Volume 15 Number 9 1987 Nucleic Acids Research

Characterization of a highly polymorphic region 5′ to J_H in the human immunoglobulin heavy chain

Alcino J.Silva, John P.Johnson[1] and Raymond L.White

University of Utah, Department of Human Genetics, Howard Hughes Medical Institute, Salt Lake City, UT 84132 and [1]Medical Genetics Division, Children's Hospital Medical Center of Northern California, 747 52nd. Street, Oakland, CA 94609, USA

Received January 28, 1987; Revised and Accepted March 23, 1987 Accession no. Y00130

ABSTRACT

A cloned DNA segment 1.25 kilobases (kb) upstream from the joining segments of the human heavy chain immunoglobulin gene revealed extensive polymorphic variation at this locus, and the polymorphic pattern was stably transmitted to the next generation. Genomic restriction analysis showed that the polymorphism was caused by insertions/deletions within an MspI/BamHI fragment. Sequencing of one allele, 848 base pairs (bp) long, revealed eleven 50-base-pair tandem repeats. A second allele, 648 bp long, was cloned from a human genomic cosmid library, sequenced, and found to contain four fewer repeats than the first allele. A survey of 186 chromosomes from unrelated individuals of primarily northern European descent revealed at least six alleles.

INTRODUCTION

The isolation of polymorphic markers with frequent multiallelic motifs, and correspondingly high heterozygosities, would provide complete information from families. Unfortunately, fewer than 10% of available markers fall into this category (1); most DNA markers recognize only point mutations that alter restriction enzyme sites. Since heterozygosity can never exceed 50% in two-allele systems (2), dimorphic markers have a special disadvantage for disease linkage studies, because the kindreds segregating a disease are often small and incomplete. Several highly polymorphic probes have been isolated, however, and wherever a molecular description exists, the variation in allele size is due to Variation in the Number of short Tandem Repeats (VNTR) of a core sequence within the polymorphic fragment (3,4,5,6). Furthermore, Uematsu *et al* (7), and Cabori *et al* (8) showed that VNTRs were present in regions containing recombinational hot spots in the Major Histocompatibility Complex locus (MHC) of mice, and it has been suggested that human VNTRs are hot

spots for recombination (9). In the present communication we describe a VNTR locus between the diversity (D_H) and the joining (J_H) segments of the human immunoglobulin heavy chain genes.

MATERIALS AND METHODS

DNA samples and isolation: Blood samples were collected in acid/citrate/dextrose tubes from 93 unrelated individuals. High molecular weight DNA was isolated from leukocyte nuclei as described by Bell et al. (10).

Enzyme digestion, electrophoresis, and hybridizations: Restriction enzymes were purchased from New England Biolabs, Bethesda Research Labs (BRL), Biotec, and Boehringer Mannheim, and used under the conditions specified by the supplier. Human DNA samples were digested with a 5- to 10-fold excess of enzyme and monitored for completion by mixing an aliquot of the digestion mixture containing 0.5 mcg of human DNA with 0.5 mcg of an appropriate marker DNA for parallel digestion. Electrophoresis in agarose was as described elsewhere (13). Nylon-66 charged Biotrace membrane was used in southern blot transfers (11). The membranes were hybridized with $8x10^6$ c.p.m. of probe per mL of hybridization solution at 42 C (50% formamide, 5xSSC, 0.02 M $NaPO_4$ (pH 6.7), 100 mcg/mL of salmon sperm carrier DNA, and 10% dextran). Labelling of DNA with ^{32}P-dCTP was done by the random primer method (12) or by nick-translation (13). Routinely, hybridization membranes were washed at 65° C for one hour with 0.1xSSC and 0.1% SDS.

Ligations and transformations: Phage T4 DNA ligase was purchased from Amersham and ligations were done as described by Maniatis et al.(14). E. coli strains JM107 and DH5 were purchased from BRL; the transformation steps followed the procedure provided by the supplier.

Colony and plaque screening was performed according to Maniatis et al (14); the transfer membrane was Pall Biodyme, 1.2 microns pore size.

DNA sequencing analysis : Sequencing of supercoiled plasmid DNA was done as described by Chen et al. (15); materials were purchased from BRL and New England Nuclear. M13 DNA sequencing (16) was done with a universal primer and with synthetic primers

to avoid further cloning. All DNA sequencing experiments were repeated at least twice to ensure accuracy, and results were processed in the Intelligenetics GEL and SEQ computer programs.

RESULTS

Restriction fragments from Ch28-6 were used in a search for polymorphisms within the human immunoglobulin heavy chain genes. The human insert in phage Ch28-6 is a 15 kb genomic DNA clone in the vector Charon 28, from a library prepared by Ravetch et al. (17). The phage encompasses the mu/alpha switch region, the heavy chain joining segments (J_H), and 5 kb of DNA 5' of J_H (fig.1) . A 3.4kb BamHI subclone of phage Ch28-6 in pBR322 (p3.4BHI) detected a polymorphic pattern (see fig.2). As figure 1 indicates, p3.4BHI contains a 3.4 kb human insert 1.25 kb 5' of a BglII site immediately 5' of the J_H segments. The Mendelian inheritance of the polymorphism detected by p3.4BHI was demonstrated in genomic Southern transfers of DNA from large three-generation families (fig.3).

The region of DNA studied is normally involved in rearrangements in B-lymphocytes. The results of Southern transfers of genomic DNA from blood were not affected by

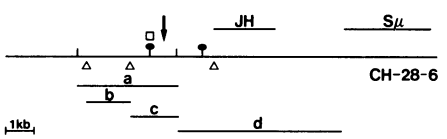

Figure 1: Ch28-6 encompasses the J_H and mu switch regions of the immunoglobulin heavy chain genes of human chromosome 14; p3.4BHI is a 3.4 kb BamHI subclone of Ch28-6 (fragment a) in pBR322. Fragments a, b and c (two restriction fragments of p3.4BHI), and d (p5.7BHI/HdIII), were labelled by the random primer method, and used as probes in southern transfers of human DNA digested with either BamHI, BglII, or MspI. The results, in conjunction with a restriction map of Ch28-6, were used to localize the insertion/deletion to a relatively small fragment of DNA to be sequenced. For the sake of simplicity we only showed relevant restriction sites between the most 5' BamHI site and the most 3' BglII site of the diagram; a detailed restriction digest map of Ch28-6 has been published elsewhere (22). Vertical line markers = BamHI sites; triangles = BglII sites; vertical lines with black ovals = MspI sites; square = SmaI site.

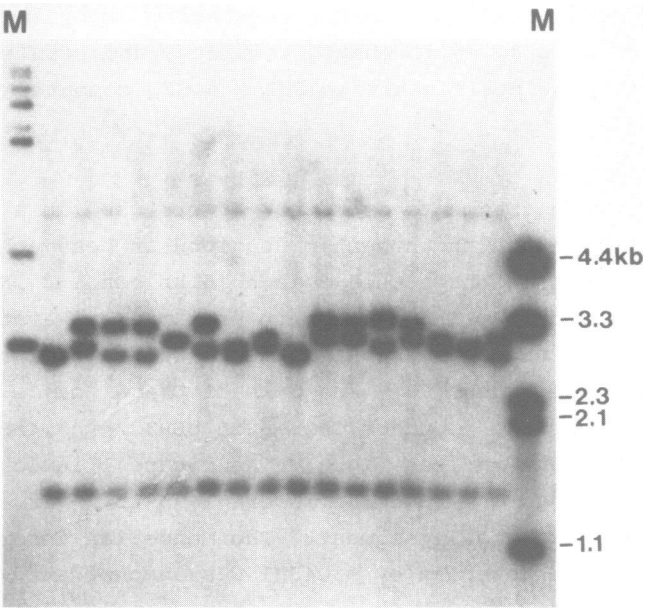

Figure 2: Autoradiograph of a southern transfer of the genomic DNA from 16 unrelated individuals, digested with BglII and electrophoresed in a 1% agarose gel; the blot was probed with labelled p3.4BHI, hybridized, washed, and exposed to X-ray film for 48 hours at -70° C. The flanking lanes (M) are molecular weight markers.

rearrangements at this locus, because B cells represent only about 20% of the white blood cell population. However, we were unable to get hybridization signals from p3.4BHI with DNA from four of 80 lymphoblastoid cell lines, even though all the lines were normal for numerous probes outside the immunoglobulin region. To avoid the problem, all DNAs used in the genomic Southern transfers described in this manuscript were derived from blood lymphocytes rather than from cell lines.

The probe p3.4BHI revealed at least four common alleles on almost all transfers of genomic DNA digested with each of the restriction enzymes (BglII, MspI, BamHI, XmaI, PstI, SinI, TaqI), except for those that gave alleles too large to resolve (SstI, EcoRI, EcoRV, and PvuII for example), or that cut frequently within the polymorphic region (HaeIII). Furthermore, even though the allele sizes varied according to which enzyme was used in the

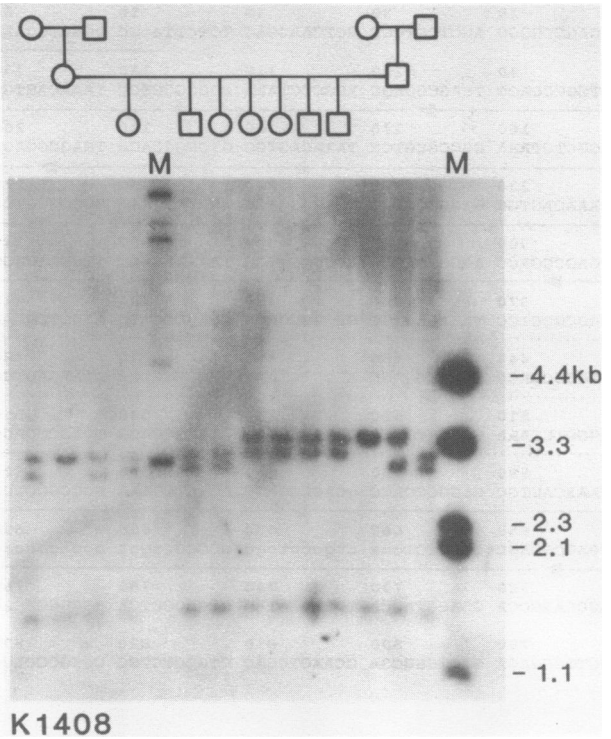

Figure 3: Autoradiograph of a southern transfer of the DNAs of kindred 1408 (Utah family) digested with BglII and electrophoresed in a 1.0% agarose gel; the blot was then probed with labelled p3.4BHI. The lanes marked with an M are molecular weight standards. This is an example of the informativeness of this marker: in a single family of 13 individuals there are at least three different alleles, and only two homozygous individuals.

preparation of the genomic southern blots, the size differences between the two alleles in any given individual tested with various enzymes was constant. Furthermore, there is simultaneous homozygosity detected in genomic southern blots prepared with four different enzymes, for 26 of 93 individuals tested. These results strongly suggest that the polymorphism is a consequence of insertion/deletion rather than the result of simultaneous mutations within restriction sites. In order to localize more precisely the insertion/deletion in the restriction map of Ch28-6, p3.4BHI subclones and another Ch28-6 subclone immediately

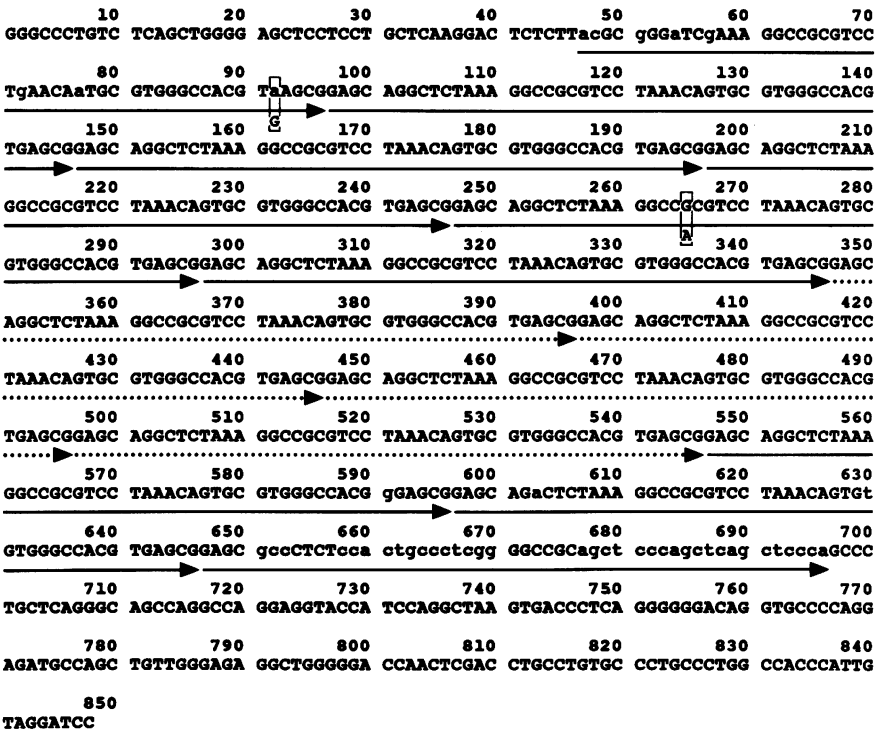

Figure 4: DNA sequence of the 848-bp MspI/BamHI fragment from Ch28-6 containing the insertion/deletion. Each arrow represents a 50-bp repeat, and mismatches between repeats are shown in lower case lettering. A second independently cloned fragment from the same area was sequenced; it is a 648-bp MspI/BamHI fragment, and it differs from the first clone in that it has four fewer repeats and two base substitutions (bases marked with a rectangle) . The two clones were aligned 5' and 3' to maximize homology, and the repeats absent in the smaller clone are marked on the first by a dotted line.

3' of p3.4BHI (p5.7BHI/HdIII) were used to probe genomic Southern blots of human DNA cut with either BamHI, BglII or Msp. The clone p3.4BHI (a in figure 1) revealed the insertion/deletion pattern when used as a probe in a Southern transfer of genomic DNA digested with BamHI. A BglII restriction fragment from p3.4BHI (fragment b in fig.1) revealed the polymorphic pattern with southern transfers of genomic DNA digested with BamHI, but not with BglII ; fragment c (fig.1), a BglII/BamHI fragment from p3.4BHI, revealed the polymorphism when used as a probe in

Southern transfers of genomic DNA digested with either BamHI or BglII; p5.7BHI/HdIII (d in figure 1) revealed the polymorphic pattern when used as a probe in Southern transfers of genomic DNA digested with either BglII or MspI, but not BamHI. These results indicate that the insertion/deletion is between the MspI (also a SmaI) site and the BamHI sites delimiting the fragment indicated by the arrow in figure 1.

DNA sequencing of the polymorphic fragment

To establish the molecular etiology of the polymorphism, the MspI/BamHI fragment containing the region of insertions/deletions was sequenced. In previously described insertion/deletion polymorphisms, the allele size variations were caused by differences in the number of small tandem repeats (VNTRs); we suspected that similar repeats might be the cause of the insertions/deletions observed in this case. Thus, p3.4BHI was digested with SmaI/BamHI. The fragment containing the polymorphism was isolated and ligated to double-digested (SmaI/BamHI) M13mp8 and M13mp9; the products were transformed into strain JM107 (recA-). The M13 clones of interest were selected using the SmaI/BamHI fragment as a probe in an M13 plaque screen; the appropriate clones were then sequenced by the dideoxy termination method. The DNA sequence (fig.4) revealed eleven 50-bp tandem repeats with very high internal sequence homology (no mismatches on the nine internal repeats). The 50-bp repeats are 64% GC and they do not show evidence of other internal repeats. Flanking the VNTR region, there are remnants of ancient repeats, but with sufficient sequence heterology (84% homology for the 5' end repeat; the 3' end repeat preserved only small stretches of homologous sequence with frequent deletions and insertions) to make identification difficult (fig.4).

In order to confirm the nature of the polymorphism, a second allele was cloned from a human genomic cosmid library, and sequenced; a 1.65 kb BglII/BamHI subclone of p3.4BHI (fig. 2) was used to probe a human genomic cosmid library; we used four genomic equivalents, or 300,000 colonies, in the search, and obtained three tentative positives. A second screening revealed that at least one was a true positive, confirmed by a limited restriction map. The positive cosmid was cut with SmaI/BamHI, and

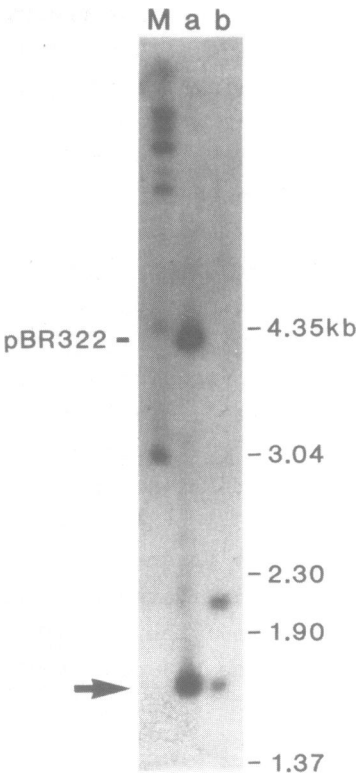

Figure 5: Cloning in recA- strains did not rearrange the tandem repeats. A pBR322 BamHI subclone of cosmid 23.3 (with the 50mer repeat) (pA 23.31) was double-digested with BglII, and BamHI : 30 picomoles of that double digested plasmid were electrophoresed side by side with 5 mcg of total genomic DNA from the same sample used to construct the cosmid library from which cosmid 23.3 was derived. The gel was transferred to Biotrace membrane, and the membrane was probed with a labelled pBR322 plasmid (p1.65BHBg) with a BamHI/BglII fragment containing the repeat (fragment c in figure 2). Hence, lane a (double digested pA23.31) has two bands: a 4.36 kb band (linear pBR322), and the BamHI/BglII 1.5 kb band with the repeat. Lane b (double-digested total genomic DNA) has two bands that correspond to the two VNTR alleles. The dashed arrows point to the band that was sequenced, which is of exactly the same size in the two lanes.

the resulting fragments were cloned into M13mp19, and M13mp18. Again, JM107 (recA-) was the host used in the cloning steps. A subclone screen, with the SmaI/BamHI fragment containing the repeat as a probe, identified the clone of interest. This 648-bp

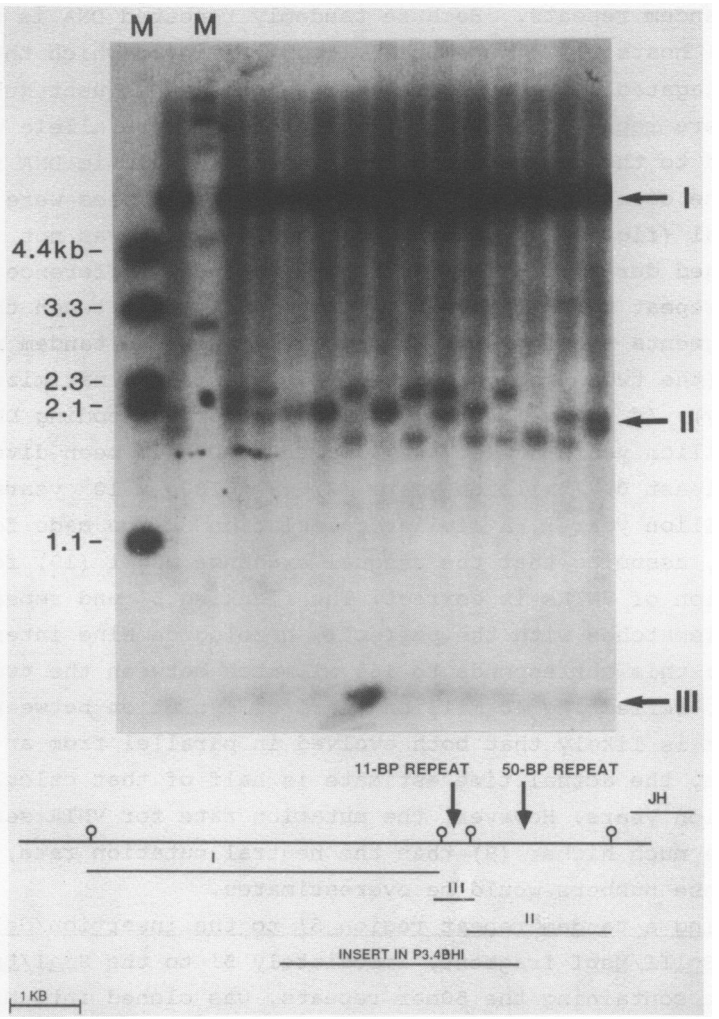

Figure 6: (Top) Autoradiograph of a southern transfer of genomic DNA from 14 unrelated individuals digested with MspI, and probed with p3.4BHI. Lanes M contain molecular weight markers. (Bottom) Diagram of a MspI digest map of the region immediately surrounding the tandem repeats; the fragments with the Roman numerals (I, II, III) in the autoradiograph correspond to the fragments with the same Roman numerals in the diagram. Fragment II contains the polymorphic 50mer VNTR, and fragment III has the 11mer repeat. Circle markers = MspI sites.

fragment was sequenced and found to contain four fewer repeats than the 848-bp fragment (fig.4). This indicates that the polymorphism is the result of a variation in the number of the

50-bp tandem repeats. Because tandemly repeated DNA is unstable in recA+ hosts (18), both the E. coli strain in which the library was propagated and the strains used in the subsequent subcloning steps were recA-. Furthermore, the cloned 648-bp allele was compared to the corresponding allele in the genomic DNA from which the cosmid library was derived, and the sizes were identical (fig.5), indicating that the fragment was not rearranged during cloning. In addition to the difference in tandem repeat number, there are two mismatches between the two DNA fragments sequenced, both of them within the tandem repeat region (the two clones sequenced were aligned to maximize homology). If the frequency of mutation for non-coding DNA is 0.4%/million years, these two alleles must have been diverging for at least 0.77 million years ($2/648 \times 100\% \times 10^6$ years/.4% = 0.77 million years). A similar calculation can be made for two repeats, assuming that the unequal exchange model (19) for the generation of VNTRs is correct. The flanking 5' end repeat has eight mismatches with the perfectly homologous nine internal repeats; this corresponds to 16% mismatch between the two repeats, reflecting 40 million years of evolution between them. Since it is likely that both evolved in parallel from an ancient sequence, the actual time estimate is half of that calculated, or 20 million years. However, the mutation rate for VNTR sequences might be much higher (9) than the neutral mutation rate, in which case these numbers would be overestimates.

Sequencing a tandem repeat region 5' to the insertion/deletion

A BglII/MspI fragment, immediately 5' to the SmaI/BamHI fragment containing the 50mer repeats, was cloned into M13mp8 and M13mp9 and partially sequenced. Approximately 250 bp 5' from the polymorphic VNTR described above, we found three 11-bp tandem repeats (5' AGAGCCAGCCC 3') with no internal mismatches. Furthermore, the first eleven base pairs of the 50mer VNTR contained only two mismatches with the complementary strand of the 11-bp repeat. Southern blotting analysis of genomic DNA from 14 unrelated individuals did not reveal any polymorphism within this second tandem repeat region (fig.6). DNA sequencing of two independently isolated clones from the region, one from Y. W. Kan's cosmid library and the other from Ch28-6, also failed to

Table 1. Results of the population study*

ALLELE SIZE (kb) (MspI)	GENE FREQUENCIES
3.0	0.005
2.4	0.005
2.2	0.178 ± 0.067
2.0	0.274 ± 0.063
1.9	0.376 ± 0.058
1.8	0.161 ± 0.067

* DNAs from 93 unrelated individuals were digested with MspI and electrophoresed in 0.8% agarose gels; the gels were transferred to Biotrace membrane, and the membranes were probed with p3.4BHI. The allele sizes were carefully estimated; frequencies were calculated as the number of alleles scored. The two largest alleles were rare (one in 186, or 0.005).

show any polymorphism: the sequence of the tandem repeat was identical in the two clones.

Population Study

Southern transfers of MspI-digested DNA from 93 unrelated individuals of primarily northern European descent, probed with a nick-translated BamHI fragment containing the repeats, revealed six alleles (table 1). There are four common alleles and two rare alleles, and both rare alleles are substantially larger than the common alleles. Careful sizing of the four smallest alleles of the immunoglobulin VNTR showed that they differ by two, two, and four repeats respectively from the smallest to the largest of the four. The two rare alleles differ from the largest of the common ones by 8 and 16 repeats. A more extensive sampling of the population could perhaps reveal other rare alleles, but would not change drastically the frequencies of the common alleles.

DISCUSSION

We have described a polymorphic locus between the D_H and J_H region of the immunoglobulin heavy chain locus on human chromosome 14. In a survey of 93 unrelated individuals from the Utah population, we have identified six alleles. Because of its high heterozygosity value (82%), this marker will be useful in studies

of genetic association in diseases of the immune system, as well as in the preparation of a genetic linkage map of chromosome 14. The cloning and sequencing of two independently derived alleles of this locus showed that the size difference between them could be explained by the difference in tandem repeat number. Careful sizing of the six alleles seen in the Utah population is consistent with this model. The high internal repeat homology breaks down at the boundaries of the VNTR; this observation is consistent with the unequal crossing over model (19). We were unable to find two alleles that differed in molecular weight by an odd number of repeats; this could either be the result of allelic fixation within the population sampled, or it could reflect the repeat-generation process; specific restrictions on the repeat alignment process could prevent the generation of certain VNTR size classes. Population data from the zeta-globin VNTR (20) reveal the same phenomenon: allele sizes seem to cluster in specific size classes; comparisons among different populations show that although the frequencies of various size classes may change, the common allele sizes do not.

In studies involving the VNTRs associated with the genes for human insulin, myoglobin, and zeta-globin, and for the MHC of mice, there are indications that they might be associated with an enhancement of recombination. It is striking that VNTRs have been found in regions which contain the MHC recombinational hot spots in each of the cases where DNA sequence information is available. Interestingly, the two tandem repeats reported in this manuscript are within a region with a high frequency of translocation breakpoints involving chromosome 14 (21), an observation perhaps consistent with the hypothesis that the tandem repeats are within a recombinational hot spot.

It is also noteworthy that we found two tandem repeats so close to each other, and it is striking that there are only two mismatches between the complementary strand of the 11-bp repeat and the first eleven base pairs of the 50-bp VNTR. This homology might be significant in the the generation and evolution of the tandem repeats.

ACKNOWLEDGMENTS

We are grateful to Ruth Foltz for editorial assistance in the preparation of this manuscript. The work was supported by National Institutes of Health grant No. GM-29789. R.W. is an investigator of the Howard Hughes Medical Institute.

REFERENCES

1. Human Gene Mapping VIII. Eighth International Workshop Human Gene Mapping, Helsinki (1985) Cytogenet. Cell Genet. 40, 387-447.
2. Botstein, D., White, R.L., Skolnick, M. & Davis, R. (1980) Am. J. Hum. Genet. 32, 314-331.
3. Proudfoot, N.J., Gill,A., Maniatis, T. (1982) Cell, 31, 553-563.
4. Capon, D.J., Chen, E.Y., Levinson, A.D., Seeburg, P.H., Goeddel, D.V.,(1983) Nature 302, 33-37.
5. Weller, P., Jeffreys, A.J., Wilson, V., Blanchetot, A.,(1984) EMBO J. 3, 439-436.
6. Stoker, N.G., Cheah, K.S.E., Griffin, J.R., Pope, F.M., Solomon, E., (1985) Nucleic Acids Res. 13, 4613-4622.
7. Uematsu, Y., Kiefer, H., Schulze, R., Fisher-Lindahl, K., Steimetz, M.,(1986) EMBO J., in press.
8. Kobori, J.A., Strauss, E., Minard, K., Hood, L. (1986) Science, 234, 173-179.
9. Jeffreys, A.J., Wilson, V., Thein, S.L.,(1985) Nature, 314, 67-73.
10. Bell, G., Karan, J. & Rutter, W.(1981) Proc. Natl. Acad. Sci. USA 78, 5759-5763.
11. Southern, E. (1975) J. Mol. Biol. 98, 503-518.
12. Feinberg, A.P., & Vogelstein, B. (1984) Anal. Biochem. 137, 266-267.
13. Barker, D., Shaffer, M. & White, R.(1984) Cell, 36, 131-138.
14. Maniatis, T., Fritsch, E.F., & Sandbrook, J. (1982) Molecular Cloning: A Laboratory Manual. Cold Spring Harbor Laboratory Press, N.Y.
15. Chen, E.Y., Seeburg, P.H. (1985) DNA 4, 2,165-170.
16. Sanger, F., Coulson, A.R., Barrel, B.G., Smith, A.J.H.,& Roe, B.A. (1980) J. Mol. Biol. 143, 161-178.
17. Ravetch, J.V., Siebenlist, U., Korsmeyer, S., Waldman, T. & Leder, P. (1981) Cell 27, 583-591.
18. Albertini, A.M., Hofer, M., Calos, M.P., Miller, J.H. (1982) Cell 29, 319-328.
19. Smith, O. (1976) Science 191, 528-535.
20. Chapman, B. S., Vincent, K.A., Wilson, A.C. (1986) Am. J. Hum. Genet. 112, 79-82.
21. Bakhshi, A., Jensen, J.P., Goldman, P., Wright, J.J., McBride, O.W., Epstein, A.L., Korsmeyer, S.L. (1985) Cell 41, 899-906.
22. Johnson, J.P., Gatti, R.A., Sears, T.S., White, R. (1986) Am. J. Hum. Genet. 39, 787-796.