Proc. Natl. Acad. Sci. USA
Vol. 80, pp. 4808–4812, August 1983
Genetics

Recombination within and between the human insulin and β-globin gene loci

(recombination hot spot/restriction fragment polymorphism/lod score)

ROGER V. LEBO*, ARAVINDA CHAKRAVARTI†, KENNETH H. BUETOW†, MEI-CHI CHEUNG*, HOWARD CANN‡, BARBARA CORDELL§, AND HOWARD GOODMAN§

*Howard Hughes Medical Institute and Department of Medicine, University of California, San Francisco, California 94143; †Department of Biostatistics, University of Pittsburgh, Pittsburgh, Pennsylvania 15261; ‡Department of Genetics, Stanford University Medical Center, Stanford, California 94305; and §Department of Molecular Biology, Massachusetts General Hospital, Boston, Massachusetts 02114

Communicated by Bruce M. Alberts, April 25, 1983

ABSTRACT **We detected a large number of polymorphic insulin restriction fragments in black Americans. These different size fragments were probably generated by unequal recombination on both sides of the human insulin gene. Population genetic analysis indicates that recombination occurred 33 times more frequently than expected to generate this large number of polymorphic fragments. Specific properties of the unique repeated 14- to 16-base-pair sequences 5′ to the insulin gene suggest that this sequence would promote increased unequal recombination. Additional pedigree analysis showed that the recombination rate between the structural insulin and β-globin gene loci was 14% with strong evidence for linkage. Since both insulin and β-globin have been mapped to the short arm of human chromosome 11, this study establishes that the genetic map distance between these genes is 14.2 centimorgans.**

Although mutation has generally been assumed to generate most of the genetic variability in natural populations, recent evidence indicates that recombination generates variability in some DNA sequences. For instance, the origin of the human hemoglobin Lepore and Kenya genotypes (1, 2), the single and triplicated α-globin gene loci (3, 4), and the variation in the number of smaller tandemly repeated nucleotide sequences $(A\text{-}T\text{-}T\text{-}T)_n$ located 1.5 kilobases (kb) 5′ to the β-globin gene (5) can all be explained by unequal crossing-over. The $(C\text{-}T\text{-}G\text{-}G)_n$ sequences in the *lac I* gene in *Escherichia coli* are sites of frequent additions and deletions of the repeat unit (6). In yeast mtDNA, variable numbers of repeated $(G\text{-}G\text{-}C\text{-}C)_n$ sequences are hot spots for illegitimate recombination between homologous nucleotide sequences at different loci (7). Similarly, we find a high degree of variability in a repeated sequence located 5′ to the human insulin gene. This variable sequence consists of 14- to 16-base-pair (bp) tandem repeats with minor variations in the repeat sequence (8–11). The yeast hot spot for recombination consists of a repeated sequence that is also palindromic in both strands (7). The 5′-β-globin and the 14- to 16-bp repeat sequences in the region 5′ to the insulin gene (which we call the 5′-*INS* locus) can be rewritten to show that each is palindromic within one strand, and this characteristic may also encourage unequal recombination (see *Discussion*). Unequal recombination at the 5′-*INS* locus must occur at a high rate as indicated by the observed heterozygosity at this locus. Furthermore, the large number of unequal restriction fragments observed in the region 3′ to the insulin gene (3′-*INS*) can also result from unequal recombination.

The insulin and β-globin polymorphisms were used to study recombination between these two gene loci. Both genes are of considerable interest because homozygosity for abnormal genes can result in diabetes (12) or β-thalassemia and sickle cell anemia (13). The measured recombination frequency can be compared with the physical map distance between the assigned gene positions (14, 15). Previous chiasma studies have revealed positive chiasma interference near the centromere (16) and negative interference at the telomere (17, 18).

MATERIALS AND METHODS

Subject Population. For insulin polymorphism frequency studies, peripheral blood was obtained from 64 black Americans from 13 families ascertained through one or more family members who carried a β-globin structural gene variant β^S, β^C, or β^{thal}. The 10 families used in the recombination study included 6 nuclear families, 3 extended pedigrees of 3 generations, and 1 family with children from two marriages by the same male.

Analysis of DNA and Protein Polymorphisms. The β-globin protein polymorphisms were determined by cellulose acetate electrophoresis of erythrocyte proteins. For restriction enzyme analysis, 7 μg of DNA from peripheral blood lymphocytes (19) was digested with *Sac* I or *Hpa* I (5 units/μg) (Bethesda Research Laboratories) in buffer recommended by the supplier. Restriction enzyme analysis was completed as described (19–21). The insulin probe was prepared from an excised nick-translated (22) *Bgl* II/*Xho* I subcloned fragment of the cloned genomic insulin gene (23) (Fig. 1). The total globin hybridization probe was prepared by reverse transcription of human globin mRNA (19).

Linkage Analysis. Insulin gene frequencies were obtained by gene counting among all originals in the families; β-globin frequencies were from previously published data (24). The four loci 5′-*INS*, 3′-*INS*, β-globin, and *Hpa* I were all codominant. Where possible, we constructed haplotypes for 5′-*INS*, 3′-*INS* (*INS*) and β-globin, *Hpa* I (*BETA*) from each family for measuring recombination between the insulin and β-globin genes. Recombination fractions for any pair of loci were estimated by quadratic interpolation from a standard lod (logarithm of the odds) score table (17, 25) after summing the lod scores over all families. A χ^2 test suggested by Morton (25) was used to detect heterogeneity in recombination fractions between families; lod scores were summed for all families since no heterogeneity was detected. All lod scores were computed by using the program

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. §1734 solely to indicate this fact.

Abbreviations: kb, kilobase(s); bp, base pair(s).

LIPED (26) provided by Jurg Ott. A source of error in linkage analysis is nonpaternity. Since individuals were typed at four loci of high heterozygosity in this population, most cases of nonpaternity would be detected. We have discovered one such case that has been discarded in our analyses.

RESULTS

Restriction Fragment Polymorphisms. The sizes of the *Sac* I human insulin gene-containing fragments were studied in American blacks. When tested with a 3.2-kb *Bgl* II/*Xho* I insulin gene-containing probe (Fig. 1), the insulin genes exhibited previously unreported *Sac* I fragment polymorphism (3.1–4.2 kb) at the 3′-*INS* locus (Fig. 2). Even greater variability in size (5.4–9.1 kb) was measured for the 5′-*INS* fragment (Fig. 2). These results showed more or less unimodal distribution of 5′-fragment sizes compared with the bimodal distribution reported for Caucasian Americans (8). Segregation of 5′-*INS* and 3′-*INS* fragments within pedigrees confirmed previous findings that a single insulin gene is inherited in Mendelian fashion from each parent (20). The 5′-*INS* and 3′-*INS* fragments appear to be associated nonrandomly, and haplotypes for 5′- and 3′-*INS* fragments display more polymorphism than either fragment alone. The 5′-*INS* and 3′-*INS* genotypes used to determine fragment frequencies were all from unrelated originals in the pedigrees and represent independent observations at the insulin locus.¶ Then, the cosegregation of these two loci within pedigrees was used to obtain the frequencies of the linked 5′-*INS* and 3′-*INS* haplotypes. Haplotypes that were ambiguous with respect to linkage phase were discarded.

From a total of 35 haplotypes constructed, one can measure the degree of nonrandom association (linkage disequilibrium) between specific 5′-*INS* and 3′-*INS* fragment lengths by using the method of Yamazaki (27). The association is 0.76 compared with a maximum value of 1.00 and thus is very high. We observed 21 different 5′-*INS* fragments out of 44 independent chromosomes but only 7 different 3′-*INS* fragments among 36 independent chromosomes.‖ From these data, we can estimate

¶ The insulin loci represent a random sample even though the linked β-globin genes were ascertained through carriers of β-globin structural gene variants. Since the recombination rate between the insulin and β-globin loci is 0.14, this will lead to a lack of association between these genes in a few generations.

‖ The number of independent chromosomes refers to the number of originals in the pedigrees. The difference in this number for 5′-*INS* and 3′-*INS* occurs because one family was not scored for the 3′-*INS* locus.

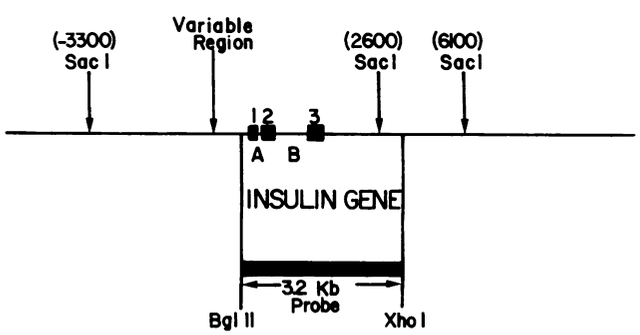

FIG. 1. Insulin restriction map. The insulin gene region is indicated, along with the three insulin gene exons 1, 2, and 3, the introns A and B, and the *Sac* I restriction enzyme sites in the structural gene region. The larger *Sac* I fragment includes the insulin gene and the region 5′ to it. The unique DNA sequence from which our 3.2-kb probe is prepared is also illustrated.

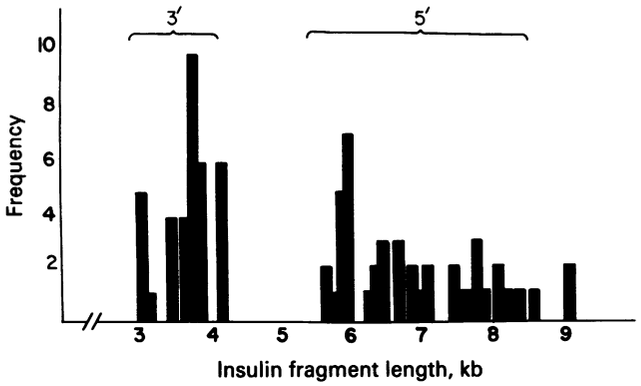

FIG. 2. Histogram of 5′ and 3′ *Sac* I insulin restriction fragment sizes. A histogram constructed for the relative frequencies of each independent fragment within the pedigrees studied indicates significant variability within the smaller size range of the 3′ fragment (mean size, 3.7 kb) in addition to the variability within the 5′ fragment containing the insulin gene (mean size, 7.0 kb).

the gene diversity (panmictic heterozygosity) h of the fragment lengths by

$$h = \frac{n}{n-1}(1 - \Sigma \chi_i^2), \qquad [1]$$

where χ_i is the frequency of the ith class of fragment; $i = 1,2, \ldots k$; and n is the sample size (28). This equation gives values for the heterozygosity of 0.95 for 5′-*INS* and 0.85 for 3′-*INS*.

Assuming that fragment lengths are selectively neutral and evolve by genetic drift in a randomly mating population, we can calculate a standardized index of genetic variability, M, from

$$E(k) = M\{(1/M) + [1/(M+1)] + \ldots + [1/(M+n-1)]\}, \qquad [2]$$

where $E(k)$ is the expected number of different restriction fragments detected in a sample of size n and k is the number of segregating fragment lengths (29). The parameter M equals $4N_e u$, where N_e is the effective population size and u is the average mutation rate per fragment and shows that evolution depends on the product of N_e and u. The M value was significantly greater for 5′-*INS* than for 3′-*INS* or the seven polymorphic restriction sites in the β-globin cluster (Table 1) (30, 31). Thus, the 7.0-kb 5′-*INS* locus was highly variable. Nonetheless, comparison with other M values must consider that this variability is being compared with those of the shorter 3.7-kb 3′-*INS* locus and the longer 50-kb β-globin gene complex as well as with the different DNA sequences that generate polymorphism (31).

Under the neutral mutation–random drift theory (32), the probability h that an individual will be heterozygous at any locus can also be written as

$$h = M/(1+M) \qquad [3]$$

Table 1. $M = 4N_e u$ values for 5′-*INS*, 3′-*INS*, and the β-globin gene cluster

Locus	Population	M value Eq. 2	M value Eq. 3	Source
5′-*INS*	U.S. black	15.1	19.1	Present study
3′-*INS*	U.S. black	2.3	5.5	Present study
β-Cluster	Italian	1.8	1.8	(30, 31)
β-Cluster	Greek	5.7	11.4	(30, 31)
5′-*INS*	Caucasian	13.8	21.0	(8–10)

for an infinite-alleles model at equilibrium. This equation can also be used to estimate M (Table 1). The M value obtained from Eq. 3 is known to generally give overestimates for small sample sizes (28). Unequal recombination of the repeated sequences 5′ to the insulin gene most likely generated the sequences reported in four alleles (9–11). Therefore, if we assume that recombination created all the different polymorphic fragments, R is the average recombination rate per fragment, and all such alleles are selectively neutral, we can write in analogy to Eq. 3

$$h = (4N_eR)/(1 + 4N_eR). \quad [4]$$

Thus, M can be identified as $4N_eR$ and can be estimated from Eqs. 2 and 3, as shown in Table 1.

The *Sac* I fragment sizes observed in American Caucasians (9) showed that the number of 14-bp repeats varied from 26 to 209, corresponding to fragment lengths of 0.36–2.93 kb. Similarly, our data in American blacks shows 5–269 repeats of the 14-bp consensus sequence (Fig. 2) with an average of 109 repeats and a standard deviation of 69 repeats. If $4N_eR$ is divided by 1,526 (14 × 109), we can obtain an estimate of the variability at the base-pair level; i.e., $M_r = 4N_er$, where r is now the average recombination rate per bp. From Table 1, $M_r = 0.0099$ if $M = 15.1$ and $M_r = 0.0125$ if $M = 19.1$ for our 5′-*INS* data. (A similar calculation of the data in ref. 8 gives $M_r = 0.0135 \rightarrow 0.0205$.) For comparison, we have used an estimate of $4N_eu = 0.0018$ for the American black population studied for restriction site polymorphisms in the β-globin cluster (31). The data in ref. 31 were used to compute the nucleotide diversity or the heterozygosity at the nucleotide level, which is approximately $4N_eu$ if the polymorphisms are due to base substitutions (28). Since both studies refer to the American black population, we assume that N_e is the same for both data sets and calculate $r/u = 5.5 \rightarrow 6.9$. Furthermore, an independent analysis of the restriction site polymorphisms in the β-globin cluster gave $u = 2 \times 10^{-9}$ per bp per yr** (30). Thus, the calculated recombination rate r in the variable insulin region is $(5.5 \rightarrow 6.9)u$ or $1.1 \rightarrow 1.4 \times 10^{-8}$ per bp per yr (mean = 1.3×10^{-8} per bp per yr).

The expected recombination rate can be calculated and compared with the observed recombination rate. The human chromosomes exhibited about 60 chiasma per meiosis (34–37), which corresponds to a genetic length of 30 morgans per 3×10^9 bp (38). If one human generation is taken as 25 years, the expected recombination rate is $30/(25 \times 3 \times 10^9)$ or 4×10^{-10} per bp per year. Therefore, the calculated value $r = 1.3 \times 10^{-8}$ is 33 times higher than the expected value. A 31-kb region with this high rate of recombination throughout would generate a map distance of 1 centimorgan in contrast to an expected DNA region of 1,000 kb. A similar calculation of r for 3′-INS showed that recombination is only 2 times the expected rate or $^1/_{16}$th as frequent as for 5′-*INS*.

Linkage Analysis. For linkage analysis, the computer program LIPED (26) was used to compute the lod score for each family (Fig. 3) at recombination fractions $\theta = 0.00$, 0.01, 0.05, 0.10, 0.20, 0.30, and 0.40. This program computes lod scores, $Z(\theta)$, for detecting linkage between any pair of loci. The lod score is

$$Z(\theta) = \log_{10}[L(\theta)/L(0.5)], \quad [5]$$

where $L(\theta)$ is the likelihood of observing the family if the recombination value is θ; $L(0.5)$ is the same under the assumption of independent segregation. Thus, $Z(\theta)$ is $\log_{10}$ of the odds for linkage versus nonlinkage assuming that the prior odds of linkage and nonlinkage are equal. A lod score of 3.0 means that the odds in favor of linkage are 10^3:1 (or 1,000:1) at the recombination fraction given. By convention, a lod score of 3.0 or greater provides evidence for linkage, values less than −2.0 at $\theta = 0.4$ provide evidence against linkage, and values in between are inconclusive (25).

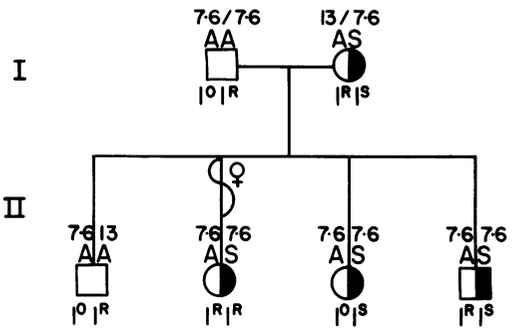

FIG. 3. Pedigree with polymorphic segregation. A typical pedigree (family 7) illustrating the segregation of the four polymorphic (I^O, I^R, and I^S) insulin alleles, the 7.6- and 13-kb *Hpa* I β-globin fragments, and the β^A- and β^S-globin proteins. The occurrence of a maternal recombination is indicated.

A lod score table was computed for each of the 10 families. Heterogeneity in the recombination value between families was tested for and, since no heterogeneity was detected, the lod scores over all families were added. The best estimate of recombination fraction θ and maximum lod score Z can then be obtained from each summed lod table for each pair of loci by quadratic interpolation. For this study, we assumed that male and female recombination values were equal and separately analyzed the two complex loci of insulin (i.e., 5′-*INS*, 3′-*INS*) and of β-globin (i.e., β-globin structural gene, *Hpa* I) as well as the four independent loci: 5′-*INS*, 3′-*INS*, β-globin structural gene, and *Hpa* I. The values of θ and Z for all pairs of loci are given in Table 2.

As expected, the 5′-*INS* and 3′-*INS* loci, which are within 5 kb of each other, displayed no recombination ($\theta = 0.00$) at a lod score of 2.95 (Table 2). Similarly, the β-globin and *Hpa* I loci, also 5 kb apart, had $\theta = 0.00$ at a lod score of 7.56. On the other hand, the best estimate of recombination between the 5′-*INS* and β-globin structural gene locus was $\theta = 0.14$ at a lod score of 2.6 (14% recombinants or 14.2 centimorgans using the mapping function in ref. 39). This does not yet conform to the conventional linkage acceptance value of 3.0. However, the odds favoring linkage are 398:1 and both genes are known to be on the short arm of chromosome 11 (14, 15, 20, 40). This does indicate a genetic map distance of 14.2 centimorgans between the two genes. Some of the other loci pairs gave values similar to a 14% recombination value (Table 2).

When the recombination value θ was estimated separately for male (θ_m) and female (θ_f), the two values were always similar and gave lod scores almost identical to those in Table 2. Furthermore, since no recombinants occurred between 5′-*INS*, 3′-*INS* and β-globin, *Hpa* I, the differences in the θ values for the different locus pairs depend on the number of informative families. In our data, there were more informative families for 5′-*INS* than for 3′-*INS* and more for *INS* than for *BETA*.

DISCUSSION

We have reported significant polymorphism in restriction fragment lengths for the 5′-*INS* and 3′-*INS* loci in the American black population. The 5′-*INS* locus appears to be the most poly-

** This value corresponds closely to 1.4×10^{-9} per bp per year in the 5′ and 3′ β-globin flanking regions given a 7% nucleotide replacement in the flanking regions in 50×10^6 years (33).

Table 2. Lod scores for the two complex loci of insulin (*INS*) and β-globin (*BETA*) as well as the four independent loci: 5′-*INS*, 3′-*INS*, β-globin, and *Hpa* I

Locus 1	Locus 2	$\hat{\theta}$	$\hat{Z}(\theta)$	θ 0	0.01	0.05	0.1	0.2	0.3	0.4
5′-*INS*	3′-*INS*	0.00	2.95	2.95	2.89	2.60	2.21	1.38	0.66	0.17
5′-*INS*	β-Globin	0.14	2.60	−∞	−1.07	1.74	2.44	2.19	1.31	0.43
5′-*INS*	*Hpa* I	0.14	1.62	−∞	−1.44	0.83	1.44	1.35	0.79	0.24
3′-*INS*	β-Globin	0.02	1.65	1.63	1.64	1.60	1.42	0.93	0.45	0.12
3′-*INS*	*Hpa* I	0.10	0.66	0.16	0.34	0.62	0.66	0.46	0.21	0.05
β-Globin	*Hpa* I	0.00	7.56	7.56	7.36	6.55	5.53	3.53	1.74	0.47
INS	β-Globin	0.14	1.76	−∞	−1.06	1.17	1.65	1.44	0.82	0.25
INS	*Hpa* I	0.14	1.24	−∞	−1.05	0.70	1.12	1.02	0.59	0.18
5′-*INS*	*BETA*	0.15	1.52	−∞	−2.00	0.67	1.43	1.46	0.92	0.32
3′-*INS*	*BETA*	0.06	0.91	0.77	0.82	0.91	0.86	0.58	0.27	0.07
INS	*BETA*	0.21	0.54	−∞	−3.03	−0.54	0.24	0.54	0.36	0.12

Results are expressed as the maximum likelihood estimate of the recombination value $\hat{\theta}$, maximum lod score $\hat{Z}(\theta)$, and lod scores for different values of recombination fraction (θ).

morphic human locus identified by molecular methods. Our mathematical analysis of the previously published Caucasian American 5′-*INS* data produced essentially the same results. In fact, the heterozygosity at this locus is higher than that for the *D14S1* locus (41, 42) or that usually obtained for the *HLA* loci. Unfortunately, restriction enzyme analysis is unable to detect variable region differences less than 0.1 kb (or seven times the 14-bp repeat length). Even if these small (14-bp) consensus sequence differences could have been measured to detect additional fragment sizes, heterogeneity in the 14-, 15-, and 16-bp nucleotide sequences (9–11) indicates that considerably greater variability remains undetected and our variability estimates are low. Our data on the frequency distribution of 5′-*INS* fragments did not exhibit the bimodality of fragment size observed in American Caucasians (8). Since longer variable 5′ insulin region sequences have not been analyzed in either population, unknown different sequences may prohibit reassociation of one segment within the variable region and prevent recombination to generate intermediate size fragments in Caucasians but not in blacks.

We note that the 14-, 15-, and 16-bp consensus sequences (8–11) that generate the variable region in 5′-*INS* are palindromic in a single strand. From Bell *et al.* (10):

$$\text{GGGACTGGG(ACAGGGTGTGGG)}_n\text{ACAGCAGCGC}$$

may be rewritten as

$$\text{GGGACTGGGACAGGGTG}_{1/2}\,(\text{G}_{1/2}\text{TGGGGACAGGGGTG}_{1/2})_{(n-1)}\,\text{G}_{1/2}\text{TGGGGACAGCAGCGC}$$

This characteristic is likely to promote unequal recombination within or between homologous sister chromosomes during meiosis or mitosis to yield new numbers of consensus sequence repeats in each daughter chromosome. Recombination between homologous chromosome regions may be catalyzed by enzymes analogous to the *E. coli* DNA synaptase or recA protein, which individually promote association of single-stranded DNA with homologous double-stranded DNA sequences and then catalyze subsequent strand fusion (43, 44). Recombination may also result during DNA replication when one palindromic repeat could hybridize with another repeat to form circular or more complex loops. Then, strand hopping by DNA polymerase could either delete a circle or replicate it twice to give daughter chromosomes of different insert size. In contrast, repeated DNA sequences that are palindromic in both strands may rehybridize with either DNA strand to form hairpin loops (45) and more complex structures (46) as well as circular loops to generate additional types of DNA rearrangement by strand hopping.

Either sister chromatid or homologous chromatid recombination in 5′-*INS* could create the large number of 5′-*INS* fragments. The relative frequency of these events may be determined by studying flanking DNA markers in pedigrees exhibiting changes in the number of repeats between parents and offspring. New fragment lengths that resulted from recombination within a single DNA strand will vary in the number of the same repeated sequences but not have recombined flanking regions. New fragments that resulted from recombination between the gene loci on homologous chromatids will contain recombined flanking DNA regions and these individuals can be tested to orient the insulin gene with respect to other polymorphic genes on the chromosomes.

It is fortunate that sequences from the 5′-*INS* variable region are unique, as shown by the fact that nick-translated DNA from this region can be used as an insulin-specific gene probe (9). If this presumptive hot spot of recombination appeared on other chromosome sites far removed from the insulin gene it might promote illegitimate recombination between homologous repeated hot spot sequences. Illegitimate homologous chromatid recombination would generate deletions [such as Wilms tumor (47)] and insertions while nonhomologous chromatid recombination would result in reciprocal chromosome translocations. On the other hand, unique palindromic sequences such as the one in 5′-*INS* may have evolved throughout all the chromosomes to promote desirable chiasma and ensure precise meiotic pairing. Chiasmata themselves are considered to regularize disjunction and ensure equal separation of chromosomes into daughter cells at meiosis (48).

The genetic map distance of 14 centimorgans between β-globin and insulin was compared with the physical map distance to determine whether negative or positive interference occurs in this region of chromosome 11. The human insulin gene has been assigned to the terminal portion of the short arm of chromosome 11 by *in situ* hybridization (15) and β-globin to band 11p12 by somatic cell hybridization (14). A chromosome 11 idiogram (49) was measured to determine that the distance between the assigned β-globin and insulin cytogenetic positions represents approximately 30% of the distance along chromosome 11. Since the entire chromosome complement has a genetic map distance of 3,000 centimorgans and the mean DNA content of each chromosome 11 bivalent is 4.76% (50), then the chromosome 11 bivalent represents 143 centimorgans. Thus,

the expected recombination frequency between the insulin and β-globin loci is approximately 43 centimorgans. Assuming the *in situ* localization may be reported with this precision, the 14% recombination rate measured in this study indicates that this region of chromosome 11 exhibits positive chiasma interference. This 14% value may not be representative of all human populations since θ can be population specific (51). Although the measured chiasma frequency is less than the chromosome gene positions predict, these values agree within a factor of 3. This agreement is in contrast to the increased rate of unequal chiasma in the 5′-*INS* gene region.

We gratefully acknowledge Andree Dozy and Leigh Nicholaisen for *Hpa* I restriction enzyme analysis.

1. Baglioni, C. (1962) *Proc. Natl. Acad. Sci. USA* **48**, 1880–1886.
2. Kendall, A. G., Ojwang, P. J., Schroeder, W. J. & Huisman, T. H. J. (1973) *Am. J. Hum. Genet.* **25**, 548–563.
3. Embury, S. H., Miller, J. A., Dozy, A. M., Kan, Y. W., Chan, V. & Todd, D. (1980) *J. Clin. Invest.* **66**, 1319–1325.
4. Goossens, M., Dozy, A. M., Embury, S. H., Zachariades, Z., Hadjiminas, M., Stamatoyannopoulos, G. & Kan, Y. W. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 518–521.
5. Spritz, R. A. (1981) *Nucleic Acids Res.* **9**, 5037–5047.
6. Farabaugh, P. J., Schmeissner, U., Hofer, M. & Miller, J. H. (1978) *J. Mol. Biol.* **126**, 847–863.
7. Faugeron-Fonty, G., Culard, F., Baldacci, G., Goursot, R., Prunell, A. & Bernardi, G. (1979) *J. Mol. Biol.* **134**, 493–537.
8. Bell, G. I., Karam, J. H. & Rutter, W. J. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 5759–5763.
9. Bell, G. I., Selby, M. J. & Rutter, W. J. (1982) *Nature (London)* **295**, 31–35.
10. Bell, G. I., Karam, J. H. & Rutter, W. J. (1983) *Progress in Clinical and Biological Research*, Proceedings of the Sixth International Congress of Human Genetics (Liss, New York), pp. 57–65.
11. Ullrich, A., Dull, T. J., Gray, A., Phillips, J. A., III, & Peter, S. (1982) *Nucleic Acids Res.* **10**, 2225–2240.
12. Craighead, J. E. (1978) *N. Engl. J. Med.* **299**, 1439–1445.
13. Maniatis, T., Fritsch, E. F., Lauer, J. & Lawn, R. M. (1980) *Annu. Rev. Genet.* **14**, 145–178.
14. Gusella, J., Varsanyi-Breiner, A., Kao, F.-T., Jones, C., Puck, T. T., Keys, C., Orkin, S. & Housman, D. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 5239–5243.
15. Harper, M. E., Ullrich, A. & Saunders, G. F. (1981) *Proc. Natl. Acad. Sci. USA* **78**, 4458–4460.
16. Mather, K. (1939) *Genetics* **24**, 413–435.
17. Bridges, C. B. & Olbrycht, T. M. (1926) *Genetics* **11**, 41–56.
18. Morton, N. E. (1978) *Cytogenet. Cell Genet.* **22**, 15–36.
19. Goossens, M. & Kan, Y. W. (1981) *Methods Enzymol.* **76**, 805–817.
20. Lebo, R. V., Kan, Y. W., Cheung, M. C., Carrano, A. V., Yu, L. C., Chang, J. C., Cordell, B. & Goodman, H. M. (1982) *Hum. Genet.* **60**, 10–15.
21. Kan, Y. W. & Dozy, A. M. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 5631–5635.
22. Rigby, P. W. J., Dieckmann, M., Rhodes, C. & Berg, P. (1977) *J. Mol. Biol.* **113**, 237–251.
23. Bell, G. I., Pictet, R. L., Rutter, W. J., Cordell, B., Tischer, B. & Goodman, H. M. (1980) *Nature (London)* **284**, 26–32.
24. Kan, Y. W. & Dozy, A. M. (1980) *Science* **209**, 388–391.
25. Morton, N. E. (1955) *Am. J. Hum. Genet.* **7**, 277–318.
26. Ott, J. (1974) *Am. J. Hum. Genet.* **26**, 588–597.
27. Yamazaki, T. (1977) *Genetics* **86**, 227–236.
28. Nei, M. & Tajima, F. (1981) *Genetics* **97**, 145–163.
29. Ewens, W. J. (1972) *Theor. Popul. Biol.* **3**, 87–112.
30. Chakravarti, A., Buetow, K. H., Antonarakis, S. E., Boehm, C. D. & Kazazian, H. H. (1981) *Am. J. Hum. Genet.* **33**, 134 (abstr.).
31. Kazazian, H. H., Chakravarti, A., Orkin, S. & Antonarakis, S. E. (1983) in *Evolution of Genes and Proteins*, eds. Nei, M. & Koehn, R. K. (Sinauer, Sunderland, MA), pp. 137–146.
32. Kimura, M. & Crow, J. F. (1964) *Genetics* **49**, 725–738.
33. Martin, S. L., Zimmer, E. A., Davidson, W. S., Wilson, A. C. & Kan, Y. W. (1981) *Cell* **25**, 737–741.
34. Hulten, M. (1974) *Hereditas* **76**, 55–78.
35. Ford, C. E. & Hamerton, J. L. (1956) *Nature (London)* **178**, 1020–1023.
36. Kurnit, D. M. (1979) *Lancet* **i**, 104.
37. Cavalli-Sforza, L. L. & Bodmer, W. F. (1971) *The Genetics of Human Populations* (Freeman, San Francisco), p. 883.
38. Strickberger, M. W. (1968) *Genetics* (Macmillan, New York), p. 54.
39. Rao, D. C., Morton, N. E., Lindsten, J., Hulten, M. & Yee, S. (1977) *Hum. Hered.* **27**, 99–104.
40. Lebo, R. V., Carrano, A. V., Burkhart-Schultz, K., Dozy, A. M., Yu, L. C. & Kan, Y. W. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 5804–5808.
41. Wyman, A. R. & White, R. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 6754–6758.
42. Skolnick, M. H. & Francke, U. (1982) *Cytogenet. Cell Genet.* **32**, 194–204.
43. Potter, H. & Dressler, D. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 2390–2394.
44. McEntee, K., Weinstock, G. M. & Lehman, I. R. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 2615–2619.
45. Muller, U. R. & Fitch, W. M. (1982) *Nature (London)* **298**, 582–585.
46. Biezunski, N. (1981) *Chromosoma* **84**, 111–129.
47. Riccardi, V. M., Sujansky, E., Smith, A. C. & Francke, U. (1978) *Pediatrics* **61**, 604–610.
48. Darlington, C. D. (1965) *Cytology* (Churchill, London), p. 127.
49. National Foundation-March of Dimes (1972) *Birth Defects*: Original Article Series VIII (National Foundation-March of Dimes, White Plains, NY), Vol. 7.
50. Mendelsohn, M. L., Mayall, B. H., Bogart, E., Moore, D. H. & Perry, B. H. (1973) *Science* **179**, 1126–1129.
51. Weitkamp, L. R. (1974) *Cytogenet. Cell Genet.* **13**, 179–182.