Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 10887–10891, November 1992
Genetics

Cloning of a human galactokinase gene (*GK2*) on chromosome 15 by complementation in yeast

(cataracts/eye disease/galactosemia)

ROBERT T. LEE*, CRAIG L. PETERSON†, ANDREW F. CALMAN*, IRA HERSKOWITZ†, AND JAMES J. O'DONNELL*‡

*Department of Ophthalmology, Kimura Laboratory of Clinical Investigation, and †Department of Biochemistry and Biophysics, University of California, San Francisco, School of Medicine, San Francisco, CA 94143-0730

Contributed by Ira Herskowitz, July 2, 1992

ABSTRACT **A human cDNA encoding a galactokinase (EC 2.7.1.6) was isolated by complementation of a galactokinase-deficient (*gal1*⁻) strain of *Saccharomyces cerevisiae*. This cDNA encodes a predicted protein of 458 amino acids with 29% identity to galactokinase of *Saccharomyces carlsbergensis*. Previous studies have mapped a human galactokinase gene (*GK1*) to chromosome 17q23-25, closely linked to thymidine kinase. The galactokinase gene that we have isolated (*GK2*) is located on chromosome 15. The relationship between the disease locus for galactokinase deficiency galactosemia, which is responsible for cataracts in newborns and possibly presenile cataracts in adults, and the two galactokinase loci is unknown.**

Galactokinase deficiency [one of three known forms of the disease galactosemia (1)] is inherited as an autosomal recessive trait with a heterozygote frequency estimated to be 0.2% in the general population (2). The major clinical manifestation of galactokinase deficiency is neonatal cataract formation (1), apparently due to galactitol accumulation (3). Presenile cataracts might also result from galactokinase deficiency (4).

The gene responsible for galactokinase deficiency galactosemia has not been determined, and thus the precise molecular alterations responsible for this disease are uncharacterized. A human galactokinase gene, *GK1*, has been previously mapped to human chromosome 17 by assaying galactokinase activity in cell extracts from mouse–human hybrid cell lines (5–8). It is not known if *GK1* is altered in patients with galactokinase deficiency.

Galactokinase (EC 2.7.1.6) serves the same function in yeast as in bacteria and humans, phosphorylation of galactose in the pathway for use of this sugar as a carbon source (9). We have therefore sought to clone a human galactokinase gene by its ability to substitute for the yeast enzyme. Several mammalian cDNAs have recently been identified by complementation or other functional activity in yeast (see refs. 10–12). We report here the cloning of a cDNA that encodes human galactokinase§ by complementation of a galactokinase-deficient (*gal1*⁻) strain of *Saccharomyces cerevisiae*. We have made the unexpected finding that this gene, *GK2*, does not map to chromosome 17. It is thus not clear whether the locus responsible for galactokinase deficiency disease affects *GK1* or *GK2*.

MATERIALS AND METHODS

Yeast Strains, Media, and Transformation. The recipient yeast strain used in transformations was *S. cerevisiae* YM20 (*MATα ura3-1 ura3-2 trp1-289 his3-532 ade2-1(ochre) gal1-del111 GAL2⁺ CRY1ˢ*) kindly provided by Mark Johnston (Washington University, St. Louis, MO). A human hepatoma cell line (HepG2) cDNA library was constructed in yeast expression vector pAB23BXN by Anthony Brake (Chiron) (12); cDNA inserts are flanked by *Bgl* II sites. This vector contains a yeast origin of replication, *URA3* gene, and glyceraldehyde-3-phosphate dehydrogenase (*GAP*) promoter for expression of the human cDNA in yeast. pBM48 (kindly provided by Mark Johnston) contains the yeast *GAL1* gene (in DNA segment Sc4918; ref. 13) in the *URA3* vector YCp19; YCp50 is a yeast low-copy vector carrying *URA3*.

Screening. A modification of the lithium acetate method was used for yeast transformation (14). Most transformants were selected on standard yeast synthetic medium lacking uracil and containing galactose as the sole carbon source (15).

Subcloning and DNA Sequencing. Plasmid pJJGK was recovered from the Gal⁺ yeast transformant by transforming *Escherichia coli* strain DH5α with total yeast DNA using electroporation, with selection for ampicillin resistance. pJJGK contained a 1.5-kilobase (kb) cDNA insert, which was cleaved into 1.0-kb and 0.5-kb fragments by *Bgl* II digestion. Each fragment was subcloned into M13mp19 (16) and propagated in JM107. DNA sequencing was carried out by the dideoxy chain-termination method (17). Four oligonucleotide primers were designed from the determined sequence to complete the sequencing on both strands. The sequence across the *Bgl* II site used for subcloning was also determined.

DNA and Protein Sequence Analysis. A FASTA program for searching amino acid sequence data bases (18) was carried out using the EUGENE sequence analysis program (Baylor University, Houston). GenBank (release 68.0, June 1991), Protein Identification Resource (PIR)-Protein (release 28.0, March 1991), PIR-Nucleic (release 36.0, March 1990), and European Molecular Biology Laboratory (release 27.0, May 1991) gene and protein sequence data bases were searched for homology to the *GK2* gene sequence (19). A phylogenetic tree was constructed by the progressive alignment method as described in ref. 20.

Polymerase Chain Reaction (PCR) Technique. PCR amplification was carried out with an Ericomp (San Diego) thermal cycler using a modification of the standard protocol provided with the Perkin–Elmer/Cetus GeneAmp DNA kit (21). The cDNA insert in pJJGK was amplified with primers hG5′-25 (5′-GCGGACAAGCTGCCCAGCTTTCTAG-3′), corresponding to positions 1260–1284, and hG3′-26 (5′-GTGGCACAGAAAGTCCTGCCAATTCC-3′), corresponding to positions 1464–1439. After 10 min at 94°C to denature genomic DNA, reactions were carried out for 35 cycles of 1 min at 94°C, 1 min at 60°C, and 2 min at 72°C. Reaction mixtures for Fig. 4 *A* and *B* contained 67 mM Tris (pH 8.8), 6.7 mM $MgCl_2$, 16.6 mM $(NH_4)_2SO_4$, 10 mM 2-mercaptoethanol, 50 pM of each primer, 0.1 μg of uncut genomic DNA or 1 ng of pJJGK,

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

‡To whom reprint requests should be addressed.

§The sequence reported in this paper has been deposited in the GenBank data base (accession no. M84443).

and 2.5 units of Ampli-Taq DNA polymerase (Perkin–Elmer/Cetus). Reaction mixtures for Fig. 4*C* contained 10 mM Tris (pH 8.3), 50 mM KCl, and 1.5 mM $MgCl_2$ and lacked $(NH_4)_2SO_4$ and 2-mercaptoethanol. A panel of human–rodent somatic cell hybrids retaining subsets of human chromosomes was obtained from the National Institute of General Medical Sciences Human Genetic Mutant Cell Repository, Coriell Institute for Medical Research (8, 22–26). Total human genomic DNA was a gift of Margit Burnmeister (University of Michigan, Ann Arbor, MI). DNA samples from human–mouse hybrid cell lines A59-3Aaz10a and MH22-6 were obtained from Huntington Willard (Stanford University, Palo Alto, CA) and Pragna Patel (Baylor University), respectively. The Hs27 (newborn human foreskin) cell line was obtained from the ATCC (CRL 1634). Primers for detecting the human β_2-microglobulin gene (*B2M*) were B2MI (5′-CACCCAGTCTAGTGCATGCCTTCT-3′) and B2M2 (5′-TGAGAAGGAAGTCACGGAGCGAGA-3′) and for detecting the human gastrin gene (*GAS*) were GAS1 (5′-ATGCTAGTCGGTGTAGAGCCATG-3′) and GAS2 (5′-TTGTACCTCATAGGGCTGCGTGA-3′) (27).

RESULTS

Isolation of a Human cDNA Clone by Complementation in Yeast. A HepG2 cDNA expression library was introduced into yeast strain YM20, which contains a deletion of the gene encoding galactokinase (*GAL1*). Transformants were plated on medium containing galactose as carbon source to select for transformants with galactokinase activity. One transformant was identified in ≈150,000 transformants tested.

We first determined whether growth on galactose medium was plasmid dependent by selecting for cells that had lost this plasmid by growth on plates containing 5-fluoroorotic-acid (28). Cells cured of the plasmid were no longer able to grow on galactose medium (data not shown). A plasmid, designated pJJGK, was isolated from the original yeast transformant by transformation into bacteria and then reintroduced into yeast strain YM20. All Ura^+ transformants grew on galactose medium, suggesting that this plasmid encoded a galactokinase.

The cDNA insert of pJJGK was sequenced, which revealed a single large open reading frame with the potential to encode a 458-amino acid polypeptide of 50,386 daltons (Fig. 1). This polypeptide (denoted GK2) shows extensive similarity to galactokinase from yeasts and bacteria: 29% overall identity (matches/458) to the *GAL1* protein of *Saccharomyces carlsbergensis* (29), 31% to *GAL1* of *Kluyveromyces lactis* (30), 22% to *galK* of *E. coli* (31), 21% to *galK* of *Lactobacillus helveticus* (32), and 16% to *GK* of *Streptomyces lividans* (33) (Fig. 2). Alignment of these sequences revealed at least six regions (A–F) with particularly striking amino acid identity. For example, the human protein is identical to the two yeast enzymes in 19 of 24 positions in region A (positions 39–62 of GK2). The human protein is identical to the yeast enzymes in 14 of 14 positions in region F (positions 396–409 of GK2). These regions contain many matches to the *E. coli* galactokinase sequence and to the yeast GAL3 sequence (34) as well. The overall identities between the human enzyme and the other galactokinases make it overwhelmingly likely that the GK2 cDNA that we have cloned complements the yeast galactokinase-deficient mutant by providing galactokinase activity.

There are several places (such as regions D, E, and F) where the human sequence is more similar to yeast than to bacterial galactokinase sequences. Presumptive phylogenetic relationships among these galactokinases are shown as a tree in Fig. 3.

Human Chromosome Mapping. We determined the chromosomal location of the segment coding for the GK2 cDNA by PCR amplification analysis of a *GK2*-specific segment. A set of PCR primers (hG5′-25 and hG3′-26) corresponding to the final 45 amino acid residues of the *GK2* open reading frame and 70 bases of the 3′ untranslated region of its mRNA was used to amplify a 205-base-pair (bp) region of the pJJGK cDNA insert; a segment of similar size was amplified also

```
                     Met Ala Thr Glu Ser Pro Ala Thr Arg Arg Val   11
AGATCTGAATTCGGCGAAAT ATG GCT ACA GAG AGC CCT GCT ACG CGT CGG GTC   53

Gln Val Ala Glu His Pro Arg Leu Leu Lys Leu Lys Glu Met Phe Asn Ser Lys Phe Gly Ser Ile Pro Lys Phe Tyr Val Arg   39
CAG GTG GCA GAA CAT CCT AGG TTA CTG AAG CTA AAG GAG ATG TTT AAC TCC AAG TTT GGA TCT ATT CCC AAG TTT TAT GTT CGA   137

Ala Pro Gly Arg Val Asn Ile Ile Gly Glu His Ile Asp Tyr Cys Gly Tyr Ser Val Leu Pro Met Ala Val Glu Gln Asp Val   67
GCA CCA GGA AGA GTC AAC ATA ATA GGA GAG CAT ATA GAT TAT TGT GGA TAT TCT GTT CTT CCT ATG GCT GTA GAA CAA GAT GTG   221

Leu Ile Ala Val Glu Pro Val Lys Thr Tyr Ala Leu Gln Leu Ala Asn Thr Asn Pro Leu Tyr Pro Asp Phe Ser Thr Ser Ala   95
CTA ATA GCT GTA GAA CCT GTG AAA ACG TAC GCT CTC CAA CTG GCC AAT ACA AAT CCC TTG TAT CCG GAC TTC AGT ACT AGT GCT   305

Asn Asn Ile Gln Ile Asp Lys Thr Lys Pro Leu Trp His Asn Tyr Phe Leu Cys Gly Leu Lys Gly Ile Gln Glu His Phe Gly   123
AAT AAC ATC CAG ATT GAT AAA ACC AAG CCT TTG TGG CAC AAC TAT TTC TTA TGT GGA CTT AAA GGA ATT CAG GAA CAC TTT GGT   389

Leu Ser Asn Leu Thr Gly Met Asn Cys Leu Val Asp Gly Asn Ile Pro Pro Ser Ser Gly Leu Ser Ser Ser Ser Ala Leu Val   151
CTT AGT AAC CTG ACT GGA ATG AAC TGC CTG GTA GAT GGA AAT ATC CCA CCA AGT TCT GGC CTC TCC AGC TCC AGT GCT TTG GTC   473

Cys Cys Ala Gly Leu Val Thr Leu Thr Val Leu Gly Arg Asn Leu Ser Lys Val Glu Leu Ala Glu Ile Cys Ala Lys Ser Glu   179
TGT TGT GCT GGC TTG GTG ACG CTC ACA GTG CTG GGA AGG AAT CTA TCC AAG GTG GAA CTT GCA GAA ATC TGT GCC AAG AGT GAG   557

Arg Tyr Ile Gly Thr Glu Gly Gly Gly Met Asp Gln Ser Ile Ser Phe Leu Ala Glu Glu Gly Thr Ala Lys Leu Ile Glu Phe   207
CGT TAC ATT GGC ACT GAA GGA GGA GGC ATG GAC CAG TCT ATA TCA TTT CTT GCA GAA GAA GGA ACT GCC AAG TTG ATA GAA TTT   641

Ser Pro Leu Arg Ala Thr Asp Val Lys Leu Pro Ser Gly Ala Val Phe Val Ile Ala Asn Ser Cys Val Glu Met Asn Lys Ala   235
AGT CCT CTG AGG GCA ACC GAT GTA AAA CTC CCA AGT GGA GCA GTG TTT GTG ATT GCC AAC AGT TGT GTG GAG ATG AAT AAG GCA   725

Ala Thr Ser His Phe Asn Ile Arg Val Met Glu Cys Arg Leu Ala Ala Lys Leu Leu Ala Lys Tyr Lys Ser Leu Gln Trp Asp   263
GCA ACT TCC CAT TTC AAT ATC AGG GTG ATG GAG TGT CGG CTG GCT GCG AAG CTC CTG GCT AAA TAC AAA AGC TTG CAA TGG GAC   809

Lys Val Leu Arg Leu Glu Glu Val Gln Ala Lys Leu Gly Ile Ser Leu Glu Glu Met Leu Leu Val Thr Glu Asp Ala Leu His   291
AAA GTA CTG AGG CTG GAG GAG GTG CAG GCT AAA CTA GGG ATT AGT CTA GAA GAA ATG CTG TTG GTC ACA GAA GAT GCC CTT CAT   893

Pro Glu Pro Tyr Asn Pro Glu Glu Ile Cys Arg Cys Leu Gly Ile Ser Leu Glu Glu Leu Arg Thr Gln Ile Leu Ser Pro Asn   319
CCT GAA CCC TAT AAC CCT GAG GAG ATC TGC AGG TGT CTG GGA ATT AGC CTG GAG GAA CTC CGA ACC CAA ATC CTG AGT CCA AAC   977

Thr Gln Asp Val Leu Ile Phe Lys Leu Tyr Gln Arg Ala Lys His Val Tyr Ser Glu Ala Ala Arg Val Leu Gln Phe Lys Lys   247
ACT CAA GAT GTG CTC ATC TTC AAA CTC TAT CAG CGG GCA AAG CAT GTG TAC AGC GAG GCT GCG CGA GTG CTC CAG TTT AAG AAG   1061

Ile Cys Glu Glu Ala Pro Glu Asn Met Val Gln Leu Leu Gly Glu Leu Met Asn Gln Ser His Met Ser Cys Arg Asp Met Tyr   375
ATA TGT GAA GAA GCA CCT GAA AAC ATG GTC CAG CTG CTG GGA GAG TTG ATG AAC CAG AGC CAC ATG AGC TGC CGG GAC ATG TAT   1145

Glu Cys Ser Cys Pro Glu Leu Asp Gln Leu Val Asp Ile Cys Arg Lys Phe Gly Ala Gln Gly Ser Arg Leu Thr Gly Ala Gly   403
GAG TGC AGC TGC CCC GAG CTG GAT CAG CTG GTG GAC ATC TGT CGG AAG TTT GGG GCT CAA GGG TCA CGA CTT ACT GGA GCA GGA   1229

Trp Gly Gly Cys Thr Val Ser Met Val Pro Ala Asp Lys Leu Pro Ser Phe Leu Ala Asn Val His Lys Ala Tyr Tyr Gln Arg   431
TGG GGA GGC TGT ACA GTA TCA ATG GTA CCT GCG GAC AAG CTG CCC AGC TTT CTA GCA AAT GTG CAC AAA GCT TAT TAC CAG AGG   1313

Ser Asp Gly Ser Leu Ala Pro Glu Lys Gln Ser Leu Phe Ala Thr Lys Pro Gly Gly Gly Ala Leu Val Leu Leu Glu Ala ★   459
AGT GAT GGA AGC TTA GCA CCG GAG AAG CAA AGT TTG TTT GCT ACC AAA CCT GGA GGT GGG GCT TTG GTT TTG CTT GAG GCC TGA   1397

AAAAATGTAAAAAGTCTGAGAGAAACTACTTAGGGCACTTAGGAATTGGCAGGACTTTCTGTGCCACAGTAAATTAATCTTCCTTCTGTTTTGTATTATGATGAACGGTTG   1508
```

FIG. 1. Nucleotide and deduced amino acid sequences of a human galactokinase cDNA.

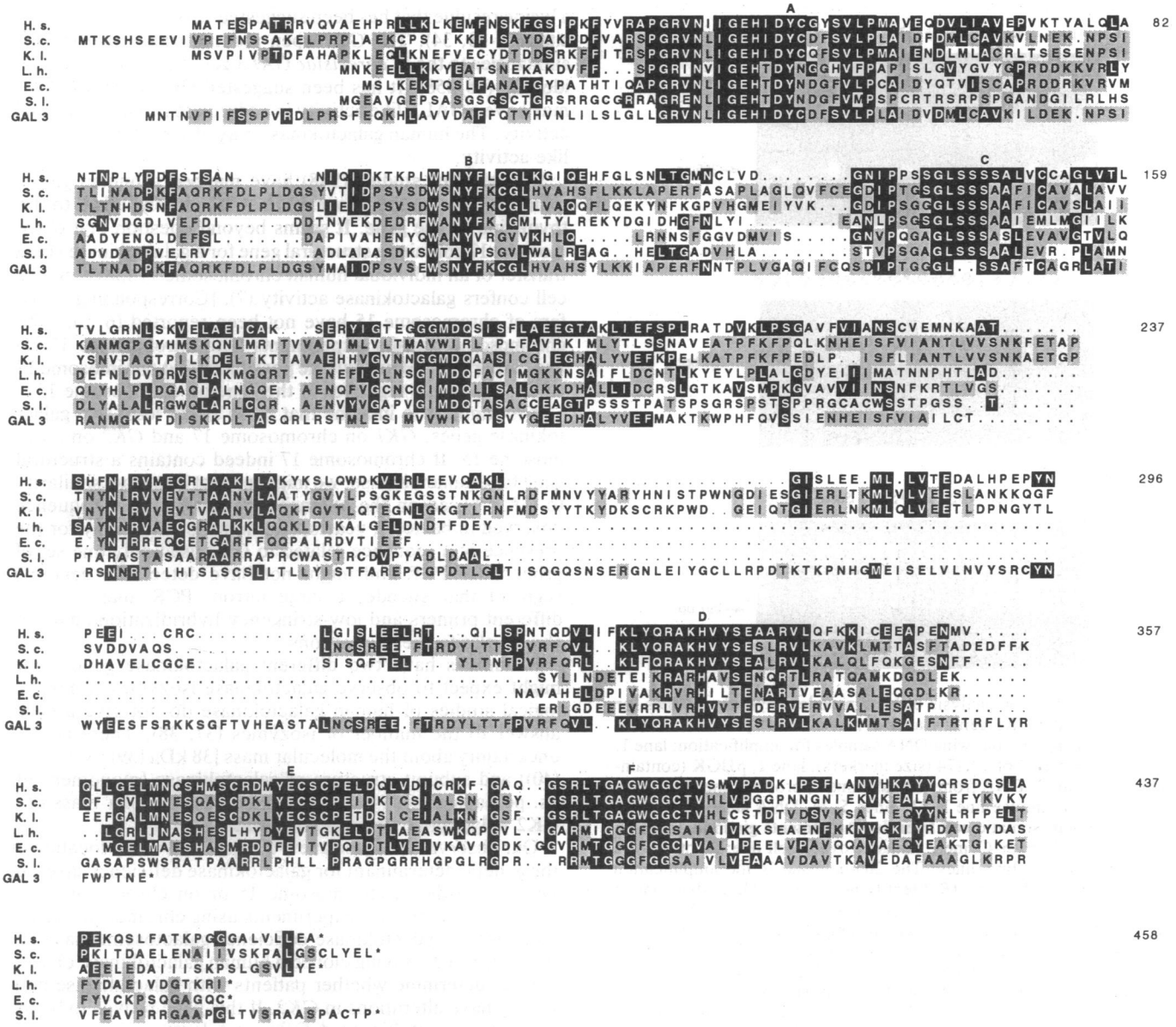

FIG. 2. Alignment of deduced amino acid sequences for human (H.s., 458 amino acid residues), *S. carlsbergensis* (S.c., 528 residues), *K. lactis* (K.l., 503 residues), *E. coli* (E.c., 382 residues), *L. helveticus* (L.h., 388 residues), and *S. lividans* (S.l., 395 residues) galactokinases and *S. cerevisiae* GAL3. Numbers indicate positions in the *GK2* open reading frame. Six regions of similarity (A–F) are indicated. Gaps (dashes) were introduced for optimal alignment. *, Translation termination codon.

from total human genomic DNA (Fig. 4*A*). The PCR products from pJJGK and from human genomic DNA were cleaved into identical fragments by *Stu* I (R.T.L., unpublished observations). Genomic DNA from a panel of well-characterized human–rodent somatic cell hybrids (Table 1) retaining

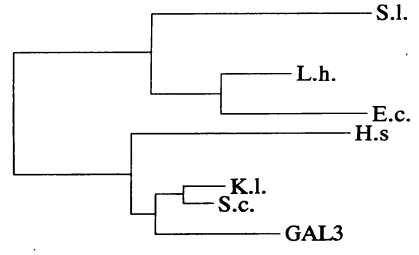

FIG. 3. Phylogenetic tree for galactokinase and galactokinase-related proteins of human, yeast, and bacteria. Symbols are as in Fig. 2. The tree was constructed by the progressive alignment method (20). The lengths of horizontal lines are a measure of divergence.

subsets of human chromosomes was used as a template for PCR (8, 22–26). The presence of an amplified PCR fragment was perfectly correlated with the presence of human chromosome 15 in the hybrid lines but not with chromosome 17, on which the *GK1* gene resides (5–8). Of seven hybrids that lack chromosome 15 but contain chromosome 17, none yielded an amplified fragment. In contrast, the one line with chromosome 15 but not chromosome 17 (A59-3Aaz10a; Fig. 4*A*, lane 23) yielded an amplified fragment. The only common chromosome in lines NA 09940 and A59-3Aaz10a is chromosome 15. These observations argue that the *GK2* gene resides on chromosome 15.

To verify the chromosome composition of our somatic cell hybrid panel, we carried out a similar PCR analysis using primers (27) for genes known to reside on either chromosome 15 or 17, β_2-microglobulin (*B2M*) and gastrin (*GAS*) genes, respectively. PCR primers for *B2M* amplified a 357-bp fragment only from DNA of cell lines reported to contain chromosome 15 (Table 1 and Fig. 4*B*). Amplification data for

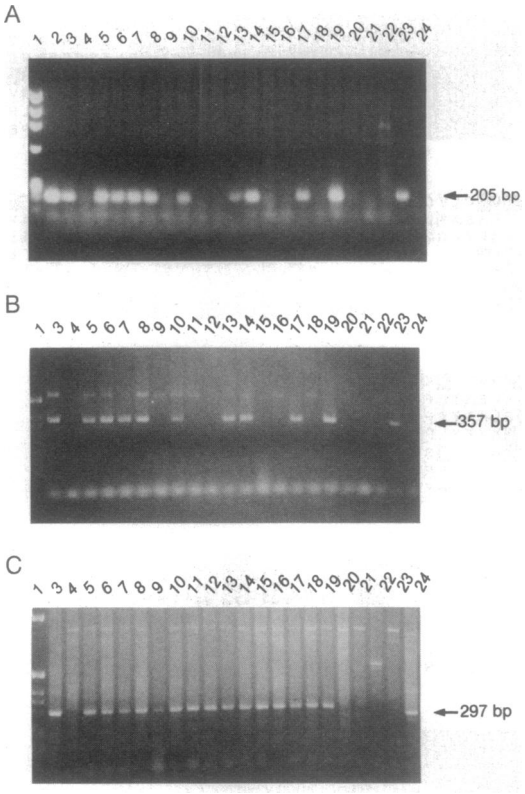

FIG. 4. (*A*) PCR analysis of human–rodent hybrid cell lines for presence of the *GK2* gene. Lanes 1–24 contained products of PCR reactions using the following DNA samples for amplification: lane 1, *Hae* III fragments of ϕX174 (size markers); lane 2, pJJGK (containing *GK2*); lane 3, DNA from normal human fibroblast line Hs27; lane 4, DNA from mouse line B-82; lanes 5–24, DNA samples from 20 human–rodent somatic cell hybrids (assignments given in Table 1). Lane 22 is a human–Chinese hamster ovary cell line; all others are human–mouse hybrid lines. The 205-bp band is the amplification product of the *GK2* gene. (*B*) Identifying hybrid cell lines that contain human chromosome 15. PCR primers (see text) for the human β_2-microglobulin gene were used to amplify a 357-bp fragment as indicated. Lane assignments are the same as in Table 1 and are identical to those in *A* except for lane 1, which contains a 1-kb DNA ladder for size markers; pJJGK (lane 2 in *A*) was not analyzed. (*C*) Identifying hybrid cell lines that contain human chromosome 17. PCR primers for the human gastrin gene amplified a 297-bp fragment as indicated. Assignments are the same as in *B*.

B2M were identical to those for *GK2*. PCR primers for *GAS* amplified a 297-bp fragment only from DNA of cell lines reported to contain chromosome 17 (Table 1 and Fig. 4*C*). These results confirm the chromosome composition of our hybrid cell panel with respect to chromosomes 15 and 17.

DISCUSSION

Two lines of evidence indicate that we have cloned a human cDNA encoding a human galactokinase. (*i*) The human cDNA complements a yeast strain with galactokinase deficiency and allows utilization of galactose. (*ii*) The amino acid sequence of the cloned cDNA has extensive similarity with galactokinases from two yeasts and three bacteria. The fact that we have isolated a functional cDNA from a hepatoma cell line indicates that the corresponding gene is not a pseudogene and is expressed at least in liver cells.

Alignment of galactokinase sequences reveals six distinct regions with identities (A–F; Fig. 2), as has been previously noted (31, 32). The greatest sequence and length variability occurs between regions C and D, making it unlikely that this region contains the catalytic site or site for binding ATP. Region F contains a G-rich sequence (GXGXXG) preceding a lysine residue that has been suggested to be an ATP-binding site (31, 33, 34). The *GK2* open reading frame also shows similarities to the *S. cerevisiae GAL3* gene (35)—19% overall identity (88/458). It has been suggested (30, 36) that GAL1 may have a GAL3-like activity in addition to its galactokinase activity. The human galactokinase may also have this GAL3-like activity.

Several prior studies (5–8) have mapped a human galactokinase gene to chromosome 17q23-25, closely linked to the thymidine kinase gene. It seems beyond question that chromosome 17 contains a structural gene for a galactokinase since transfer of an individual human chromosome 17 into a mouse cell confers galactokinase activity (7). [Corresponding transfers of chromosome 15 have not been reported (6, 7).] The chromosome transfer experiments with chromosome 17 (7) and earlier data (6) rule out the possibility that chromosome 17 carries a positive regulator of the gene on chromosome 15.

It is simplest to imagine that there are at least two galactokinase genes, *GK1* on chromosome 17 and *GK2* on chromosome 15. If chromosome 17 indeed contains a structural gene for a galactokinase, we might expect it to have similarity to *GK2*, based on the extensive conservation of sequence observed in bacteria, yeasts, and humans. Our assay for the presence of a sequence related to *GK2* used only one set of primers for PCR and might not have detected a genomic segment that encodes a large intron. PCR analysis with different primers and low-stringency hybridization analyses should clarify these questions.

If humans have two different galactokinase genes, one might expect to observe galactokinase isozymes. Enzymological studies of human galactokinase do not give a clear answer to the number of isozymes (37, 38). There is also uncertainty about the molecular mass [38 kDa (39) vs. 55 kDa (40)] and subunit structure of galactokinase [monomer (39) vs. perhaps dimer (40)]. The predicted molecular mass of a GK2 polypeptide chain is 50.4 kDa.

Our observations raise anew the question of the location of the genetic determinant for galactokinase deficiency disease: Does it reside on chromosome 15 or on chromosome 17? Chromosome transfer experiments using chromosomes from humans with galactokinase-deficiency galactosemia have not been reported. Having cloned a human galactokinase cDNA, we can determine whether patients with galactokinase deficiency have alterations in *GK2*. If this gene proves to be the locus for galactokinase deficiency galactosemia, it can be used to determine definitively the basis for presenile cataracts, in particular, whether this disease results from heterozygosity or other genetic alterations at this locus. The strategy used here to clone a human galactokinase gene could be used to clone various human disease genes affecting intermediary metabolism (41) for which yeast mutants are known (42).

We gratefully acknowledge Geoffrey Duyk for helpful discussions and encouragement and Susan R. Hopkins for expert technical assistance. We also thank Flora Banuett, Joachim Li, other members of the Herskowitz laboratory, and Margit Burnmeister for useful comments. We thank J. Fernando Bazan for alignment of the galactokinase amino acid sequences, Anthony Brake for the cDNA library, Hunt Willard for somatic cell hybrid DNA samples, Mark Johnston for yeast strains, and Hunt Willard, Charles Epstein, Mark Johnston, Geoffrey Duyk, Jim Hopper, David Cox, William Sly, and Ed Anderson for comments on the manuscript. R.T.L. is a Howard Hughes Medical Institute Medical Student Fellow. This work was supported by the Howard Hughes Medical Institute, the Kimura Memorial Fund, That Man May See, Inc., the Stulsaft and Weingart Foundations, Sam Zakessian, and a research grant from the National Institutes of Health (to I.H.).

1. Gitzelmann, R. (1965) *Lancet* **ii**, 670–671.
2. Levy, H. L. & Hammersen, G. (1978) *J. Pediatr.* **92**, 871–877.
3. Kinoshita, J. H. (1974) *Invest. Ophthalmol. Visual Sci.* **13**, 713–724.

Table 1. Mapping of a human galactokinase gene in a panel of human–rodent somatic cell hybrids

PCR lane assignments	DNA	Human chromosome																							Amplified PCR fragment		
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	X	Y	*GK2*	*β2M*	*GAS*
1	φX174 RF DNA/HaeIII																											
2	pJJGK																									+		
3	Human Fibroblast	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
4	B-82 mouse L cell	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
5	NA 09925	+	+	–	+	+	+	+	+	–	–	–	+	–	+	+	+	+	+	+	+	–	+	–	–	+	+	+
6	NA 09926	+	+	+	+	–	+	+	+	–	+	–	–	+	+	+	–	+	+	+	+	–	–	–	–	+	+	+
7	NA 09927	+	+	+	+	–	+	+	+	–	+	–	–	+	+	+	–	+	+	+	+	–	–	–	–	+	+	+
8	NA 09928	–	+	+	–	+	+	–	+	–	–	–	–	–	+	+	–	+	–	+	–	+	+	–	+	+	+	+
9	NA 09929	–	–	+	+	–	+	–	+	–	–	+	+	–	+	–	–	+	–	–	+	–	–	–	–	–	–	+
10	NA 09930A	–	+	+	–	+	–	+	–	–	–	–	+	+	+	+	–	+	+	–	+	+	+	–	–	+	+	+
11	NA 09931	–	–	–	–	+	–	+	–	–	+	–	+	–	+	–	–	+	–	–	+	+	–	–	+	–	–	+
12	NA 09932	–	–	–	+	+	+	–	+	–	–	+	+	–	–	–	–	+	–	–	–	+	–	–	–	–	–	+
13	NA 09933	+	–	+	+	+	+	+	+	–	–	–	+	+	+	+	–	+	+	+	+	+	+	–	+	+	+	+
14	NA 09934	–	+	–	–	+	+	–	+	–	–	+	+	–	–	+	–	+	+	–	+	+	–	–	–	+	+	+
15	NA 09935A	–	–	+	+	+	+	–	–	–	–	–	+	+	+	–	–	+	+	–	–	+	+	–	–	–	–	+
16	NA 09936	–	–	–	+	–	+	+	+	–	+	+	–	–	+	–	–	+	–	+	+	–	+	–	–	–	–	+
17	NA 09937	–	–	+	+	–	+	+	+	–	–	–	+	–	+	+	–	+	+	–	–	–	–	–	–	+	+	+
18	NA 09938	–	–	–	+	+	+	+	–	–	–	+	+	–	+	–	–	+	–	–	+	+	+	–	–	–	–	+
19	NA 09940	–	–	+	–	–	–	+	+	–	–	–	–	–	–	+	–	+	–	–	–	–	–	–	–	+	+	+
20	NA 10324	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	–	–	–
21	NA 10567	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–
22	NA 10611	–	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–
23	A59-3Aaz10a	+	–	–	–	–	–	–	–	–	+	–	–	+	–	+	–	–	+	–	+	–	–	+	–	+	+	–
24	MH-22.6	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	–	+	–	–	–	–	–	–	–	–	–	+

Human–rodent cell lines (lanes 5–24) were analyzed for *GK2* sequences by PCR analysis. Other lanes are described in the legend to Fig. 4. The presence of a known chromosome 15 gene (*B2M*) and a known chromosome 17 gene (*GAS*) was similarly analyzed. +, The percentage of cells containing the particular chromosome is >9% (as determined by karyotypic analysis performed by Coriell Institute, Camden, NJ). RF, replicative form.

4. Monteleone, J. A., Beutler, E., Monteleone, P. L., Utz, C. L. & Casey, E. C. (1971) *Am. J. Med.* **50**, 403–407.
5. Orkwiszewski, K. G., Tedesco, T. A. & Croce, C. M. (1974) *Nature (London)* **252**, 60–62.
6. Elsevier, S. M., Kucherlapati, R., Nichols, E. A., Creagan, R. P., Giles, R. E., Ruddle, F. H., Willecke, K. & McDougall, J. K. (1974) *Nature (London)* **251**, 633–636.
7. Xu, W., Gorman, P. A., Rider, S. H., Hedge, P. J., Moore, G., Prichard, C., Sheer, D. & Solomon, E. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 8563–8567.
8. VanTuinen, P., Rich, D. C., Summers, K. M. & Ledbetter, D. H. (1987) *Genomics* **1**, 374–381.
9. Douglas, H. C. & Hawthorne, D. C. (1964) *Genetics* **49**, 837–844.
10. Lee, M. G. & Nurse, P. (1987) *Nature (London)* **327**, 31–35.
11. Colicelli, J., Nicolette, C., Birchmeier, C., Rodgers, L., Riggs, M. & Wigler, M. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 2913–2917.
12. Schild, D., Brake, A. J., Kiefer, M. C., Young, D. & Barr, P. H. (1990) *Proc. Natl. Acad. Sci. USA* **87**, 2916–2920.
13. St. John, T. P. & Davis, R. W. (1981) *J. Mol. Biol.* **152**, 285–315.
14. Ito, H., Fukuda, Y., Murata, K. & Kimura, A. (1983) *J. Bacteriol.* **153**, 163–168.
15. Sherman, F. (1991) *Methods Enzymol.* **194**, 14–15.
16. Norrander, J., Kempe, T. & Messing, J. (1983) *Gene* **26**, 101–106.
17. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
18. Lipman, D. J. & Pearson, W. R. (1985) *Science* **22**, 1435–1441.
19. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
20. Feng, D.-F. & Doolittle, R. F. (1990) *Methods Enzymol.* **183**, 375–387.
21. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. & Erlich, H. A. (1988) *Science* **239**, 487–491.
22. Taggart, R. T., Mohandas, T. K., Shows, T. B. & Bell, G. I. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 6240–6244.
23. Mohandas, T., Heinzmann, C., Sparkes, R. S., Wasmuth, J., Edwards, P. & Lusis, A. J. (1986) *Somatic Cell Mol. Genet.* **12**, 89–94.
24. Warburton, D., Gersen, S., Yu, M. T., Jackson, C., Handelin, B. & Housman, D. (1990) *Genomics* **6**, 358–366.
25. Callen, D. F. (1986) *Ann. Genet.* **29**, 235–239.
26. Willard, H. F. & Riordan, J. R. (1985) *Science* **230**, 940–942.
27. Theune, S., Fung, J., Todd, S., Sakaguchi, A. Y. & Naylor, S. L. (1991) *Genomics* **9**, 511–516.
28. Boeke, J. D., LaCroute, F. & Fink, G. R. (1984) *Mol. Gen. Genet.* **197**, 345–346.
29. Citron, B. A. & Donelson, J. E. (1984) *J. Bacteriol.* **158**, 269–278.
30. Meyer, J., Walker-Jonah, A. & Hollenberg, C. P. (1991) *Mol. Cell. Biol.* **11**, 5454–5461.
31. Debouck, C., Riccio, A., Schumperli, D., McKenney, K., Jeffers, J., Hughes, C., Rosenberg, M., Heusterspreute, M., Brunel, F. & Davison, J. (1985) *Nucleic Acids Res.* **13**, 1841–1853.
32. Mollet, B. & Pilloud, N. (1991) *J. Bacteriol.* **173**, 4464–4473.
33. Adams, G. W., Fornwald, J. A., Schmidt, F. J., Rosenberg, M. & Brawner, M. E. (1988) *J. Bacteriol.* **170**, 203–212.
34. Saraste, M., Sibbald, P. R. & Wittinghofer, A. (1990) *Trends Biochem. Sci.* **15**, 430–434.
35. Bajwa, W., Torchia, T. E. & Hopper, J. E. (1988) *Mol. Cell. Biol.* **8**, 3439–3447.
36. Bhat, P. J., Oh, D. & Hopper, J. E. (1990) *Genetics* **125**, 281–291.
37. Srivastava, S. K., Blume, K.-G., Van Loon, C. & Beutler, E. (1972) *Arch. Biochem. Biophys.* **150**, 191–198.
38. Stambolian, D., Scarpino-Myers, V. & Harris, H. (1984) *Exp. Eye Res.* **38**, 231–237.
39. Stambolian, D., Scarpino-Myers, V. & Harris, H. (1985) *Biochim. Biophys. Acta* **831**, 306–312.
40. Blume, K.-G. & Beutler, E. (1971) *J. Biol. Chem.* **246**, 6507–6510.
41. McKusick, V. A. (1990) *Mendelian Inheritance in Man: Catalogs of Autosomal Dominant, Autosomal Recessive, and X-linked Phenotypes* (Johns Hopkins Press, Baltimore), 9th Ed.
42. Jones, E. W. & Fink, G. R. (1981) *The Molecular Biology of the Yeast Saccharomyces: Metabolism and Gene Expression*, eds. Strathern, J. N., Jones, E. W. & Broach, J. R. (Cold Spring Harbor Lab., Cold Spring Harbor, NY).