

# Recombination Breakpoints in the Human $\beta$ -Globin Gene Cluster

By Rachelle A. Smith, P. Joy Ho, John B. Clegg, Judith R. Kidd, and Swee Lay Thein

The human  $\beta$ -globin gene complex spans a region of 70 kb and contains numerous sequence variants. These variant sites form a 5' cluster (5'  $\beta$ -haplotype) and a 3' cluster (3'  $\beta$ -haplotype) with strong linkage disequilibrium among the sites within each cluster, but not between the two clusters. The 9-kb region between the 5' and 3' clusters has been estimated to have rates of recombination that are 3 to 30 times normal, and the region has therefore been proposed as a 'hotspot' of recombination. We describe three families with evidence of meiotic recombination within this 'hotspot' of the  $\beta$ -globin gene cluster and in which the cross-over breakpoints have been defined at the sequence level. In one

family, the recombination has occurred in the maternal chromosome within a region of 361 bp between positions -911 and -550 5' to the  $\beta$ -globin gene. In the other two families, the recombination has occurred in the paternal chromosome within a region of approximately 1,100 bp between positions -542 and +568 relative to the  $\beta$ -globin gene cap site. Both regions occur within the 2-kb region of replication initiation (IR) in the  $\beta$ -globin gene domain with no overlap. The IR region contains a consensus sequence for a protein (*Pur*), which binds preferentially to single-stranded DNA, a role implicated in recombination events.

© 1998 by The American Society of Hematology.

**R**ECOMBINATION BETWEEN homologous DNA sequences plays an important role in generating genetic diversity in all organisms. Although meiotic recombination occurs throughout the human genome, it does not occur randomly but appears to be concentrated in specific regions.<sup>1</sup> Such areas of relatively increased recombination frequency are present in the major histocompatibility complex (MHC),<sup>2,3</sup> where they are responsible for Ig class switching, near or within the Duchenne muscular dystrophy, insulin, collagen, and the  $\beta$ -globin gene loci.<sup>4-8</sup> The human  $\beta$ -globin gene region on chromosome 11p is one of the most intensively studied of all human loci; mutations here cause  $\beta$ -thalassemia and sickle cell anemia, which are among the most common genetic diseases in the world.<sup>9</sup> The earliest prenatal diagnosis of these disorders using DNA analysis relied on linkage disequilibrium of  $\beta$ -globin gene mutations with adjacent restriction fragment length polymorphisms (RFLPs).<sup>10,11</sup> This procedure always carries some degree of risk of recombination between the mutant allele and the RFLP. Indeed a meiotic recombination in the paternal  $\beta$  chromosome in one family has caused an error in the prenatal diagnosis of  $\beta$ -thalassemia using this approach.<sup>12</sup>

Over 20 RFLPs have now been identified in the  $\beta$ -globin locus.<sup>13</sup> These polymorphisms fall into two groups: a 34-kb 5' cluster that includes the *HindII*- $\epsilon$ , *HindIII*- $G\gamma$ , *HindIII*- $A\gamma$ , *HindII*- $\Psi\beta$ , *HindII*-3' $\Psi\beta$ , and *Taq I*-5' $\delta$  polymorphic sites, and a 19-kb 3' cluster that includes the *Hgi* AI- $\beta$ , *AvaII*- $\beta$  and *BamHI*- $\beta$  sites<sup>14</sup> (see Fig 1). Population genetic analysis showed that while sites within the two clusters show strong linkage disequilibrium to each other, no linkage disequilibrium exists between these two clusters and the polymorphic sites within the 9-kb region separating the two clusters are randomly associated.<sup>15</sup> Indeed, a recombination rate of 3 to 30 times greater than expected was found in the region, which is the site of 75% of the recombination events in the entire  $\beta$ -globin cluster.<sup>8</sup> Thus, this region has been implicated as a 'hotspot' for genetic recombination. Studies have identified four separate families with evidence of recombination events within this 'hotspot'.<sup>12,16-18</sup> The localization of the crossover events in these families has been limited by the number of informative RFLPs available, and the narrowest localization was approximately 10 kb.<sup>18</sup>

Several sequence elements have been proposed as candidate recombination signals, some of which have been identified in the 9-kb 'hotspot.' The 'hotspot' encompasses the 2-kb region of replication origin in the  $\beta$ -globin gene domain within which

is a 16-bp consensus sequence for the *Pur* element.<sup>19,20</sup> In addition, upstream of the initiation replication region, but still within the 'hotspot,' lies a 21-bp consensus sequence (with two base mismatches) that is unique to potential initiation regions.<sup>21</sup> It is proposed that both the *Pur* and the 21-bp element facilitates DNA replication and recombination by initiating the unwinding of duplex DNA or maintaining an open duplex. Here we report on three families with evidence of recombination events within the 'hotspot' in the  $\beta$ -globin cluster. Two (Greek Cypriot and Asian Indian) are previously unreported, while the third family of Amish origin was previously reported by Gerherd et al.<sup>16</sup> In an attempt to address the role of the various putative recombination signals in the 'hotspot,' we have undertaken direct sequence analysis to delineate the crossover region. Two distinct regions of crossover within the 9-kb 'hotspot' were defined, which suggests that recombination sites in the  $\beta$ -globin complex show a clustering pattern also observed in the human HLA class II region.<sup>3</sup> The breakpoint in each family occurs within the 2 kb of replication initiation in the  $\beta$ -globin gene domain.

## FAMILIES AND METHODS

**Recombinant families.** Three families (Greek Cypriot, Asian Indian, and Amish) were previously identified by typing of the RFLPs ( $\beta$ -haplotype analysis) in the  $\beta$ -globin gene cluster. All families analyzed included both parents and at least three offspring. The Amish family is part of a Centre d'Etude du Polymorphisme Humain (CEPH)

From MRC Molecular Haematology Unit, Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford, UK; Duke University School of Medicine, Durham, NC; and the Department of Genetics, School of Medicine, Yale University, New Haven, CT.

Submitted June 1, 1998; accepted July 28, 1998.

P.J.H. was a Nuffield Dominions Fellow; R.A.S. was supported by the Medical Research Fund, Oxford; and J.R.K. is supported in part by a grant (SBR-96 32509) from the National Science Foundation (USA). We thank the MRC, UK, for support.

Address reprint requests to Swee Lay Thein, MD, Medical Research Council Molecular Haematology Unit, Institute of Molecular Medicine, John Radcliffe Hospital, Headington, Oxford, OX3 9DS UK; e-mail: swee.thein@imm.ox.ac.uk.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. section 1734 solely to indicate this fact.

© 1998 by The American Society of Hematology.

0006-4971/98/9211-0037\$3.00/0

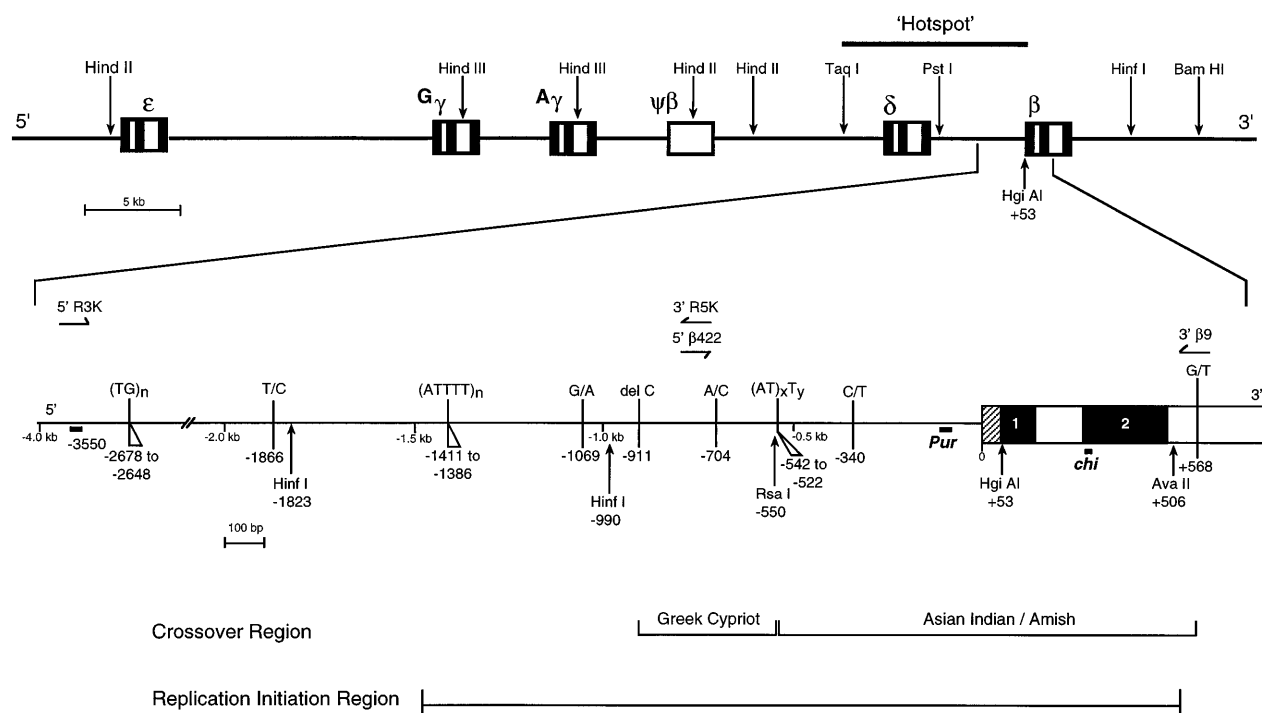
family. Informed consent was obtained in all cases before the collection of blood samples.

**$\beta$ -Haplotype analysis.** DNA was extracted from peripheral blood leukocytes or lymphoblastoid cell lines using standard procedures and analyzed by restriction enzyme digestion.  $\beta$  Haplotypes were derived from the following RFLPs in the  $\beta$ -globin gene cluster—*HindII- $\epsilon$* , *HindIII- $G\gamma$* , *HindIII- $A\gamma$* , *HindIII- $\Psi\beta$* , *HindI-3'  $\Psi\beta$* , *AvaII- $\beta$* , *BamHI- $\beta$* , *Taq I-5'  $\delta$* , *Pst I-3'  $\delta$* , *HinfI-5'  $\beta$* , *Rsa I-5'  $\beta$* , and *HgiAI- $\beta$* . The RFLPs were analyzed using a combination of standard Southern blot hybridization<sup>15</sup> and restriction enzyme analysis of DNA that was specifically amplified by the polymerase chain reaction (PCR) (see Fig 1). PCR amplification of the various polymorphic sites was performed in a total volume of 100  $\mu$ L containing 10 mmol/L Tris-HCl pH 8.3, 50 mmol/L KCl, 1.5 to 2.5 mmol/L MgCl<sub>2</sub>, 0.2 mmol/L each of dTTP, dGTP, dCTP, and dATP, 10 pmol of each primer, and 2.5 U of *Taq* polymerase (Cetus, Perkin Elmer, Warrington, UK). The sequences of the primers used and details of the conditions of amplification are available on request.

**DNA sequence analysis.** A 4.5-kb DNA fragment encompassing the 3'  $\delta$ - $\beta$  globin gene region, exons 1 and 2, and intron 1 of the  $\beta$ -globin gene was enzymatically amplified using two sets of primers, 5'R3K and 3'R5K and 5' $\beta$ 422 and 3' $\beta$ 9. Primers 5' R 3K (5'-AAT CTG TAC ATC AAG ACC CAG TGA TAT G-3') and 3' R5K (5'-GAC ATC TAA CTG TTT CTG CCT GGA CT-3') corresponding to GenBank (HUMHBB U01317) coordinates 58299-58326 and 61315-61290, respectively, direct the amplification of a 3,016-bp fragment in the  $\delta$ - $\beta$  region (Fig 1). PCR amplification was performed in a 100- $\mu$ L reaction volume containing 0.2 mmol/L of each dNTP, 50 mmol/L KCl, 10 mmol/L Tris-HCl pH 8.3, 2.0 mmol/L MgCl<sub>2</sub>, 2.5 U *Taq* polymerase (Cetus), and 20 pmol of each primer 5' R3K and 3' R5K. After an initial

denaturation of 4 minutes at 95°C, 30 cycles of denaturation at 95°C for 1 minute, annealing at 64°C for 2 minutes, and extension at 72°C for 3 minutes were performed, the last extension reaction was prolonged for 8 minutes. An aliquot of the PCR product was examined on a 1.0% agarose gel. The 3.0-kb amplification product was isolated and purified using a QIAEX II Gel extraction kit (Qiagen Ltd, Surrey, UK) after electrophoresis in a 1.0% agarose gel. The purified PCR product was resuspended in 40  $\mu$ L 10 mmol/L Tris-HCl pH 8.3. A small aliquot (3/40  $\mu$ L) of the DNA was examined in a 1% agarose gel to check for purity and concentration. Fifty to 500 ng (generally 6 of the 40  $\mu$ L) of the product was directly sequenced using the thermal cycle sequencing technique with <sup>33</sup>P-labeled terminators and *Thermosequenase* (Amersham, Buckinghamshire, UK). A total sequence of 3,016 nucleotides was determined using a series of forward and reverse primers.

Primers 5'  $\beta$ 422 (5'-TCC AGG CAG AAA CAG TTA GAT GTC-3') and 3'  $\beta$ 9 (5'-CAT TCG TCT GTT TCC CAT TCT A-3') corresponding to GenBank coordinates 61292-61315 and 62745-62724, respectively, direct the amplification of a 1,453-bp fragment encompassing the 5' end of the  $\beta$ -globin gene from position -845 through to position 609 relative to the cap site including exon 1, intron 1, and exon 2 of the  $\beta$ -globin gene (Fig 1). PCR amplification was performed in a 100- $\mu$ L reaction volume containing 0.2 mmol/L of each dNTP, 10 mmol/L Tris-HCl pH 8.3, 50 mmol/L KCl, 2.5 mmol/L MgCl<sub>2</sub>, 2.5 U of *Taq* polymerase (Cetus), and 20 pmol of each primers 5' $\beta$ 422 and 3' $\beta$ 9. After an initial denaturation of 4 minutes at 94°C, 30 cycles of denaturation at 94°C for 1 minute, annealing at 59°C for 2 minutes, and extension at 72°C for 3 minutes were performed, the last extension reaction was prolonged for 7 minutes. After checking for amplification, the PCR product was isolated using a QIAEXII Gel extraction kit



**Fig 1.** Map of the  $\beta$ -globin gene cluster with the 9-kb 'hotspot' delineated by the *TaqI*-5'  $\delta$  (5') and the *HgiAI*- $\beta$  (3') sites. RFLP sites are indicated by the arrows. The 4.5-kb fragment encompassing the 5' flanking  $\beta$  region is amplified using two sets of primers: 5'R3K versus 3'R5K and 5' $\beta$ 422 and 3' $\beta$ 9. The other polymorphic sites are indicated by vertical lines. The *pur*, *chi*, and the 21-bp element at position -3550<sup>21</sup> are represented by thick horizontal lines. The crossover regions and the region of initiation replication (IR) are indicated below. A total of 11 and 14 primers were used to sequence the forward and reverse strand of the 4.5-kb region, respectively. The sequences and locations of these primers are available on request.

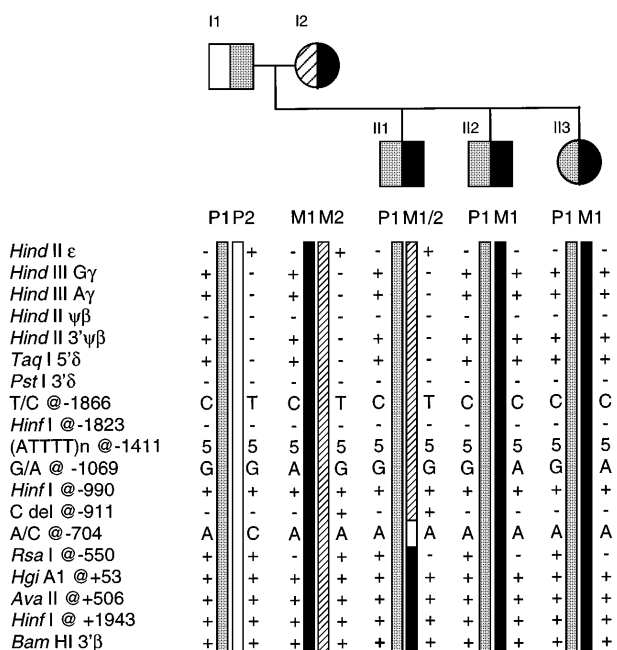
(Qiagen Ltd, Surrey, UK) as before. The purified DNA fragment was directly sequenced by the thermal cycle sequencing reaction using <sup>33</sup>P-labeled terminators and a series of forward and reverse primers.

**DNA fingerprinting.** To exclude false parentage, DNA from each member of the families was digested with *Hinf*I, Southern blotted, and hybridized with a panel of seven probes known to be specific for hypervariable minisatellites (MS1, MS8, MS29, MS31, MS43, MS51, P $\lambda$ g3).<sup>22</sup> These hypervariable loci are extremely variable with heterozygosities ranging from 90% for MS8 to 99% for MS31 and are dispersed over four autosomes. The locus-specific minisatellites act as very sensitive hybridization probes for these loci and can be pooled to detect the hypervariable (HVR) loci simultaneously giving rise to multilocus Southern blot patterns that are highly individual-specific (DNA fingerprints). Based on the heterozygosities and the average number of DNA fragments (8.4) resolved using a mixture of five probes, the chance that all of the fragments in one individual are present in a second randomly selected individual has been estimated at  $<6 \times 10^{-7}$ .<sup>22</sup> Therefore, these patterns provide a high level of individual specificity.

RESULTS

**Greek Cypriot family.** In this family, both parents were heterozygous for different  $\beta$ -thalassaemia mutations ( $\beta^+ 33 C \rightarrow G$  in I1 and  $\beta^o 39 C \rightarrow T$  in I2), and all three offspring were compound heterozygotes<sup>23</sup> (Fig 2).

The  $\beta$  haplotypes for each member of the family were derived from the results of 14 RFLPs (*Hind*II- $\epsilon$ , *Hind*III- $G\gamma$ ,



**Fig 2.** The Greek Cypriot family illustrating segregation of the  $\beta$  haplotypes and informative recombination events. Both parents are heterozygotes ( $\beta^+ 33 C \rightarrow G$  in I1 and  $\beta^o 39 C \rightarrow T$  in I2) and all three offspring, compound heterozygotes for  $\beta$ -thalassaemia.  $\beta$  Haplotypes P1 and M1 are associated with the paternal and maternal  $\beta$ -thalassaemia alleles, respectively. While individuals II2 and II3 have inherited the P1 and M1 haplotypes intact, individual II1 has inherited P1 and a maternal recombinant chromosome, as identified by the haplotype M1/M2. The site of recombination is located within the 361-bp interval between positions -911 and -550 relative to the  $\beta$ -globin gene cap site.

*Hind*III- $A\gamma$ , *Hind*II- $\Psi\beta$ , *Hind*II- $3'\Psi\beta$ , *Ava*II- $\beta$ , *Bam*HI- $\beta$ , *Taq* I- $5'\delta$ , *Pst* I- $3'\delta$ , *Hinf*I- $5'\beta$  at positions -1823 and -990, *Rsa* I- $5'\beta$ , *Hgi*AI- $\beta$ , and *Hinf*I- $3'\beta$  (see Fig 2). There are four  $\beta$  haplotypes in this family, the phase of the RFLPs was established using the two virtually homozygous offspring, II2 and II3. Both II2 and II3 have inherited the  $\beta$ -thalassaemia alleles from the father (I1) and the mother (I2) associated with haplotypes P1 and M1, respectively. III1, who is also a compound heterozygote for  $\beta$ -thalassaemia, has inherited an intact paternal chromosome, P1. However, while III1 has inherited the maternal  $\beta$ -thalassaemia mutation, the  $\beta$ -thalassaemia allele in III1 is associated with the 5' haplotype of M2 and the 3' haplotype of M1. On the basis of RFLP analysis, the M2 chromosome is intact 5' up to the *Taq* I- $5'\delta$  site, and the M1 chromosome is intact 3' up to the *Rsa* I- $5'\beta$  site, but the origin of the variant sites in the 10.4-kb segment between these two sites, including the *Pst* I- $3'\delta$  and *Hinf*I- $5'\beta$ , is not clear.

Direct sequence analysis of the area between the *Taq* I- $5'\delta$  and *Rsa* I- $5'\beta$  site was undertaken to search for additional sequence polymorphisms. The analysis identified several polymorphisms; the T/C polymorphism at position -1866, G/A polymorphism at position -1069, deletion of a single C at position -911 (a novel polymorphism), and A/C polymorphism at position -704 from the cap site of the  $\beta$ -globin gene (see Fig 2). Analysis of the (TG)<sub>n</sub> repeat<sup>24</sup> was not attempted as it lay outside the area of interest.

The direct sequence analysis established that the proband, III1, had inherited an intact maternal M2 chromosome from *Hind*II- $\epsilon$  to the C deletion at position -911 (5' to 3'), and an intact M1 chromosome from *Bam*HI- $\beta$  to the *Rsa* I- $5'\beta$  site at position -550 relative to the  $\beta$ -globin gene cap site (3' to 5') (Fig 3). The segment bordered by these two sites is 361 bp; no informative polymorphisms were found within this segment.

**Asian Indian family.** The  $\beta$  haplotypes of both parents and the three offspring were generated from 13 RFLPs: *Hind*II- $\epsilon$ , *Hind*III- $G\gamma$ , *Hind*III- $A\gamma$ , *Hind*II- $\Psi\beta$ , *Hind*II- $3'\Psi\beta$ , *Taq* I- $5'\beta$ , *Hinf*I- $5'\beta$  at position -1823, *Hinf*I- $5'\beta$  at position -990, *Rsa* I- $5'\beta$ , *Hgi*AI- $\beta$ , *Ava*II- $\beta$ , *Hinf*I- $3'\beta$ , and *Bam*HI- $\beta$  (see Fig 4). In this family, the inheritance of the  $\beta$ -alleles is not associated with thalassaemia. Phase of the RFLPs was established in the children based on the phase in the mother, I2, who was homozygous for most of the RFLPs. Two of the children, II2 and II3, have inherited intact chromosomes, P2 and M2, from their father (I1) and mother (I2), respectively. The proband, II1, has inherited an intact chromosome (M1) from the mother but the paternally inherited chromosome comprised of the 5' end of P1 (up to *Hind*I- $3'\Psi\beta$ ) and the 3' end of P2 (up to *Hinf*I- $3'\beta$ ). The origin of the chromosome in the 15-kb segment between the *Hind*II- $3'\Psi\beta$  and *Hinf*I- $3'\beta$  sites is not clear from RFLP analysis because the sites within it are uninformative both in the father and proband.

Sequence analysis of the 5' of the  $\beta$ -globin gene showed several polymorphisms, including the T/C polymorphism at position -1866, (ATTTT)<sub>n</sub> at position -1411, G/A polymorphism at position -1069, A/C polymorphism at position -704, (AT)<sub>x</sub>T<sub>y</sub> at positions -542 to -522, and G/T polymorphism in intron 2 (at position +568 from the  $\beta$ -globin gene cap site).

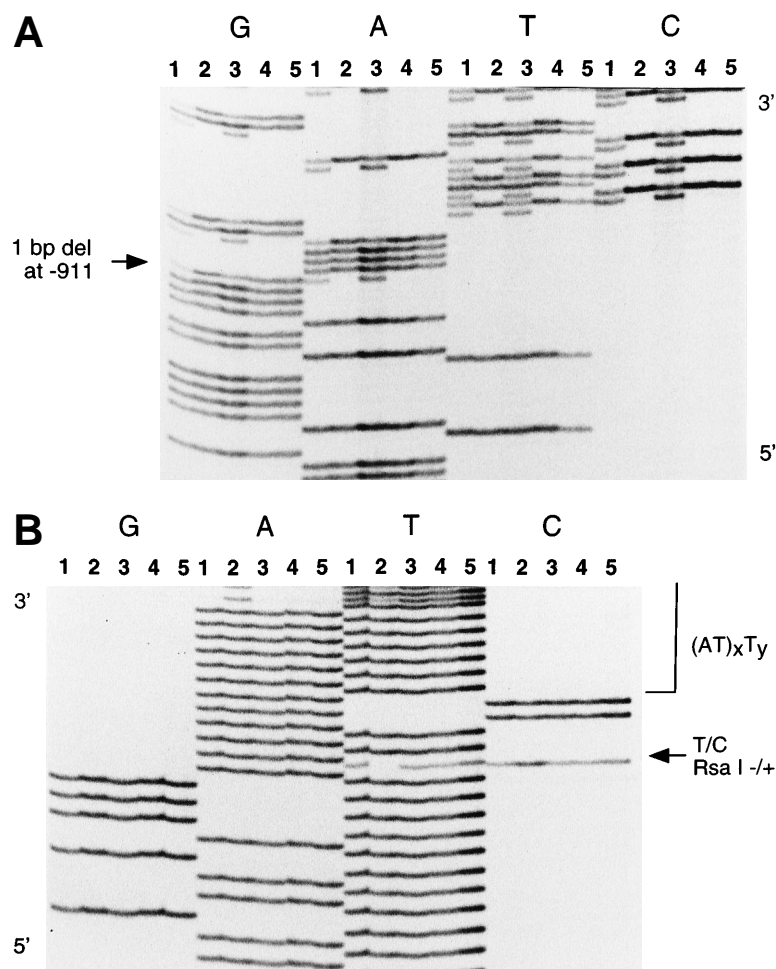


Fig 3. DNA sequence immediately 5' of the  $\beta$ -globin gene in the Greek Cypriot family. The individuals II1, I1, I2, II2, and II3 (as shown in Fig 2) are represented by lanes 1, 2, 3, 4, and 5. The termination reactions are loaded in consecutive lanes in a block, ddGTP (G), ddATP (A), ddTTP (T), and ddCTP (C). (Panel A) The 5' breakpoint indicated by the 1-bp deletion (arrowed) at position -911; (Panel B) the 3' breakpoint indicated by the arrow (*Rsa* I site) at position -550.

The sequence analysis established that III had inherited the 5' P1 chromosome intact up to the  $(AT)_xT_y$  region and that the 3' P2 chromosome was inherited intact up to the G/T polymorphism in intron 2 of the  $\beta$ -globin gene (Fig 5).

**Amish family.** Previous study on the third family, of Amish descent, by Gerhard et al,<sup>16</sup> had established that a single recombination event occurred in the paternal chromosome of II2 between the *Taq* I-5' $\delta$  and *Bam*HI-3' $\beta$  sites, within a region of 15 kb (see Fig 6), which was confirmed by our RFLP analysis of the same sites. Our analysis of additional RFLPs suggested that the putative recombination occurred within a 2.5-kb segment between the *Rsa* I-5' $\beta$  and the *Hinf*I-3' $\beta$  sites. There are three haplotypes in this family; the parents I1 and I2 share a common  $\beta$  haplotype (M1 and P1).

Direct sequence analysis proved that the family was informative at the same sequence polymorphisms as in the Indian family with similar recombination breakpoints, ie, the  $(AT)_xT_y$  site at positions -542 to -522 and the G/T polymorphism at position +568.

There are four possible explanations for the recombinant chromosomes in the three families: (1) a meiotic recombination event has occurred within the  $\beta$ -globin cluster; (2) false parentage; (3) point mutations at multiple sites; or (4) gene

conversion. False parentage had been previously excluded in the Amish family.<sup>16</sup> In the Greek-Cypriot family, Southern blot hybridization of *Hinf*I digest with a pool of seven hypervariable minisatellite probes showed 10 bands in the father, 14 in the mother, 13 in II1, 9 in II2, and 11 in II3. Of the 13 bands in II1, 6 were shared with father and 7 with mother. Of the 9 bands in II2, 5 were shared with father and 4 with mother. Of the 11 bands in II3, 6 were shared with father and 5 with mother. Similarly, in the Asian Indian family, 12 bands could be scored in the father, 10 in the mother, 11 in II1, 10 in II2, and 12 in II3. In II1, 5 of the 11 bands were shared with father, 6 with mother; in II2, 5 of the 10 bands were shared with the father and 5 with the mother, while in II3, 6 of the 12 bands were shared with father and the other 6 with mother. In all cases, all the alleles present in the offspring could be traced to either of the parents, indicating that the genetic relationships have been correctly assigned with no evidence of false paternity or maternity. The third explanation is unlikely because it requires several independent point mutations at the different RFLP sites and, in the case of the Greek family, including the C-T mutation at codon 39 of the  $\beta$ -globin gene to create a  $\beta$ -thalassemia mutation identical to the one in the mother. The fourth explanation is not ruled out but seems unlikely as it would involve gene conversion of a large stretch

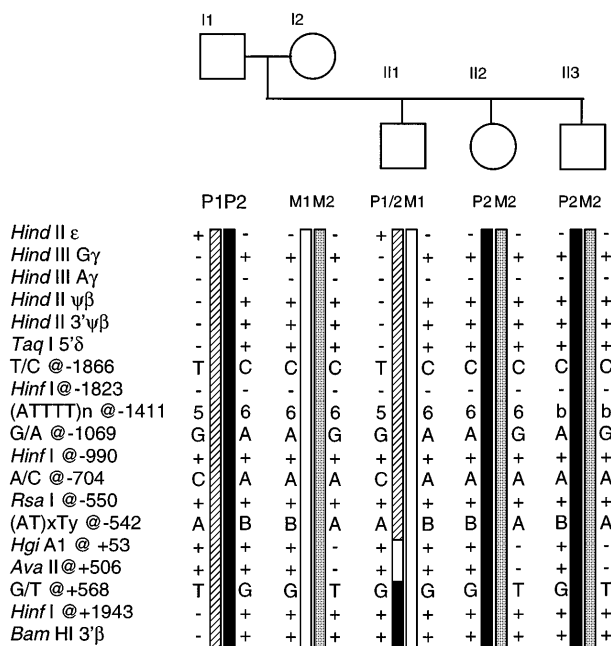


Fig 4. The Asian Indian family showing segregation of the β haplotypes and informative recombination events. Individual II1 has inherited a paternal recombinant chromosome, as identified by the haplotype P1/P2. The site of recombination is located within a region of ~1,100 bp between the (AT)<sub>x</sub>T<sub>y</sub> repeats at position -542 and the G/T polymorphism at position +568. (AT)<sub>x</sub>T<sub>y</sub> at position -542 is represented by A: (AT)<sub>7</sub>T<sub>7</sub> and B: (AT)<sub>9</sub>T<sub>5</sub>.

of DNA (~22 kb) in the Amish and Asian Indian families. In the Greek Cypriot family a conversion of ~1 kb of M2 by M1, from 3' of the A/C at position -704 to 3' of the thalassemia mutation at codon 39, could explain the composite maternal chromosome in III.

The most likely explanation is that a meiotic recombination has occurred within the β-globin gene cluster, between the C deletion at position -911 and the *Rsa* I site at position -550 in the Greek family, and between the (AT)<sub>x</sub>T<sub>y</sub> repeats at position -542 and the G/T site at position +518 in the Asian Indian and Amish families.

DISCUSSION

β Haplotypes generated from RFLPs have led to the identification of a recombination event in the β-globin gene cluster in two families, bringing the total reported to six. We then used DNA sequence analysis to delineate the region of crossover to 361 bp in one family and 1,100 bp in the other family. The region of crossover was also sequenced in a previously reported family and this corresponded to the same region of 1,100 bp. These represent the most precisely defined region of recombinations within the human β-globin gene cluster.

The 9-kb region immediately 5' to the β-globin gene has been proposed as a recombination 'hotspot' with recombination rates of 3 to 30 times higher than those of the surrounding regions. The 'hotspot,' as identified by Chakravarti et al,<sup>8</sup> extends from the *Taq* I-5'δ site to the *Hgi*AI-β site (Fig 1). Although the crossover region falls within this 'hotspot' in one

family (Greek Cypriot), the 3' boundary of the crossover region in the other two families extends ~500 bp 3' beyond the 'hotspot.' However, in the absence of any further polymorphisms within the 1,100-bp segment, it is quite possible that the cross-over region in the Amish and Asian Indian families falls within the boundaries of the 9 kb 'hotspot.'

Previous studies have suggested several potential signal sequences that could enhance recombination within the 'hotspot.'<sup>25</sup> These included the (ATTT)<sub>n</sub>, (TG)<sub>n</sub>, and (AT)<sub>x</sub>T<sub>y</sub> repeats. In all three families, the crossover regions are located 3' of the (TG)<sub>n</sub><sup>24</sup> and (ATTTT)<sub>n</sub> repeats which lie ~1,800 bp and ~500 bp, respectively, upstream of position -911 (the 5' crossover breakpoint in the Greek family) which makes these repeats as unlikely candidates initiating recombination, at least in these three families. Interestingly, a previous study of sequence-specific meiotic recombination in *Saccharomyces cerevisiae* comparing three adjacent restriction fragments immediately 5' of the human β-globin gene showed that deletion of the (TG.CA)<sub>n</sub> sequences at the -2678 to -2648 interval does not significantly reduce the high frequency of genetic recombination in this region.<sup>26</sup>

It is of note that in all three families, the crossover region lies within the small well-defined 2-kb fragment (between positions -1461 and +476) which contains the replication origin of the β-globin cluster.<sup>19,27</sup> This initiation region (IR) contains a 16-bp consensus sequence (5'-GGNNGAGGGAGARRRR-3' at positions -63 to -48) for the *Pur* protein, which has been shown to have preferential binding for single-stranded DNA.<sup>20</sup> On this basis it has been suggested that *Pur* would be expected to function as a helix-destabilizing protein, especially at the position of the *Pur* element, a role implicated in initiation of replication and recombination. The *Pur* element is located within the 1,100-bp crossover region. Although the *Pur* element lies outside the crossover region in the Greek family, it is still possible that it is involved in the initiation of recombination/replication as initial duplex opening is not restricted to a single site but can occur throughout within each initiation zone.<sup>28</sup> Evidence to support this possibility includes, firstly: in many mammalian loci where the *Pur* element has been identified, it is located at least 1 kb away from the actual start site of replication.<sup>20</sup> Secondly, it has been observed in primate cells that the binding of replication initiator protein *dnaA*, for *oric* of *Escherichia coli*, destabilizes the DNA helix to promote duplex opening at a region removed from the original recognition site.<sup>29</sup> Frequently, this region is a reiterated AT-sequence,<sup>30</sup> such as the (AT)<sub>x</sub>T<sub>y</sub> repeats, which forms the 5' breakpoint of the 1,100-bp crossover regions. Furthermore, the 3' breakpoint of the 361-bp crossover region also lies within an AT-rich sequence.

Another consensus DNA sequence implicated in the initiation of DNA replication and recombination is the 21-bp 5'-WAWTTDDWWDHWGWHMAWTT-3' where M = A or C, W = A or T, D = A or G or T, and H = A or C or T.<sup>21</sup> This consensus is found, with two base mismatches, within the 'hotspot' at position -3550 upstream of the β-globin gene. However, it is located ~1 kb outside the IR region and certainly outside the crossover regions in the three families. The 1,100-bp crossover region also contains a chi (χ) sequence (5'-

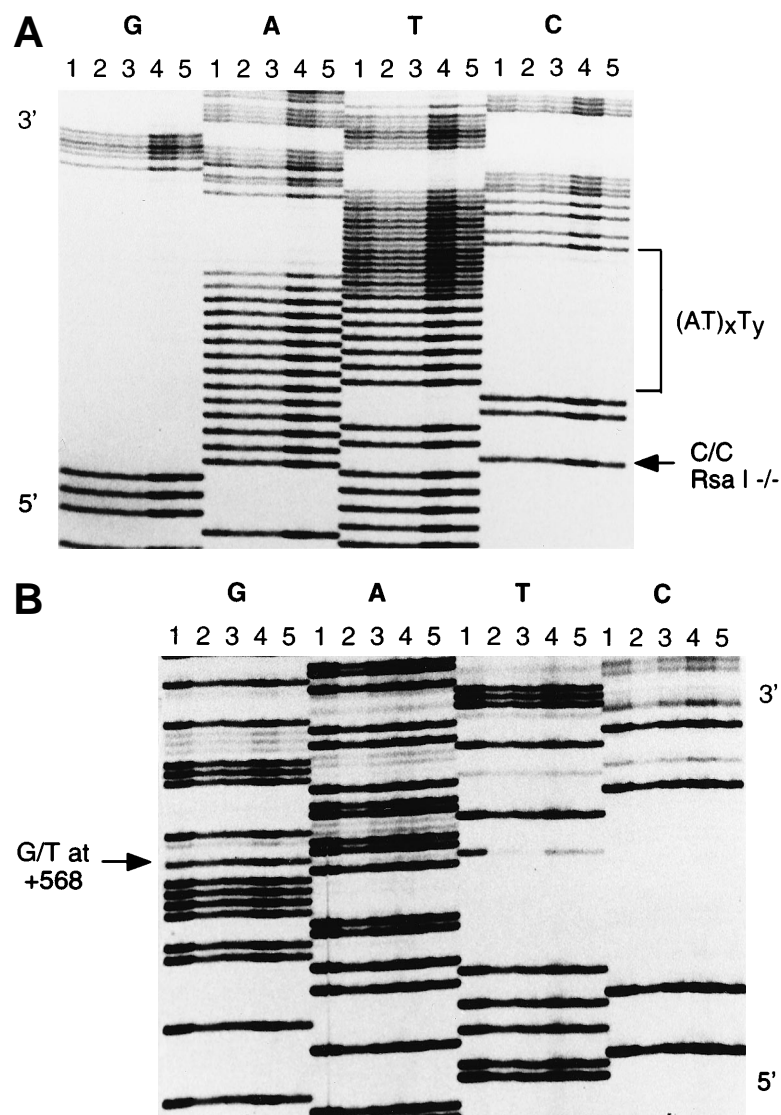


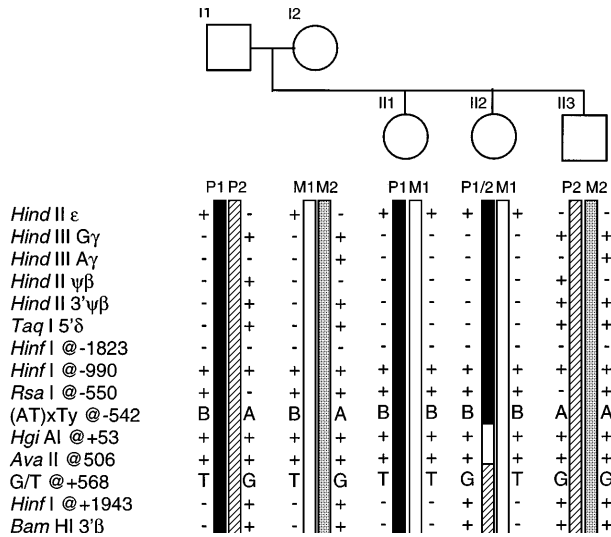
Fig 5. DNA sequence of the 5' region of the  $\beta$ -globin gene in the Asian Indian family. The individuals I1, I2, II1, II2, and II3 (as in Fig 4) are represented by lanes 1, 2, 3, 4, and 5, respectively. The termination reactions are loaded in consecutive lanes in a block, ddGTP (G), ddATP (A), ddTTP (T), and ddCTP (C). (a) The 5' breakpoint of the crossover region is located in the region of the (AT)<sub>x</sub>T<sub>y</sub> repeats; (b) the 3' breakpoint is located at the G/T polymorphic site at position +568 (arrowed).

GCTGGTGG-3') at ~300 bp 5' of the G/T polymorphic site in exon 2 of the  $\beta$ -globin gene. The  $\chi$  sites are known to be hotspots for homologous recombination in *E coli*, increasing recombination by 5- to 10-fold.<sup>25,31,32</sup> However, the exact role of these elements in recombination remains unclear.

Although all the three crossover regions fall within the broad 9-kb 'hotspot,' sequence analysis indicates that the crossover events occurred at distinct sites within this area. It is possible that this 9-kb hotspot contains several smaller areas with increased rates of recombination that are nonrandomly distributed. This suggests a clustering of recombination breakpoints such as has been observed in the MHC complex where the crossover breakpoints occurred within several defined areas associated with an increased frequency of recombination, all located within the larger 'hotspot.'<sup>13</sup> Despite sequence analysis, only an estimate of the minimal crossover regions can be obtained, the limiting factor being the presence of informative polymorphisms. In one family this region is 361 bp and 1,100 bp in the other two families. In contrast, sequence analysis of a

crossover site was defined to a 138-bp segment in a recombination event in the HLA class II region.<sup>33</sup>

Although only six families with recombination events within the  $\beta$ -globin cluster have been reported to date, in all six the crossovers occurred within the region immediately 5' to the  $\beta$ -globin gene, which suggests that the frequency of recombination in this area is higher than those in the rest of the  $\beta$  complex. These observations support the results of a recent study of allelic sequence diversity at the human  $\beta$ -globin locus which concluded that recombination and gene conversion in the 5' region of the  $\beta$ -globin gene has contributed to the  $\beta$ -haplotype diversity, which was higher than expected on the basis of the observed nucleotide polymorphisms.<sup>34</sup> However, to fully appreciate the frequency of recombination, ie, to establish if this is truly a 'hotspot' or just a 'warm patch' in cold surroundings, a systematic study of the  $\beta$ -globin cluster compared with its flanking regions on chromosome 11p involving sperm DNA typing or a larger number of families (such as the other CEPH pedigrees) is needed.



**Fig 6.** The Amish family showing segregation of the  $\beta$  haplotypes and informative recombination events. Individual II2 has inherited a paternal recombinant chromosome as identified by the haplotype P1/P2. The site of recombination is located within a region of  $\sim$ 1,100 bp between the (AT) $x$ T $y$  repeats at position -542 and the G/T polymorphic site at position +568. (AT) $x$ T $y$  is represented by A: (AT) $_{7}$ T $_{7}$  and B: (AT) $_{9}$ T $_{5}$ .

ACKNOWLEDGMENT

We thank Dr John Old for the gift of some of the PCR primers; Prof Sir D.J. Weatherall for encouragement and support; and Liz Rose and Milly Graver for preparation of the manuscript.

REFERENCES

- Fisher-Lindahl K: His and hers recombinational hot-spots. *Trends Genet* 7:273, 1991
- Bouissou C, Pontarotti P, Crouau-Roy B: A precise meiotic map in the class I region of the human major histocompatibility complex. *Genomics* 30:486, 1995
- Cullen M, Noble J, Erlich H, Thorpe K, Beck S, Klitz W, Trowsdale J, Carrington M: Characterisation of recombination in the HLA Class II region. *Am J Hum Genet* 60:397, 1997
- Grimm T, Müller B, Dreier M, Kind E, Bettecken T, Meng G, Müller CR: Hot spot of recombination within DXS164 in the Duchenne muscular dystrophy gene. *Am J Hum Genet* 45:368, 1989
- Oudet C, Hanauer A, Clemens P, Caskey T, Mandel JL: Two hot spots of recombination in the DMD gene correlate with the deletion prone regions. *Hum Mol Genet* 1:599, 1992
- Lebo RV, Charavarti A, Buetow KH, Cheung M-C, Cann H, Cordell B, Goodman H: Recombination within and between the human insulin and  $\beta$ -globin gene loci. *Proc Natl Acad Sci USA* 80:4808, 1983
- Bowcock AM, Hebert JM, Wijsman E, Gadi I, Cavalli-Sforza LL, Boyd CD: High recombination between two physically close human basement membrane collagen genes at the distal end of chromosome 13q. *Proc Natl Acad Sci USA* 85:2701, 1988
- Chakravarti A, Buetow KH, Antonarakis SE, Waber PG, Boehm CD, Kazazian HH: Nonuniform recombination within the human  $\beta$ -globin gene cluster. *Am J Hum Genet* 36:1239, 1984
- Weatherall DJ, Clegg JB: Thalassemia—a global public health problem. *Nat Med* 2:847, 1996
- Old JM, Ward RHT, Petrou M, Karagözü F, Modell B, Weatherall DJ: First-trimester fetal diagnosis for haemoglobinopathies: Three cases. *Lancet* 2:1413, 1982
- Boehm CD, Antonarakis SE, Phillips JA III, Stetten G, Kazazian HH Jr: Prenatal diagnosis using DNA polymorphisms. Report on 95 pregnancies at risk for sickle-cell disease or  $\beta$ -thalassaemia. *N Engl J Med* 308:1054, 1983

- Camaschella C, Serra A, Saglio G, Tiziana Bertero M, Mazza U, Terzoli S, Brambati B, Cremonesi L, Travi M, Ferrari M: Meiotic recombination in the  $\beta$  globin gene cluster causing an error in prenatal diagnosis of  $\beta$  thalassaemia. *J Med Genet* 25:307, 1988
- Antonarakis SE, Kazazian HHJ, Orkin SH: DNA polymorphism and molecular pathology of the human globin gene clusters. *Hum Genet* 69:1, 1985
- Collins FS, Weissman SM: The molecular genetics of human hemoglobin. *Prog Nucleic Acid Res Mol Biol* 31:315, 1984
- Antonarakis SE, Boehm CD, Giardinia PJV, Kazazian HHJ: Non-random association of polymorphic restriction sites in the  $\beta$ -globin gene cluster. *Proc Natl Acad Sci USA* 79:137, 1982
- Gerhard DS, Kidd KK, Kidd JR, Egeland JA, Housman DE: Identification of a recent recombination event within the human  $\beta$ -globin gene cluster. *Proc Natl Acad Sci USA* 81:7875, 1984
- Old JM, Heath C, Fitches A, Thein SL, Jeffreys AJ, Petrou M, Modell B, Weatherall DJ: Meiotic recombination between two polymorphic restriction sites within the  $\beta$  globin gene cluster. *J Med Genet* 23:14, 1986
- Hall GW, Sampietro M, Barnetson R, Fitzgerald J, McCann S, Thein SL: Meiotic recombination in an Irish family with beta-thalassaemia. *Hum Genet* 92:28, 1993
- Kitsberg D, Selig S, Keshet I, Cedar H: Replication structure of the human  $\beta$ -globin gene domain. *Nature* 366:588, 1993
- Bergemann AD, Johnson EM: The HeLa Pur factor binds single-stranded DNA at a specific element conserved in gene flanking region and origins of DNA replication. *Mol Cell Biol* 12:1257, 1992
- Dobbs DL, Shaiu W-L, Benbow RM: Modular sequence elements associated with origin regions in eukaryotic chromosomal DNA. *Nucleic Acids Res* 22:2479, 1994
- Wong Z, Wilson V, Patel I, Povey S, Jeffreys AJ: Characterization of a panel of highly variable minisatellites cloned from human DNA. *Ann Hum Genet (Lond)* 51:269, 1987
- Ho PJ, Rochette J, Fisher CA, Wonke B, Jarvis MK, Yardumian A, Thein SL: Moderate reduction of  $\beta$  globin gene transcript by a novel mutation in the 5' untranslated region: A study of its interaction with other genotypes in two families. *Blood* 87:1170, 1996
- Miesfeld R, Krystal M, Arnheim N: A member of a new repeated sequence family which is conserved throughout eucaryotic evolution is found between the human  $\delta$  and  $\beta$  globin genes. *Nucleic Acids Res* 9:5931, 1981
- Steinmetz M, Uematsu Y, Lindahl KF: Hotspots of homologous recombination in mammalian genomes. *Trends Genet* 3:7, 1987
- Treco D, Thomas D, Arnheim N: Recombination hot spot in the human  $\beta$ -globin gene cluster: Meiotic recombination of human DNA fragments in *Saccharomyces cerevisiae*. *Mol Cell Biol* 5:2029, 1985
- Aladjem MI, Groudine M, Brody LL, Dieken ES, Fournier REK, Wahl GM, Epher EM: Participation of the human  $\beta$ -globin locus control region in initiation of DNA replication. *Science* 270:815, 1995
- Vaughan JP, Dijkwel PA, Hamlin JL: Replication initiates in a broad zone in the amplified CHO dihydrofolate reductase domain. *Cell* 61:1075, 1990
- Bramhill D, Kornberg A: Duplex opening by dnaA protein at novel sequences in initiation of replication at the origin of the *E. coli* chromosome. *Cell* 52:743, 1988
- Bramhill D, Kornberg A: A model for initiation at origins of DNA replication. *Cell* 54:915, 1988
- Lam ST, Stahl MM, McMilin KD, Stahl FW: Rec-mediated recombinational hot spot activity in bacteriophage lambda. *Genetics* 77:425, 1974
- Smith GR, Kunes SM, Schultz DW, Taylor A, Triman KL: Structure of chi hotspots of generalized recombination. *Cell* 24:429, 1981
- Cullen M, Erlich H, Klitz W, Carrington M: Molecular mapping of a recombinant hotspot located in the second intron of the human TAP2 locus. *Am J Hum Genet* 56:1350, 1995
- Fullerton SM, Harding RM, Boyce AJ, Clegg JB: Molecular and population genetic analysis of allelic sequence diversity at the human  $\beta$  globin locus. *Proc Natl Acad Sci USA* 91:1805, 1994