Perspectives in Diabetes
Glucokinase and NIDDM
A Candidate Gene That Paid Off
M. ALAN PERMUTT, KEN C. CHIU, AND YUKIO TANIZAWA

Glucokinase, the major enzyme that phosphorylates glucose upon entry into liver and islet β-cells, has been considered a prime candidate for inherited defects predisposing to NIDDM. Now that the human gene has been isolated, this question has been addressed directly. Polymorphic markers flanking the gene were identified. These markers (microsatellites) are composed of variable numbers of dinucleotide repeats that vary in size, resulting in different alleles. Variably sized alleles can be typed rapidly from genomic DNA of individuals by the PCR. Studies of inheritance of glucokinase genes have revealed significant linkage in families with early-onset NIDDM, or MODY, and mutations have been identified within the coding region of the gene in some families. These studies are extremely encouraging, as they indicate that genes can be identified even in this heterogeneous genetic disorder. This study considers the phenotypes that result from glucokinase defects and the relationship of MODY to NIDDM, and it estimates the role of glucokinase defects in NIDDM in general. Diabetes 41:1367–72, 1992

A defect in pancreatic islet β-cell glucose sensing has been suggested to be the primary abnormality in the pathogenesis of NIDDM (1,2,3). Meglasson and Matschinsky (4) initially proposed the concept that the enzyme glucokinase (ATP:c-glucose 6-phosphotransferase [EC 2.7.1.1]) is the key component in glucose sensing by pancreatic islet β-cells. The following is quoted from a review of their work (4):

"Glucokinase occupies an important role in controlling glucose phosphorylation and metabolism both in the liver and in pancreatic islets. In the β-cells glucokinase functions as a pacemaker of gluconeogenesis at physiological glucose levels. It determines the unique characteristic of islet glucose usage...Because metabolism controls hexose induced insulin release, glucokinase is considered the best qualified candidate for the elusive glucose sensor of the pancreatic β cells. A deficiency of glucokinase would disturb glucose homeostasis. Decreased islet glucokinase would diminish islet glycolysis and would result in a higher set point of β-cells for glucose induced insulin release. Decreased liver glucokinase would cause less efficient hepatic glucose disposal."

For numerous reasons, it was not easy to study glucokinase in pancreatic islets (4), so defects in man could only be hypothesized. With the introduction of modern molecular genetic methods, however, it became possible to evaluate the potential role of defects in glucokinase in human diabetes.

A breakthrough occurred with the cloning of the glucokinase cDNA from rat liver by Andreone et al (5). Shortly thereafter, an islet-specific glucokinase was isolated from a rat insulinoma cDNA library, which differed from the liver cDNA on the 5'-end (6). The two tissue-specific cDNAs predicted proteins that differed by 15 amino-terminal residues. Finally, Southern blot analysis and isolation of genomic clones showed that the two cDNAs originate from the same gene (6,7). Both forms of glucokinase are encoded by 10 exons, with shared exons 2–10, differing only in the first exons. Presumably, separate promoter regions for islet and liver glucokinases allow for tissue-specific regulation (8). Alternatively, spliced forms of rat glucokinase mRNAs have been found in islets (6), liver (9), and pituitary (10,11), although no evidence exists to suggest that these minor variant
glucokinase mRNAs encode functional enzymes. In experimental animals, insulin appears to be the major regulator of the liver enzyme (7,12), whereas plasma glucose concentrations control the level of the islet enzyme (13-15). As a major rate-limiting factor in both hepatic and islet glucose metabolism, and hence glucose-stimulated insulin release from islets, the level of glucokinase activity would be anticipated to play a critical role in glucose homeostasis. Glucokinase levels are known to vary with altered hormonal and nutritional conditions in experimental animals, although these changes are established more firmly for the liver enzyme (16).

ISOLATION OF THE HUMAN GLUCOKINASE GENE

The rat islet glucokinase cDNA was used to screen a human liver cDNA library, and two different forms of the human liver cDNA were isolated (17). The more abundant form predicted a protein of 464 amino acids, with >97% identity at the amino acid level with rat liver glucokinase. A less abundant splice variant was found that included a 124-bp cassette exon between exons 1 and 2 (exon 2A in Fig. 1), similar to the variant described in rat liver (9). This cassette exon altered the reading frame of the human gene and predicted a protein that differs at the amino-terminus. Both human liver glucokinase proteins were synthesized in vitro and shown to have comparable biological activity. Specific antibodies will be required to determine whether both predicted forms of human liver glucokinase exist.

Next, an islet glucokinase cDNA clone was isolated from a human islet cDNA library (18). Like the rat liver and rat islet enzymes, the human liver and islet glucokinase cDNAs differ at their 5'-ends, and thus differ in the predicted 15 amino-terminal residues. The human liver and islet glucokinases also are encoded by a single gene, as in the rat, as became apparent from the following studies. Southern blot analysis of genomic DNA from individuals, hybridized with either liver or islet cDNAs, revealed only a single glucokinase gene (19). When human genomic clones were isolated (19), the arrangement of the gene (Fig. 1) was found to be quite similar to that of the rat gene (7). The total number of exons is 12, with exons 2-10 common to all forms of the gene. In addition, different first exons were identified for islet (1B) and liver (1H). A cassette exon (2A) between exons 1 and 2 was found that was expressed in the variant liver form of glucokinase (17), accounting for the additional 124 nucleotides found in the variant liver glucokinase mRNA. Thus, the two different glucokinase mRNAs in liver and islets appear to be the result of transcription from tissue-specific promoters, with different first exons followed by common splicing of exons 2-10. Analysis of glucokinase mRNA from human islets and liver confirmed the tissue specificity of this processing (18). Different 5'-ends of the mRNAs predict different amino-terminal ends of the enzymes in islets and liver, but the biological consequences of this have not been determined.

IDENTIFICATION OF MICROSATELLITE REPEATS IN
GLUCOKINASE GENOMIC CLONES

Over the last 12 yr, candidate genes have been evaluated in people with diabetes by identifying polymorphisms in DNA, which serve as genetic markers (20). The frequencies of the polymorphisms have been compared in populations of unrelated diabetic patients and control subjects (21). If the frequency differs, it is presumed that the polymorphism is in linkage disequilibrium with a mutation in the gene. DNA polymorphisms also serve to distinguish parental alleles in families, and provide markers to evaluate cosegregation of the gene with the diabetes phenotype.

The polymorphisms previously identified were in restriction enzyme sites (RFLPs), or, more rarely, in a hypervariable region of DNA consisting of tandem repeats of 10-25 bp, called minisatellites (22). These polymorphisms were evaluated by the tedious and time-consuming Southern blot technique. Recently, regions of di-, tri-, and tetra-nucleotide repeats have been found to occur every 1-30 kb in genomic DNA (23-25). These microsatellite repeats, if they contain more than 5-10 repeat units, are often polymorphic. Once the sequence around the microsatellite has been determined, then oligonucleotide primers unique to the repeats can be used to amplify, by the PCR, the region of DNA from individuals. This technique is illustrated schematically in Fig. 2A. Alleles of different size then are defined by electrophoresis on DNA sequencing gels, where alleles differing by two or more nucleotides can be identified easily. Compared with RFLPs, microsatellite analysis is less expensive, less tedious, and less time consuming. Most important, microsatellite repeats are far more polymorphic, and thus considerably more informative in family studies.

We screened the glucokinase genomic clones and identified a region of microsatellite repeats --8 kb 3' to the gene (26) (Fig. 1). PCR amplification with 32P-labeled primers revealed extensive polymorphism. One nomenclature for the polymorphism designates the common allele Z or 0. The other alleles differ from the common allele by the number of nucleotides. For example, the Z + 2 allele has an additional [CA] relative to the Z allele. A total of 5 alleles were identified in American blacks, 6 alleles in whites, and 4 alleles in Pima Indians, with heterozygosity varying from 0.48 to 0.55 (i.e., in 48-55% of individuals, the glucokinase genes on each of their two chromosomes can be distinguished; thus, inheritance of these genes can be followed in their offspring). An
The glucokinase gene, involved in diabetes, was mapped to chromosome 7p in unrelated individuals. Examples of microsatellite repeat polymorphisms at the 3'-end of the gene are illustrated in Fig. 3. This microsatellite repeat was found in 38% of NIDDM patients and 23% of control subjects. Another allele was associated with NIDDM in Mauritians. In contrast to the findings in these two black populations, we recently found the frequency of alleles to be the same in Welsh NIDDM patients and control subjects. One limitation of population studies, however, is the requirement that a substantial number of markers already mapped in these CEPH families are as a result of the Human Genome Project. Mapping the glucokinase gene required typing all of the members of the pedigrees, and determining the linkage disequilibrium between these two markers has been found (Y.T., K.C.C., M. Province, A. Rees, and M.A.P., unpublished observations). Linkage disequilibrium indicates that the frequency of the observed haplotypes differed from the frequencies of the individual repeats predicted if the two regions were associated randomly. This finding of linkage disequilibrium predicts that the two flanking microsatellites will be in linkage disequilibrium with mutations in the gene, if they exist, and thus more predictive of the risk of diabetes than either marker alone.

**ROLE OF GLUCOKINASE IN NIDDM**

**Population studies.** Using the rapid PCR assays for the microsatellite repeats at the glucokinase locus, populations of unrelated individuals of various racial groups were evaluated. The frequency of one allele (3': Z + 4) was found to be associated significantly with NIDDM in American blacks, occurring in 38% of NIDDM patients and 23% of control subjects. Another allele (3': Z + 2) was associated significantly with NIDDM in Mauritian Creole NIDDM patients. In contrast to the findings in these two black populations, we recently found the frequency of alleles to be the same in Welsh white NIDDM patients and control subjects (Y.T., K.C.C., M. Province, A. Rees, and M.A.P., unpublished observations). More recently, we have typed NIDDM patients at both the 5' and 3' repeats and developed haplotypes, which may be better predictors of the risk for NIDDM than either repeat alone. One limitation of population studies, however, is the requirement that a substantial number of NIDDM patients have the same mutation.

**Family studies.**

Inheritance of specific glucokinase genes in a family is illustrated in Fig. 3. This microsatellite repeat, and thus the glucokinase gene, was mapped to chromosome 7p in somatic cell hybrids and in CEPH (Paris, France) pedigrees (26,43). Mapping the glucokinase gene was accomplished readily because of the large number of markers already mapped in these CEPH families as a result of the Human Genome Project. Mapping the glucokinase gene required typing all of the members of the CEPH families at this locus, and determining the recombination frequency with the previously mapped gene markers.

Subsequent to the identification of the microsatellite in the 3'-region, additional genomic clones were found that encoded another microsatellite, located about 6 kb 5' to islet exon 1 (Fig. 1) (19). These two microsatellites thus flank opposite ends of the gene, and are separated by >50 kb of DNA. By typing individuals for both repeats, haplotypes could be defined (a haplotype can be described by typing an individual at two or more loci on a single chromosome). Significant linkage disequilibrium between these two flanking markers has been found (Y.T., K.C.C., M. Province, A. Rees, and M.A.P., unpublished observations). Linkage disequilibrium indicates that the frequency of the observed haplotypes differed from the frequencies of the individual repeats predicted if the two regions were associated randomly. This finding of linkage disequilibrium predicts that the two flanking microsatellites will be in linkage disequilibrium with mutations in the gene, if they exist, and thus more predictive of the risk of diabetes than either marker alone.

**Linkage to glucokinase is found in families with early-onset NIDDM.** The microsatellite repeats were used to evaluate linkage in NIDDM pedigrees from France (29). A remarkable degree of linkage was identified, but only in families with so-called early-onset NIDDM, or MODY, with at least two individuals diagnosed <25 yr of age. The significance of this finding will be more clear from the following discussion. Sixteen families were studied, and a maximum lod score of 11.6 was found for combined analysis of all 16 families, for \( \Theta = 0.01 \) (lod is log of the odds of linkage versus non-linkage).
nonlinkage, thus the odds of linkage was $10^{11.6}: 1$). Within the 16 families, evidence for genetic heterogeneity was observed, as linkage to glucokinase was not evident in all 16 families. MODY previously had been mapped to the ADA locus on chromosome 20q in one large pedigree (30). Twelve of the 16 French MODY families had negative lod scores at the ADA locus, but one family had a lod score of 1.24 at $\Theta = 0.05$. Interestingly, this was one of the two French families with significant negative ($<-2$) lod scores at the glucokinase locus. Thus, even within this restricted subphenotype of NIDDM, heterogeneity was observed. One of the early-onset NIDDM families subsequently was shown to have a G to T substitution in codon 279 (using the amino-acid numbering of the islet form), changing a GAG to TAG, a stop codon, in one allele of exon 7 (31). This mutation encodes a truncated glucokinase mRNA, and thus a presumed nonfunctional glucokinase. As the individuals in this family are heterozygotes, they would be predicted to have 50% of normal levels of hepatic and islet $\beta$-cell glucokinase, provided the normal allele could not compensate.

Confirmation of the linkage of glucokinase to MODY was seen when an independent study evaluated two large pedigrees in England and found a positive lod score of $4.6 (\Theta = 0)$ in one (32). Many of the members of this pedigree were found to have classical NIDDM, suggesting that glucokinase defects may be common in NIDDM.

The importance of the linkage results, and phenotypic characterization of the families. The results of the two linkage studies (29,32), along with the identification of a point mutation in the gene in one family (31), provided compelling evidence that mutations in glucokinase genes result in NIDDM in affected family members. Now, for the first time, we have a genetic marker to identify affected individuals, and to test Meglasson and Matschinsky’s (4) predictions at the phenotypic level, even before the onset of overt symptoms of diabetes. As the mutations were described only recently, few studies have been done.

Additional glucokinase mutations in exon 8 have been found recently in 3 of the French MODY families (33). These 3 mutant glucokinase proteins resulted in different phenotypes among the 3 families. In general, they were characterized by either glucose intolerance, mild fasting hyperglycemia, or mild diabetes. In the large BX pedigree from England, affected individuals were identified by glucokinase allele, regardless of symptoms (32: R.C.L. Page, A.T. Hattersley, B. Barrow, R. Spivey, P. Patel, M.A.P., J.S. Wainscoat, R.C. Turner, unpublished observations). Diabetes was usually asymptomatic at diagnosis, and was treated by diet in 15 of 18 diabetic members. One subject, age 18 yr and lean, was normoglycemic and not clinically affected. Gestational diabetes had been diagnosed in 5 of 11 women. Fasting plasma glucose was 4.3–12.6 mM at the time of study. Islet $\beta$-cell function and insulin sensitivity were assessed by a constant infusion of glucose with model assessment (CIGMA; 34). A ~50% decrease in $\beta$-cell function was noted in affected members, with no difference in insulin sensitivity.

The glucokinase gene paid off, but how much? Identifying a mutant gene that is responsible for NIDDM in some families is certainly a generous reward for years of effort evaluating candidate genes in this disorder. The glucokinase results in MODY families clearly indicate that specific genes predisposing to diabetes can be identified, even in this heterogeneous genetic disorder. Yet the full extent of defects at the glucokinase locus in NIDDM has not been defined. We are now faced with a number of obvious questions:

1. What fraction of MODY patients have glucokinase defects, and how do we approach those families with other genetic defects? Further analysis of the MODY pedigrees from France (Ph. Froguel, M. Vaxillaire, S. Lesage, G. Velho, M.O. Butel, H. Zouali, M. Stoffel, N. Vionnet, Ph. Passa, M.A.P., J.S. Beckmann, G.I. Bell, G.M. Lathrop, D. Cohen, unpublished observations) revealed two distinct phenotypes. Of the MODY pedigrees, ~55% were linked to glucokinase. Age of onset in linked families tended to be younger, before puberty in most cases, than that in MODY families not linked. These findings suggest other genes involved in the later onset form of MODY. As the glucokinase gene appears to have been excluded in these families, assessment of other candidate genes, or positional cloning (35) will be likely strategies. Positional cloning uses random markers scattered throughout the genome to identify regions of DNA likely to be involved in families. Later, one specific gene (perhaps previously unknown) can be identified within this region of DNA.

2. What is MODY, and what is its relationship to NIDDM? MODY appears to be a well-characterized form of NIDDM with early onset, and an autosomal dominant mode of inheritance (36). Yet age of onset is really a poorly defined and qualitative criterion for diagnosis, as the early stages of the disease often are asymptomatic and only disclosed when younger family members are tested. Furthermore, some cases of early-onset NIDDM may be phenocopies of MODY, with two diabetic parents contributing multiple diabetes susceptibility genes to their offspring (37). Perhaps the best definition of MODY, and the delineation of its relationship to NIDDM, comes from the recent studies of the French families (29). In a campaign to identify multigenerational pedigrees with NIDDM, >200 were examined. Of those families studied, 32 fit the criteria of MODY, defined by a 2-h post–oral glucose load plasma glucose concentration >7.8 mM (140 mg/dl) or a fasting plasma glucose >6.1 mM (110 mg/dl) on two separate measurements, with at least two individuals diagnosed <25 yr of age. Using these moderately stringent criteria for diagnosis, MODY may comprise as much as 10–15% of familial NIDDM, but less of general or later-onset NIDDM.

3. What role do glucokinase defects play in NIDDM in
general? The results of the French studies indicated that perhaps 5–6% of familial NIDDM in whites could be attributable to glucokinase defects. Our preliminary data in whites indicate no association in population studies, and lack of linkage in family analysis (S. Elbein, K.C.C., Y.T., and M.A.P., unpublished observations). These results suggest that, like other candidate genes such as insulin (38,39), the insulin receptor (40), and the muscle/adipose tissue glucose transporter (GLUT4) (41,42), only rare mutations (<1%) will be found in the general white NIDDM population. The results in racial groups other than whites are less certain. We have found highly significant associations in blacks with glucokinase, as noted above (26,27). The results of molecular scanning with methods that detect single-base changes in DNA are in progress in blacks and other ethnic groups.

**THE IMPORTANT MESSAGE**

Finally, although it is still too early to predict anything with much confidence, the results of studies of affected families so far suggest that glucokinase defects result in mild, predominantly early-onset NIDDM, characterized by impaired β-cell function. Clearly, the clinical picture will be complicated by the presence of additional diabetes-predisposing genes in some families, and by environmental factors, such as obesity and age. In this respect, we have noted in American blacks the risk for NIDDM in those individuals with the high risk Z + 4 allele progresses with age (26, Fig. 4).

Success in this field was attributable to choosing the right candidate gene in the right population of people with diabetes. We can anticipate many more positive findings in the near future, but our chances of success will be enhanced by studying well-characterized families. In this regard, the American Diabetes Association is endeavoring to facilitate the initiation of Family Acquisition Centers for the purpose of establishing permanent cell lines on 200 families with NIDDM. Efforts will be made to characterize all family members with respect to islet β-cell function and insulin action. The DNA from these family members will be an invaluable resource to all interested researchers.

**ACKNOWLEDGMENTS**

This work was supported in part by National Institutes of Health Grants DK-16746 (M.A.P.) and DK-07120 (K.C.C.), RR-00036 from the Diabetic Research and Training Center at Washington University and Grant DK-20575 from the U.S. Public Health Service. Y.T. was the recipient of a mentor-Based Fellowship Award from the American Diabetes Association.

The authors wish to thank Robert Turner, and Philippe Froguel for providing data before publication; and Philip Cryer, Mike Mueckler, and Steven Elbein for commenting on the manuscript. The laboratory assistance of Cris Welling and manuscript preparation by Ms. Jeannie Wokurka also are gratefully acknowledged.

**ADDENDUM**

Sixteen different glucokinase structural gene mutations in 18 of 32 French white MODY pedigrees recently have been described (Ph. Froguel, unpublished observations). It is now clearer that the phenotypes of these MODY pedigrees, i.e., early-onset mild glucose intolerance, as described above, differ from that of general late-onset NIDDM. Some older MODY patients with glucokinase defects remain mildly glucose intolerant, while other members of these pedigrees develop more severe glucose intolerance characteristic of NIDDM. The penetrance for glucokinase mutations resulting in the early-onset mild glucose intolerant phenotype is very high, yet we do not know whether glucokinase mutations alone are sufficient for development of full-blown NIDDM. It appears likely that other gene defects and/or environmental factors are required.

The contribution of glucokinase structural mutations to late-onset NIDDM in whites and American blacks has recently been further clarified by SSCP analysis. No structural mutations were found (K.C.C., Y.T., M.A.P., S. Elbein, unpublished observations). In white NIDDM patients, the combined negative results of population studies, linkage in families, and SSCP analysis suggest that glucokinase structural gene mutations are rare. In the American blacks, where a positive association between glucokinase alleles and NIDDM was found, failure to find mutations in the structural region does not rule out the possibility of mutations in regulatory regions of the gene.

**REFERENCES**

6. Magnuson MA, Shelton KD: An alternate promoter in the glucoki-
GLUCOKINASE AND NIDDM