Protein disulfide isomerase (Pdi) is reported to be an insulin-regulated gene whose expression level is increased in the livers of rats with streptozotocin-induced diabetes. We found that Pdi mRNA is ~20-fold more abundant in the diabetes-susceptible BTBR mouse strain relative to the diabetes-resistant C56BL/6 (B6) strain. A genetic analysis was carried out to determine whether there is a causal relationship between elevated Pdi expression and diabetes phenotype in BTBR-ob/ob mice. We mapped Pdi mRNA abundance as a quantitative trait in 108 (B6 × BTBR)F₂-ob/ob mice segregating for diabetes. We detected a single linkage at the telomeric end of chromosome 11, where the Pdi gene itself resides (logarithm of odds score >30.0). No linkage was detected for the Pdi mRNA trait in the regions where we have previously identified quantitative trait loci for diabetes traits. Sequencing of the Pdi promoter and cdNA revealed several single nucleotide polymorphisms between these two mouse strains. We conclude that in our experimental model, elevated Pdi expression is cis-regulated and is not linked to diabetes susceptibility. Genetic analysis is a powerful tool for distinguishing covariation from causation in expression array studies of disease traits. Diabetes 53:240-244, 2004

Gene array technology has made it possible to simultaneously interrogate the expression of virtually all expressed genes in a cell. This has prompted numerous studies of gene expression in disease. The chief rationale is that altered patterns of gene expression might shed light on links between gene expression and the disease phenotype. However, a major difficulty in the interpretation of gene expression data is distinguishing covariation from causation. Genes whose expression levels change in a disease state may or may not be related to the etiology or pathology of the underlying disease. This is particularly evident in diabetes research. If tissue samples are collected from diabetic animals, it is difficult to know whether differences in gene expression are a cause or consequence of diabetes. If the samples are collected before the onset of diabetes, it is still difficult to distinguish generic strain differences from strain differences causally related to diabetes. Similarly, when carrying out analyses of genes that covary in comparison groups, it is difficult to infer causation from covariation.

Genetic analysis establishes a one-way chain of causation, from genomic variation to differential gene expression. Gene expression as measured through mRNA abundance is affected by both genetic variations and environmental factors; therefore, it can be a heritable trait (1). Genetic contribution to the expression variation is a reflection of DNA sequence variations in the expression control elements. In eukaryotes, the abundance of an mRNA transcript is controlled either locally by cis-regulatory elements, such as sequence polymorphisms in the promoters and in the 3′-untranslated regions (UTRs), or remotely by trans-acting elements, such as transcription factors and RNA-binding proteins, or simultaneously by both. In a segregating population, a trait controlled by a cis-acting gene will be mapped to a locus where the gene itself resides, whereas a trait controlled by a trans-acting gene may be linked to one or more loci. Recent studies in model organisms have shown that one can map mRNA abundance traits as quantitative trait loci (QTL) (2,3).

Our laboratory has been studying gene expression in a diabetes-susceptible congenic mouse strain, BTBR-ob/ob. The strain was derived by introgression of the leptin<sup>ob</sup> allele from the donor C57BL/6 (B6)-ob/ob mutant strain. In contrast to B6-ob/ob mice, BTBR-ob/ob mice are severely diabetic (4,5). In a survey for genes that are differentially expressed in diabetic BTBR-ob/ob mice, we observed an ~20-fold increase in the mRNA abundance of the protein disulfide isomerase (Pdi) gene. PDI (the mouse gene is also called P4hb, Thbp, and Erp59; locus ID: 18453) is an abundant luminal endoplasmic reticulum protein with diverse functions in the endoplasmic reticulum (6). PDI is also a component of the enzymes prolyl 4-hydroxylase and microsomal triglyceride transfer protein. Pdi mRNA is

From the 1Department of Biochemistry, University of Wisconsin, Madison, Wisconsin; and the 2Departments of Statistics and Horticulture, University of Wisconsin-Madison, Madison, Wisconsin.

Address correspondence and reprint requests to Alan D. Attie, Department of Biochemistry, University of Wisconsin-Madison, 433 Babcock Dr., Madison, WI 53706. E-mail: attie@biochem.wisc.edu.

Received for publication 25 July 2003 and accepted in revised form 24 September 2003.

PDL, protein disulfide isomerase; qRT-PCR, quantitative RT-PCR; QTL, quantitative trait loci; SNP, single nucleotide polymorphism; UTR, untranslated region.

© 2004 by the American Diabetes Association.
expression ranges from 6- to 40-fold. As in islets, the strain difference in \textit{Pdi} expression was unrelated to body weight; it occurred in both lean and obese mice.

In contrast to the observations in the livers of diabetic rats (9), insulin does not seem to be responsible for the differential \textit{Pdi} expression in our models. The B6-\textit{ob/ob} mice are known to be hyperinsulinemic, with fasting plasma insulin levels >20-fold higher than in the lean mice (4,5). However, in adipose tissue as well as liver and skeletal muscle (all insulin-responsive tissues), \textit{Pdi} mRNA levels are comparable between lean and obese mice. Also, glucose levels do not correlate with \textit{Pdi} expression. The BTBR-\textit{ob/ob} mice have a fasting glucose level >400 mg/dl, but the \textit{Pdi} mRNA levels in the diabetic BTBR-\textit{ob/ob} mice are comparable with those in the lean BTBR mice in all the tissues surveyed.

\textit{PDI} can be induced by stress-inducing agents, such as heat shock, tunicamycin, cycloheximide, dithiothreitol, etc., but PDI levels seldom change more than three- or fourfold (11). Thus, changes in physiological conditions between these two mouse strains are unlikely to explain the ~20-fold constitutive elevation of \textit{Pdi} mRNA in the BTBR strain. The data in Tables 1 and 2 strongly suggest that the elevated \textit{Pdi} mRNA in the BTBR mice is due to a genetic difference between these two mouse strains. However, the data cannot address whether the strain difference responsible for the \textit{Pdi} expression is causally related to the diabetes susceptibility in the BTBR-\textit{ob/ob} mice. Studying the segregation of \textit{Pdi} mRNA abundance in a population segregating diabetes susceptibility alleles is a powerful tool to distinguish causation from an unrelated strain difference.

\textbf{Linkage mapping.} To identify the expression control elements, we genetically mapped their loci through linkage mapping. We previously mapped gene expression phenotypes in a population of (B6 × BTBR) F\textsubscript{2}-\textit{ob/ob} mice (12). We used the same system to map the \textit{Pdi} mRNA determinants in the liver. The mapping panel consists of 108 F\textsubscript{2}-\textit{ob/ob} mice. The mice were genotyped for 191 microsatellite markers covering the 19 mouse autosomes with an average spacing of 20 cm.

The genome-wide scan revealed a single linkage peak on chromosome 11 (Fig. 1) with a maximum logarithm of odds (LOD) score >30.0 (Fig. 1A). This locus explains 88% of the phenotypic variance as estimated by MAPMAKER/QTL. The strikingly high LOD score and the lack of linkage elsewhere suggest a single-gene control of the liver \textit{Pdi} mRNA phenotype. The linkage peak is located near the telomeric end of mouse chromosome 11, where the \textit{Pdi}
gene itself resides (Fig. 1B). The linkage was confirmed with multiple interval mapping (data not shown). Figure 1C shows the values of the F2 mice according to their genotypes at the closest marker, D11Mit48, which is ~3 cM upstream of the Pdi locus. Mice with homozygous B6/B6 or homozygous BTBR/BTBR alleles have Pdi mRNA levels similar to the respective parental strains (P > 0.05). Mice with heterozygous genotypes are closer to the BTBR strain than to the B6 strain, implying a dominant effect of the BTBR allele. The degree of dominance (dominance/additivity) is ~60%.

**Polymorphisms in promoter and cDNA sequences between B6 and BTBR mice.** As a cis-regulated gene, the higher Pdi mRNA abundance in BTBR mice could be due to either a higher rate of transcription or higher mRNA stability. In an attempt to determine the molecular basis of this phenomenon, we searched the promoter regions of both strains for sequence variations. No difference was found in the region within 1.0 kb of the transcription site. There was one single nucleotide polymorphism (SNP) in the promoter region between B6 and BTBR at nucleotide 1341 (Mouse Genome Browser, February 2003 Assembly, coordinate at chromosome 11:121454350), a T→G transition (B6→BTBR). We used the TFSEARCH (13) program, which uses the TRANSFAC databases (14), to search for transcription factor binding sites. The results show that this SNP may affect one of the several binding sites for the GATA transcription factor (15). Pdi gene transcription is largely controlled by elements in the promoter within 600 nucleotides of the transcription site (8). Deletion of all the six CCAAT boxes in the Pdi promoter still retains 10% transcriptional activity (7). Since the mutation in the BTBR strain only affects one of the several potential GATA sites, it is unlikely this single point mutation would be responsible for an ~20-fold constitutional elevation of Pdi gene transcription.

Sequence changes in the 3’-UTR can affect RNA stability. We sequenced the full-length Pdi cDNA sequences in B6 and BTBR strains. There are seven SNPs, four of which are in the coding region (ORF) and the other three in 3’-UTR. None of the SNPs in the ORF would change the encoded amino acids. To assess the possible impacts of these SNPs on mRNA stability, we checked whether these changes would be predicted to affect RNA secondary structure. We used the Zuker minimum free energy algorithm (16), implemented in the software mFOLD (version 3.1), to predict the mRNA secondary structures of the B6 and BTBR transcripts. Among the seven SNPs, six are predicted to have no or a minor effect on RNA secondary structure. A C→U substitution at position 1509 in the
BTBR strain is predicted to form a new stem, which is absent in the B6 strain (Table 3). The predicted secondary structures of *Pdi* mRNA for B6 and BTBR strain look very similar (data not shown). It is not known whether this mutation would cause an ~20-fold reduction of mRNA degradation in the BTBR mice. It is puzzling that the BTBR phenotype is dominant over the B6 phenotype, even though it is a cis-acting difference.

In conclusion, we have used genetic analysis to distinguish covariation from causation in a gene expression study. Previously, we mapped QTLs of the insulin trait to chromosomes 2, 16, and 19 and the glucose trait to chromosomes 16 and 19 (4). *Pdi* mRNA abundance is independent of insulin or glucose traits in the same F2 population. Here, we show that *Pdi* mRNA maps to chromosome 11. Therefore, we conclude that the *Pdi* mRNA abundance trait of the BTBR strain is not associated with insulin or glucose and is unrelated to the strain’s diabetes susceptibility.

### RESEARCH DESIGN AND METHODS

**Quantitation of *Pdi* mRNA.** The mRNA abundance in B6 and BTBR mice was estimated using the quantitative real time RT-PCR (qRT-PCR) assay. The primers were designed against the mouse full-length *Pdi* mRNA sequence (J051855) in the GeneBank. The primer sequences are forward 5'-ttcctctgactttgctggtgg. PCR sequencing was performed as described previously (12).

**Linkage mapping.** The 108 F2-ob/ob mice were a subset of the F2 population derived from the B6 and BTBR strains that we previously used to study QTLs associated with obesity and diabetes (4). There were 101 microsatellite markers spanning the 19 mouse autosomes that were genotyped and assembled into a framework map using MAPMAKER/EXP (17). Interval mapping of liver *Pdi* mRNA traits in MAPMAKER/QTL was used to reveal the statistical significance of the linkages. Multiple interval mapping (18) implemented in WinQTLCart v2.0 was used to confirm the results from the interval mapping.

**Sequencing.**

**Promoter.** A genomic DNA segment of 2,000 bp, indexed at chromosome 11:121453010–121455009 in the February 2003 Assembly, was retrieved with the Mouse Genome Browser (19) and used as a reference sequence for the *Pdi* promoter region. The 5’ transcription start site of the mouse *Pdi* gene is unknown. The published human *PDI* gene promoter (7), M22803, may in fact contain some mRNA sequences. BLAST search and pairwise sequence alignment showed that the first 61 nucleotides of the mouse *Pdi* mRNA (J051855) are mapped into the human “promoter” sequence, including the proposed TATA box, suggesting some of the promoter sequence is actually exon 1 of the human *PDI* gene. Among the available *Pdi* mRNA and EST sequences, J051855 has the most extended 5’ sequence, so we treated the first base of the J051855 sequence as the transcription start site. The *Pdi* mRNA was transcribed from the reverse genomic DNA strain. Our template sequence starts two nucleotides before the published sequence J051855, whose first nucleotide aligned to chromosome 11:121453008 in the genome assembly. The *Pdi* mRNA sequence XM_129743.3 has more 5’ sequence than J051855, in which the first 397 nucleotides match the 5’ end of our promoter template sequence. However, this 397-nucleotide “mRNA” sequence was based on automatic computer prediction. The most 5’ sequence was not supported by any experimental mRNA or EST data. We therefore treat it as nontranscriptional sequence. We designed four pairs of PCR primers to amplify overlapping fragments in the *Pdi* promoter region in both the B6 and the BTBR genomic DNA. PCR products were purified using Qiagen columns before they were subject to sequencing. Sequencing was performed using Big-Dye reagents from Applied Biosystems. Sequences are assembled and analyzed using Sequencher 4.1.4 (Gene Codes).

**cDNA.** We designed four pairs of PCR primers based on the *Pdi* mRNA sequence J051855 to amplify overlapping fragments from cDNA made from total RNA isolated from both B6 and BTBR livers. The fragments cover ~2.5 kb *Pdi* transcript of 5’-UTR, coding region, and 3’-UTR. PCR sequencing was performed and analyzed as with promoter sequencing.

**RNA secondary structure analysis.** The RNA secondary structure prediction and the affects of mutations were performed using the Zuker minimum free energy algorithm implemented in mFOLD version 3.1 (16).

### ACKNOWLEDGMENTS

This work was supported by the National Institutes of Health Grant DK-58037 and the American Diabetes Association Innovation Grant 7-03-1G-01.

We thank Drs. Ann Palmenberg, Jean-Yves Sgro, and Mark Craven for their help on RNA secondary structure prediction. We also thank Dr. Susanne Clee for her assistance in DNA sequencing analysis.

### REFERENCES


## Table 3

<table>
<thead>
<tr>
<th>Position (B6-BTBR)</th>
<th>Region</th>
<th>Base</th>
<th>Paired with</th>
<th>Structure</th>
<th>Base</th>
<th>Paired with</th>
<th>Structure</th>
<th>Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>831(U-C) ORF</td>
<td>ORF</td>
<td>831-U</td>
<td>952-A</td>
<td>start of a 9-bp stem</td>
<td>831-C</td>
<td>none</td>
<td>next to a 8-bp stem</td>
<td>Decrease stem size by 1</td>
</tr>
<tr>
<td>1077(U-C) ORF</td>
<td>ORF</td>
<td>1077-U</td>
<td>none</td>
<td>loop</td>
<td>1077-C</td>
<td>none</td>
<td>loop</td>
<td>No change</td>
</tr>
<tr>
<td>1209(A-G) ORF</td>
<td>ORF</td>
<td>1209-A</td>
<td>2487-U</td>
<td>start of a 15-bp stem</td>
<td>1209-G</td>
<td>2487-U</td>
<td>start of a 15-bp stem</td>
<td>No change</td>
</tr>
<tr>
<td>1509(C-U) ORF</td>
<td>ORF</td>
<td>1509-C</td>
<td>none</td>
<td>loop</td>
<td>1509-U</td>
<td>2170-A</td>
<td>10-bp stem</td>
<td>Structure change</td>
</tr>
<tr>
<td>2049(C-U) 3’-UTR</td>
<td>2049-C</td>
<td>none</td>
<td>loop</td>
<td>2049-U</td>
<td>none</td>
<td>loop</td>
<td>2049-U</td>
<td>No change</td>
</tr>
<tr>
<td>2098(C-U) 3’-UTR</td>
<td>2098-C</td>
<td>none</td>
<td>loop</td>
<td>2098-U</td>
<td>2127-G</td>
<td>5-bp stem</td>
<td>2098-G</td>
<td>No change</td>
</tr>
<tr>
<td>2240(A-G) 3’-UTR</td>
<td>2240-A</td>
<td>none</td>
<td>loop</td>
<td>2203-U</td>
<td>2203-G</td>
<td>14-bp stem</td>
<td>2204-G</td>
<td>No change</td>
</tr>
</tbody>
</table>

The table shows the most probable structure predicted by mFOLD. The positions in column 1 are indexed with coordinates of the *Pdi* mRNA sequence in the genebank (J051855).


