich, A., T. S. Dull, A. Gray, J. A. Phillips, and S. Peter. 1982. Variation in the sequence and modified state of the human insulin gene flanking regions. Nucleic Acids Res. **10**:2225–2240.

Wong, Z., V. Wilson, A. J. Jeffreys, and S. L. Thein. 1986. Cloning a selected fragment from a human DNA 'fingerprint': isolation of an extremely polymorphic minisatellite. Nucleic Acids Res. **14**:4605–4616.

Zerial, M., J. Salinas, J. Filipski, and G. Bernardi. 1986. Gene distribution and nucleotide sequence organization in the human genome. Eur. J. Biochem. **160**:479–485.