Perspectives in Diabetes
A Practical Approach to Identification of Susceptibility Genes for IDDM
JOHN A. TODD AND STEPHEN C. BAIN

Of all the common diseases that have a genetic component, IDDM is probably the most tractable to the experimentalist. Large numbers of nuclear multiplex families are available, which can be stored as permanent cell lines; diagnosis is relatively unambiguous; and a mouse strain, the NOD, spontaneously develops autoimmune IDDM similar to the human disorder. In addition, the resolution and accessibility of the human genome map has been revolutionized by the discovery and widespread application of the PCR, particularly the amplification of short, tandemly repeated segments of DNA called microsatellites, which display high levels of allelic polymorphism. With these reagents, the stage is set for dissection of the genetic factors that control the pathophysiology of IDDM. Diabetes 41:1029–34, 1992

S
ignificant familial clustering of IDDM exists: up to 15% of IDDM patients have a first-degree relative with the disease (1). The risk for siblings is much greater than the population prevalence, and this ratio, defined by Risch (2) as \( \lambda_{sr} \), has been calculated to be 15 (average lifetime sibling risk of 6% divided by the population prevalence, 0.4%). The higher the value of \( \lambda_{sr} \), the more familial the disease. The aggregation of IDDM in families probably is attributable both to shared genetic factors and shared environmental factors. The importance of the environment is demonstrated by the 64% discordance of genetically identical twins (3) and suggests that environment plays a dominant role in disease development. It is possible that genetically susceptible individuals are protected by environment (4), which is consistent with the observed full penetrance of NOD mouse susceptibility genes when all sources of infection are removed (5). Important changes in susceptibility factors in the environment are evident by the increase in incidence of IDDM in several countries (6,7).

HLA REGION
The best evidence that genes are important in susceptibility comes from the established associations and linkages of IDDM with chromosomes 6p21 at the HLA region (8) and 11p15 at the INS region (9-11). Both of these were discovered using a candidate gene approach. However, the linkage disequilibrium that exists and that created the opportunity for detection of the association in the first place has plagued attempts to identify the primary disease mutations. At least one polymorphic site, at position 57 of the HLA-DQ \( \beta \)-chain, has been correlated with susceptibility (12) and also has been shown to influence the functions of class II molecules in T-cell recognition and, presumably, peptide binding (13). It seems likely that a component of HLA-encoded susceptibility is determined by various allelic conformations of the heterodimeric DQ molecule in antigen presentation. Some evidence indicates that other class II molecules, such as HLA-DR, also influence susceptibility (14), and the significant association of certain haplotypes extending from the HLA-A locus to HLA-DP indicates that genes outside the class II region may play a role (1,15,16). In 130 multiplex families that we collected from the United Kingdom, the A1,B8,DR3 haplotype accounted for >60% of DR3 haplotypes present in diabetic probands compared with ~25% of DR3 haplotypes in the nondiabetic population. All coding sequences of the MHC are steadily being identified and characterized (17). Poly-

From the Nuffield Department of Surgery, University of Oxford; and the Department of Medicine, University of Birmingham, United Kingdom.

Address correspondence and reprint requests to John A. Todd, Nuffield Department of Surgery, University of Oxford, John Radcliffe Hospital, Headington, Oxford, UK, OX3 9DU.

Received for publication 29 April 1992 and accepted 14 May 1992.

IDDM, insulin-dependent diabetes mellitus; PCR, polymerase chain reaction; HLA, human leukocyte antigen; INS, insulin gene; MHC, major histocompatibility complex; IBD, identity-by-descent; df, degree of freedom; NIDDM, non-insulin-dependent diabetes mellitus; MODY, maturity onset diabetes of the young; bp, base pair.
morphisms in these newly identified genes, for example the peptide transporter loci in the class II region (18), will help continued understanding of the association of HLA with IDDM.

Nevertheless, it will be difficult to separate the primary effects of newly characterized polymorphisms from established HLA-DR and -DQ effects. This will necessitate reagents, such as very large groups of proband and sporadic IDDM patients (isolated cases without a first-degree relative with IDDM) and ethnically matched controls that are of sufficient size to allow subgrouping, for example, into DR3- or DR4-negative and DR3- or DR4-positive individuals. Currently, we have collected 135 multiplex families (19), defined as two-generation families with at least two affected children, and with both parents available and at least one of the affected children diagnosed <17 yr of age. Cell lines and DNA from all of these families are available from the British Diabetic Association. In addition, we have collected 300 sporadic IDDM patients (diagnosed <17 yr of age with no IDDM siblings) and 300 control subjects (with no IDDM siblings), all whites with grandparents born in the UK (11). Lymphoblastoid cell lines are being established from all these individuals to create permanent repositories that will provide continuous sources of DNA. These resources can be studied by several laboratories, allowing accumulation of genetic information on the same individuals. Cell lines and DNAs from the sporadic IDDM patients and control subjects will be available from the European Collection of Animal Cell Cultures.

The criterion of ethnic homogeneity is stressed because, clearly, the frequencies of certain HLA haplotypes differ significantly between European countries. For example, a Dw15-like subtype of DR4 (DRB1*0405) is present in 20% of French IDDM patients (20) but <1% of UK or US IDDM patients (14). Also, the B18,DR3 haplotype is much more common in French IDDM patients (15) compared with cases from the UK (1). Collections of IDDM families in the United States (21) therefore may be subject to such intraEuropean ethnic differences in haplotype and genotype frequencies. This should be taken into account before excluding association or linkage of a marker locus to disease.

Comparative studies in different racial groups also will continue to be useful in the search for additional susceptibility genes within and without the MHC (22–25). Allelic associations with IDDM, such as the protective effect of DQB1*0602 that is maintained in all racial groups studied so far (26), are likely to have primary influences on disease susceptibility. Recently, we characterized diabetogenic HLA alleles in Indian-Hispanic IDDM patients and control populations collected from one area in Colorado (27). Most strikingly, the inferred haplotype (parental typing was unavailable) with the greatest relative risk was A1,B8,DR3, which was present in 11/65 Indian-Hispanic IDDM patients, even though this combination of alleles was not found in 47 Indian-Hispanic control subjects (compared with 9/56 non-Hispanic subjects). This inferred haplotype was present at approximately the same frequency in non-Hispanic and Indian-Hispanic IDDM patients (18/73 vs. 11/65, respectively). The data suggest that racial admixture has introduced the diabetogenic A1,B8,DR3 haplotype from white populations into this Indian-Hispanic group, and that the reduction in frequency of DR3 haplotypes in this ethnic group might be responsible for the observed twofold lower incidence of IDDM within this population.

Just how class II molecules predispose and protect from IDDM is currently unknown. In the NOD mouse, the immune system can be dissected experimentally, for example, by the creation of allogeneic chimeric mice (28). Allogeneic mice are created by pairing 8 cell-stage embryos in culture, followed by transfer into the uteri of pseudopregnant foster mothers and selection of progeny with coat-color chimerism. NOD thymic epithelium expressing the unique I-A<sup>δ</sup> molecule is required for the development of insulins, and it has been suggested that the I-A<sup>δ</sup> molecule causes insulins by positively selecting autoreactive T-cells (28). Breeding experiments in mice have led to the replacement of the NOD H-2<sup>δ</sup> region by the H-2<sup>d</sup> haplotype (29). These congenic mice develop extensive periductal, perivascular, and peri-insulitis T-cell infiltrates, but rarely develop destructive insulins. It appears that the NOD H-2<sup>δ</sup>, of which the I-A gene is likely to be critical disease determinants, provides a necessary stimulus to infiltrating T-cells to invade and destroy β-cells. Part of this stimulus probably is provided by antigen-specific T-cell recognition and activation, with as yet unidentified peptide fragments derived from β-cells bound to NOD I-A. The NOD H-2 is therefore essential but not sufficient for the development of insulins and diabetes. Loci not linked to the MHC on chromosome 17 are also necessary.

**INS REGION**

A second susceptibility region for IDDM occurs at the tyrosine hydroxylase–INS-insulinlike growth factor II gene cluster in 11p15. Fine mapping by linkage disequilibrium of polymorphisms in this region has shown that the association of the region with IDDM peaks around the INS gene (10). We have confirmed these associations in a study of 126 probands, 235 sporadic IDDM patients, and 218 control subjects from the UK (19). The precise mutations that cause the susceptibility determined by the INS gene region remain to be determined, but they are in strong linkage disequilibrium with polymorphisms at the minisatellite at the 5' end of the gene (9,10) and variant Dralll, Pstl, and Fokl sites at positions 805, 1127, and 1428 of INS (10). It is possible that these polymorphisms themselves may be involved.

The relative risk of the INS +/+ genotype (where + is a disease-associated DNA polymorphism in INS, for example a variant Fokl site at position 1428) relative to +/− or −/− genotypes is ~2.5–3 in France, the UK, and the US (9–11). This compares with the relative risk for the DR4/X,4/4 genotype relative to DRXX (X is not DR3 or DR4) of ~8–16 in the UK (11). Previous, in sporadic and proband IDDM patients (of multiplex families) from France, it was found that the predisposing effects of DR4 are only present in INS +/+ individuals and not in INS +/− or −/− individuals (10). In the UK, however, the
relative risk for the DR4/X,4/4 genotypes is ~8 in INS +/+ or −/− individuals, indicating that in these proband and sporadic IDDM patients no significant interaction occurs between DR4 and INS (11). Nevertheless, in both French and UK multiplex families, INS + variants from heterozygous parents (+/+ or −/−) segregate more frequently to DR4/X,4/4-positive diabetic offspring than DR3/X,3,3-positive offspring, which is consistent with a previous study in the US (30).

We estimate that the genotype HLA-DR3,4,DQB*0302,INS +/+ is present in ~0.7% of the nondiabetic UK population compared with 22% of IDDM patients. If we assume that the population prevalence of IDDM in the UK is 0.2% up to 17 yr of age, then the absolute risk of this genotype is (0.30 + 0.002)/0.007 = 0.06 or ~6%. Therefore, UK population screening for genetic risk is not justified at present, but would be valuable in families with an affected member, particularly in combination with immunological markers, such as autoantibodies (31).

One of the most important conclusions to come from the study of INS and IDDM is that the power of family studies is restricted if the disease allele frequency is high, as originally discussed by Risch (2). In 75 IDDM families (with both siblings diagnosed <17 yr of age, no IDDM parents) that were fully informative at the INS region, the IBD probabilities are 0.10, 0.52, and 0.32 for sharing 0, 1, and 2 haplotypes IBD, respectively. The significance of the deviation from expected just fails to reach \( P < 0.05 \), but clearly is not convincing evidence of linkage. Compare HLA in the same data set: In 85 fully informative families, 0.08, 0.39, and 0.53 siblings share 0, 1, and 2 alleles IBD at HLA (\( \chi^2 = 39, df = 2, P < 0.001; \text{ lod} > 3 \)). The UK population frequency \( P \) of the INS FokI + variant = 0.78. If \( P = 0.2 \), and assuming that the relative penetrances for the INS genotypes +/+ or +/−, and −/− are 3, 1, and 0, respectively, and that the dominance variance is 0, the expected IBD probabilities can be calculated (2.32) to be 0.11, 0.5, and 0.39. This deviation from expected (no linkage) for 75 fully informative sibling pairs gives \( \chi^2 = 11.8 (df = 2, \text{ lod} = 2.6) \).

If a disease-associated polymorphism is known, then increased power to detect linkage in families can be achieved by selecting families whose parents are heterozygous for the polymorphism (that is, a parent possesses a susceptibility haplotype and nonsusceptibility haplotype) (10,30). Stronger evidence of linkage was obtained using this approach by analysis of the sharing of the INS variants in affected sibling pairs from parental meioses (\( P < 0.01 \)). Interestingly, in French IDDM families, it was observed that all the linkage came from male meioses, suggesting that maternal imprinting was operating. We have not observed any significant difference between male and female meioses using equivalent numbers of informative families from the UK: 18 sibing pairs shared 1 INS FokI allele IBD from paternal meioses compared with 8 sharing 0 (\( P < 0.05 \)); and 16 shared 1 from maternal meioses versus 6 sharing 0 (\( P < 0.05 \); combined evidence of linkage, \( P < 0.005 \)). These differences in results from independent data sets stress the importance of further replication using additional, inde-
β-cell function or the operation of the immune system is a potential candidate. The magnitude of the task of testing candidate loci has been reduced by the application of PCR to direct sequencing and other techniques (46–50). Most genes are close to segments of tandemly repeated simple sequences or microsatellites that display high levels of allelic polymorphism by variation in the number of repeats (51,52). Microsatellites are <100 bp in length and, therefore, are easy to clone, sequence and PCR. The presence of several alleles, perhaps at different closely linked loci provides a powerful way to test for linkage disequilibrium between a haplotype and disease (53–55). The Z-16 allele of a tetranucleotide repeat microsatellite in the 5‘ half of the tyrosine hydroxylase gene is in linkage disequilibrium with the mutation(s) that cause IDDM in this INS gene region (52). Failure to detect linkage disequilibrium does not mean that it is absent, so several polymorphisms should be tested to eliminate a candidate gene.

It is possible that genes that influence susceptibility to NIDDM also affect the development of IDDM (56). The region near the adenosine deaminase gene on chromosome 20q, which has been linked to MODY in one American pedigree and one French family, also can be considered as a candidate region for IDDM susceptibility. In addition, any regions or genes that have been linked or associated with susceptibility to other autoimmune diseases are candidates. For example, genetic overlap between autoimmune thyroiditis and IDDM is likely in humans (57) and in mice (58).

Other regions of the human genome that are candidate gene regions are those chromosomal segments that are homologous to regions of the mouse genome (59). We have mapped several susceptibility loci to the NOD mouse (37). One of the most interesting comparative mapping results to come from these studies involves mouse chromosome 1. We mapped a gene that causes diabetes and insulinitis in NOD mice to the centromeric portion of mouse chromosome 1 within a region of 40M that is highly homologous or syntenous with human chromosome 2q (38). Because genes for complex traits will be almost impossible to map with the methods used for the characterization of monogenic diseases, such as cystic fibrosis and muscular dystrophy, a combination of the candidate gene approach and linkage analysis is required. On mouse chromosome 1, the Lsh/Ily/Bcg locus controls susceptibility to a number of infectious diseases (60). It is a good candidate for IDDM because it affects macrophage function (60), and macrophages are essential for the development of NOD insulinitis and diabetes (61). It is known that NOD carries the disease resistance allele at Lsh/Ily/Bcg, whereas C57BL/10, the diabetes-resistant parent used to map the chromosome 1 diabetes gene, is susceptible at Lsh/Ily/Bcg (62). The cloning and characterization of this locus is now of even wider interest. Interestingly, a second locus that controls sialitis in NOD mice on chromosome 1 maps close to Bcl-2 (63). The Bcl-2 gene is itself a candidate locus, because mice transgenic for Bcl-2 expressed with an immunoglobulin gene promoter possess B-cells that are resistant to apoptosis and produce abnormal levels of autoantibodies (64).

The strongest effect we have observed so far in mice, outside the H-2 region, is encoded by chromosome 3, designated ldd-3 (37). ldd-3 may be composed of more than one gene because the region of linkage to diabetes and insulinitis extends over a 40cM region of chromosome 3, including the II-2 and Tshb genes. Wicker and Peterson (unpublished observations) have bred NOD mice using these microsatellite markers that are congeneric for the II-2-Tshb interval, and these mice show reduced susceptibility to diabetes. These mice can now be used to study the functional defects specifically determined by this region of chromosome 3, and also to fine map the susceptibility loci.

**SUBPHENOTYPES**

As pointed out by Cox et al. (65), the great advantage of studying MODY as a way into the more complex disease classified as NIDDM is that MODY is inherited as a single-gene dominant trait. Although MODY is uncommon, the high penetrance of MODY genes leads to multiple affected members in several generations. It is much easier to map genes in large pedigrees than in many nuclear, two-generation families, but in IDDM we have no choice. To further increase the degree of difficulty, different genes may be acting in different families (reflected as locus heterogeneity), as has been found in MODY (43, 66). We have evidence of genetic heterogeneity in NOD mice backcrossed to diabetes-resistant C57BL/10 mice because the diabetic first backcross progeny have different genotypes at the different susceptibility loci. About 15% of the diabetic progeny were heterozygous at the most likely location of ldd-3, whereas other mice were heterozygous at different susceptibility loci. None of the susceptibility loci we have mapped therefore are fully recessive, and NOD homozygosity at ldd-3 and at the other individual loci is not essential for the development of diabetes (37). Whether or not any of these non-MHC genes is absolutely essential for the development of diabetes can be determined by making congenic strains. Locus heterogeneity is likely to be present in outbred human IDDM families, and therefore every effort must be made to identify families that might be atypical clinically or genetically. For example, a mitochondrial mutation has been identified that causes a form of diabetes that is inherited maternally and resembles IDDM in that the patients can develop the disease <30 yr of age, and some take insulin (67). We have identified at least one family with these characteristics.

Continued investigation of biochemical defects in IDDM should reveal characteristics that are less complex genetically. Because NOD mice can be manipulated experimentally, Wicker and Peterson (unpublished observations) identified such a single-gene trait in NOD. In our laboratory, Jan-Bas Prins has located a mutation in a gene encoding an Fc receptor that co-segregates exactly with the phenotype. Another example of this approach is the recent finding by segmentation analysis that a major locus may control fasting insulin levels in NIDDM pedigrees (68).
EXCLUSION MAPPING

The scanning of the genome with mapped marker loci was successful in NOD mice. It was, compared to humans, a defined experimental problem. Mice are the animal of choice to study mammalian genetics because the parental strains are inbred and therefore homozygous at every locus, so that only two sets of alleles are present: thus, the environment variation can be controlled to some extent, and large pedigrees can be generated. Exclusion mapping by Bell et al. (66) led to the first mapping of a MODY locus on chromosome 20, but this trait is monogenic. From studies by Risch (41) and Hyer et al. (42), we know that we can detect genes with effects equivalent to the insulin gene region. At least 200 families with 2 affected siblings (see above) and fully informative marker loci (or clusters of polymorphic markers) at at least every 200Cm are required. The latter is now feasible with the development of large numbers of micro-satellites (52). Areas of the genome not covered adequately can be made informative by isolation of microsatellites from YAC clones. If weak evidence of linkage is obtained, then further markers can be developed in the region, the study can be extended to other collections of families, and candidate genes and the most closely linked polymorphisms in the region analysed to test for linkage disequilibrium.

Given current availability, carefully assembled resources, and genetic mapping reagents and techniques, the genetics of diabetes is entering an extremely exciting phase. Ultimately, this research will lead to better prediction and open up new areas in the search for preventative treatments.

ACKNOWLEDGMENTS

We thank the Medical Research Council, the British Diabetic Association, the Juvenile Diabetes Foundation, and the Wellcome Trust for financial support. We also thank our colleagues, particularly Neil Risch, Mark Lathrop, Larry Peterson, and Linda Wicker for valuable discussions.

REFERENCES

34. Cox NJ, Bell GI: Disease associations: chance, artifact, or suscept-


30. 2