RESEARCH

Mouse Galactokinase: Isolation, Characterization, and Location on Chromosome 11

Yunjun Ai, Nancy A. Jenkins, Neal G. Copeland, Debra J. Gilbert, Derk J. Bergsma, and Dwight Stambolian

Departments of Ophthalmology and Genetics, School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104; Mammalian Genetics Laboratory, ABL-Basic Research Program, National Cancer Institute–Fredrick Cancer Research and Development Center, Frederick, Maryland 21702; Department of Molecular Genetics, Smith Kline Beecham Pharmaceutical, King of Prussia, Pennsylvania 19406

Elevated galactose levels can be caused by several enzyme defects, one of which is galactokinase. Galactokinase deficiency causes congenital cataracts during infancy and presenile cataracts in the adult population. We have isolated the mouse cDNA for galactokinase, which shares extensive amino acid sequence homology, 88% identity, with a recently cloned human galactokinase. It is expressed in all tissues examined. In an interspecific backcross analysis galactokinase maps to the distal region of mouse chromosome 11, a region that is homologous to human chromosome 17q22-25. The availability of the mouse gene provides an opportunity to make a knockout model for galactokinase deficiency.

Genetic defects of galactose metabolism constitute a class of genetic disorders termed galactosemia. One form of galactosemia in humans is caused by an enzyme deficiency of galactokinase (Segal 1989). Galactokinase acts on the sugar substrate galactose to convert it to galactose-1-PO4. Individuals with homozygous galactokinase deficiency become symptomatic in the early infantile period showing galactosemia, galactosuria, increased galactitol levels, cataracts, and in a few cases, mental retardation (Segal et al. 1979). Heterozygotes for galactokinase deficiency are prone to presenile cataracts ~20–50 years of age (Stambolian et al. 1986).

Galactokinase activity has been found in a variety of mammalian tissues, including liver, kidney, brain, lens, placenta, erythrocytes, and leukocytes. The isolation and characterization of the galactokinase gene has been completed in Escherichia coli, Saccharomyces, and Streptomyces lividans (Citron et al. 1984; Debouck et al. 1985; Adams et al. 1988). A potential human galactokinase, denoted GK2, with 29% homology to the galactokinase gene of Saccharomyces, was cloned by complementation in yeast and mapped to chromosome 15 (Lee et al. 1992). However, the role of GK2 in galactose metabolism remains unclear because active enzyme activity has not been demonstrated (beyond yeast complementation experiments) and mutations in GK2 have not been identified in patients with galactokinase deficiency. Recently, the cDNA encoding human galactokinase, termed GALK1, was cloned, functionally characterized, and mapped to chromosome 17q24 (Stambolian et al. 1995). Additionally, two distinct mutations were identified within GALK1 gene-coding sequences of two unrelated families that exhibited galactokinase deficiency and cataracts.

As a first step toward the generation of a murine model for galactokinase deficiency, we have cloned and characterized the murine galactokinase cDNA. The cDNA obtained shows 88% and 22% protein sequence identity to GALK1 and GK2, respectively. The mouse galactokinase gene was mapped to the distal portion of mouse chromosome 11, a region that is homologous to human chromosome 17q.
RESULTS
Isolation and Characterization of Mouse Galactokinase cDNA Clones

Three cDNA clones that hybridized strongly with human galactokinase GALK1 cDNA (Stambolian et al. 1995) were isolated from a cDNA library constructed from mouse (strain 129) liver RNA. These clones were all found to contain partial cDNA inserts that encoded a portion of mouse galactokinase but lacked the amino-terminal domain of the protein. Clone pGlk012 contained the largest cDNA insert of 662 bp in length, which terminated with a poly(A) tract of nucleotides (not shown). The missing 5' end of the galactokinase cDNA was obtained from mouse liver RNA by RT–PCR techniques, using primers corresponding to DNA sequences present within the mouse coding region of clone pGlk012 and the 5'-untranslated region (UTR) of the human GALK1 cDNA (including the ATG initiation codon) (Stambolian et al. 1995). A 900-bp RT–PCR product was found to contain the missing 5' end of the mouse galactokinase cDNA. Independent RT–PCR reactions were performed, and the DNA sequences of these multiple PCR products were determined to confirm the DNA sequence of the mouse galactokinase cDNA and exclude the possibility of PCR-generated mutations occurring within the sequence.

The 900-bp RT–PCR product was combined with the insert of pGlk012 to generate a mouse galactokinase (Glkl) cDNA of 1316 bp in length following removal of the overlapping nucleotide sequence. This cDNA is terminated with a poly(A) tract of nucleotides that is preceded 12 bp upstream by a canonical eukaryotic polyadenylation signal, AATAAA (Fig. 1) (Proudfoot and Brownlee 1976). The cDNA encodes a peptide of 392 amino acids with a calculated molecular mass of 42,271 daltons.

Alignment of the predicted mouse galactokinase protein with human GALK1, GK2, and E. coli galactokinase shows a low to high degree of overall sequence similarity, ranging from ~22% to 88% amino acid sequence identity (Fig. 2). The mouse galactokinase sequence shares the greatest level of sequence conservation, 88% identity, with the human GK17 galactokinase protein. The mouse Glkl protein also shares 42% amino acid sequence identity with E. coli galactokinase. In contrast, the human GK2 and mouse Glkl proteins are only 22% identical. Importantly, three regions containing a galactokinase signature sequence (Debouck et al. 1985) and two ATP-binding motifs (Tsay et al. 1991) are identically conserved in the human and mouse galactokinases.

To examine the size and pattern of expression of mouse galactokinase mRNA in different mouse tissues, Northern blot analysis was performed using the 5' end of the galactokinase cDNA as a hybridization probe. As shown in Figure 3, a single mRNA transcript of ~1.35 kb was detectable in all mRNA samples. This size is comparable to that of the Glk1 cDNA, indicating that the isolated cDNA likely represents the entire mRNA transcript of the gene (with the exception of some 5' UTR sequence).

Expression of Galactokinase Activity

The full-length Glk1 cDNA was subcloned into the mammalian cell expression vector pCDN, forming vector pCDNGlk1, and transformed into COS cells. After 3 days of incubation, cell extracts were prepared and analyzed for galactokinase activity. COS cells transfected with the pCDN vector alone exhibited very low levels of galactokinase activity, amounting to a mean activity of 0.02 U/μg of protein. In contrast, galactokinase activity increased >10-fold to a mean activity of 0.80 U/μg of protein when COS cells were transfected with the pCDNGlk1 vector. This result functionally confirms that Glk1 cDNA encodes mouse galactokinase.

Interspecific Backcross Mapping of Glk1

The murine chromosomal location of Glk1 was determined by interspecific backcross analysis using progeny derived from matings of [(C57BL/6J × Mus spretus) F1 × C57BL/6J] mice. This interspecific backcross mapping panel has been typed for >1800 loci that are well distributed among all mouse autosomes and the X chromosome (Copleland and Jenkins 1991, and unpubl.). C57BL/6J and M. spretus DNAs were digested with several restriction enzymes and analyzed by Southern blot hybridization for informative restriction fragment-length polymorphisms (RFLPs) using a mouse Glk1 cDNA probe (see Methods). A 6.1-kb M. spretus-specific BglII RFLP was used to follow the segregation of the Glk1 locus in backcross DNAs. The mapping results indicated that Glk1 is located in the distal region of mouse chromosome 11 (Fig. 4). Although 146 mice were analyzed for every marker shown in the haplotype...
MOUSE GALACTOKINASE
gene order are centromere-
Pka-13/149-Glk1-6/176-
Thbp, where Pka is protein
kinase C-α and Thbp is prolyl
4-hydroxylase-β polypep-
tide. The recombination fre-
quencies (expressed as ge-
etic distances in cM ± the
s.e.) are centromere–Pka–
8.7 ± 2.3–Glk1–3.4 ± 1.4–
Thbp.

DISCUSSION
We have described the isola-
tion of a mouse cDNA that
encodes a protein that
shares 88% sequence iden-
tity with human galactoki-
rase GALK1. Evidence that
the isolated cDNA encodes
mouse galactokinase fol-
loows: (1) The encoded pro-	ein shares significant se-
quence homology with all
known galactokinases and,
more notably, contains sev-
eral sequence motifs, includ-
ing a galactokinase signature
sequence and a putative
ATP-binding domain that is
highly conserved among all
galactokinases; (2) the cDNA
produced a 10-fold increase
of galactokinase enzyme ac-
tivity when expressed in
COS cells; (3) the cognate
gene of the cDNA was
mapped to a distal location
on chromosome 11, a region
consistent with the location
of the human homolog on
chromosome 17q. Collect-
ively, these results establish
that we have cloned the
mouse galactokinase gene,
which we designate Glk1.

The mouse galactokinase gene had been
mapped previously by enzyme activity on starch
gel electrophoresis to chromosome 11 in a panel
of Chinese hamster/mouse somatic cell hybrids
(Mishkin et al. 1976; Kozak and Ruddle 1977;
McBreen et al. 1977). Subsequently, its location
was positioned on the distal region of mouse chro-

Figure 1 Nucleotide and deduced amino acid sequence of the Glk1 cDNA.
Amino acids (represented by the one-lettered code) are indicated below their
respective codons beginning with the initiation methionine codon (M). Nucleo-	ide positions are numbered at right. The underlined nucleotide sequence
indicates canonical eukaryotic polyadenylation signals. The poly(A) tract of the
cDNA is not shown. Nucleotides preceding the first ATG are from the 5' UTR.
al analysis (Fig. 4), up to 176 mice were typed for
some pairs of markers. Each locus was analyzed in
pairwise combinations for recombination fre-
quencies using the additional data. The ratios of
the total number of mice exhibiting recombinant
chromosomes to the total number of mice ana-
lized for each pair of loci and the most likely
mosome 11 based on (1) its syntenic relationship with \textit{Umph}-2 (Wilson et al. 1987), and (2) the homology between distal chromosome 11 and human chromosome 17q21-q25, where the human galactokinase gene had been mapped by somatic cell hybrid analysis. Our placement of \textit{Gk1} on the distal region of mouse chromosome 11 (Fig. 4) agrees with this previous mapping data. Recently, the human galactokinase gene was mapped by in situ hybridization to chromosome 17q24 (Stambolian et al. 1995), which is consistent with our location of the mouse gene on chromosome 11.

Finally, we have aligned our interspecific linkage map of distal chromosome 11 with a composite linkage map that reports the map location of many uncloned mouse mutations (compiled from the Mouse Genome Database, a computerized data base of mouse linkage information maintained at The Jackson Laboratory, Bar Harbor, ME). None of the mutations close to \textit{Gk1}, that is, \textit{cod}, \textit{tn}, and \textit{js} (Lyon and Searle 1989), have a phenotype that might be expected for a mutation (i.e., cataracts) at this locus. However, it is still possible that one of these mutations is caused by a defect in \textit{Gk1}. Additional studies will be required to rule out this possibility.

The isolation of the mouse galactokinase cDNA now makes it possible to use it as a probe to clone its genomic counterpart. This gene will then be used to create mouse knockout strains, thus providing a mouse model that will be useful for evaluating the role of galactokinase in cataract formation.

### METHODS

cDNA Cloning

A cDNA library was constructed from liver poly(A)\(^{+}\) RNA extracted from a female mouse (129 strain), primed with

![Figure 3](image)
MOUSE GALACTOKINASE

with 0.2x SSPE, 0.1% SDS, at 50°C and exposed to X-ray film at -70°C overnight. Positive clones were plaque purified. Three partial cDNAs were isolated from the initial screening.

RT-PCR

Poly(A)⁺ RNA from strain 129 liver was used. The first-strand cDNA was synthesized using primer 1949 (5'-CCTTGTAAGTCCTCGGCTCATA-3') which was designed to the 3' end of the cloned mouse galactokinase cDNA. To generate the specific S' cDNA product, the PCR reaction was performed using a nested primer, 1950 (5'-CTCCGGATCTCGCTTACCACAT-3'), located 5' to primer 1949, and another primer, DJ4 (5'-CAGAGCTGCAGGCGCGCTCA-3'), designed to the human galactokinase 5' UTR. The thermal cycler (MJ) was programmed for 35 cycles at 94°C for 1 min, 50°C for 1 min, and 72°C for 1 min. The DJ4/1950 primer pair gave a specific band of ~900 bp. The RT-PCR product was directly subcloned into the TA vector (Invitrogen) for sequencing.

Sequence Analysis

Sequencing of both strands was performed from isolated plasmid. Five separate subclones of the RT-PCR product were sequenced to rule out the possibility of any Taq polymerase errors. Sequencing was done on an automated ABI 373A sequencer. Alignments were done using MacVector software (Kodak).

Expression in COS Cells

An EcoRI-KpnI cDNA fragment encompassing the entire coding region of the mouse galactokinase gene was subcloned into the COS cell (Gluzman 1981) expression vector pCDN (Aljar et al. 1994) generating the clone pCDNGLk1. COS cells were grown to 60%-80% confluency and then transfected with 50 µg of DNA per flask of pCDNGLk1 or pCDN control as described previously (Caltabiano et al. 1989). After incubation at 37°C for 3 days, the cells were washed with phosphate-buffered saline, scraped from the plate with a rubber policeman, pelleted by low-speed centrifugation, and frozen on dry ice. Cell lysates were prepared for the galactokinase assay (Stambolian et al. 1985) by dissolving in 10 mM phosphate buffer (pH 7.2).

Northern Blot Analysis

A Northern blot containing 2 µg of poly(A)⁺ RNA per lane from different mouse tissues was purchased from Clontech Laboratories (Palo Alto, California) and probed with a 32P-labeled 474-bp cDNA fragment from the 5' end of the mouse galactokinase cDNA according to the recommendations of the manufacturer. Following hybridization, the filter blot was washed in 0.1x SSPE, 0.1% SDS, at 65°C and exposed to Kodak X-ray film at -70°C with an intensifying screen overnight.

Interspecific Mouse Backcross Mapping

Interspecific backcross progeny were generated by mating
AI ET AL.

(C57BL/6J × M. spretus) F1 females and C57BL/6J males as described (Copeland and Jenkins 1991). A total of 205 backcross mice were used to map the galactokinase locus (Gkl). DNA isolation, restriction enzyme digestion, agarose gel electrophoresis, Southern blot transfer, and hybridization were performed essentially as described (Jenkins et al. 1982). All blots were prepared with Hybond N+ nylon membrane (Amersham). The probe, a 600-bp EcoRI–XhoI mouse cDNA fragment, was labeled with [γ-32P]dCTP by nick translation (Boehringer Mannheim); washing was done to a final stringency of 1.0x SSCP, 0.1% SDS, at 65°C. A 7.5-kb fragment was detected in BgllI-digested C57BL/6J DNA, and a 6.1-kb fragment was detected in BgllI-digested M. spretus DNA. The presence or absence of the 6.1-kb M. spretus-specific BgllI fragment was followed in backcross mice.

A description of the probes and RFLPs for the loci linked to Gkl including Pkca and Thbp have been reported previously (Buchberg et al. 1989; Morishige et al. 1993). Recombination distances were calculated as described by Green (1981), using the computer program SPRETUS. Gene order was determined by minimizing the number of double and multiple recombination events across the chromosome.

ACKNOWLEDGMENTS

We thank Debbie Barnhart for excellent technical assistance. This work was supported by National Institutes of Health grant RO1 EY09404 (D.S.), National Cancer Institute, Department of Health and Human Services, under contract NO1-CO-46000 with ABL, and the Pennsylvania Lions Sight Conservation and Eye Research Foundation. D.S. is a Research to Prevent Blindness James S. Adams Scholar.

The publication costs of this article were defrayed in part by payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 USC section 1734 solely to indicate this fact.

REFERENCES


MOUSE GALACTOKINASE


Received May 30, 1995; accepted in revised form July 6, 1995.
Mouse galactokinase: isolation, characterization, and location on chromosome 11.

Y Ai, N A Jenkins, N G Copeland, et al.

*Genome Res.* 1995 5: 53-59
Access the most recent version at doi:10.1101/gr.5.1.53