Structure of the human hexokinase type I gene and nucleotide sequence of the 5′ flanking region

Annamaria RUZZO, Francesca ANDREONI and Mauro MAGNANI[1]

'G. Fornaini' Institute of Biological Chemistry, University of Urbino, Via Saffi 2, 61029 Urbino, Italy

This study reports the precise intron/exon boundaries and intron/exon composition of the human hexokinase type I gene. A yeast artificial chromosome containing the hexokinase type I gene was isolated from the yeast artificial chromosome library of the Centre d'Étude du Polymorphisme Humaine. A cosmid sublibrary was created and direct sequencing of the individual cosmids was used to provide the exon/intron organization. The human hexokinase type I gene was found to be composed of 18 exons ranging in size from 63 to 305 bp. Intron 1 is at least 15 kb in length, whereas intron 2 spans at least 10 kb. Overall, the length of the 17 introns ranges from 104 to greater than 15 kb. The entire coding region is contained in at least 75 kb of the gene. The structure of the gene reveals a remarkable conservation of the size of the exons compared with glucokinase and hexokinase type II. Isolation of the 5′ flanking region of the gene revealed a 75–90 % identity with the rat sequence. Direct evidence of an alternative red-blood-cell-specific exon 1 located upstream of the 5′ flanking region of the gene is also provided.

INTRODUCTION

Hexokinase (ATP:D-hexose 6-phosphotransferase; EC 2.7.1.1) catalyses the first step of glucose metabolism, utilizing ATP for the phosphorylation of glucose to glucose 6-phosphate. Mammalian cells express four hexokinase isoenzymes, types I–IV, with different tissue distributions and kinetic properties [1,2]. The hexokinase type I–III isoenzymes in mammalian tissues consist of a single polypeptide chain of approx. 100 kDa, have a high affinity for glucose and are subject to inhibition by the reaction product glucose 6-phosphate [2,3]. Hexokinase type IV (approx. 50 kDa) (commonly known as glucokinase) differs from the other members of the hexokinase family in that it is not inhibited by physiological concentrations of glucose 6-phosphate, shows a lower affinity for glucose and is expressed in mammalian liver and pancreas [2,3].

Studies comparing the N- and C-terminal halves of mammalian hexokinase have revealed similarities between the two halves of the different isoenzymes and with yeast hexokinase, probably as a result of gene duplication and fusion of an ancestral gene resembling the actual yeast form [4–7]. Tsai and Wilson [8] suggested that, among all mammalian isoenzymes, hexokinase type II most closely resembles the ancestral 100 kDa enzyme because the N- and C-terminal halves of hexokinase type II exhibit 60 % identity in their amino acid sequences (compared with 51 % and 47 % respectively for the type I and III isoenzymes) and because both halves of this isoenzyme contain a catalytic site. Thus hexokinase types I and III evolved from hexokinase type II, whereas hexokinase type IV might be the product of either the ancestral 50 kDa enzyme [8–10] or of a re-splitting of the ancestral 100 kDa enzyme gene [7,8,11]. Determination of the structure of the genes of hexokinase types II [9,10,12,13], IV [14–17] and rat hexokinase type I [18], as well as evidence for the conservation of the intron/exon structure provide valuable direct support for this previous hypothesis.

Unique hexokinase type I mRNA species have been described in murine [19,20] and recently in human spermatogenic cells (GenBank accession numbers U38226, U38227, U38228) and in human red blood cells [21]. These mRNA species result in distinct forms of hexokinase type I by an unknown mechanism.

The present study reports the structure of the human hexokinase type I gene. The precise intron/exon boundaries and intron/exon composition of the gene were determined in the region corresponding to the human hexokinase type I cDNA previously reported by Nishi et al. [22]. The human hexokinase type I gene and its 5′ flanking region were then isolated and sequenced from a yeast artificial chromosome (YAC) library.

Direct evidence is provided that testis- and red blood cell-specific hexokinase type I differ in their 5′ upstream regions starting from exon 2, thus suggesting that these isoenzymes might have arisen by means of an alternative splicing mechanism.

EXPERIMENTAL

Preparation of probes from human hexokinase type I cDNA

Three probes spanning the entire coding sequence of the human hexokinase type I cDNA [22], HK-ter, HK-Nα and HK-Nβ, were used in this study. Probe HK-ter was prepared from the pJHK502 clone [23], containing a 2155 bp fragment corresponding to the C-terminal region of human hexokinase type I cDNA, after digestion with restriction enzymes *Nco*I and *Eco*RI [24]. The N-terminal coding region (nt +20 to +1715 of the cDNA) was amplified with a pair of primers (HK 1 forward, 5′-CGCCAGGGCTGCGGAGGACCGA-3′ and HK 10 reverse, 5′-ATTTTCACCAGCAGCACACGGA-3′) [25]. The 1696 bp PCR product was cloned in the TA Cloning vector pCR II (Invitrogen, San Diego, CA, U.S.A.) and named pHK1-10. After digestion of pHK1-10 with *Eco*RI, we obtained the HK-Nα (nt +20 to +353) and HK-Nβ (nt +354 to +1715) probes. The probes were labelled with the random primer DNA labelling kit (Bio-Rad, Hercules, CA, U.S.A.) and 30 μCi of [α-^{32}P]dCTP with 50–100 ng of DNA as template.

Hybridizations were performed at 65 °C overnight in 6 × SSC buffer (20 × SSC = 3 M NaCl/0.3 M trisodium citrate)/5 × Denhardt's solution [100 × Denhardt's = 2 % (w/v) BSA/

Abbreviations used: YAC, yeast artificial chromosome; UTR, untranslated region.

[1] To whom correspondence should be addressed.

The nucleotide sequence data reported in this paper will appear in DDBJ, EMBL and GenBank Nucleotide Sequence Databases under the accession numbers AF016349–AF016365.

2% (w/v) Ficoll, 2% (w/v) poly(vinylpyrrolidone)]/0.5% (w/v) SDS/20 μg/ml sonicated salmon sperm. The filters were washed down to 0.1 × SSC at 65 °C for 15 min and autoradiographed on X-ray film (Kodak X-OMAT AR) [24].

Isolation of the human hexokinase type I gene from the YAC library

We used 22 exonic oligonucleotide primers (forward and reverse), designed with the nucleotide sequence of hexokinase type I cDNA [22], previously employed in our laboratory to study hexokinase type I cDNA mutations [25], to perform several PCR amplifications on human genomic DNA template. PCR amplifications were performed in a 25 μl volume with the Perkin Elmer GeneAMP kit (1.5 mM $MgCl_2$/1 unit of AmpliTaq) with 1 μg of human genomic DNA, 5% (v/v) DMSO and 20 pmol of each primer. PCR reactions were performed for 30–35 cycles of 94 °C for 60 s, 58–65 °C for 30–60 s and 72 °C for 30–120 s, with a final extension at 72 °C for 10 min. Products were analysed by electrophoresis on agarose gel, blotted and hybridized with one of the three probes described above containing the corresponding sequence.

Positive PCR products were cloned in the TA cloning vector pCR II (Invitrogen) and sequenced by the Sanger method [26] (Sequenase Version 2.0 Sequencing kit; USB, Amersham, Bucks., U.K.).

The isolated DNA fragments contained introns that were later identified as introns 10, 11 and 17. These intronic sequences were used to design more selective oligonucleotide primers with which to screen two YAC libraries, the CEPH 'mega-YAC' (Centre d'Étude du Polymorphisme Humaine) [27] and the ICI YAC (ICI Diagnostics) libraries [28] by the PCR method: the HK 27 forward primer (5′-AGGGGGCTGTCTGTGCTTTGGT-3′) (Table 1) complementary to intron 17 and the HK 16 reverse primer (5′-CGTTACATTTTGGTGACAGTTC-3′) [25] complementary to exon 18. Hot Start PCR [29] amplification (141 bp product) was performed with 1 μl of melted YAC plug in a reaction volume of 25 μl with the Perkin Elmer GeneAMP kit (1.5 mM $MgCl_2$/1 unit of AmpliTaq), 5% (v/v) DMSO and 10 pmol of each primer. The reaction was performed for 40 cycles of 94 °C for 25 s, 62 °C for 25 s and 72 °C for 25 s, with a final extension of 10 min at 72 °C in a Perkin Elmer thermal cycler.

Four positive clones were isolated and analysed by Southern blot hybridization. The yeast DNA was isolated after growing cells for 3 days at 30 °C in AHC broth [30]. The cells were then resuspended in 0.5 ml YRB [1.2 M sorbitol/10 mM Tris/HCl (pH 7.5)/20 mM EDTA/14 mM 2-mercaptoethanol] and the yeast DNA was prepared in solution by standard protocols [31]. Yeast DNA (250 ng) was digested with *Eco*RI and separated on 0.8% (w/v) agarose gel in a Tris/acetic acid/EDTA buffer. DNA fragments were transferred to a nylon membrane by the Southern blot method and hybridized with the probes described above [24].

Subcloning of hexokinase type I YAC into cosmids

Yeast cells were harvested after 3 days of culturing at 30 °C in AHC broth [30]. YAC DNA was isolated by the 'lithium method' [28] as intact yeast chromosome, embedding the yeast cells in 2% (w/v) agarose plugs containing 1 M sorbitol, 20 mM EDTA, 14 mM 2-mercaptoethanol, 20 units/ml lyticase. In brief, yeast spheroplasts were prepared by incubating the plugs at 37 °C for 2 h in 1 M sorbitol/20 mM EDTA/14 mM 2-mercaptoethanol/10 mM Tris/HCl (pH 7.5)/20 units/ml lyticase, then lysed in 1% (w/v) lithium dodecyl sulphate/100 mM EDTA/10 mM Tris/HCl(pH 8.0) for 60 min at 37 °C and overnight at 37 °C with fresh yeast lysis solution. Plugs were stored at 4 °C in 0.5 M EDTA and, before use, washed three times in 10 mM Tris/HCl (pH 8.0)/0.1 mM EDTA for 30 min at room temperature.

A partial *Mbo*I digestion of yeast DNA plugs was performed to prepare DNA with an average size of approx. 50 kb [24]. After removal of agarose with β-agarase (New England Biolabs, Beverly, MA, U.S.A.), the DNA was dephosphorylated with calf intestinal alkaline phosphatase (Promega, Madison, WI, U.S.A.). After the addition of both trinitriloacetic acid, pH 8.0, to a final concentration of 0.015 M and Dextran T40 as DNA carrier and purification by phenol/chloroform extraction, the DNA fragments were ligated to the *Xba*I–*Bam*HI-digested Supercos 1 cosmid vector (Stratagene, La Jolla, CA, U.S.A.). The ligation mix was then packaged with the Gigapack II XL (Stratagene) λ phage packaging extract. XL1 blue MR host cells were infected with the packaged phage and cultured at 37 °C overnight on Hybond-N nylon membranes (Amersham, Little Chalfont, Bucks., U.K.) placed on 10 cm × 10 cm diameter Luria–Bertani agar plates containing 50 μg/ml ampicillin [24]. Replica nylon filters were prepared from master plates and screened with HK-ter and HK-Nα. In this first screening, 34 cosmid clones were found to be positive.

Restriction mapping of cosmids

Cosmid clones were digested with *Eco*RI and separated on 0.8% (w/v) agarose gel in a Tris/acetic acid/EDTA buffer. DNA fragments were transferred to a nylon membrane with the vacuum blot technique and hybridized with the three cDNA probes [24].

Sequencing of hexokinase type I exon/intron junctions and determination of intron sizes

The structure of the hexokinase type I gene, including the exon/intron junction sequences, was determined by PCR amplification and direct sequencing of cosmids. All primers used to amplify and sequence the cosmids are listed in Table 1. To verify the exon/intron junction sequences obtained with exonic primers, intron-specific primers (primers HK 23, HK 24, HK 26, HK 27, HK 49, HK 50 and HK 54–HK 78) (Table 1) were used for reverse or second-strand sequencing.

Cosmid DNA was isolated from each clone with Qiagen-tip 500 columns and buffers (Qiagen, Chatsworth, CA, U.S.A.) in accordance with the supplier's instructions. Nucleotide sequence analysis was performed with the Sanger dideoxy chain termination method [26]: 7 μg of cosmid DNA and 10 pmol of sequencing primer were incubated, in the presence of 10% (v/v) DMSO, at 100 °C for 5 min and transferred immediately into a bath of solid CO_2. DNA template-annealed primer was used for the sequencing reaction in accordance with the instructions for use of the Sequenase Version 2.0 kit (USB). Direct sequencing was also performed by cycle sequencing with the 'fmol DNA sequencing system' (Promega) and the Thermo Sequenase cycle sequencing kit (Amersham) in accordance with the manufacturer's protocols.

PCR amplifications of DNA fragments containing introns 5, 6, 9, 10, 11, 12, 13, 14, 15 and 17 were performed in a 25 μl volume with the Perkin Elmer GeneAMP kit (1.5 mM $MgCl_2$/1 unit of AmpliTaq) with 25 ng of cosmid DNA, 20 pmol of each primer and 5% (v/v) DMSO. The PCR conditions used were: 30–35 cycles of denaturation at 95 °C for 15–30 s, annealing at 58–65 °C for 30–60 s and extension at 72 °C for 30–180 s with a

Table 1 Primers used in PCR amplifications and sequencing of the human hexokinase type I gene

The primers used in this study are shown; exonic primers HK 1–HK 22 are listed in Bianchi et al. [25]. The location and orientation of each primer is shown.

Primer number	Location	Orientation	Primer sequence
HK 53	Exon 1R*	Sense	5′-CCACAACCTGACACTGGGCAAGAT-3′
HK 64	Intron 1R	Anti-sense	5′-CTGACCAAGCATCCCCTCTCAT-3′
HK 51	Intron 1R†	Anti-sense	5′-AGCCGACCCACCTCCTCCAC-3′
HK 52	Intron 1R†	Sense	5′-GAGGAGGAGGAGCCGCCGAGCAG-3′
HK 47	Exon 1‡	Anti-sense	5′-AGCCCTGGCGAGCCGTGGTCCT-3′
HK 74	Intron 1	Anti-sense	5′-GGAGGATGGAGGCAGCGGAGGC-3′
HK 65	Intron 1	Sense	5′-GGTGACATGGATGACAGCAGTT-3′
HK 28	Exon 2	Sense	5′-CAAGTATCTGTATGCCATGCGG-3′
HK 48	Exon 2	Anti-sense	5′-TTGACTGTGGCTGTTGGATT-3′
HK 29	Exon 2	Anti-sense	5′-GAATGGACCTTACGAATGTTGG-3′
HK 66	Intron 2	Anti-sense	5′-TTGATGGAATGGTGAATGAATG-3′
HK 67	Intron 2	Sense	5′-GCTTCCCCTTAACATTTGAATC-3′
HK 79	Exon 3	Anti-sense	5′-TCACTTGCACCCGCAGAATTCGA-3′
HK 68	Intron 3	Anti-sense	5′-TTCCAGCAACCCTCTTCC-3′
HK 75	Intron 3	Sense	5′-TTTATATTTTTCTATGAAATGCTA-3′
HK 36	Exon 4	Sense	5′-GTTGCTGAGTGCCTGGGAGATT-3′
HK 37	Exon 4	Anti-sense	5′-CGTGAATCCCACAGGTAACTTC-3′
HK 69	Intron 4	Anti-sense	5′-AGTTGTGGAGAAAAGGGCTTGT-3′
HK 70	Intron 4	Sense	5′-GGATGATCGGGAGATGGAAAT-3′
HK 38	Exon 5	Sense	5′-GAGTGGAAGGAGCAGATGTGGT-3′
HK 39	Exon 5	Anti-sense	5′-TGACCACATCTGCTCCTTCCAC-3′
HK 71	Intron 5	Anti-sense	5′-GGAAGGGCACTGGTCTGTTTTA-3′
HK 72	Intron 5	Sense	5′-CACTCACAAAAGCAGGGGTTCT-3′
HK 43	Intron 5	Sense	5′-GACAATGACACCCCGTTATCTG-3′
HK 73	Intron 7	Anti-sense	5′-TTTTCATCTCCCATCCCAAATC-3′
HK 54	Intron 7	Sense	5′-CCATTCCTTTTATGTCTGTCCC-3′
HK 55	Intron 8	Anti-sense	5′-TGAAGGAAAAGCCACCAAACT-3′
HK 56	Intron 8	Sense	5′-ACATGGTAAGTGGGGCTGTC-3′
HK 44	Exon 9	Anti-sense	5′-ACCAAGTTGGCTGAGCGAAATG-3′
HK 57	Intron 9	Anti-sense	5′-ATTGTCTTGTCTGCCGAAAAAC-3′
HK 58	Intron 9	Sense	5′-CGGAGGTCCCCAATAAATGCTC-3′
HK 23	Intron 10	Anti-sense	5′-TGGCAGGAGCAGGACAGGCATG-3′
HK 24	Intron 10	Sense	5′-GGGTTCTCCCCTTGAAAGTCTG-3′
HK 30	Exon 12	Sense	5′-GGACTACATGGGGATCAAAGGC-3′
HK 32	Exon 12	Anti-sense	5′-GGCAGGGAAATGAGAACGTGAA-3′
HK 76	Intron 12	Anti-sense	5′-CCACAGGGGACTCAGTGCTG-3′
HK 59	Intron 12	Sense	5′-TTCTCCCCTTCCAGGTT-3′
HK 33	Exon 13	Sense	5′-CGTGGGCCACGATGTAGTCACC-3′
HK 31	Exon 13	Anti-sense	5′-TCTCTCCTTTTTATCGCATCCC-3′
HK 45	Exon 13	Sense	5′-GATGCGATAAAAAGGAGAGAGG-3′
HK 60	Intron 13	Anti-sense	5′-TTCTTTACAACAGGCAAATACA-3′
HK 49	Intron 13	Sense	5′-GATGATGACCAGTGGCTCTCT-3′
HK 35	Exon 14	Sense	5′-ACCTGGACGTGGTGGCTGTGGT-3′
HK 34	Exon 14	Anti-sense	5′-AACCTCACAGGTGGGCTCCTCA-3′
HK 77	Intron 14	Anti-sense	5′-ACCCCATACAGAGGACC-3′
HK 78	Intron 14	Sense	5′-TGCAACGACCCCCAAAT-3′
HK 40	Exon 15	Sense	5′-ACATGGAGTGGGGGGCCTTTGG-3′
HK 50	Intron 15	Anti-sense	5′-CTCTGTTCAGACCTCTCCTCCTA-3′
HK 61	Intron 15	Sense	5′-GCGCTTGAGGGGCAGTAGGAGA-3′
HK 41	Exon 16	Sense	5′-ATCGTCCGCAACATCTTAATCG-3′
HK 42	Exon 16	Anti-sense	5′-GTTGCGGACGATTTCACCCAGG-3′
HK 62	Intron 16	Anti-sense	5′-CACCCCCATCACAATCCACCAG-3′
HK 63	Intron 16	Sense	5′-GGCTGCTCTTGTGGGTCCTG-3′
HK 46	Exon 17	Anti-sense	5′-TGTCATCGCAGGTGCTATTCAG-3′
HK 26	Intron 17	Anti-sense	5′-CAAGCTCTGCGCATCCCAAC-3′
HK 27	Intron 17	Sense	5′-AGGGGGCTGTCTGTGCTTTGGT-3′

* The primer is located on the 5′ untranslated region (UTR) of exon 1R.
† The primer is located on the 5′ flanking region of the hexokinase type I gene.
‡ The primer is located on the 5′UTR of exon 1.

final extension at 72 °C for 7 min. AmpliTaq was added to the reaction after preheating at 97 °C for 3 min [29].

DNA fragments containing introns 3, 4, 8 and 16 were amplified by the Long PCR method [32] in a 25 μl volume with the Perkin Elmer GeneAmp XL PCR kit (1.5 mM magnesium acetate/ 2 units of rTth DNA polymerase, XL) with 25 ng of cosmid DNA and 25 pmol of each primer. The Long PCR was performed in two steps: 20 cycles at 93 °C for 1 min and 66–68 °C

for 8–10 min followed by an additional 16 cycles at 93 °C for 1 min and 66–68 °C for 8–10 min with 15 s increments in the annealing/extension step.

Cloning and sequencing of exon 1 and the 5′ flanking region

Sequencing and Southern blot hybridization revealed that selected cosmid clones did not contain exon 1 and the promoter region of the hexokinase type I gene. Therefore the cosmid library was screened again with a probe selective for exon 1. The probe (nt +20 to +143 of the cDNA) was prepared by pHK1-10 digestion with *Eco*RI, the fragment (nt +20 to +353) was purified from agarose gel (Prep-A-Gene; Bio-Rad) and digested with *Fok*I [24]. The sticky ends were transformed into blunt ends with Klenow polymerase and the fragment (nt +20 to +143) was cloned in pBluescript KS+ (Stratagene). One cosmid clone was isolated from the cosmid library by hybridization with this new probe.

Direct sequencing of the positive clone allowed us to confirm the presence of exon 1 and to determine the 5′ sequence of intron 1. Because of the difficulty in performing direct sequencing on the 5′ flanking region of the gene, it was necessary to construct a 'vectorette' library by using *Rsa*I and following the procedure as outlined by Riley et al. [33]. A 25 μl PCR reaction included 3 μl of the library, 25 pmol of primer HK 47 (Table 1) and 224 [33], the Perkin Elmer GeneAMP kit (1.5 mM $MgCl_2$/1 unit of AmpliTaq) and 5% (v/v) DMSO. Cycling conditions were as follows: 40 cycles of 95 °C for 30 s, 62 °C for 60 s and 72 °C for 180 s with a final extension of 72 °C for 7 min. The 2.3 kb fragment was gel-purified (Prep-A-Gene; Bio-Rad), subcloned in the TA Cloning vector pCR II (Invitrogen) and sequenced.

Isolation of red-blood-cell-specific exon 1

The position of red-blood-cell-specific exon 1 [21] was determined by the Long PCR method with the HK 53 forward primer complementary to red-blood-cell-specific exon 1 and the HK 51 reverse primer complementary to the 5′ flanking region of the gene (Table 1). The amplification reaction was performed in two steps: 20 cycles at 94 °C for 20 s and 68 °C for 7 min followed by an additional 18 cycles under the same conditions with 15 s increments in the annealing/extension step. The PCR was conducted in a 25 μl volume with the Perkin Elmer GeneAmp XL PCR kit (1.5 mM magnesium acetate/3 units rTth DNA Polymerase, XL), 25 ng of cosmid clone containing the 5′ flanking region of the gene and 25 pmol of each primer. The 2.9 kb fragment was gel-purified (Prep-A-Gene; Bio-Rad) and sequenced.

RESULTS

Exon/intron organization of the human hexokinase type I gene

The working hypothesis from which we began this study assumed that the structure of the human hexokinase type I gene could be similar to that of the hexokinase type II gene [9,10,12,13]. Guided by the exon/intron structure of the human hexokinase type II gene, oligonucleotide primers complementary to the human hexokinase type I coding sequences previously reported in Bianchi et al. [25] were used to amplify across sequences where introns were expected to be located, with total human genomic DNA as template. The exonic oligonucleotide primers allowed us to isolate and sequence introns 10, 11 and 17. The identified intronic sequences were used to design new primers for a more selective PCR to screen the CEPH and the ICI YAC libraries. This strategy was necessary to exclude the other hexokinase genes present in the human genome.

Four positive YACs were identified: three of these clones were found in the ICI (YAC 34D/F1, YAC 29D/E9 and YAC 6A/D12) (results not shown) and one in the CEPH (YAC 908/E9) library.

The CEPH clone 908/E9 was then subcloned into the cosmid vector. Five cosmid clones (cos 9, cos 16, cos 28, cos 34 and cos 35) containing the entire coding region of the human hexokinase type I gene were used in sequencing and mapping the exon/intron junctions of the gene. The position of each cosmid clone on the structure of the gene was determined by *Eco*RI and *Eco*RV restriction analysis and the alignment of restriction fragments, as shown in Figure 1.

Sequencing of the five cosmid clones revealed that the gene contains 18 exons and 17 introns (Figure 1). The exons range in size from 63 to 305 bp. The sequences of the splice junctions were found to conform strictly to the GT/AG rule of splice junction sequences [34]. The introns were shown to be mostly class 0 and class II, although introns 2, 6, 10 and 14 were class I. Intron sizes ranged from 104 bp to more than 15 kb (Table 2). The lengths of introns 3–5, 8–10 and 12–17 were determined by gel electrophoresis of amplified fragments that contained an intron flanked by coding sequences. The lengths of introns 6 (104 bp) and 11 (314 bp) were determined by sequencing, whereas those of introns 1, 2 and 7 were estimated by restriction mapping. Introns 1 and 2 were longer than 15 and 10 kb respectively. Intron 7, 1.6 kb in length, was established by restriction mapping because this segment of the gene was difficult to amplify by PCR with the primers available.

The data therefore showed that the entire coding region was contained in a gene longer than 75 kb.

Nucleotide sequence of the 5′ flanking region of the human hexokinase type I gene

To isolate clones containing exon 1, the cosmid library was next screened with a probe corresponding to exon 1 (nt +20 to +143 of the cDNA). Cosmid cos 35 was isolated and examined by direct sequencing. This cosmid, with a 35 kb insert, in addition to exon 1 contained the 5′ flanking region, intron 1, exon 2 and partial intron 2. To investigate the 5′ flanking region, cos 35 was subcloned in a 'vectorette' library and a set of PCR was

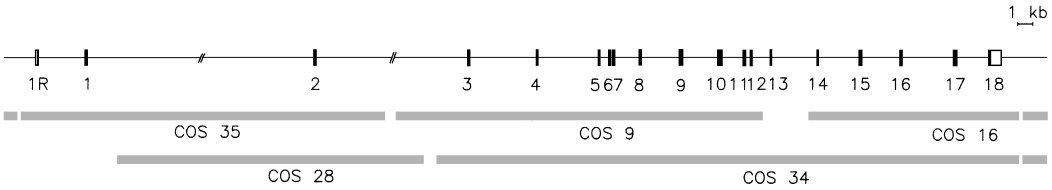

Figure 1 Structure of the human hexokinase type I gene

Exons are represented by boxes and introns by a thin line. Coding sequences are represented by black boxes; non-coding sequences are represented by white boxes. Box 1R represents the alternative red-blood-cell-specific exon 1. The breaks in introns 1 and 2 indicate that the length was not determined exactly. Overlapping YAC-derived cosmid clones are represented by horizontal bars.

Table 2 Characteristics of the exon/intron junctions of the human hexokinase type I gene

The sizes of each exon and intron, along with the exon/intron boundary sequences, are shown. Exons 1 and 1R represent respectively the first alternative exon for the ubiquitous isoform and for the red-blood-cell-specific isoform. Introns 1 and 1R represent the intron between exon 1 and exon 2 and the intron between red blood cell exon 1R and exon 2 respectively. In exon/intron boundaries, exon sequences are given in capitals and introns in lower cases. Note that each intron begins with a GT and ends with an AG. Amino acid residues are relative to the initiation codon [22] except for the amino acids at the junction between exon 1R and exon 2 [21]. The length of the coding sequences only is given for exons 1R, 1 and 18.

Exon			Intron				
No.	Size (bp)	5′ splice donor	No.	Size (kb)	Class	3′ splice acceptor	Amino acid(s) at junction
1R	60	GGG **gt**gagtagctgtggtccagggaagg	1R	> 18	0	ctccttccttctcatccccctcc**ag** ATT	Gly-Ile
1	63	AAG **gt**gagcccccgcccgcgccgccgct	1	> 15	0	ctccttccttctcatccccctcc**ag** ATT	Lys^{21}-Ile^{22}
2	163	CTG **gt**aagtctgtcacccagagattgaa	2	> 10	I	tgatacctgttgtctccctaaac**ag** AAA	Glu^{76}
3	149	CAG **gt**gggtccctgctccctccgggtca	3	4.4	0	ctgactgctctcatgtttccttc**ag** CTT	Gln^{125}-Leu^{126}
4	120	GAG **gt**aaggatgttctgggattatcggg	4	4	0	cagccccatccattcttctttgc**ag** GCC	Glu^{165}-Ala^{166}
5	96	GGG **gt**aatttctcctgggccctctgcct	5	0.6	0	atgacaccccgttatctgtcccc**ag** GAC	Gly^{197}-Asp^{198}
6	100	TCG **gt**aatgcattcccctttgcccatcc	6	0.104	I	tagcttctgatcttgtcctgccc**ag** GCA	Gly^{231}
7	184	GCT **gt**gagtccttgacttttgcttctaa	7	1.6	II	agtgccggtgtgcctttctccac**ag** GTT	Leu^{292}
8	156	AAA **gt**aggtaccatcccctcaaggcttt	8	2.5	II	tcagtattggcttctaacttcac**ag** GAA	Lys^{344}
9	234	ACA **gt**gagtctgccctttgctatcattg	9	2.3	II	ggaatgtccccccctgcccccata**ag** GTA	Gln^{422}
10	305	CCG **gt**gagggcctgctgggggctgacat	10	1.4	I	ttctttacgcttttgactgcaac**ag** AGA	Glu^{524}
11	149	GAG **gt**gagattacaaaaccatagtgcat	11	0.314	0	ggcgatcctttcttttcccctgc**ag** CTG	Glu^{573}-Leu^{574}
12	120	GCG **gt**gagtctctgttcttagggctcag	12	1.2	0	tctcctggtgctttttatgttgt**ag** GGA	Ala^{613}-Gly^{614}
13	96	GAG **gt**aactattaaaagaatgtttttta	13	3	0	aagctgtgtccctttctttgcaa**ag** GAA	Glu^{645}-Glu^{646}
14	100	TTG **gt**gagtgtcctggaaggtctctttc	14	2.7	I	ctcattgccccctcgtgtgttcc**ag** GGA	Gly^{679}
15	184	AAG **gt**aaccccgcctggtggagaggaca	15	2.5	II	ccccaggcccctcctcctgtctc**ag** GTA	Arg^{740}
16	156	GAG **gt**gagtgggcagtgtcttccctgcc	16	3.4	II	gccctctggctcttgtcccccac**ag** TGA	Ser^{792}
17	234	ACA **gt**gagtgggccttccagttgggatg	17	2.1	II	ccttctgttttctcgtccttttt**ag** CTT	His^{870}
18	142	– –	–	–	–	– –	–

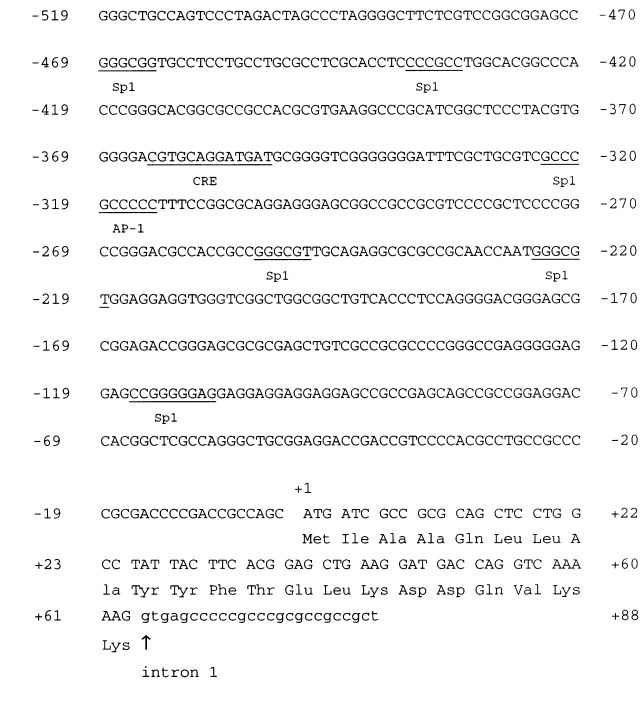

Figure 2 Sequence of hexokinase type I exon 1 and 5′ flanking region

Sequences of the 5′ flanking region of exon 1, exon 1 and 5′ splice junction of intron 1 are shown; +1 refers to the translation start point. The location of intron 1 is indicated by the vertical arrow. Underlined sequences refer to several potential Sp1-binding sites, AP-1 and CRE *cis*-acting elements.

performed to isolate the upstream region of the hexokinase type I gene, as described in the Experimental section. This strategy was necessary because of the difficulty involved in performing direct sequencing. A 2.3 kb PCR product was isolated and partly sequenced. The 438 nt fragment sequenced (Figure 2) had a base composition of 77.4% G/C and 22.6% A/T.

Red-blood-cell-specific exon 1

The human ubiquitous hexokinase type I cDNA [22] and human red blood cell-specific cDNA [21] differ at the 5′ untranslated sequence (UTR) and in the coding region respectively at the first 63 and 60 nt. Comparison between the red blood cell-specific cDNA and the structure of the gene revealed that the tissue-specific cDNA displays an alternative exon 1. This alternative exon 1 was located approx. 3.1 kb upstream from the somatic exon 1 and the splice junction was determined by direct sequencing (Figure 1).

DISCUSSION

The structures of the rat and human hexokinase type II [9,10,12,13] and glucokinase [14–17] genes are known. Furthermore, after the completion of the studies reported here, the structure of the rat hexokinase type I gene was also reported [18]. In addition, the cDNA sequences of bovine [11], murine [35], rat [36] and human [22] hexokinase type I, as well as rat and human types II and III, are available [37–40].

In the present study the structural organization of the human hexokinase type I gene was determined by analysis of cosmid clones and by PCR amplification across introns. It was found that the gene spans at least 75 kb and comprises 18 exons.

A comparison between the human hexokinase type I gene and human hexokinase type II gene [12,13] shows that the intron/exon structures are identical. In fact, the coding sequences of both genes are interrupted at identical nucleotide positions (Table 3), confirming and supporting the prediction of Printz et al. [10] that all the genes for the mammalian isoenzymes of hexokinase have

Table 3 Comparison of exon sizes between human glucokinase, hexokinase type I and hexokinase type II

Exon sizes for the coding sequence of glucokinase (hGK) [14,16,17], hexokinase type I (hHKI) and hexokinase type II (hHKII) [12,13] genes are depicted. The hHKI and hHKII genes are divided into the half encoding the N-terminus of the enzyme (hHKI N and hHKII N) and the half encoding the C-terminus of the enzyme (hHKI C and hHKII C).

	Exon size (nt)					
Exon	hGK	hHKI N	hHKII N	hHKI C	hHKII C	Exon
1	45/48*	63†	63			
2	163	163	163	305 (142+163)	305 (142+163)	10
3	155	149	149	149	149	11
4	120	120	120	120	120	12
5	96	96	96	96	96	13
6	100	100	100	100	100	14
7	184	184	184	184	184	15
8	156	156	156	156	156	16
9	234	234	234	234	234	17
10	142	305 (142+163)	305 (142+163)	142	142	18

* 45 and 48 nt refer to the sizes of the islet-specific exon 1β and to the liver-specific exon 1L respectively [14].
† 63 nt refers to the ubiquitous exon 1.

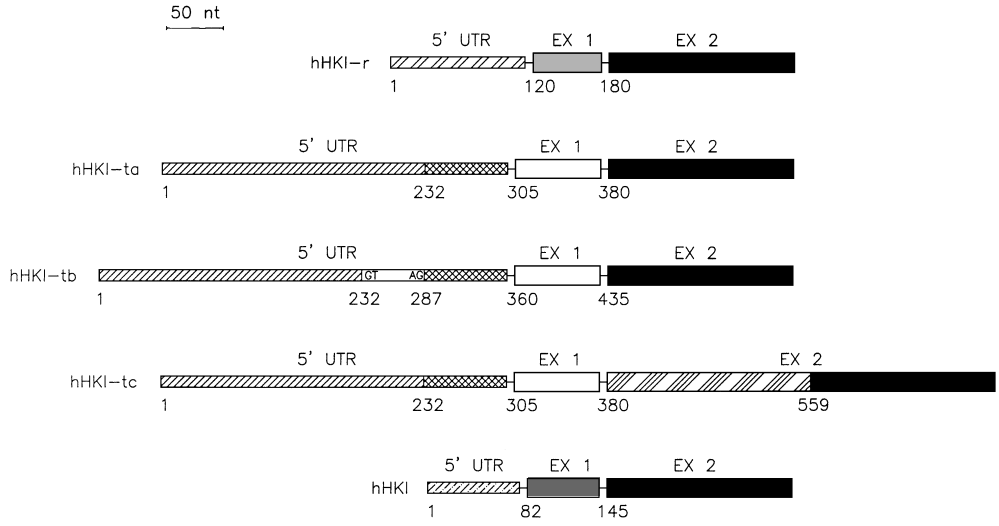

Figure 3 Comparison of somatic human hexokinase type I gene with human testis-specific and human red-blood-cell-specific hexokinase type I cDNA species

5′UTR sequences are represented by narrow boxes: the same hatching represents the same sequence. Wide boxes represent coding sequences: exon 1 (white), common to the testis-specific isoforms (hHKI-ta, hHKI-tb and hHKI-tc), and exon 1 (light grey) of the red-blood-cell-specific (hHKI-r) isoform differ from exon 1 (dark grey) of the somatic hexokinase type I (hHKI). Exon 2 (black) is common to all the isoforms, but in hHKI-ta, hHKI-tb and hHKI-r this sequence is read as in hHKI, whereas in hHKI-tc it is not read in frame because of a 179 bp (striped hatch) insertion before it. Numbers refer to the first nucleotide of each segment. In hHKI-tb the sequence 232–286 follows the starting (GT) and ending (AG) rule as an unspliced intron.

similar organizations. In terms of the intron size, it can be observed that introns 1 and 2 of hexokinase types I and II are very large. Most of the other introns differ in size. With the exception of introns 7 and 12, which are longer in the hexokinase type II gene, hexokinase type I introns are similar in length or longer than those of hexokinase type II.

Comparison of the structure of the rat hexokinase type I gene [18,41] and the structure of the human hexokinase type I gene showed that there are no significant differences between these genes.

Our studies also support the hypothesis [8] that hexokinase type I evolved from hexokinase type II, the form that most closely resembles an ancestral 100 kDa hexokinase that arose as a result of the duplication and fusion of a 50 kDa precursor. In fact, the hexokinase type I gene probably evolved by means of gene mutations resulting in an increase in intron size and in a functional differentiation between the two halves of the enzyme, namely a regulatory domain in the N-terminus and a catalytic domain in the C-terminus [2,42–47].

Cosmid 35 contains the 5′ flanking region of the hexokinase type I gene; the preliminary data reported here for this region will be further investigated in the future. Computer analysis with the BLASTN program [48] revealed identities ranging from 75% (nt −251 to −135) to 90% (nt −381 to −350) with the rat hexokinase type I promoter (results not shown). The 5′ flanking region sequences line up starting from nt −135 (human hexokinase type I) and nt −137 (rat hexokinase type I), assuming as +1 the translational start point. No TATA motif was found in this region, but this is not surprising because the absence of the TATA motif has been reported [49,50] to be a characteristic of 'housekeeping genes'. Analysis of the 5′ flanking sequence with TESS software [51] showed various hypothetical sites for Sp1,

AP-1 (activator protein-1) and a cAMP response element-binding protein (CRE-binding protein) (Figure 2).

The recent description of a red-blood-cell-specific hexokinase isoform lacking the porin-binding domain [21] led us to further investigate the structure of the gene for the presence of an alternative exon 1 in order to confirm the presence of an alternative splicing to produce this isoform. This alternative exon 1 was found to be upstream from the 5′ flanking region (Figure 1).

Three human testis-specific hexokinase type I mRNA species (hHKI-ta, hHKI-tb and hHKI-tc) lacking the porin-binding domain were recently submitted to GenBank (accession numbers U38226, U38227 and U38228). These isoforms have in common the 5′UTR and the first 75 bp of the coding region, whereas the hHKI-tb form differs from the others by an additional 55 bp in the 5′UTR. Downstream of the 75 bp of the coding region, hHKI-ta and hHKI-tb have in common with the somatic human hexokinase type I cDNA the sequence starting from nt +145 of cDNA reported by Nishi et al. [22]. From nt +145 onwards, this sequence is also present in hHKI-tc but is preceded by a 179 bp segment unique to that form (Figure 3).

By comparing these testis-specific mRNA species with the structure of the hexokinase type I gene, it can be observed that the sequence shared by hHKI-ta (from nt +380), hHKI-tb (from nt +435) and the somatic hexokinase type I (from nt +145) corresponds to exon 2 of the gene. Thus the first 75 bp of the coding region of hHKI-ta and hHKI-tb represent an alternative exon 1. hHKI-tc also presents the same alternative exon 1 but, because of an additional 179 bp, exon 2 is not read in frame as in the other isoforms. Mori et al. [52] suggest that this isoform is not translated because of stop codons.

The study of the exon/intron organization and the sequences of their splice junctions permits an understanding of potential alternative splicing events that are likely to occur during expression of the hexokinase type I gene. The possibility of finding an alternative promoter region upstream from the 5′UTR of these isoforms could be helpful in understanding the regulation of tissue-specific gene expression. With regard to red blood cells, a number of erythroid-specific enzymes are produced by the use of alternative promoter/alternate exons [53–55].

The results reported in this paper should be useful in the diagnosis of human hexokinase type I mutations at the DNA level, which was hitherto impossible given the lack of information on the gene organization.

We thank Cinzia Sala and Daniela Toniolo of IGBE – CNR, DIBIT – HSR (Milan) for providing CEPH and ICI libraries and for helpful discussion. This work was supported by funding from CNR and M.U.R.S.T. F.A. was supported by an E.N.E.A. fellowship.

REFERENCES

1 Katzen, H. M. and Schimke, R. T. (1965) Proc. Natl. Acad. Sci. U.S.A. **54**, 1218–1225
2 Wilson, J. E. (1995) Rev. Physiol. Biochem. Pharmacol. **126**, 65–198
3 Grossbard, L. and Schimke, R. T. (1966) J. Biol. Chem. **241**, 3546–3560
4 Easterby, J. S. and O'Brien, M. J. (1973) Eur. J. Biochem. **38**, 201–211
5 Rose, I. A., Warms, J. V. B. and Kosow, D. P. (1974) Arch. Biochem. Biophys. **164**, 729–735
6 Holroyde, M. J. and Trayer, I. P. (1976) FEBS Lett. **62**, 215–219
7 Ureta, T. (1982) Comp. Biochem. Physiol. **71B**, 549–555
8 Tsai, H. J. and Wilson, J. E. (1996) Arch. Biochem. Biophys. **329**, 17–23
9 Kogure, K., Shinohara, Y. and Terada, H. (1993) J. Biol. Chem. **268**, 8422–8424
10 Printz, R. L., Koch, S., Potter, L. R., O'Doherty, R. M., Tiesinga, J. J., Moritz, S. and Granner, D. K. (1993) J. Biol. Chem. **268**, 5209–5219
11 Griffin, L. D., Gelb, B. D., Wheeler, D. A., Davison, D., Adams, V. and McCabe, E. R. B. (1991) Genomics **11**, 1014–1024
12 Malkki, M., Laakso, M. and Deeb, S. S. (1994) Biochem. Biophys. Res. Commun. **205**, 490–496
13 Printz, R. L., Ardehali, H., Koch, S. and Granner, D. K. (1995) Diabetes **44**, 290–294
14 Iynedjian, P. B. (1993) Biochem. J. **293**, 1–13
15 Magnuson, M. A., Andreone, T. L., Printz, R. L., Koch, S. and Granner, D. K. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 4838–4842
16 Stoffel, M., Froguel, P., Takeda, J., Zouali, H., Vionnet, N., Nishi, S., Weber, I. T., Harrison, R. W., Pilkis, S. J., Lesage, S. et al. (1992) Proc. Natl. Acad. Sci. U.S.A. **89**, 7698–7702
17 Tanizawa, Y., Matsutani, A., Chiu, K. C. and Permutt, M. A. (1992) Mol. Endocrinol. **6**, 1070–1081
18 White, J. A. and Wilson, J. E. (1997) Arch. Biochem. Biophys. **343**, 207–214
19 Mori, C., Welch, J. E., Fulcher, K. D., O'Brien, D. A. and Eddy, E. M. (1993) Biol. Reprod. **49**, 191–203
20 Visconti, P. E., Olds-Clarke, P., Moss, S. B., Kalab, P., Travis, A. J., de las Heras, M. and Kopf, G. S. (1996) Mol. Reprod. Dev. **43**, 82–93
21 Murakami, K. and Piomelli, S. (1997) Blood **89**, 762–766
22 Nishi, S., Seino, S. and Bell, G. I. (1988) Biochem. Biophys. Res. Commun. **157**, 937–943
23 Magnani, M., Bianchi, M., Casabianca, A., Stocchi, V., Daniele, A., Altruda, F., Ferrone, M. and Silengo, L. (1992) Biochem. J. **285**, 193–199
24 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., Cold Spring Harbor Laboratory, Cold Spring Harbor, NY
25 Bianchi, M. and Magnani, M. (1995) Blood Cells Mol. Dis. **21**, 2–8
26 Sanger, F., Nicklen, S. and Coulson, A. R. (1977) Proc. Natl. Acad. Sci. U.S.A. **74**, 5463–5467
27 Albertsen, H. M., Abderrahim, H., Cann, H. M., Dausset, J., le Paslier, D. and Cohen, D. (1990) Proc. Natl. Acad. Sci. U.S.A. **87**, 4256–4260
28 Anand, R., Riley, J. H., Butler, R., Smith, J. C. and Markham, A. F. (1990) Nucleic Acids Res. **18**, 1951–1956
29 Bassam, B. J. and Caetano-Anolles, G. (1993) Biotechniques **14**, 30–34
30 Brownstein, B. H., Silverman, G. A., Little, R. D., Burke, D. T., Korsmeyer, S. J., Schlessinger, D. and Olson, M. V. (1989) Science **244**, 1348–1351
31 Nelson, D. L. and Brownstein, B. H. (1994) YAC Libraries: A User's Guide (Nelson, D. L. and Brownstein, B. H., eds.), pp. 69–70, W. H. Freeman, New York
32 Cheng, S., Chang, S. Y., Gravitt, P. and Respess, R. (1994) Nature (London) **369**, 684–685
33 Riley, J., Butler, R., Ogilvie, D., Finniear, R., Jenner, D., Powell, S., Anand, R., Smith, J. C. and Markham, A. F. (1990) Nucleic Acids Res. **18**, 2887–2890
34 Mount, S. M. (1982) Nucleic Acids Res. **10**, 459–472
35 Arora, K. K., Fanciulli, M. and Pedersen, P. L. (1990) J. Biol. Chem. **265**, 6481–6488
36 Schwab, D. A. and Wilson, J. E. (1989) Proc. Natl. Acad. Sci. U.S.A. **86**, 2563–2567
37 Thelen, A. P. and Wilson, J. E. (1991) Arch. Biochem. Biophys. **286**, 645–651
38 Deeb, S. S., Malkki, M. and Laakso, M. (1993) Biochem. Biophys. Res. Commun. **197**, 68–74
39 Schwab, D. A. and Wilson, J. E. (1991) Arch. Biochem. Biophys. **285**, 365–370
40 Furuta, H., Nishi, S., Le Beau, M. M., Fernald, A. A., Yano, H. and Bell, G. I. (1996) Genomics **36**, 206–209
41 White, J. A., Liu, W. and Wilson, J. E. (1996) Arch. Biochem. Biophys. **335**, 161–172
42 Nemat-Gorgani, M. and Wilson, J. E. (1986) Arch. Biochem. Biophys. **251**, 97–103
43 Schirch, D. M. and Wilson, J. E. (1987) Arch. Biochem. Biophys. **254**, 385–396
44 White, T. K. and Wilson, J. E. (1989) Arch. Biochem. Biophys. **274**, 375–393
45 Arora, K. K., Filburn, C. R. and Pedersen, P. L. (1993) J. Biol. Chem. **268**, 18259–18266
46 Baijal, M. and Wilson, J. E. (1992) Arch. Biochem. Biophys. **298**, 271–278
47 White, T. K. and Wilson, J. E. (1987) Arch. Biochem. Biophys. **259**, 402–411
48 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. and Lipman, D. J. (1990) J. Mol. Biol. **205**, 403–410
49 Gardiner-Garden, M. and Frommer, M. (1987) J. Mol. Biol. **196**, 261–282
50 Lowy, D. R. and Willumsen, B. M. (1993) Annu. Rev. Biochem. **62**, 851–891
51 Schug, J. and Overton, G. C. (1997) Transcriptional Element Search Software. Technical Report CBIL-TR-1997-1001-v0.0, Computational Biology and Informatics Laboratory, School of Medicine, University of Pennsylvania, U.S.A.
52 Mori, C., Nakamura, N., Welch, J. E., Shiota, K. and Eddy, E. M. (1996) Mol. Reprod. Dev. **44**, 14–22
53 Iacronique, V., Boquet, D., Lopez, S., Kahn, A. and Raymondjean, M. (1992) Nucleic Acids Res. **20**, 5669–5676
54 Kanno, H., Fujii, H. and Miwa, S. (1992) Biochem. Biophys. Res. Commun. **188**, 516–523
55 Pietrini, G., Agguajaro, D., Carrera, P., Malyszko, J., Vitale, A. and Borgese, N. (1992) J. Cell. Biol. **117**, 975–986

Received 3 November 1997/22 December 1997; accepted 12 January 1998