Identification of a Structurally Distinct CD101 Molecule Encoded in the 950-kb Idd10 Region of NOD Mice

Carlos Penha-Gonçalves,1* Carolyn Moule,1* Luc J. Smink,1 Joanna Howson,1 Simon Gregory,2 Jane Rogers,2 Paul A. Lyons,1 Joseph J. Suttie,1 Christopher J. Lord,1 Laurence B. Peterson,3 John A. Todd,1 and Linda S. Wicker1

Genes affecting autoimmune type 1 diabetes susceptibility in the nonobese diabetic (NOD) mouse (Idd loci) have been mapped using a congenic strain breeding strategy. In the present study, we used a combination of BAC clone contig construction, polymorphism analysis of DNA from congenic strains, and sequence mining of the human orthologous region to generate an integrated map of the Idd10 region on mouse chromosome 3. We found seven genes and one pseudogene in the 950-kb Idd10 region. Although all seven genes in the interval are Idd10 candidates, we suggest the gene encoding the EWI immunoglobulin subfamily member EWI-101 (Cd101) as the most likely Idd10 candidate because of the previously reported immune-associated properties of the human CD101 molecule. Additional support for the candidacy of Cd101 is the presence of 17 exonic single-nucleotide polymorphisms that differ between the NOD and B6 sequences, 10 causing amino acid substitutions in the predicted CD101 protein. Four of these 10 substitutions are nonconservative, 2 of which could potentially alter N-linked glycosylation. Considering our results together with those previous reports that antibodies recognizing human CD101 modulate human T-cell and dendritic cell function, there is now justification to test whether the alteration of CD101 function affects autoimmune islet destruction. Diabetes 52:1551–1556, 2003

The NOD mouse is increasingly recognized as an accurate and informative model of type 1 diabetes in humans. The diabetes phenotype in the NOD mouse, as found in the human disease, is under the control of numerous genetic and environmental factors. At least 17 insulin-dependent diabetes susceptibility (Idd) loci have been mapped in the NOD mouse through the analysis of genetic crosses involving the diabetes-resistant C57BL/6 (B6) or C57BL/10 strains (1–3). Except for the major histocompatibility complex (MHC) class II genes encoding the IA and IE molecules in Idd1 (2) and the β2 microglobulin locus in Idd13 on chromosome 2 (4); no other Idd genes have been identified with even a moderate level of certainty.

Idd10 is located on mouse chromosome 3 and was genetically dissected from neighboring Idd loci using a panel of congenic strains that carry B6 donor DNA (5–8). The diabetes protective effect of the B6 allele at the Idd10 locus in the chromosome interval as currently defined is only apparent when combined with resistance alleles at other loci. Thus, the cumulative frequency of diabetes in congenic mice bearing resistant alleles at the Idd3 and Idd18 loci (18% type 1 diabetes at 7 months) is decreased to 8% in combination with the Idd10 resistance allele (8).

Previous mapping studies using NOD.B6 congenic strains defined the Idd10 locus in the NOD genome within a 1.3-cM interval defined by the microsatellite markers D3Mit10 and D3Nds35 (8). Here we report on the construction and sequencing of an Idd10 B6 bacterial artificial chromosome (BAC) clone contig and the ascertainment of its gene content. Consequently, the Idd10 interval was more precisely defined by fine mapping the recombination break points of key congenic strains using polymorphic microsatellite markers obtained from the sequence of the region. Of the eight genes found in the Idd10 region, the CD101 gene was selected as a candidate for Idd10 based on its function in humans and its polymorphism in the NOD strain.
Genotyping. Microsatellite genotyping was performed as previously described (9).

Idd10 BAC contig construction. To seed a BAC clone contig across the Idd10 interval RPCI-23 library (10) (www.chori.org/bacpac), filters were hybridized with radioactively labeled probes derived from sequence-tagged sites (STSs) known to map to the region. Hybridization-positive clones were confirmed by PCR. Hybridization-positive BAC clones from the RPCI-23 library were identified within assembled fingerprint contigs (FPCs) and extracted into an Idd10-specific FPC database (11). The BCGSC (British Columbia Genome Sequence Centre) whole genome fingerprint database (http://www.bcgsc.bc.ca/projects/mouse_mapping/) was used as the source of initial fingerprint and contig data. Incorporation of marker data and further analyses of the fingerprint data resulted in the construction of a contiguous BAC map across the Idd10 critical interval. Where gaps existed in the contig, end sequences of STS-positive clones were obtained from the BAC end database (www.tigr.org/tdb) and used to generate new probes to rescreen the library as above. This procedure was performed iteratively until all gaps were closed.

Genomic sequencing. To obtain the genomic sequence of the Idd10 region, a minimal tile path of B6 BAC clones was shotgun-sequenced at the Wellcome Trust Sanger Institute.

Sequencing of mouse CD101 cDNA. RNA was extracted from unstimulated NOD spleens using RNAeasy columns (Qiagen). Poly(A) RNA isolated using Genelute-mRNA miniprep kit (Sigma) was reverse transcribed (Superscript II; Invitrogen). The putative exon structure of Cd101 was predicted using est2genome (Emboss GUI, http://bioinfo.pbi.nrc.ca:8090/EMBOSS/index.html) with the human CD101 mRNA sequence NM_004258 and the B6 BAC clone RPCI-23–357D5 as input. Primers for intron-spanning PCR products were designed and the resulting PCR products sequenced (ABI). Sequence trace data were analyzed using the Staden package. Est2genome with the NOD clone DN-358O3 and the obtained cDNA sequence was used to identify the beginning and ends of exons. The coordinates of the exons were extracted using extractseq and translated using transeq (Emboss).

Gene identification. The repeat masked B6 BAC contig sequence was subjected to blast analysis against vertebrate mRNA, dbEST, and exoBlast databases. Expressed sequence tag hits were used to identify genes within the Idd10 region. Mouse ensembl and the orthologous human region in ensembl (www.ensembl.org/) confirmed the gene content of Idd10.

RESULTS
Physical delimitation of the Idd10 genetic interval. Physical mapping of Idd10 was undertaken by constructing a B6 BAC clone contig covering the congenic interval previously defined as the region located between D3Mit10 and D3Nds35 (8). The completed minimal tile path of Idd10 spanning and including D3Mit10 and D3Nds35 contained 19 BAC clones representing a size of ~3–4 Mb (data not shown). Shotgun sequence provided the opportunity to harvest potentially polymorphic microsatellites useful in refining the proximal (R323) and distal (R93) boundaries of the Idd10 interval (Fig. 1). This iterative process narrowed the congenic interval to 950 kb and localized the Idd10 locus distal of RP23–70A7SSR18 and proximal to D3Mit342. The recombination event defining the proximal boundary of Idd10 occurs in the R323 strain and is in the 9 kb of DNA between RP23–70A7SSR18 and RP23–434L7. The distal boundary as defined by the
TABLE 1
The gene content of Idd10 and the orthologous region on human 1p12

<table>
<thead>
<tr>
<th>Human</th>
<th>Mouse</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q9NXK0</td>
<td>AK015259</td>
<td>Ensembl predicted gene. The predicted gene is comprised of two exons. There are no domains listed, and the protein is a member of an unknown protein family.</td>
</tr>
<tr>
<td>MAN1A2</td>
<td>Man1b</td>
<td>Mannosyl-oligosaccharide 1,2-alpha-mannosidase B. Ensembl has predicted 13 exons. The protein contains eight glycoside hydrolase domains and one coiled coil. The protein is involved in N-glycan maturation.</td>
</tr>
<tr>
<td>Q9H6B2</td>
<td>FLJ22418</td>
<td>Human ensembl predicted gene. There are seven predicted exons in human and the gene is not annotated in mouse ensembl, although FLJ22418 shows sequence similarity to the Idd10 BAC contig. The protein contains one Ig-MHC domain, two transmembrane domains, and a signal peptide.</td>
</tr>
<tr>
<td>Q9HSW5</td>
<td>FLJ13181</td>
<td>Human ensembl predicted gene. Weakly similar to transcription intermediary factor 1β. In human, six exons have been predicted and in mouse there are five exons predicted. The proteins contains a B-Box zinc finger, a Filamin/ABP280 repeat, an ATP/GTP-binding site motif (P-loop), and a RING finger.</td>
</tr>
<tr>
<td>TTF2</td>
<td>Tgf2</td>
<td>RNA polymerase II termination factor. Twenty-three exons have been predicted in human and only 18 exons have been predicted in mouse. The protein contains a SNF2 related domain, a helicase COOH-terminal domain, and an ATP-dependent helicase DEAH-box.</td>
</tr>
<tr>
<td>CD101</td>
<td>Cd101</td>
<td>Leukocyte surface protein. There are nine predicted exons in human and mouse. The protein contains seven Ig-MHC domains, a transmembrane domain, and a signal peptide.</td>
</tr>
</tbody>
</table>

The human gene content is taken from ensembl on 28 June 2002 (human build 29). The mouse gene list is as found in the clones sequenced in this study. All genes documented in the human analysis are supported by EST evidence. All genes in the mouse analysis are supported by EST evidence.

R93 strain is in the 82 kb between RP23–434L7SSR1 and D3Mit342. Overall, the narrowing of the Idd10 interval based on the new polymorphic markers decreased the number of BAC clones in the minimal tile path from 19 to 7.

**Idd10 gene content.** We constructed a gene map of the Idd10 region aligning the sequence data from the seven B6 BACs in the tile path with 5.9 Mb of the human genomic sequence orthologous to Idd10 (Fig. 1). This map provides strong evidence that the human region orthologous to Idd10 lies in an ~900-kb region on human chromosome 1p12 (Fig. 1). Comparative analysis of sequence from the B6 tile path with the human sequence revealed eight orthologous genes within the Idd10 interval (summarized in Table 1). Six of the seven genes with intact open reading frames within the Idd10 interval (Gapd is a pseudogene) have not been reported to have particular functions in the immune system or in islet β-cells; these are Man1b, three genes of unknown function (AK015259, FLJ22418 and FLJ13181), Tgf2, and Fprp (an EWI immunoglobulin subfamily member) (12,13). In contrast, Cd101, also an EWI subfamily member, is present on immune cells and has been partially characterized in the human immune system. CD101 is highly expressed on monocytes, granulocytes, dendritic cells (DCs), and activated T-cells (14). In contrast to the lack of CD101 on most resting peripheral blood lymphocytes, CD101 is expressed on the surface of nearly all CD8+, CD3+ intestinal intraepithelial lymphocytes (IELs) (15). Interestingly, very few of these cells express CD28. Using IELs as responders, anti-CD101 mAbs were shown to induce proliferation in conjunction with suboptimal concentrations of anti-CD3 mAb. Therefore, CD101 could be a costimulatory receptor functioning in this subpopulation of CD28− mucosal T-cells (15). Additional studies also suggest a direct function of CD101 on DCs (16,17). Cutaneous DCs incubated with anti-CD101 mAb produced IL10 that led to the inhibition of T-cell proliferation (17). A direct effect of anti-CD101 mAb on T-cell signaling has also been reported (18,19). Thus, from the gene content analysis, CD101 is our favored candidate for further study to test its potential in mediating the effect of Idd10 in autoimmune diabetes. However, all seven genes within the interval remain candidates until and unless it is proven that Cd101 is Idd10 by transgenic or knock-in approaches or other highly specific functional analyses.

**NOD allotype of CD101.** We tested for variants in the
FIG. 2. The cDNA sequence of B6 Cd101 showing the amino acid translation below. SNPs are highlighted; the NOD variants are shown above the B6 nucleotide sequence. The codon on either side of the codon affected by the polymorphism is underlined. Amino acid coding changes are highlighted, the B6 residue first followed by the NOD.

**Cd101 IS A CANDIDATE GENE FOR TYPE 1 DIABETES**
primary sequence of the CD101 protein by comparing the NOD and B6 genomic sequence. Sequencing cDNA derived from CD101 mRNA obtained from NOD spleen cells was used to establish exon/intron boundaries (because mouse CD101 had not been previously characterized) and confirm sequence variants found in the genomic analysis. Seventeen single nucleotide polymorphisms (SNPs) between NOD and B6 are found in the coding region of CD101, 10 of which resulted in amino acid changes (Fig. 2). Of these 10 changes, 4 are nonconservative and 2 of the 4 nonconservative changes could potentially alter N-linked glycosylation of the CD101 molecule by their presence in a N-X-S/T consensus sequence.

**DISCUSSION**

The congenic breeding strategy is a powerful method (2) to identify genes, such as Idd10, controlling a complex disease process exemplified in the present study by the NOD mouse model of type 1 diabetes. To reduce the number of candidate genes that must be considered, the size of the congenic region is minimized, a process that takes many years of selective breeding and phenotypic assessment. We have now reached the limit of resolution of recombination mapping that is practical for the Idd10 interval. Considerable narrowing of the interval was gained from the sequence of the B6 BAC clones, which allowed the experimental identification of new B6/NOD polymorphic microsatellite markers. Once the informative resolving power of homologous recombination was exhausted, we were faced with the possibility of having a small physical interval that was very gene rich (20). However, only seven candidate genes were found in the Idd10 interval (Table 1). As expected (21), nothing is known about the expression or function of half of the genes in the interval (three of the seven).

Our recent unpublished results identifying both the human type 1 diabetes locus IDDM12 as CTLA4 and the orthologous NOD mouse locus Idd5.1 as Cita4 confirm the critical contribution of immune-related loci to the genetic control of autoimmune disease, a role that was first established by the MHC class II genes. While all genes within an interval are potential candidates, we use a triage process in which the known immune-related genes in an Idd interval are those that are first considered as candidates. Because definitive proof of an Idd gene ultimately requires a knock-in of the putative causative SNP or SNPs or a sufficiently controlled transgenic-based experiment, circumstantial evidence must be obtained for the candidacy of one of the genes within the interval and the critical nucleotide variants responsible for the disease phenotype. Hopefully, such evidence would be persuasive enough to justify the time and expense of a definitive "proof experiment."

Reports that the human CD101 molecule has costimulatory function for a subset of CD8 cells that are found primarily in the mucosal immune system of the intestine placed CD101 at the top of our Idd10 candidate gene list. Although CD101 is essentially uncharacterized in mice, these findings in the human immune system triggered the sequencing of the CD101 gene from the NOD strain. The striking level of polymorphism found in the B6 and NOD CD101 proteins indicates it is likely that the CD101 molecule is under selective pressure. These results suggest that a functional polymorphism exists for CD101 in the mouse population and that this phenotypic variation could account for the activity of Idd10. Sequencing of other strains of mice to determine the distribution of the CD101 variants and their haplotypic categorization is in progress. It is also important to determine the level of allelic variation of the other seven genes in the NOD Idd10 interval by extending the NOD BAC clone shotgun sequencing or by targeted PCR product resequencing. The pattern of ancestral haplotype breakpoints in the Idd10 region could also help to further map the causal variants (22,23). In addition, the potential association of the ortholog with human type 1 diabetes is being tested, and a knock-out of the mouse CD101 gene is underway to aid in its characterization and potentially in transgenesis experiments that may provide evidence toward the identification of Idd10.

CD101 is also of interest because of its potential therapeutic approachability. As has been done with other Ig superfamily members expressed at the cell surface (24,25), an engineered, soluble immunoglobulin-fusion version of CD101 designed to prevent productive engagement of CD101 or, alternatively, an agonistic antibody to augment the function of CD101 could have therapeutic effects in autoimmune disorders or in other immune-mediated processes. Pursuit of these experiments would also be an attractive way of building a case that CD101 might be Idd10.

**ACKNOWLEDGMENTS**

The laboratories of L.S.W., P.A.L., and J.A.T. are supported by a joint grant from the Juvenile Diabetes Research Foundation (JDRF) and the Wellcome Trust. The availability of NOD congenic mice through the Taconic Farms Emerging Models Program has been supported by grants from the Merck Genome Research Institute, National Institute of Allergy and Infectious Diseases, and the JDRF.

**REFERENCES**


Cd101 is a candidate gene for type 1 diabetes


