Am. J. Hum. Genet. 51:1103–1110, 1992

# Linkage Disequilibrium among RFLPs at the Insulin-Receptor Locus despite Intervening Alu Repeat Sequences

Steven C. Elbein

Division of Endocrinology and Metabolism, Veterans Affairs Medical Center and University of Utah School of Medicine, Salt Lake City

## Summary

**Multiple mutations of the insulin receptor (INSR) gene have been identified in individuals with extreme insulin resistance. These mutations have included recombination events between Alu repeat units in the tyrosine kinase–encoding β-chain region of the gene. To evaluate the influence of Alu and dinucleotide repetitive sequences on recombination events within the insulin receptor gene, I examined the degree of linkage disequilibrium between RFLP pairs spanning the gene. I established 228 independent haplotypes for seven RFLPs (two each for *Pst*I, *Rsa*I, and *Sst*I and one for *Msp*I and 172 independent haplotypes which included an additional RFLP with *Bgl*II) from 19 pedigrees. These RFLPs span >130 kb of this gene, and my colleagues and I previously demonstrated that multiple Alu sequences separate RFLP pairs. Observed haplotype frequencies deviated significantly from those predicted. Pairwise analysis of RFLP showed high levels of linkage disequilibrium among RFLP in the β-chain region of the insulin receptor, but not between α-chain RFLPs and those of the β-chain. Disequilibrium was present among β-chain RFLPs, despite separation by one or more Alu repeat sequences. The very strong linkage disequilibrium which was present in sizable regions of the INSR gene despite the presence of both Alu and microsatellite repeats suggested that these regions do not have a major impact on recombinations at this locus.**

## Introduction

Although the mode of inheritance for non-insulin-dependent diabetes mellitus (NIDDM) is uncertain (Rotter and Rimoin 1981), the disease is clearly inherited. Insulin resistance appears to be central to this inherited predisposition (DeFronzo et al. 1992), and the insulin receptor (INSR) gene has thus been a prime candidate in genetic studies. A large number of RFLPs have been described in >130 kb of genomic DNA (Elbein et al. 1986; Cox et al. 1988*c;* Elbein and Sorenson 1990; Sten-Linder et al. 1991*b*). Furthermore, a number of INSR gene mutations have been described in syndromes of extreme insulin resistance, including insulin-resistant (type A) diabetes (Taylor et al. 1991). These mutations have included single nucleotide substitutions and deletions resulting from recombination between Alu repeat units in the β-subunit. However, neither linkage studies (Elbein et al. 1987, 1988; Cox et al. 1988*b;* O'Rahilly et al. 1988; Elbein et al. 1992) nor direct analysis of the β-chain region (Elbein and Sorensen 1991; O'Rahilly et al. 1991) have implicated the INSR mutations in common NIDDM.

Several regions of the INSR gene have physical properties which suggest that they might promote recombination and lead to low levels of linkage disequilibrium. These sequences include both a large number of highly homologous Alu repeat units among the introns of the β-chain region of the receptor (Elbein 1989) and at least one long microsatellite (CA repeat) region upstream of exon 10 (author's unpublished data). Both regions result in common RFLPs, one of which results from recombination between Alu units (Elbein 1989) and a second of which results from variable numbers of CA repeats. The latter region is highly unstable in cloning vectors (Seino et al. 1989; author's

Received January 24, 1992; final revision received June 17, 1992.

Address for correspondence and reprints: Steven C. Elbein, M.D., Assistant Professor of Medicine, Endocrinology 111E, Veterans Affairs Medical Center, 500 Foothill Boulevard, Salt Lake City, UT 84148.

© 1992 by The American Society of Human Genetics. All rights reserved.
0002-9297/92/5105-0020$02.00

unpublished data), while the Alu repeats have been responsible for several reported deletions in patients with extreme insulin resistance (Taylor et al. 1991).

I hypothesized that these two types of repetitive sequences represent regions in which recombination events are markedly more frequent than in the surrounding (nonrepetitive) sequences and, consequently, that markers spanning these regions would show very little linkage disequilibrium. To address this hypothesis, I examined 228 independent haplotypes constructed from marker segregation in 444 individuals in 19 pedigrees, for seven INSR RFLPs (Elbein et al. 1992). An additional RFLP was examined in 172 haplotypes. I compared the observed and expected frequencies for full haplotypes and calculated the standardized disequilibrium statistic for each of the 28 RFLP pairs.

## Subjects and Methods

### *Subjects*

I used RFLP data from 444 individuals representing 19 pedigrees, to establish 228 independent haplotypes. In general, these pedigrees contained only two full generations, and parental haplotypes were inferred by segregation of RFLPs in a large number of offspring. For each pedigree, up to four parental (founder) haplotypes were identified. Additional haplotypes were contributed by each second- and third-generation spouse. Sixteen pedigrees were ascertained for at least two NIDDM subjects, as described in detail elsewhere (Elbein et al. 1991, 1992). An additional three pedigrees were ascertained for a mixture of type 1 (IDDM) and type 2 diabetes. All pedigree members and spouses of pedigree members whose children were sampled were included in the analysis. Since the INSR locus is not linked to NIDDM, we did not distinguish between "diabetic" and "normal" haplotypes, for purposes of calculating the linkage disequilibrium statistic.

### *DNA Analysis*

RFLPs were determined according to a method described elsewhere (Elbein et al. 1986, 1992). In brief, two *Pst*I RFLPs (Cox et al. 1988*c*), which I have designated "P1" and "P2," were detected with a 1-kb α-chain *Eco*RI fragment of the INSR cDNA (Ullrich et al. 1985). A *Rsa*I RFLP, which I call "R1" and which I demonstrated to represent variable numbers of CA repeats, is located in intron 9. Although I detected this RFLP with an 800-bp genomic clone which includes exon 10 and extends ≈600 bp upstream, it is also detectable with cDNA clones and enzymes *Hin*dIII and *Eco*RI (Elbein et al. 1986; Sten-Linder et al. 1991*b;* author's unpublished data). A second *Rsa*I polymorphism, which I call "R2," represents a site mutation in intron 15. I detected R2 with a genomic fragment containing exon 15 and part of intron 15. Another insertion/deletion polymorphism is present in intron 14 and is detectable by multiple enzymes. I characterize this RFLP, which results from recombination between Alu sequences and which I call "S1," with the enzyme *Sst*I and probes from exon 14/intron 14 or exon 15/introns 14 and 15 (Elbein 1989). A second *Sst*I polymorphism, designated "S2," represents a site polymorphism in intron 20 (Elbein et al. 1986; Elbein 1989; Sten-Linder et al. 1991*b*) and, in the present study, was detected with a 700-bp clone which included exon 17 and ≈500 bp of downstream sequence (Elbein and Sorensen 1990). A colleague and I used this same clone to identify, with *Msp*I, a recently described RFLP in intron 17 (Elbein and Sorenson 1990). Finally, we detected, with probes for either exon 14/intron 14 or exon 15/introns 14 and 15, a *Bgl*II site polymorphism in intron 14. In all cases except that in which *Msp*I was used, cDNA probes described elsewhere (Elbein et al. 1986) detect the same RFLP. The RFLPs and their frequencies are listed in table 1, and locations are shown in figure 1; detailed methods are described elsewhere (Elbein et al. 1992).

For each RFLP, I have designated the smaller allele as "1" and the larger allele (upper band) as "2." For each digest, 5 μg of DNA was digested overnight with a fivefold excess of restriction buffer according to the

**Table 1**

**Allele Frequencies of INSR RFLPs**

| RFLP Name | Minor Allele | Minor-Allele Frequency |
|---|---|---|
| P1 (*Pst*I) | 1 | .140 |
| P2 (*Pst*I) | 2 | .224 |
| R1 (*Rsa*I) | 1 | .421 |
| R2 (*Rsa*I) | 2 | .197 |
| S1 (*Sst*I) | 2 | .097 |
| S2 (*Sst*I) | 1 | .110 |
| M (*Msp*I) | 1 | .496 |
| B (*Bgl*2) | 1 | .250 |

NOTE.—Frequencies of the less common allele are shown. For each RFLP, the smaller restriction fragment is labeled 1 (presence of site, except for R1 and S1 polymorphisms, which are of the insertion/deletion type). Locations of the RFLPs are shown in Fig. 1.

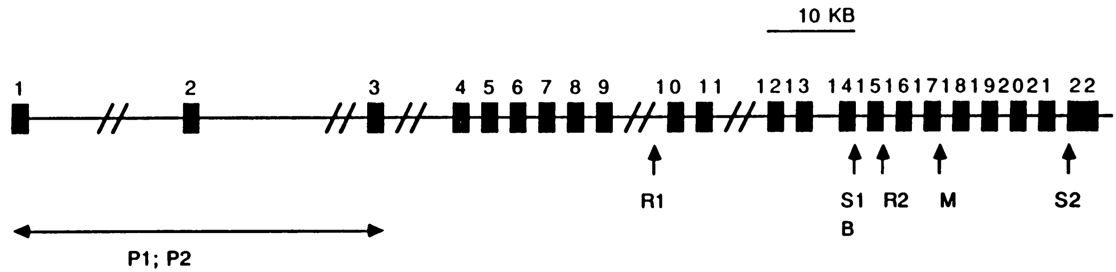


**Figure 1** Location of each RFLP, relative to the genomic structure of the insulin receptor gene. Exons and introns are shown in their approximate sizes, by scale shown on the upper right. Double slashes represent uncloned regions and unknown distances. RFLPs are designated as in table 1 and Subjects and Methods: R1 and R2, *Rsa*I; S1 and S2, *Sst*I; P1 and P2, *Pst*I; and M and B, *Msp*I and *Bgl*II, respectively. For P1 and P2, exact genomic location is unknown, and the arrows denote the range of possible locations.

## Table 2

**Number of INSR Haplotypes**

| P1 | P2 | R1 | R2 | S1 | S2 | M | B | Total No. Observed | Total No. Expected |
|---|---|---|---|---|---|---|---|---|---|
| 2 | 1 | 2 | 1 | 1 | 2 | 2 | 2 | 38 | 16.4 |
| 2 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 16 | 11.8 |
| 2 | 1 | 1 | 2 | 1 | 2 | 1 | 1 | 16 | 1.0 |
| 2 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 9 | 1.3 |
| 2 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 8 | 15.8 |
| 2 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 8 | 3.4 |
| 2 | 1 | 1 | 1 | 2 | 2 | 1 | 2 | 7 | 1.2 |
| 2 | 1 | 2 | 1 | 1 | 1 | 1 | 2 | 6 | 1.9 |
| 2 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 5 | 4.6 |
| 1 | 1 | 2 | 1 | 1 | 2 | 2 | 2 | 5 | 2.7 |
| 1 | 1 | 1 | 1 | 1 | 2 | 2 | 2 | 5 | 1.9 |
| 2 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 5 | 1.4 |
| 1 | 2 | 2 | 2 | 1 | 2 | 1 | 1 | 5 | 0 |
| 2 | 1 | 1 | 1 | 1 | 2 | 1 | 2 | 4 | 11.4 |
| 2 | 1 | 2 | 1 | 2 | 2 | 1 | 2 | 4 | 1.8 |
| 2 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 3 | .5 |
| 2 | 2 | 2 | 2 | 1 | 2 | 1 | 1 | 3 | .4 |
| 2 | 2 | 1 | 2 | 1 | 2 | 1 | 1 | 3 | .3 |
| 2 | 1 | 1 | 2 | 1 | 2 | 1 | 2 | 2 | 2.9 |
| 1 | 1 | 2 | 1 | 1 | 2 | 1 | 2 | 2 | 2.6 |
| 1 | 2 | 2 | 1 | 1 | 2 | 2 | 2 | 2 | .8 |
| 1 | 2 | 1 | 1 | 1 | 2 | 2 | 2 | 2 | .5 |
| 1 | 1 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | .2 |
| 2 | 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 0 |
| 2 | 2 | 1 | 1 | 1 | 2 | 1 | 2 | 1 | 3.2 |
| 2 | 1 | 1 | 2 | 1 | 2 | 2 | 1 | 1 | 1 |
| 2 | 2 | 2 | 1 | 1 | 1 | 1 | 2 | 1 | .5 |
| 2 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | .4 |
| 2 | 2 | 2 | 1 | 1 | 2 | 2 | 1 | 1 | .4 |
| 1 | 1 | 1 | 1 | 2 | 2 | 1 | 2 | 1 | .2 |
| 1 | 2 | 1 | 1 | 1 | 1 | 1 | 2 | 1 | .2 |
| 1 | 1 | 2 | 2 | 1 | 2 | 1 | 1 | 1 | .2 |
| 1 | 2 | 2 | 1 | 2 | 2 | 1 | 2 | 1 | 0 |

NOTE.—Haplotypes are for the 33 chromosomes typed at all eight loci. Each RFLP is designated as in table 1 and Subjects and Methods. The total number of alleles identified was 172; the additional 56 alleles included in the pairwise comparisons were not typed for the B RFLP. To compare the observed number of haplotypes and the expected number of haplotypes, all haplotypes with an expected number less than five were pooled. The $\chi^2$ for this comparison (4 df) was 39.5 ($P < .001$).

manufacturer's instructions, was separated on agarose gels, and was blotted to Nylon 6.6 membrane according to a method described elsewhere (Elbein et al. 1986, 1992). Blots were hybridized to the above-described probes labeled with $^{32}P$ dCTP by random-priming, were washed at high stringency, and were exposed to XAR-5 X-ray film, for 3–7 d, with an intensifier screen.

### *Linkage Disequilibrium*

To detect deviation from random assortment of RFLPs, frequencies were calculated for the 33 complete haplotypes (table 2), and were compared with the frequencies predicted under complete linkage equilibrium. Pooling of all haplotypes with an expected number of less than five left five categories for comparison by $\chi^2$ analysis. For further analysis, the standardized linkage disequilibrium statistic $D'$ was calculated for each RFLP pair, according to a method described by Thompson et al. (1988). The $D'$ statistic is more revealing of the lack of power to detect disequilibrium between rare haplotypes than are other statistics (Hedrick 1987; Thompson et al. 1988), including the "Δ" statistic proposed by Chakravarti et al. (1986) and used in our previous studies (Elbein et al. 1986). I followed the conventions of Cox et al. (1988*a*) and Thompson et al. (1988), such that $D'$ is positive when the rarer alleles at each locus are associated and is negative when a common allele is associated with a rare allele. In brief, $D'$ is calculated from the disequilib-

**Table 3**

**Pairwise Haplotype Frequencies for INSR RFLPs**

| | No.[a] | | | | | |
|---|---|---|---|---|---|---|
| RFLP Pair | 11 | 12 | 21 | 22 | Total | $\chi^2$ hr[b] |
| P1-P2 | 19 | 13 | 158 | 38 | 228 | 7.2 |
| P1-R1 | 13 | 19 | 83 | 113[c] | 228 | 0 |
| P1-R2 | 24 | 8 | 159 | 37 | 228 | .7 |
| P1-S1 | 29 | 3 | 177 | 19 | 228 | 0 |
| P1-S2 | 2 | 30 | 23 | 173 | 228 | .8 |
| P1-M | 16 | 16 | 97 | 99 | 228 | 0 |
| P1-B | 8 | 19 | 35 | 110 | 172 | 1.5 |
| P2-R1 | 73 | 104[c] | 23 | 28 | 228 | .2 |
| P2-R2 | 145 | 32 | 38 | 13 | 228 | 1.4 |
| P2-S1 | 160 | 17 | 46 | 5 | 228 | 0 |
| P2-S2 | 17 | 160 | 8 | 43 | 228 | 1.4 |
| P2-M | 86 | 91 | 27 | 24 | 228 | .3 |
| P2-B | 29 | 104 | 14 | 25 | 172 | 4.3 |
| R1-R2 | 72 | 24 | 110 | 22[c] | 228 | 2.9 |
| R1-S1 | 83 | 13 | 122 | 10[c] | 228 | 2.8 |
| R1-S2 | 11 | 85 | 14 | 118[c] | 228 | 0 |
| R1-M | 55 | 41 | 58 | 74[c] | 228 | 3.9 |
| R1-B | 22 | 53 | 21 | 76 | 172 | 3.2 |
| R2-S1 | 161 | 22 | 45 | 0 | 228 | 6.0 |
| R2-S2 | 23 | 160 | 2 | 43 | 228 | 2.4 |
| R2-M | 69 | 114 | 44 | 1 | 228 | 52.4 |
| R2-B | 1 | 127 | 42 | 2 | 172 | 152 |
| S1-S2 | 25 | 181 | 0 | 22 | 228 | 3.0 |
| S1-M | 91 | 115 | 22 | 0 | 228 | 23.9 |
| S1-B | 43 | 113 | 0 | 16 | 172 | 5.7 |
| S2-M | 25 | 0 | 88 | 115 | 228 | 27.7 |
| S2-B | 2 | 14 | 41 | 115 | 228 | 1.9 |
| M-B | 41 | 47 | 2 | 82 | 172 | 47.7 |

[a] Data are number of alleles containing each pairwise haplotype. Alleles are designated 1 or 2 as per table 1; the approximate location of each RFLP is shown in fig 1.

[b] Same as that calculated for the $D'$ values in table 4.

[c] A single haplotype contained an unusual RFLP at locus R1; this haplotype was included with the "2" allele for all analyses.

rium measure $D = h_{11} - pq$, where $h_{11}$ is the frequency of the haplotype with the rarer allele at each locus, and where $p$ and $q$ are frequencies of the rarer alleles at loci 1 and 2, respectively. $D'$ is calculated as $D/D_{max}$, where $D_{max} = \min[pq, (1-p)(1-q)]$, for $D < 0$, or $\min[p(1-q), q(1-p)]$, for $D > 0$. The significance of differences of $D'$ from 0 is calculated from $D$, as $D^2N/[p(1-p)q(1-q)]$, where $N$ is the number of chromosomes, and where the statistic is distributed as $\chi^2$ with 1 df (Cox et al. 1988*a;* Thompson et al. 1988; Sten-Linder et al. 1991*a*). I also examined the deviation from expectation for the data in table 3, by simple $\chi^2$ analysis. Since the resulting $\chi^2$ value was nearly identical to that calculated from $D$, I only show the latter.

## Results

Haplotypes were established unambiguously by segregation in large pedigrees ascertained for familial NIDDM, as described above and in detail elsewhere (Elbein et al. 1992). I examined eight RFLPs, as described in Subjects and Methods: two for *Pst*I (P1 and P2), two for *Rsa*I (R1 and R2), two for *Sst*I (S1 and S2), and one each for *Msp*I (M) and *Bgl*II (B). Five of these (R1, R2, S1, S2, and B) have been described in a previous, smaller study which included individuals from a similar but independent population (Elbein et al. 1986). A number of other RFLPs have been described (Xiang et al. 1989; Sten-Linder et al. 1991*a*, 1991*b*) but are not included in the present study. Of 256 haplotypes predicted from the eight RFLPs, 33 were observed from the 19 pedigrees examined. Because of early evidence for nearly complete linkage disequilibrium between B and R2 RFLPs, not all individuals were examined with *Bgl*II. Thus, analysis of the complete haplotype was only possible for 172 chromosomes. Among those 56 chromosomes are typed for *Bgl*II, I noted additional haplotypes which are not represented in the completely typed chromosomes. The frequencies of the 172 complete haplotypes clearly differed from those expected (table 2), and this difference was highly significant after pooling into one group all haplotypes with an expected number of less than five ($P < .001$). The four most common haplotypes (table 2), which would be predicted to account for <18% of all INSR chromosomes, in fact accounted for 46% of INSR alleles. Heterozygosity for this locus was 91.7%.

For subsequent analyses, 228 chromosomes were analyzed for all RFLP pairs except those including the *Bgl*II polymorphism, for which 172 pairs were analyzed. The haplotype data are shown in table 3. I show the $\chi^2$ value as calculated from $D$ (see Subjects and Methods). Disequilibrium statistic $D'$ was calculated for each pair, as shown above the diagonal in table 4. Although the D′ statistic is relatively independent of allele frequencies, disequilibrium is often very difficult to prove when a rare allele is associated with a common allele (negative $D$ in the present calculations; Thompson et al. 1988).

Although the genomic structure of the INSR has been recently described (Seino et al. 1989), and although a number of the INSR RFLPs have been accu-

**Table 4**

**Pairwise Distance and Standardized Disequilibrium**

| | P1 | P2 | R1 | S1 | B | R2 | M | S2 |
|---|---|---|---|---|---|---|---|---|
| P1 ...... | | .24** | 0 | 0 | .13 | .07 | 0 | −.42 |
| P2 ...... | <50 | | .05 | 0 | .16[1] | .08 | 0 | .12 |
| R1 ..... | 40–90 | 40–90 | | .29 | .21 | .19 | .15* | .03 |
| S1 ...... | 60–110 | 60–110 | >22 | | −1.00* | −1.00* | .98** | −1.00 |
| B ....... | 60–110 | 60–110 | >25 | 3.3 | | .97** | .90** | −.54 |
| R2 ..... | 63–113 | 63–113 | >25 | 3.6 | 3.3 | | .96** | −.58 |
| M ...... | 67–117 | 67–117 | >29 | 6.5 | 5.6 | 2.3 | | −1.00** |
| S2 ...... | 72–122 | 72–122 | >34 | 12 | 11.8 | 8.5 | 6.2 | |

NOTE.—Pairwise comparison of $D'$ is shown above the diagonal, and estimated distances between RFLP pairs are shown below the diagonal. $D'$ values which are significantly different from zero are noted. Where the exact RFLP location is unknown or where gaps exist in the reported sequence of the insulin receptor gene, distances are estimated.

* $P < .05$.

** $P < .01$.

rately mapped both by my colleagues and I (Elbein 1989; Elbein and Sorenson 1990) and by others (Sten-Linder et al. 1991*a,* 1991*b*), the genomic sequence contains several gaps of unknown size. These gaps include (*a*) regions which separate the P1 and P2 polymorphisms from the R1 polymorphism and (*b*) all three of the most 5′ RFLPs from the polymorphisms of the tyrosine kinase region (fig. 1). Furthermore, available data localize the two *Pst*I polymorphisms (P1 and P2) only to a broad region of the α-chain, not to specific introns. These *Pst*I polymorphisms, which are in low but significant disequilibrium, may be physically close. I found very low levels of linkage disequilibrium between the CA-repeat polymorphism R1 (intron 9) and all other RFLPs. Only the intron 17 *Msp*I (M) RFLP (≥30 kb distant) was in significant linkage disequilibrium with R1, and even here *D*′ was low (.15).

Unlike the three 5′ RFLPs, the five RFLPs which span introns 14–21 showed degrees of linkage disequilibrium which approached the maximum for most pairs. The notable exception was the most 3′ of the RFLPs tested in the present analysis, S2. Although disequilibrium between M and S2 was highly significant and maximal ( – 1.0), I did not find statistically significant disequilibrium in any other pair involving S2. The M and S2 polymorphisms are relatively close (6 kb), but the high minor-allele frequency of the M RFLP (nearly .5) was a more likely reason for finding significant disequilibrium in this pair. In contrast, the other RFLPs in this region have minor-allele frequencies of ≤.25. The power to demonstrate significant linkage disequilibrium is low when a rare allele is associated with a common allele (Thompson et al. 1988).

To increase the power of the present analysis, I calculated the disequilibrium statistics for pooled haplotypes among the RFLPs shared with both the Caucasian sample of an earlier study by my colleagues and I (Elbein et al. 1986) and the sample in a recent study by Sten-Linder et al. (1991*a*). Since the Utah Caucasian population of the present study is generally representative of the northern European population, it should be similar to the Scandinavian population studied by Sten-Linder et al. I pooled data for the four RFLPs studied by both Sten-Linder et al. (P2, R1, S1, and S2 in the present study's nomenclature) and an earlier study by my colleagues and I (R1, R2, S1, and S2; Elbein et al. 1986). I did not distinguish between "diabetic" chromosomes and normal chromosomes in pooling these data, since linkage studies (Cox et al. 1988*b;* Elbein et al. 1992) and association studies by

**Table 5**

**Linkage Disequilibrium from Published Caucasian Data**

| RFLP Pair | No. of Alleles Scored | *D*′ | $\chi^2$ (1 df) |
|---|---|---|---|
| P2R1 ..... | 398 | .08 | .9 |
| P2S1 ...... | 402 | .13 | 3.5 |
| P2S2 ...... | 402 | .05 | .4 |
| R1R2 ..... | 317 | .21 | 4.2* |
| R1S1 ..... | 478 | .33 | 9.6*** |
| R1S2 ..... | 478 | .18 | 2.4 |
| R2S1 ..... | 314 | –.85 | 6.8** |
| R2S2 ..... | 273 | –.393 | 1.0 |
| S1S2 ...... | 508 | –1.0 | 8.56*** |

NOTE.—Analysis of pooled data from three published Caucasian studies. For pairs involving R1, four alleles containing R1 polymorphisms of sizes different than 6.3 kb (allele 1) and 6.8 kb (allele 2) were discarded from analysis. Both normal and NIDDM alleles were pooled from all studies.

* $P < .05$.

** $P < .01$.

*** $P < .005$.

Sten-Linder et al. (1991*a,* 1991*b*) suggested no differences between NIDDM and nondiabetic haplotypes. Scoring of haplotypes was converted to the present study's nomenclature (e.g., the smaller allele was termed "1"). The results are shown in table 5. Significant ($P < .05$) linkage disequilibrium was evident in pairs R1-R2, R1-S1, and S1-S2, although levels for pairs involving R1 were very low. Each of these pairs showed a trend toward significant disequilibrium in the present study's population, but *D*′ was not statistically different from zero. On the other hand, the trend toward linkage disequilibrium between R2 and S2 was lessened with the addition of data from the earlier study by my colleagues and I.

## Discussion

In a previous study (Elbein et al. 1986) my colleagues and I were unable to find evidence for linkage disequilibrium among insulin receptor RFLPs in Caucasians, except for *Rsa*I (R2) and *Bgl*II (B) polymorphisms. I still find nearly complete linkage disequilibrium between the rare alleles for these two loci but, in addition, now find substantial linkage disequilibrium among the five RFLPs of the insulin receptor β-chain in the region of exons 13–22. The most likely explanation for this discrepancy is the difference in sample size

between these studies. The earlier study probably did not have the power to detect significant disequilibrium. In the present study, I also added three RFLPs which permit more thorough analysis of this region. The results of the present study are similar to those obtained by Sten-Linder et al. (1991*a*) using a different set of RFLPs and a similar population.

Although both the present study and that of Sten-Linder et al. (1991*a*) show very high levels of linkage disequilibrium in the region of the β-chain exons, disequilibrium outside this region is less certain. Sten-Linder et al. demonstrated significant and maximal disequilibrium between a common *Xba*I RFLP in the region of exons 4–7 and the *Pst*I RFLP which I have called "P2." The distance between these RFLPs is unknown. My laboratory has not evaluated any RFLP in the region of exons 4–8. I also found low levels of disequilibrium between the two *Pst*I RFLPs of the α-chain region. Again, the precise distance between these RFLPs is unknown. In contrast, both the studies and the combined data failed to demonstrate significant disequilibrium between the *Rsa*I RFLP in intron 9 and either upstream or downstream RFLPs. A single exception was the finding of disequilibrium between R1 and S1 (present study's nomenclature) in the diabetic subgroup of Sten-Linder et al.'s study ($D' = .65$). In contrast, pooled data suggest significant but low levels of disequilibrium for R1-S1 and R1-R2 pairs.

In general, linkage disequilibrium will be affected by several factors. Under the hypothesis of uniform recombination, the likelihood of recombination will increase in proportion to physical distance. The number of recombination events between two RFLPs should also increase in proportion to RFLP age (number of generations). Thus, increasing age or distance should independently decrease the linkage disequilibrium between two RFLPs. Finally, recombination is probably not uniform, and linkage disequilibrium will be decreased between RFLP which span regions of increased recombination. Although the general lack of disequilibrium in the region coding for extracellular (α-chain) regions of the INSR locus might result from any of these factors, Sten-Linder et al. (1991*a*) and I (unpublished data) have shown the R1 RFLP to result from an insertion/deletion polymorphism. I have sequenced this region and demonstrated a long (>500 bp) region of CA repeats. Both the experience of my colleagues and I and, apparently, that of Seino et al. (1989) suggest that this region is highly unstable in vitro and thus may represent a recombinational hot spot. On the other hand, the R1 RFLP is separated from both upstream and downstream RFLPs by gaps of unknown size in the genomic DNA sequence, and low levels of linkage disequilibrium may also reflect age or distance.

My colleagues and I have previously demonstrated highly homologous Alu repeats, by DNA sequence analysis in introns 14 and 15 but not in exon 16 (Elbein 1989) and by hybridization in lambda clones containing exons 17–21. Thus, each of the β-chain RFLPs is separated by at least one Alu sequence, and four or more such sequences separate all pairs except R2 and M, which are separated by only 1 repeat. In examining the influence of Alu repeats on recombination, one can advance at least three hypotheses: (1) the recombination rate is uniformly low among nonrepetitive DNA but is markedly increased in regions of repetitive DNA (Alu or microsatellite); (2) recombination occurs at a uniformly low level when viewed over large DNA regions but occurs mostly or entirely within repetitive sequences; and (3) repetitive sequences increase recombination in some but not all genomic regions. With hypothesis 2 one might anticipate that recombination would be proportional to the density of repetitive sequences in any DNA region. The data of the present study do not fully distinguish between these hypotheses. Nonetheless, the striking linkage disequilibrium among β-chain RFLPs, despite intervening Alu sequences, suggests that these sequences do not markedly increase recombination in this region of the genome, either because of markedly increased recombination at a single Alu sequence or because of the large number of clustered Alu repeat units. This result is surprising, since recombination events certainly occur around exon 17. Both the S1 RFLP (Elbein 1989) and gene deletions (Taylor et al. 1991) represent such events. If Alu repeats promote recombination, that recombination must occur with low frequency, more in support of hypothesis 2 or 3 than hypothesis 1. Linkage-disequilibrium measures suggest that the microsatellite/dinucleotide repeat region in intron 9 may play a larger role in promoting recombination, but we cannot be certain about the role of distance between RFLP pairs spanning that polymorphism.

The combination of the present study's data and those of Sten-Linder et al. suggests the most informative combination of RFLPs for future linkage studies. This combination might include the P1 RFLP, which shows minimal disequilibrium with other RFLPs, a recently described microsatellite polymorphism of intron 2 (Xiang et al. 1991), the *Xba*I RFLP of the exon 4–7 region, the R1 RFLP of intron 9, and either the

*Msp*I or *Dra*I polymorphisms of exon 17. The remaining RFLPs are less informative and demonstrate significant disequilibrium and should be reserved for pedigrees which are uninformative with the other RFLPs. However, with this combination, the heterozygosity for the INSR locus should considerably exceed the 90% found both in the present study and in the earlier study by my colleagues and I.

## Acknowledgments

This work was supported by National Institutes of Health grant DK39311 to S.C.E. and by the Research Service of the Veterans Affairs Medical Center. Lise Sorensen and Michelle Taylor provided excellent technical assistance. The manuscript also benefited from the comments and suggestions of the anonymous reviewers.

## References

Chakravarti A, Elbein SC, Permutt MA (1986) Evidence for increased recombination near the insulin gene: implication for disease association studies. Proc Natl Acad Sci USA 83:1045–1049

Cox NJ, Bell GI, Xiang K-S (1988*a*) Linkage disequilibrium in the human insulin/insulin-like growth factor II region of human chrmosome 11. Am J Hum Genet 43:495–501

Cox NJ, Epstein PA, Spielman RS (1988*b*) Linkage studies on NIDDM and the insulin and insulin receptor genes. Diabetes 38:653–658

Cox NJ, Spielman RS, Kahn CR, Muller-Wieland D, Kriauciunas KM, Taub R (1998*c*) Four RFLPs of the human insulin receptor gene: PstI, KpnI, RsaI (2 RFLPs). Nucleic Acids Res 16:8204

DeFronzo RA, Bonadonna RC, Ferrannini E (1992) Pathogenesis of NIDDM: a balanced overview. Diabetes Care 15:318–355

Elbein SC (1989) Molecular and clinical characterization of an insertional polymorphism of the insulin receptor gene. Diabetes 38:737–743

Elbein SC, Borecki I, Corsetti L, Fajans SS, Hansen AT, Nerup J, Province M, et al (1987) Linkage analysis of the human insulin receptor gene and maturity onset diabetes of the young. Diabetologia 30:641–647

Elbein SC, Corsetti L, Ullrich A, Permutt MA (1986) Human insulin receptor polymorphisms permit linkage analysis with diabetes. Proc Natl Acad Sci USA 83:5223–5227

Elbein SC, Maxwell TM, Schumacher MC (1991) Insulin and glucose levels and prevalence of glucose intolerance in pedigrees with multiple diabetic siblings. Diabetes 40: 1024–1032

Elbein SC, Sorensen L (1990) MspI and SstI RFLPs at the human insulin receptor locus on chromosome 19. Nucleic Acids Res 18:209

——— (1991) Genetic variation in insulin receptor β-chain exons among members of familial type 2 (non-insulin dependent) diabetic pedigrees. Diabetologia 34:742–749

Elbein SC, Sorensen L, Taylor M (1992) Linkage analysis of the insulin receptor gene in familial non-insulin dependent diabetes mellitus. Diabetes 41:648–656

Elbein SC, Ward WK, Beard JC, Permutt MA (1988) Familial NIDDM: molecular-genetic and assessment of insulin action and pancreatic β-cell function. Diabetes 37:377–382

Hedrick PW (1987) Gametic disequilibrium measures: proceed with caution. Genetics 117:331–341

O'Rahilly S, Choi WH, Patel P, Turner RC, Flier JS, Moller DE (1991) Detection of mutations in the insulin receptor gene in NIDDM patients by analysis of single stranded conformation polymorphisms. Diabetes 40:777–782

O'Rahilly SO, Trembath RC, Patel P, Galton DJ, Turner RC, Wainscoat JS (1988) Linkage analysis of the human insulin receptor gene in type 2 (non-insulin-dependent) diabetic families and a family with maturity onset diabetes of the young. Diabetologia 31:792–797

Rotter JI, Rimoin DL (1981) The genetics of the glucose intolerance disorders. Am J Med 70:116–126

Seino S, Seino M, Nishi S, Bell GI (1989) Structure of the human insulin receptor gene and characterization of its promoter. Proc Natl Acad Sci USA 86:114–118

Sten-Linder M, Olsson M, Iselius L, Efendic S, Luthman H (1991*a*) DNA haplotype analysis suggests linkage disequilibrium in the human insulin receptor gene. Hum Genet 87:469–474

Sten-Linder M, Vilhelmsdotter S, Wedell A, Stern I, Pllare T, Arner P, Efendic S, et al (1991*b*) Screening for insulin receptor gene DNA polymorphisms in a Scandanavian population. Diabetologia 34:265–270

Taylor SI, Cama A, Accili D, Barbetti F, Imano E, Kadowaki H, Kadowaki T (1991) Genetic basis of endocrine disease I: molecular genetics of insulin resistant diabetes mellitus. J Clin Endocrinal Metab 73:1158–1163

Thompson EA, Deeb S, Walker D, Motulsky AG (1988) The detection of linkage disequilibrium between closely linked markers: RFLPs at the AI-CIII apolipoprotein genes. Am J Hum Genet 42:113–124

Ullrich A, Bell JR, Chen EY, Herrera R, Petruzzelli LM, Dull TJ, Gray A, et al (1985) Human insulin receptor and its relationship to the tyrosine kinase family of oncogenes. Nature 313:756–761

Xiang K-S, Cox NJ, Sanz N, Huang P, Karam JH, Bell GI (1989) Insulin-receptor and apolipoprotein genes contribute to development of NIDDM in Chinese Americans. Diabetes 38:17–23

Xiang K, Granzvist M, Seino M, Seino S, Bell GI (1991) Microsatellite polymorphism in human insulin receptor gene (INSR) on chromosome 19. Nucleic Acids Res 19: 5094