Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A

Delia Lakich[1], Haig H. Kazazian, Jr.[2], Stylianos E. Antonarakis[2,3] & Jane Gitschier[1]

Mutations in the factor VIII gene have been discovered for barely more than half of the examined cases of severe haemophilia A. To account for the unidentified mutations, we propose a model based on the possibility of recombination between homologous sequences located in intron 22 and upstream of the factor VIII gene. Such a recombination would lead to an inversion of all intervening DNA and a disruption of the gene. We present evidence to support this model and describe a Southern blot assay that detects the inversion. These findings should be valuable for genetic prediction of haemophilia A in ~45% of families with severe disease.

[1]*Howard Hughes Medical Institute and Department of Medicine, University of California, San Francisco, California 94143-0724, USA*
[2]*Center for Medical Genetics and Departments of Pediatrics and Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland 21287, USA*
[3]*Department of Genetics, University of Geneva Medical School Geneva, Switzerland*

Correspondence should be addressed to J.G.

Haemophilia A is one of the most common X-linked diseases in man. With an incidence of about 1 in 5,000 males, it follows fragile-X syndrome and Duchenne/Becker muscular dystrophy in frequency. For each of the latter two disorders, the majority of cases are caused by a single type of mutation: amplification of a trinucleotide repeat for the fragile-X syndrome[1] and partial gene deletions for Duchenne/Becker muscular dystrophy[2]. Consequently, diagnostic assays that directly test for the presence of the responsible mutations are now used routinely for genetic prediction of these two disorders.

Haemophilia A, by contrast, is characterized by many different mutations of the factor VIII gene. A variety of nonsense and missense mutations, frameshifts, deletions, duplications and insertions have been described[3]. While the same mutation has been observed occasionally in unrelated individuals, these data suggest that generally each family bears its own mutation.

By analysing the coding and adjoining intron sequences in genomic DNA, Higuchi *et al.*[4,5] sought to define the mutation in every haemophilia A patient studied. Mutations were found in almost all mild and moderate patients, but were identified in only about half of the severe patients. This surprising finding has led to the idea that the remaining mutations giving rise to the severe form of the disease may lie in sequences regulating factor VIII, for example within the promoter or introns, or possibly in a tightly linked gene that serves as a transcription factor required for factor VIII activity[4,5].

Naylor *et al.*[6], using a different approach, recently made a discovery that began to clarify this issue. They assayed for mutations in factor VIII cDNA rather than in genomic DNA by PCR amplification of reverse-transcribed RNA (RT-PCR). For 10 of 24 severely affected patients, they found that PCR amplification across the boundary of exon 22 and exon 23 was not possible, whereas amplification across any combination of exons 1–22 or exons 23–26 yielded the expected products. No other mutations were found in these patients. Their findings suggested that changes in intron 22 are responsible for this class of severe haemophilia A[6].

In many respects, intron 22 is unusual. At 32 kilobases (kb), it is the largest intron in the factor VIII gene[7]. It also contains a CpG island, located about 10 kb downstream of exon 22 (Fig. 1*a*). This island appears to serve as a bidirectional promoter for two genes referred to as factor VIII-associated genes A and B[8,9]. Gene A, which is transcribed in the opposite direction to factor VIII, is intronless and completely nested within intron 22. Two additional transcribed copies of gene A also lie ~500 kb upstream of the factor VIII gene[8,10,11]. Gene B is transcribed in the same direction as the factor VIII gene; its first exon lies within intron 22 and is spliced to exons 23–26. Both gene A and gene B are expressed ubiquitously.

Here, we present a model to account for the unusual factor VIII cDNA PCR results[6]. In this model, one of the upstream A genes homologously recombines with the intron-22 A gene on the same chromosome (Fig. 1*b*). If the orientation of the upstream A gene is opposite to that of the intron-22 copy, then the homologous mispairing and a single crossover event will result in an inversion of the intervening DNA sequences. Consequently, the factor VIII gene would be divided into two parts, with an intact promoter and exons 1–22 greatly separated from, and in opposite orientation to, exons 23–26 (Fig. 1*c*).

This model is appealing for a number of reasons. First,

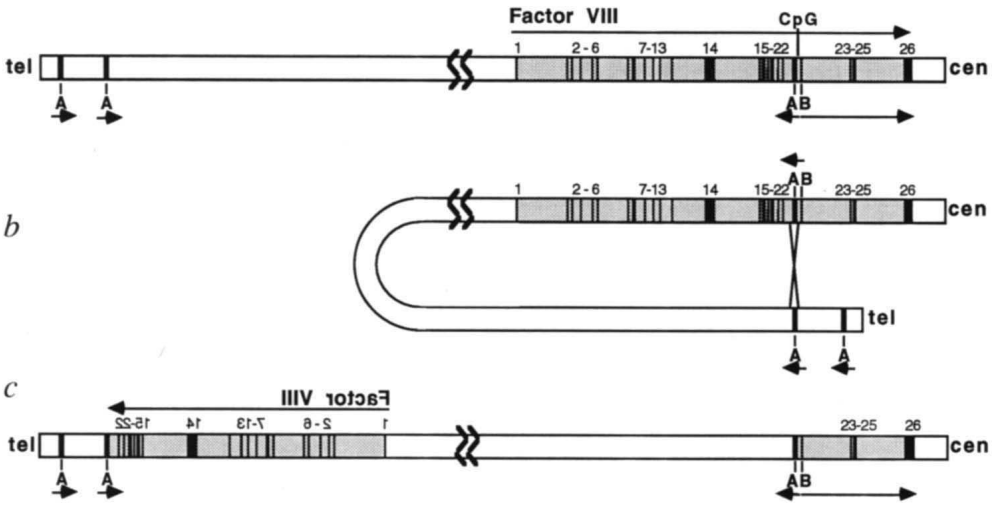

Fig. 1 Diagram of the factor VIII gene and illustration of the inversion model. *a*, Region of Xq28 that includes the factor VIII gene, oriented with the telomere at the left, is depicted. Three copies of the A gene are indicated, two lying upstream of factor VIII and one inside intron 22. The location of the B transcript is also shown. The arrows indicate the direction of transcription of the factor VIII and internal A and B genes. The direction of the upstream A genes is hypothesized to be as shown. *b*, Proposed homologous recombination between the intron-22 copy of gene A and one of the two upstream copies. A crossover between these two identical regions, oriented as shown, would result in an inversion of sequence between the two recombined A genes (*c*). A recombination could involve either of the upstream A genes, but only one is presented. The crossover could occur anywhere in the region of homology which includes the A genes.

this disruption would explain the inability to amplify across exons 22–23 by RT-PCR. Second, as the factor VIII promoter would not be affected, a transcript could still be produced from the 5' portion of the truncated factor VIII gene, accounting for the RT-PCR amplification of exons 1–22. Third, the B gene would not be disrupted by the inversion, and the B transcript, which includes exons 23–26, could be detected by RT-PCR as well, giving the false impression that factor VIII mRNA is being assayed. Fourth, homologous recombination could provide a common mechanism for independently occurring mutations in a large number of individuals. We present evidence to support this model as the cause of severe haemophilia A in about 45% of cases.

PCR Identification and analysis of test samples

To test this model, we first had to identify patient samples that exhibit the RT-PCR peculiarity[6]. Lymphoblastoid cell lines from five previously unstudied, unrelated boys with severe haemophilia A were used in this study. RT-PCR analysis indicated that amplification across the exon 22–23 boundary is not possible for two of the five samples (Fig. 2). In this experiment, cDNA synthesis was primed with 16.2A, a primer based on exon 24 sequence. Two sets of nested PCR reactions were performed on the cDNAs: a control set involving amplification between exons 18 and 21, and an experimental set involving amplification between exons 22 and 24. Samples 1, 4 and 5 and a non-haemophilic control sample gave a product of the expected size for both amplifications, whereas samples 2 and 3 gave the control product, but not the experimental product. Thus samples from two of the five patients presented the features described previously[6] and could be further analysed.

We hypothesize that primer 16.2A primes cDNA synthesis at an imperfect target sequence in the mRNA of samples 2 and 3 because, according to our model, the target sequence in exon 24 would not be adjacent to the 5' portion of the factor VIII gene. The low temperature (42 °C) of the reverse transcriptase (RT) reaction would permit this lack of specificity. We have achieved identical results with two additional "exon 24-specific" primers (RT24A, ER), each differing from that used by Naylor *et al.*[6], as well as with random primers (results not shown).

We also tested the hypothesis that the B transcript would not be disrupted by the proposed inversion and could therefore serve as a template for RT-PCR amplification between different combinations of exons 23–26 as seen by Naylor *et al.*[6]. Amplification with a primer (P26) in the B exon[9] and a primer (ER) in exon 24 yielded the expected product in samples 2 and 3 (data not shown), indicating that the B gene is intact.

Southern blot analysis

The predicted DNA inversion should lead to changes in length of restriction fragments that include the two breakpoints. To detect such a rearrangement, we needed

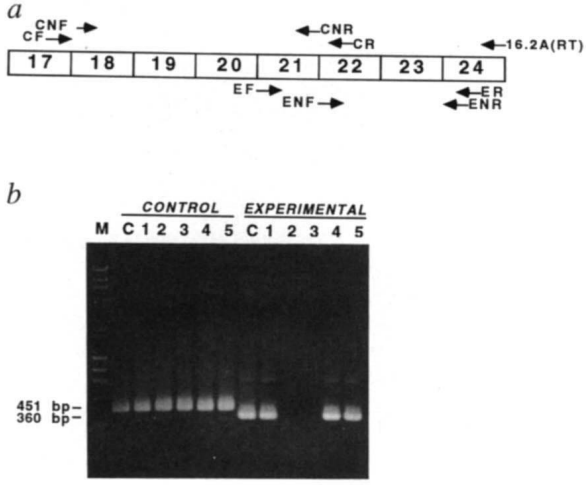

Fig. 2 RT-PCR of RNA from patient and control lymphoblastoid cell lines. The locations of exons for a portion of the factor VIII cDNA are shown in *a*. The positions of primers used for reverse transcription (16.2A), nested control PCR (CF and CR followed by CNF and CNR), and nested experimental PCR (EF and ER followed by ENF and ENR) are indicated by the arrows. Resulting PCR products from samples of a control are shown in *b*. Non-haemophilic individual (lane C) and five males with severe haemophilia A (samples 1–5 in lanes 1–5, respectively) are shown. Lane M contains *Hin*dIII-digested λ and *Hae*III-digested ϕX DNAs. The PCR reactions were performed with an extension time of 1 min.

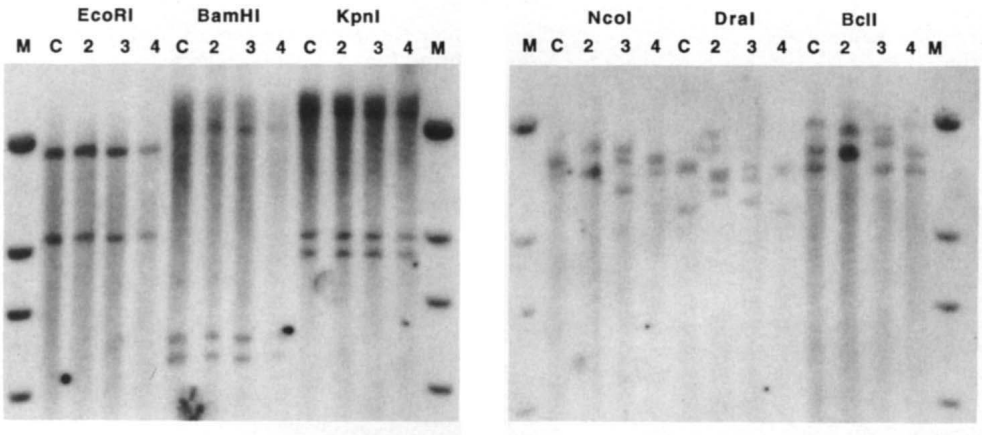

Fig. 3 Southern blots showing altered fragments in patient samples 2 and 3. DNA from a control individual (lane C) and haemophilic patients 2, 3 and 4 (lanes 2, 3 and 4) were digested with *Eco*RI, *Bam*HI, *Kpn*I, *Nco*I, *Dra*I, or *Bcl*I. The blot was hybridized to the gene A probe. Lane M contains λ DNA digested with *Hin*dIII.

to identify restriction enzymes whose sites lie outside the region of homology, estimated to be ~5–10 kb (unpublished observations). As expected, use of enzymes such as *Eco*RI, *Bam*HI and *Kpn*I that have sites within the region of homology[7], did not reveal the postulated rearrangements in samples 2 and 3 (Fig. 3). However by digestion with *Nco*I, *Dra*I or *Bcl*I, two of the three bands are altered in these samples. This result suggests that some form of rearrangement has affected the sizes of two gene A-containing fragments. In addition, the patterns of samples 2 and 3 differ from each other. An explanation for this finding is that the recombination event for samples 2 and 3 occurred with different copies of the upstream A gene.

By inspection of each digest in Fig. 3, one can also infer which fragment contains the intron 22-copy of gene A. For example, as seen in the *Bcl*I digest of Fig. 4*a*, the largest fragment (21.5 kb) is altered in both samples 2 and 3 and should contain the intron-22 copy of gene A if the model is correct. Indeed, Fig. 4*b* shows that this band is found only in YACs that encompass intron 22 of the factor VIII gene (yWXD924 and yWXD527[10]; Fig. 4*c*). The 16-kb band is found only in YACs yWXD250 and yWXD37, which include the most distal copy of gene A. This band is altered in sample 3, suggesting recombination with the most distal A gene. The 14-kb band is observed only in YACs yWXD506 and yWXD348, which contain the more proximal copy. In sample 2, the 14-kb band is altered in size, suggesting recombination with the proximal upstream A gene. A smaller *Bcl*I fragment that is not present in genomic DNA is found in yWXD348; this fragment probably corresponds to overlap of yWXD348 with the most distal A gene. These results are consistent with the inversion hypothesis and suggest that a crossover between the gene A copy in intron 22 can occur with either of the upstream A genes.

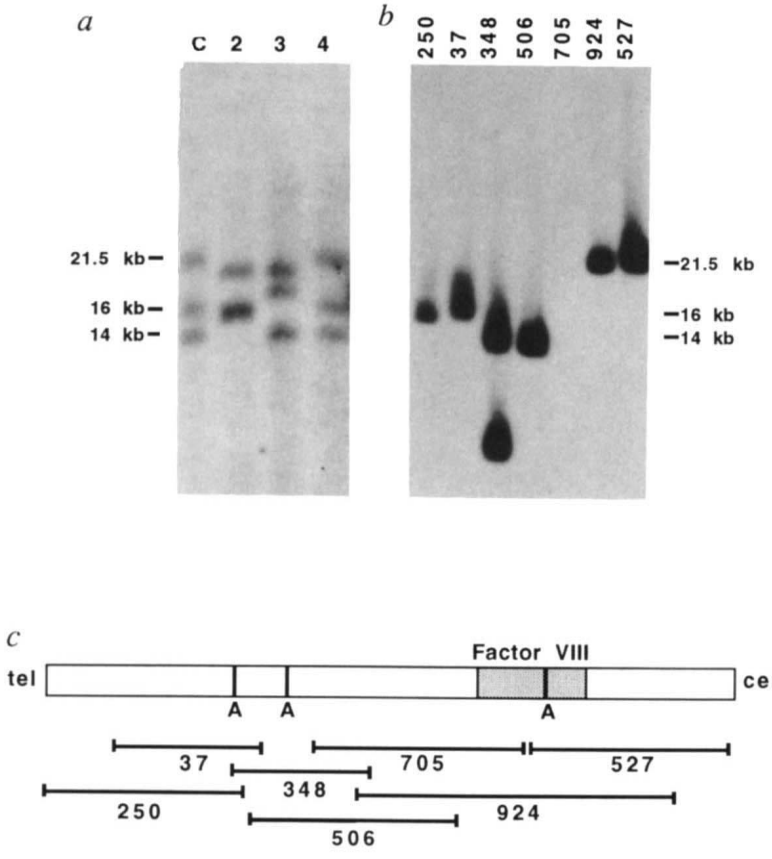

Fig. 4 Southern blot analysis of *Bcl*I fragments. Genomic DNAs (*a*) from a control individual (lane C), and from three haemophilic patients (lanes 2, 3 and 4, as described above) and YAC DNAs (*b*) were digested with *Bcl*I and hybridized to the gene A probe. *c* shows the regions encompassed by the YAC DNAs for the normal Xq28 region as described[10].

The Southern blots of Figs 3 and 4 also suggest that no DNA is lost or gained in samples 2 and 3. For example, the sum of the sizes of the altered *Bcl*I bands in sample 2 (20 and 15.5 kb) equals the sum of the sizes of the corresponding bands in the control sample (21. 5 and 14 kb). Likewise, the sum of the sizes of the altered *Bcl*I bands in sample 3 (20 and 17.5 kb) equals the sum of the sizes of the corresponding bands in the control sample (21.5 and 16 kb). These findings are consistent with an inversion generated by homologous recombination.

Identification of sequence adjacent to exon 22

Another prediction of the inversion

hypothesis is that the novel factor VIII transcript might include sequences from the newly juxtaposed region. Specifically, it is possible that exon 22, separated from its usual splicing partner, exon 23, splices instead to an exon located near the upstream A genes.

To test this possibility, we endeavoured to obtain the sequence adjacent to exon 22 in the transcript by RT-PCR (Fig. 5). As stated above, we reasoned that the oligonucleotide used for reverse transcription (16.2A) in the initial series of RT-PCR reactions must have primed at an imperfect target sequence in samples 2 and 3. But, once cDNA is synthesized, the 16.2A sequence is incorporated into the cDNA sequence. This sequence then provides a perfectly matched target sequence for subsequent PCR reactions involving 16.2A as the primer. Thus, cDNAs generated by RT priming with 16.2A were amplified in two rounds of hemi-nested PCR using EF and ENF in combination with primer 16.2A (Fig. 5*a*). A Southern blot of the products of these reactions was hybridized with an internal oligonucleotide (CR in exon 22) and a 700 basepair (bp) fragment containing exon 22 sequence was identified for sample 3 (results not shown). The sequence of this fragment includes the intact exon 22 sequence up to its normal 3' splice junction, juxtaposed to a sequence of unknown origin (GenBank accession No. U00684). This sequence does not correspond to any other portion of factor VIII, nor is it the adjacent intron 22 sequence. This finding suggested that in sample 3, exon 22 is spliced at the appropriate splice donor to an exon from a different gene or to a cryptic splice acceptor site.

Southern blot analysis revealed that the new sequence is located 5' to the factor VIII gene, in proximity to the upstream copies of the A genes (Fig. 5). A fragment containing only the unknown sequence was generated by PCR with specific primers 3F and TB (Fig. 5*a*). This probe, 3NA, was hybridized to the same blot of *BclI*-digested DNAs from YAC clones as used in Fig. 4*b*. Figure 5*b* shows that two hybridizing fragments (4.0 and 5.2 kb) are present in YACs containing the upstream A genes (yWXD250, yWXD37, yWXD348 and yWXD506), but absent in YACs that cover the factor VIII gene (yWXD705, yWXD924 and yWXD527). This result indicates that the sequence adjacent to exon 22 in the cDNA lies in the vicinity of the upstream A genes. The presence of two *BclI* bands, distinct from those containing the A genes, further indicates that the new sequence corresponds to at least one new exon. The observation of the same hybridizing bands in non-overlapping YACs (yWXD250 and yWXD506) suggests an extensive region of duplication in the vicinity of the upstream A genes, as described previously[10].

Analysis of other patient samples

Further studies showed that the abnormal *BclI* patterns in patients 2 and 3 are found in other severely affected haemophilia A patients. The gene A probe was hybridized to *BclI*-digested DNA from nine severe patients for whom mutations had not been found in a previous study after extensive analysis[5]. We anticipated that the hypothesized inversions would be present in most of these patients. Two of the samples exhibited the same pattern as patient 2 and four others exhibited the same pattern as patient 3 (data not shown). Two of the samples appeared to be the same as the control. The hybridization pattern in the ninth patient consisted of the pattern seen in patient 3, but with the addition of the normal 16-kb band. We have found previously that the number of A genes varies occasionally among individuals (unpublished observations), and we suggest that this extra band may represent an additional copy of the A gene.

Analysis of these patients also showed that each of the abnormal *BclI* patterns is associated with different four-site haplotypes for polymorphisms in the factor VIII gene[5]. For example, three different haplotypes (1, 3 and 4) were found in the four individuals with the patient 3 pattern, and the haplotypes for the two patients with the patient 2 pattern differ (types 3 and 4). These results suggest that each mutation arose as an independently occurring event.

To achieve a better estimate for the prevalence of the proposed inversions, a further 14 previously unstudied, unrelated, severe haemophilia A patients were examined by Southern blot analysis. An aberrant *BclI* pattern was observed in 7 of these. Surprisingly, all of these patients exhibited the patient 3 pattern, suggesting that recombination with the most distal A gene occurs more frequently than recombination with the more proximal A gene. These results, combined with the finding of abnormalities in two of the five severe patients analysed above, give an overall estimate of the frequency of the proposed inversions as 9 in 19, or 47% of all severe haemophilia A. This value agrees with the 42% found by Naylor *et al.*[6], implicating inversions as the basis for severe haemophilia A in most, if not all patients exhibiting the RT-PCR abnormality.

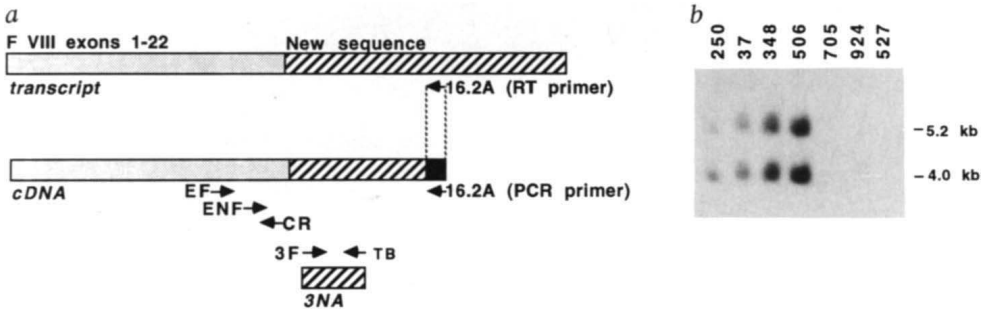

Fig. 5 Localization of new sequence in the novel factor VIII transcript. *a*, Strategy used to isolate the new sequence adjacent to exon 22 in the disrupted factor VIII transcript. The stippled region represents spliced exons 1–22, while the hatched region represents the adjacent sequence of interest. cDNA synthesis was primed with oligonucleotide 16.2A, and hemi-nested PCR was then conducted with primers EF and ENF in conjunction with 16.2A, each with an elongation time of 3 minutes. Primers (3F and TB) corresponding to the new sequence (3NA) were prepared to produce a smaller fragment containing the adjacent sequence only. *b*, Hybridization of the 3NA fragment to the same Southern blot of *BclI*-digested YAC DNAs shown in Fig. 4.

Discussion

We have described a model to account for the observation of a disruption of the exon 22/exon 23 junction in the factor VIII mRNA of many severe haemophilia A patients. We propose that the region encompassing gene A in intron 22 inappropriately pairs with one of the two homologous regions upstream of the factor VIII gene on the same chromosome. We suggest that this pairing is sometimes followed by homologous recombination,

resulting in a large inversion that disrupts the factor VIII gene. We have tested several predictions made by this model, and in each case the results are consistent with the hypothesis: (i) two of the three fragments containing the A genes are altered; (ii) one of the two altered fragments always contains the A gene in intron 22; (iii) no DNA appears to be lost or gained; (iv) the altered factor VIII transcript includes sequences from the vicinity of the upstream A genes. One untested prediction is that the upstream A genes are transcribed in the direction opposite to that of the intron-22 copy of gene A (Fig. 1). Only this arrangement would lead to an inversion by homologous, intrachromosomal recombination[12]. If the upstream and intron-22 A genes were oriented in the same direction, then an intrachromosomal recombination would lead to two products: a chromosome deleted for all material between the upstream and intron-22 A genes, and an excised closed circle containing the deleted material. These products were not observed.

Interchromosomal recombination between misaligned A genes was also rejected as the basis for the exon 22/23 disruption. Again, two orientations must be considered. If the genes are oriented in opposite directions, then a homologous interchromosomal recombination would generate one dicentric chromosome and one acentric fragment. If the genes are oriented in the same direction, then the two recombination products would be a chromosome deleted for all material between the upstream gene A and intron-22 gene A, and a reciprocal chromosome with a duplication of the intervening material, leaving one factor VIII gene untouched (the so-called "Lepore-like" and "anti-Lepore-like" products[13]). Such unequal crossing over due to misalignment of dispersed repeats is the probable basis for deletions causing steroid sulphatase deficiency[14] and duplications in Charcot-Marie-Tooth disease, type 1A[15]. None of the products of such an interchromosomal crossover are compatible with the RT-PCR phenomenon and the BclI patterns we describe. Although one can construct more elaborate models to account for an inversion, invoking misalignment and multiple interchromosomal crossovers, the intrachromosomal single-crossover model we propose in Fig. 1 is the most parsimonious explanation for the observations.

Although inversions have been described as the basis for some inherited diseases, we have not been able to find a precedent for the situation we describe herein. Inversions studied previously do not appear to use homologous sequences and are not typically found in more than one family. The most well-studied examples of these inversions also show that they are accompanied by deletions of flanking DNA[16–18]. Thus we suggest that it is the presence of the dispersed homologous sequences that makes the factor VIII gene particularly vulnerable to the proposed recombination. Sequence comparison of control and patient DNA at the inversion junctions will be needed for further support of this model. We also speculate that the proximity of the factor VIII gene to the telomere may augment the frequency of an intrachromosomal recombination due to possible lack of steric constraints in the region.

Our results show that a Southern blot assay can be used to detect the haemophilia-causing mutation for many cases (~45%) of severe haemophilia A. Our findings also suggest that the inversion detected by this assay is likely to occur *de novo* in many patients without a family history of haemophilia. As severe cases account for ~50% of haemophilia A[19], we suggest that inversions are likely to underlie at least 20% of all haemophilia A cases. Thus the use of this assay should greatly increase the ease, frequency, and accuracy with which carrier detection and prenatal diagnosis can be made. It offers two advantages for diagnosis over the RT-PCR procedure: previously stored blood or DNA samples can be assayed instead of RNA, and carrier status can be established definitively by the presence of heterozygous abnormal *BclI* bands, rather than by quantitative RT-PCR.

Note added in proof. As stated in the discussion, our model predicts that the upstream A genes are orientated in the direction opposite to that of the intron-22 A gene. We have been able to test this prediction for the most distal A gene by taking advantage of two independent YACs (yWXD250 and yWXD348) that fortuitously contain this gene near one of their arms. Southern blot analysis with a variety of restriction enyzymes and probes has confirmed this orentation in both YACs.

Methodology

Samples. Blood was collected from patients following informed consent. Lymphocytes were immortalized by transformation with Epstein-Barr virus (a modification of ref. 20). DNA was prepared in agarose blocks[21] and polyA$^+$ RNA was obtained using a Pharmacia Quick-Prep RNA extraction kit. Agarose blocks containing YAC DNA were prepared according to Carle & Olson[22].

Oligonucleotide sequences. Sequences of oligonucleotides used for RT-priming, PCR amplification, hybridization and sequencing are as follows: 16.2A (exon 24) 5'–CCTGAGGTCTCCAGGC-ATTACTCC; CF (exon17) 5'–CTCGAATATCCAGATGGAAG; CR (exon 22) 5'–GCCGTGAATAATCATTGGTG; CNF (exon 18) 5'–CAATGGCTACATAATGGATAC; CNR (exon 21) 5'–GATCC-AAGAAAAGGGCTCC; EF (exon 20) 5'–CTGGCCAGACTT-CATTATTC; ER (exon 24) 5'–GGTAAAGTAGGATGAAGCAG; ENF (exon 22) 5'–GGATCTGTTGGCACCAATG; ENR (exon 24) 5'–GAAGCAGTAATCTGTGCATC; RT24A (exon 24) 5'–CATTAC-TCCTCCCTTG; P26 (exon B) 5'–CTGTGAGCGGCGTATG-CAAATC; 3F (new sequence) 5'–CACGGTGTCTTGGAAAATG; TB (new sequence) 5'–GGCTGGCAACAGGTCTG.

RT-PCR. cDNA was obtained using an Invitrogen cDNA cycle kit, with primer 16.2A, RT24A, ER or random primers (supplied with the kit). PCR was carried out with AmpliTaq (Perkin Elmer-Cetus) in the buffer recommended by the manufacturer at 1.5 mM $MgCl_2$. Samples were denatured at 94 °C for 30 s. The primers were annealed at 53 °C for 30 s; elongation was carried out at 72 °C for varying times according to the length of the sequence being amplified (see Fig. legends). All PCR reactions were carried out for 35 cycles.

Southern blots. Agarose blocks were digested with restriction enzymes overnight and electrophoresed on either 0.4% or 0.6% agarose gels, which were then UV-nicked in a BioRad GS Gene Linker, and denatured and neutralized twice for 20 min each. The DNA was transferred to Hybond N or N$^+$ (Amersham) in 10 × SSC. Filters were then UV-crosslinked.

Gene A probe. The gene A probe is a 0.9 kb *EcoRI/SstI* fragment described previously[8]. It can be prepared from plasmid p482.6, obtainable from ATCC, catalogue no. 57203.

Hybridizations. Southern blots were prehybridized and hybridized at 65 °C in Church buffer (0.5 M $NaPO_4$, 7% SDS, 1 mM EDTA, 10% dextran sulphate and 100 mg ml^{-1} boiled salmon sperm DNA)[23]. The gene A and 3NA probes were labelled using an Amersham Megaprime labeling kit. One Southern blot that was probed with an oligonucleotide (CR, see Results) was prehybridized and hybridized

Acknowledgements

We thank C. Diamond, A. Metzenberg, P. de Moerloose, B. Levinson, S. Das, D. Freija, D. Schlessinger, M. Higuchi, P. Hutter, M. Morris, M. Gunthorpe, L. Kasch, C. Boehm, C. Krebs, M. Young, B. Jing and S. Whitney for their help and contribution to this work. This work was funded by grants from the NIH to J.G. and to H.H.K., Jr. and S.E.A. and a Swiss FNRS grant to S.E.A. J.G. is an assistant investigator with the Howard Hughes Medical Institute.

at 40 °C in 35% formamide, 6 × SSC, 1× Denhardt's, 0.5% SDS, 10% dextran SO_4, and 100 mg ml^{-1} boiled salmon sperm DNA. The CR oligo was end-labelled with T4 polynucleotide kinase. All probes were used at 2×10^6 cpm ml^{-1}.

Received 16 July; accepted 16 August 1993.

1. Verkerk, A.J.M.H. *et al.* Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **65**, 905–914 (1991).
2. Kunkel, L.M. Analysis of deletions in DNA from patients with Becker and Duchenne muscular dystrophy. *Nature* **322**, 73–77 (1986).
3. Tuddenham, E.G.D. *et al.* Haemophilia A: Database of nucleotide substitutions, deletions, insertions and rearrangements of the factor VIII gene. *Nucl. Acids Res.* **19**, 4821–4833 (1991).
4. Higuchi, M. *et al.* Molecular characterization of mild-to-moderate hemophilia A: Detection of the mutation in 25 of 29 patients by denaturing gradient gel electrophoresis. *Proc. natn. Acad. Sci. U.S.A.* **88**, 8307–8311 (1991).
5. Higuchi, M. *et al.* Molecular characterization of severe hemophilia A suggests that about half the mutations are not within the coding regions and splice junctions of the factor VIII gene. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7405–7409 (1991).
6. Naylor, J.A., Green, P.M., Rizza, C.R. & Giannelli, F. Analysis of factor VIII mRNA defects in everyone of 28 haemophilia A patients. *Hum. molec. Genet.* **2**, 11–17 (1993).
7. Gitschier, J. *et al.* Characterization of the human factor VIII gene. *Nature* **312**, 326–330 (1984).
8. Levinson, B., Kenwrick, S., Lakich, D., Hammonds, G. & Gitschier, J. A transcribed gene in an intron of the human factor VIII gene. *Genomics* **7**, 1–11 (1990).
9. Levinson, B., Kenwrick, S., Gamel, P., Fisher, K. & Gitschier, J. Evidence for a third transcript from the human factor VIII gene. *Genomics* **14**, 585–589 (1992).
10. Freije, D. & Schlessinger, D. A 1.6-Mb contig of yeast artificial chromosomes around the human factor VIII gene reveals three regions homologous to probes for the DXS115 locus and two for the DXYS64 locus. *Am. J. hum. Genet.* **51**, 66–80 (1992).
11. Patterson, M. *et al.* An intronic region within the human factor VIII gene is duplicated within Xq28 and is homologous to the polymorphic locus DXS115 (767). *Am. J. hum. Genet.* **44**, 679–685 (1989).
12. Smithies, O. Chromosomal rearrangements and protein structure. *Cold Spring Harb. Symp. Quant. Biol.* **29**, 309–319 (1964).
13. Efremov, G.D. Hemoglobins Lepore and anti-Lepore. *Hemoglobin* **2**, 197–233 (1978).
14. Yen, P.H. *et al.* Frequent deletions of the human X chromosome distal short arm result from recombination between low copy repetitive elements. *Cell* **61**, 603–610 (1990).
15. Pentao, L., Wise, C.A., Chinault, A.C., Patel, P.I. & Lupski, J.R. Charcot-Marie-Tooth type 1A duplication appears to arise from recombination at repeat sequences flanking the 1.5 Mb monomer unit. *Nature Genet.* **2**, 292–300 (1992).
16. Jones, R.W., Old, J.M., Trent, R.J., Clegg, J.B. & Weatherall, D.J. Major rearrangement in the human β-globin gene cluster. *Nature* **291**, 39–44 (1981).
17. Karathanasis, S.K., Ferris, E. & Haddad, I.A. DNA inversion within the apolipoproteins AI/CIII/AIV-encoding gene cluster of certain patients with premature atherosclerosis. *Proc. natn. Acad. Sci. U.S.A.* **84**, 7198–7202 (1987).
18. Kulozik, A.E., Bellan-Koch, A., Kohne, E. & Kleihauer, E. A deletion/inversion rearrangement of the β-globin gene cluster in a Turkish family with λβo-Thalassemia Intermedia. *Blood* **79**, 2455–2459 (1992).
19. Rizza, C.R. & Spooner, R.J.D. Treatment of haemophilia and related disorders in Britain and Northern Ireland during 1976–1980: report on behalf of the directors of haemophilia centres in the United Kingdom. *Brit. Med. J.* **286**, 929–933 (1983).
20. Katz, B.Z., Raab-Traub, N. & Miller, G. Latent and replicating forms of Epstein-Barr virus DNA in lymphomas and lymphoproliferative disease. *J. inf. Dis.* **160**, 589–598 (1989).
21. Kenwrick, S. & Gitschier, J. A contiguous, 3-Mb physical map of Xq28 extending from the color-blindness locus to DXS15. *Am. J. hum. Genet.* **45**, 873–882 (1989).
22. Carle, G.F. & Olson, M.V. An electrophoretic karyotype for yeast. *Proc. natn. Acad. Sci. U.S.A.* **82**, 3756–3760 (1985).
23. Church, G. & Gilbert, W. Genomic sequencing. *Proc. natn. Acad. Sci. U.S.A.* **81**, 1991–1995 (1984).