GENOMICS **44,** 8–14 (1997)
ARTICLE NO. GE974822

Genomic Organization of Two Novel Genes on Human Xq28: Compact Head to Head Arrangement of *IDHγ* and *TRAPδ* Is Conserved in Rat and Mouse

Volker Brenner, Gerald Nyakatura, André Rosenthal,[1] and Matthias Platzer

Institut für Molekulare Biotechnologie, Genomanalyse, Beutenbergstrasse 11, 07745 Jena, Germany

Received December 23, 1996; accepted May 16, 1997

In this paper we present the entire genomic sequence as well as the cDNA sequence of two new human genes encoding the γ subunit of the NAD$^+$-dependent isocitrate dehydrogenase (H-IDHγ) and the translocon-associated protein δ subunit (TRAPδ). These genes are located on region q28 of the human X chromosome, approximately 70 kb telomeric to the *adrenoleukodystrophy* locus (*ALD*). The sequences of the transcripts of both genes were obtained by searching the EST database with genomic data. Identified ESTs were completely sequenced and assembled to cDNAs comprising the entire coding region. For *IDHγ*, several EST clones indicate differential splicing. *IDHγ* and *TRAPδ* are arranged in a compact head to head manner. The nontranscribed intergenic region represents only 133 bp and is embedded in a CpG island. The CpG island obviously functions as a bidirectional promoter to initiate the transcription of both functionally unrelated genes with quite distinct expression patterns. This exceptional gene arrangement prompted us to clone and sequence genomic DNA fragments containing the homologous intergenic regions of rat and mouse. We show that in both species this area is similarly compact and represents less than 249 bp in rat and not more than 164 bp in mouse. In both cases this intergenic region is embedded in a CpG island and is highly conserved with nucleotide identity values ranging from 70.1% between human and rat to 92.6% between mouse and rat.

INTRODUCTION

The human genome contains approximately 45,000 CpG islands (Antequera and Bird, 1993). In both human and mouse species all known housekeeping genes are associated with CpG islands. A large proportion of genes with a tissue-restricted expression pattern also contain CpG islands. An interesting feature of CpG islands is that at least some are capable of initiating transcription in both directions (Faranda *et al.*, 1996; Shinya and Shimada, 1994; Wright *et al.*, 1995).

The two human genes described in this article encode the γ subunit of the NAD$^+$-dependent isocitrate dehydrogenase (H-IDHγ) and the δ subunit of the translocon-associated protein (TRAPδ). The gene products show different cellular localizations and are involved in different biochemical pathways. H-IDHγ is a subunit of a heterotetrameric enzyme that is located in mitochondria and presumed to play a major role in the oxidative decarboxylation of isocitrate in the tricarboxylic acid cycle (Nichols *et al.*, 1993). TRAPδ forms with three other subunits a protein complex, which is located at a specific site on the membrane of the endoplasmic reticulum, the so-called translocon, where nascent secretory proteins enter the ER lumen (Hartmann *et al.*, 1993).

We have obtained the complete genomic sequence of human IDHγ and TRAPδ in the course of a large-scale sequencing project covering about 400 kb of the region q28 on the X chromosome (M. Platzer *et al.*, in preparation). This region is one of the areas of the human genome with the highest gene density, approximately 1 gene every 20 kb, and is extremely G+C-rich (De Sario *et al.*, 1996). We present in this paper the genomic organization of the human *IDHγ* and *TRAPδ* genes as an example of an exceptional compact genomic arrangement. We have identified the corresponding cDNA clones by exploiting the EST database. The full-length coding sequences of both human genes and of mouse *IDHγ* were obtained by the resequencing and assembling of different EST clones. For comparative analysis of the intergenic region the syntenic loci in mouse and rat were amplified by genomic PCR and sequenced.

MATERIALS AND METHODS

Genomic sequencing. The human cosmid clones 10E8 and E153 were sequenced using a combination of the shotgun approach and

Sequence data from this article have been deposited with the EMBL Data Library under Accession Nos. U52111, U52112, Z68907, Z69043, U68564, U63009, U73205, U69268, and U69269.

[1] To whom correspondence should be addressed at Institut für Molekulare Biotechnologie, Abteilung für Genomanalyse, Beutenbergstrasse 11, 07745 Jena, Germany. Telephone: 0049-3641-656242. Fax: 0049-3641-656255. E-mail: arosenth@imb-jena.de.

0888-7543/97 $25.00

TABLE 1

Primers Used for Nested and Seminested PCR

Name	Sequence	Application
Rat1.fwd	5′-GAA CAT GCG GAG CAG AC	Outer forward primer for rat and mouse
Rat2.fwd	5′-ACT GAC GTG TCC GTT TC	Inner forward primer for rat
Rat3.rev	5′-TAT TGC CTT TGC AGC AC	Inner reverse primer for rat and outer reverse primer for mouse
Rat4.rev	5′-GGA CGG CAG AGG AGA GC	Outer reverse primer for rat
Mouse1.fwd	5′-TCG CTT CCC TAG CCA GGG TC	Inner forward primer for mouse
Mouse2.rev	5′-GCA CGT ATC GCC ACC TTC AGC	Inner reverse primer for mouse

primer walking strategy essentially as described by Craxton (Griffin and Griffin, 1993). The data were collected on Applied Biosystems 373A automated sequencers (Perkin–Elmer) and assembled using the XGAP program (Bonfield *et al.,* 1995).

Resequencing of EST clones. Plasmid DNA of I.M.A.G.E. cDNA clones (IMAGE Consortium), which were acquired from the Reference Library–Database (Lehrach *et al.,* 1990; Lennon *et al.,* 1996), was prepared with QIAwell 8 (Qiagen) according to the manufacturer's instructions. The sequence of the plasmid inserts was determined with the flanking primers T7, T3, and Sp6 and additional custom-made walking primers as described above.

Primer extension analysis. The primer extension experiments were carried out with radioactively labeled oligonucleotides of the sequence 5′-GCA AAC GAG GAA GAG CGA CG and 5′-CCA TCG CCT CTT CTC TGC to determine the transcription start site of *TRAPδ* and with the primers 5′-CGA AGT TTC GGG GAC AGG CG and 5′-CAC GCA GAT ACC GCT CTC GC designed for *IDHγ*. Superscript reverse transcriptase (GIBCO BRL) was used on total RNA at 40 and 42°C (Ausubel *et al.,* 1988). The products were loaded on a 6% polyacrylamide/8 *M* urea sequencing gel. The extended products were detected by a Fujix Imaging Analyzer BAS1000.

RT-PCR. The human cDNA library panel was obtained from Clontech. PCR was performed according to the manufacturer's protocol with primer A (5′-TAG GAC ACA GGC AGG CTC GG) located in the alternatively spliced region and primer B (5′-TGG CCA TGT GTA CGC GGT G) specific for exon 10 of human *IDHγ*. For the seminested PCR, primer C (5′-GCA AGA GTA TCG CCA ATA AG) complementary to *IDHγ* exon 11 was used in combination with primer A. Obtained products were extracted from agarose gels and purified with micropure separators (Amicon) and sequenced directly with the PCR primers using dye-terminator chemistry (Perkin–Elmer).

Northern blot analysis. The human multiple Northern blot obtained from Clontech was successively probed with ^{32}P-labeled cDNAs of *TRAPδ*, *IDHγ*, and *β-actin* generated by random priming according to the manufacturer's protocol (Boehringer Mannheim). The signal intensities were quantified on a Fujix Imaging Analyzer BAS1000 with TINA 2.08e software. To account for the variations of the amounts of RNA in individual lanes of the blot, the ratio of probe-specific signal to the *β-actin* control for each lane was used for evaluating the relative signal intensities. In each case minimal expression levels of the respective gene were arbitrarily defined as unity. The muscle-specific cross-hybridization of *β-actin*, which appears as a thick additional band on the muscle lane (Fig. 2A), was excluded from the evaluation as it lies beyond the region of the specific signal.

PCR amplification. Genomic template DNA was extracted from mouse and rat whole blood or buffy coat using the QIAamp Blood kit (Qiagen) according to the manufacturer's protocol. The PCR was performed with 25–100 ng of genomic DNA, 10 units of Thermosequenase (Amersham), 25 pmol of each primer (Table 1), 4 m*M* $MgCl_2$, and each dNTP at a final concentration of 0.35 m*M* in a total volume of 50 μl. Several independent PCR products were directly sequenced as described above.

RESULTS

Genomic Sequencing in Human Xq28 Revealed Two New Genes 70 kb Telomeric to ALD

Large-scale genomic sequencing is a fast and efficient approach to identify genes in chromosomal regions between several hundred kilobases and 1–2 Mb. We have sequenced a 400-kb cosmid contig (Rogner *et al.,* 1994) spanning the region between the *adrenoleukodystrophy* locus (*ALD*) and the locus encoding the methyl-CpG-binding protein (*MeCP2*) in Xq28. The annotated sequence was submitted to EMBL Database (Accession Nos. U52111 and U52112). Gene prediction using the GRAIL program revealed among others two exon clusters in particularly compact head to head arrangement. The predicted exons were used to search the EMBL Database using BLAST/FASTA programs. The exon clusters were identified as the human *IDHγ* and *TRAPδ* genes on the basis of the EST hits and homologous rat cDNAs. Alignments of the rat cDNA sequences against the human genomic sequences enabled the preliminary identification of the exon–intron boundaries in both genes. The boundaries were subsequently confirmed by human cDNA sequences. The entire transcribed regions of *IDHγ* and *TRAPδ* span 13,032 bp on the cosmid clones 10E8 and E153 (Fig. 1) and have a G+C content of 59%. This value is very high when compared to the genome-wide average, but is characteristic of this particularly G+C-rich region of Xq28 (M. Platzer, in preparation).

cDNA Sequences of IDHγ and TRAPδ Obtained from EST Database Entries

Searching the public databases with predicted exons of the human *IDHγ* gene revealed 42 human EST database entries, 16 mouse ESTs, 3 rat ESTs, and the rat *IDHγ* cDNA sequence (Accession No. X74125). The predicted *TRAPδ* exons matched 65 human ESTs, 21 mouse ESTs, 1 rat EST, and the mouse and rat *TRAPδ* cDNA sequences (Accession Nos. X90582 and Z19087). To obtain cDNA sequences covering the complete coding region of human and mouse *IDHγ*, we completely sequenced and assembled 3 human EST clones (Accession Nos. R37343, R60496, and T49923) and 3 mouse EST clones (Accession Nos. W11042, W12212, and

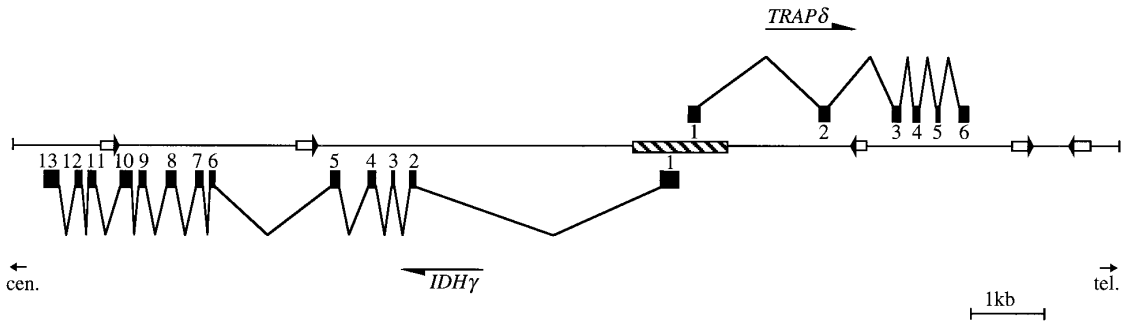


FIG. 1. Schematic representation of the genomic organization of the human *IDHγ* and *TRAPδ* genes. Exons are indicated by black boxes, the CpG island as a striped box, and repetitive elements as arrows.

W36405). Similarly for *TRAPδ*, we sequenced and assembled 3 human EST clones (Accession Nos. H04881, T54362, and T85291). In this way three cDNA sequences comprising the full-length coding sequence were obtained and deposited in GenBank (Accession Nos. Z68907, U68564, and Z69043). The final consensus sequences of human *IDHγ* and *TRAPδ* showed no discrepancies with the genomic sequence. The EST database entries contained between 0.38 and 10.8% errors when compared to the final consensus sequences.

Genomic Structure and Differential Splicing of Human IDHγ

The human *IDHγ* gene is transcribed in centromeric direction. The genomic organization of the human *IDHγ* gene was determined by aligning the cDNA sequence with the genomic sequence using the global alignment program GAP (Huang, 1994) (Table 2). The gene has 13 exons with a coding sequence of 1182 nt and spans 8.7 kb of genomic DNA (Fig. 1). The exon–intron boundaries of *IDHγ* are in accordance with the GT/AG rule.

TABLE 2

Genomic Organization of Human *IDHγ*

	Exon		Intron		
No.	First base	Length	First base	Length	Repetitive elements/length
1	107543	259	107284	3394	
2	103890	42	103848	234	
3	103614	12	103602	271	
4	103331	98	103233	370	
5	102863	113	102750	1576	*Alu*-Sg/288
6	101174	61	101113	119	
7	100994	133	100861	225	
8	100636	134	100502	310	
9	100192	103	100089	103	
10	99986	147	99840	351	L1MB7/215
11	99488	95	99393	82	
12	99311	61	99250	238	
13	99012	205			
Average		112.5		606	

Note. The positions (second and fourth columns) are as in the database entry Accession No. U52111. The repetitive elements were identified using the CENSOR program (Jurka *et al.,* 1996).

We confirmed the completeness of the submitted human cDNA sequence by determining the transcription start site for *IDHγ* gene using primer extension assays (data not shown). The rat *IDHγ* mRNA database entry seems to lack the 5′ end, including 15 bp of the coding sequence, and the start codon. This region and the 13-nt 5′ UTR are covered by an EST clone of *Rattus* sp. (Accession No. H34632), the sequence of which is 100% identical to the genomic sequence we derived from *Rattus norvegicus* (Accession No. U63009).

To date no alternative splicing of the *IDHγ* gene transcript has been reported. To establish conclusively the absence or presence of alternative splicing events of the *IDHγ* gene product, we performed a database search using exclusively intronic sequences. We detected three alternatively spliced ESTs originating from the same human cDNA library 1NFLS constructed by Soares and Bonaldo from liver and spleen of a 20-week-postconception male fetus. In one clone (Accession No. H69258) an alternative splice acceptor site 36 bp upstream of the major acceptor site of exon 13 is used. The deduced protein contains 12 additional amino acids. In two other clones (Accession Nos. H60000 and H47588) intron 12 is not spliced out. In the deduced protein 33 amino acids of the C-terminus of H-IDHγ are replaced by 20 other residues. The first 360 amino acids are identical in all three deduced protein sequences. Using RT-PCR we could observe transcripts showing an unprocessed intron 12 but with a processed intron 11 in all examined tissues (heart, brain, placenta, lung, liver, skeletal muscle, kidney, and pancreas). Utilization of the alternative splice acceptor site of exon 13 was detected only in heart, placenta, lung, and skeletal muscle.

Modeling of the three-dimensional structure of IDHγ using the Swiss-Model ProServer, Version 1.1 (Peitsch *et al.,* 1996), showed that both the NAD-binding domain (residue 164 to 309; Branden and Tooze, 1991) and the putative ATP/GTP-binding site (residue 309 to 316; Devereux *et al.,* 1984) are not affected by the alternative splice events.

TABLE 3

Genomic Organization of *TRAPδ*

Exon			Intron		
No.	First base	Length	First base	Length	Repetitive elements/length
1	107677	117	107794	1679	
2	109473	119	109592	905	*Alu*-Sxzy/157
3	110497	75	110572	192	
4	110764	90	110854	256	
5	111110	66	111176	192	
6	111368	170			*Alu*-Jb, *Alu*-Sq, and *Alu*-Sz/267, 292, and 291
Average		106.2		644.8	

Note. The positions (second and fourth columns) are as in the database entry Accession No. U52111. The repetitive elements indicated in line 6 are located 0.5 to 3 kb distal to the gene.

Genomic Structure of TRAPδ

The genomic structure of the human *TRAPδ* gene was determined using methodology analogous to that described above for the *IDHγ* gene. The gene extends over 3.8 kb of genomic DNA, consists of six exons, and has a coding sequence of 522 nt (Fig. 1). The gene is transcribed in the direction opposite to that of *IDHγ*. All six exons are protein coding and vary in size from 66 to 170 bp (Table 3). The donor and acceptor splice sites show the consensus dinucleotides GT and AG. The first intron (1679 bp), analogous to the first intron of human *IDHγ*, is the largest intronic region of the gene and is also free of known interspersed repetitive elements. We were unable to map a distinct transcription start site for *TRAPδ*. This negative result might reflect secondary structures of the mRNA due to the high G+C content of this area or might be an indication of variable start sites. The polyadenylation signal is located 34 nt downstream of the stop codon and 32 nt upstream of the major polyadenylation site. An alternative polyadenylation site of human *TRAPδ*, which is located 27 nt downstream of the polyadenylation signal, was found in the EST clone with Accession No. T54362.

Alternatively spliced cDNAs of *TRAPδ* missing the region homologous to the human exon 5 have been described for *Xenopus laevis* and *Mus musculus* (Accession No. X90584 and X90582; Holthuis *et al.*, 1995). We were unable to detect alternative splicing of exon 5 in 42 human EST clones.

Human IDHγ and TRAPδ Show Different Tissue-Specific Expression Patterns

To examine the expression pattern of *IDHγ* and *TRAPδ* genes we performed Northern blot analysis with RNA from eight different human tissues (Fig. 2). The estimated sizes of 1.5 kb for *IDHγ* and 0.7 kb for *TRAPδ* are in agreement with the cDNA sequence data. Both genes are expressed in all examined tissues. The tissue-specific expression levels vary considerably (Figs. 2C and 2D).

Compact Head to Head Arrangement of IDHγ and TRAPδ Is Conserved in Rat and Mouse

To find out whether this strikingly compact intergenic region between the human *IDHγ* and *TRAPδ* genes is a unique feature of these human loci or if the same arrangement is conserved in other mammals, we examined the homologous regions in rat and mouse and submitted them to GenBank (Accession Nos. U73205 for mouse and U63009 for rat). The results show that both the gene order and the compact head to head arrangement are conserved in all three organisms (Fig. 3A). The nontranslated intergenic region is a stretch of 361 bp in human, 392 bp in rat, and 384 bp in mouse. Homology is less pronounced in the noncoding region, but still exceeding identity values of 70% (Fig. 3B). Only 6 bp of the genomic sequence (position 292 to 298 in Fig. 3A) are not represented in the cDNA sequences of any of the three species.

CpG Islands and Predicted Transcription Factors in Human, Rat, and Mouse

Computer analysis of the human genomic sequence using the GRAIL network client for the detection of CpG islands (Überbacher and Mural, 1991) revealed a region of 1317 bp (position 106863 to 108179) as CpG island. The island contains the first exons, the start sites for transcription and for translation, as well as parts of the first introns of both the *IDHγ* and *TRAPδ* genes (Fig. 1). The G+C content reaches 68% in human, 62% in rat, and 61% in mouse. There are no obvious TATA-box and CAAT-box promoter elements in human, rat, or mouse, as is expected for CpG island regions of housekeeping genes (Sehgal *et al.*, 1988). In the human 509-bp fragment containing the first exons of both genes as well as the intergenic region no Sp1 site (Dynan *et al.*, 1985; Dynan and Tjian, 1983) was detected. In the homologous region of mouse 1 Sp1 site and in rat 2 Sp1 sites were found. Screening of intronic sequences revealed 12 Sp1 sites in the first intron of the human *IDHγ* gene and 3 Sp1 sites in the first intron

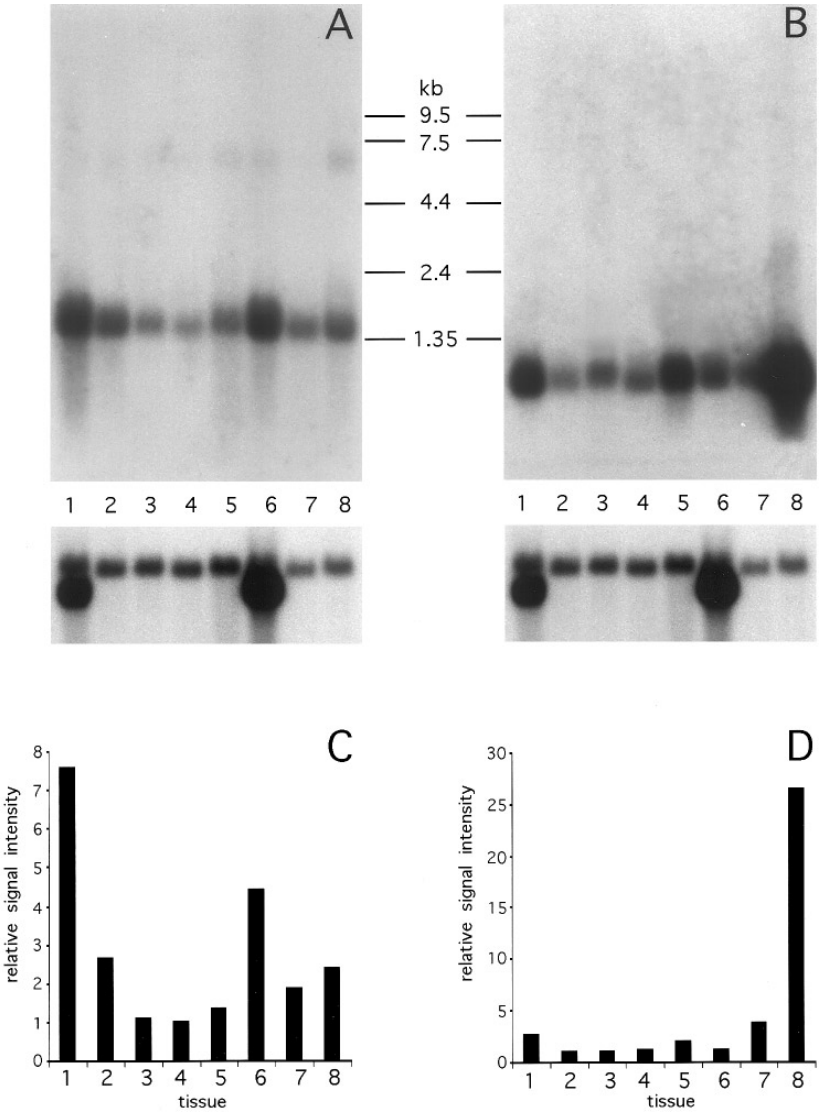


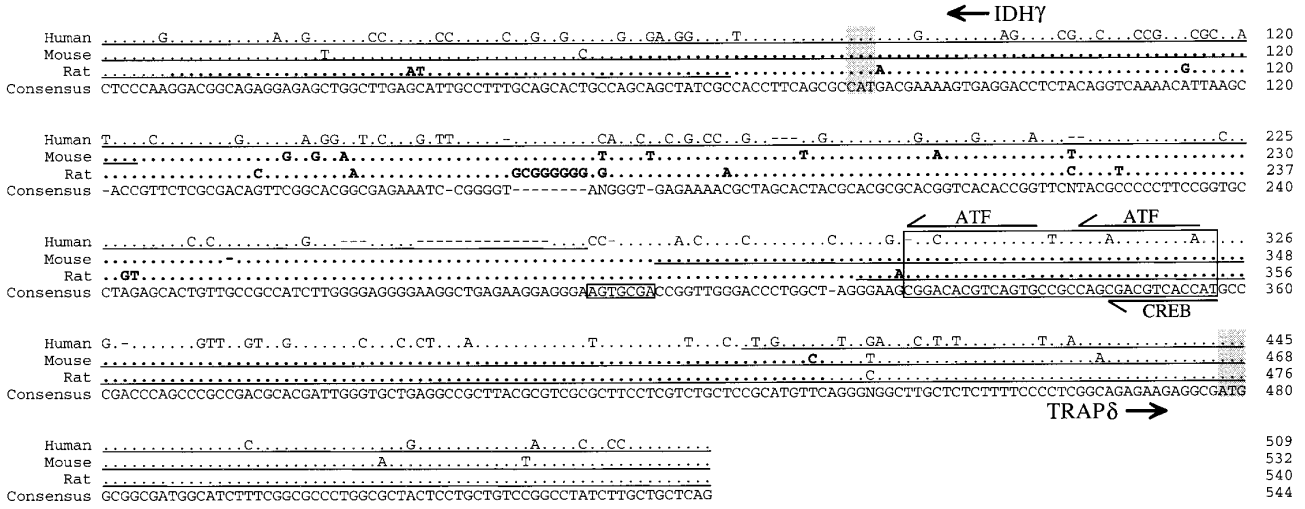
FIG. 2. Northern blot analysis of human *IDHγ* and *TRAPδ*. Ubiquitous expression of *IDHγ* (**A**) and *TRAPδ* (**B**). The *β-actin* control is indicated at the bottom. Lane **1,** heart; lane **2,** brain; lane **3,** placenta; lane **4,** lung; lane **5,** liver; lane **6,** skeletal muscle; lane **7,** kidney; and lane **8,** pancreas. Marker sizes are indicated in the middle. The histograms of normalized signal intensities are shown for *IDHγ* (**C**) and *TRAPδ* (**D**).

of human *TRAPδ* gene. Computer search for conserved transcription factor binding sites using the programs TSSG, TSSW (Solovyev *et al.*, in preparation), TFSEARCH (http://www.tokyo-center.genome.ad.jp/SIT/TFSEARCH.html), and Signal Scan (Prestridge and Stormo, 1993) revealed several putative mammalian motifs. In all three species the sequence of a putative cAMP response element (CRE) was detected within stretches of 32 or 33 nt with two predicted binding sites for the activating transcription factor (ATF) (Fig. 3A).

DISCUSSION

In this paper for the first time two genes that are driven by a bidirectional promoter region and that encode proteins that are involved in unrelated biochemical pathways located in different compartments of the cell are reported. Our results from Northern blot analysis show that their expression levels vary considerably in individual tissues. TRAP*δ* is assumed to be involved in the secretion of proteins. The highest concentration of *TRAPδ* transcripts was observed in pancreas, where large quantities of lipases, nucleases, and proteases are synthesized and secreted. The *IDHγ* gene product is involved in the energy metabolism of the cell. Correspondingly, the highest expression levels are in tissues with increased energy turnover, like heart, skeletal muscle, and brain. Upstream of the human *IDHγ* gene the intergenic region contains a CRE together with two putative binding sites for an ATF conserved through all three species. The cAMP-second messenger system could be a tool to regulate H-IDH at the transcriptional level to the energy needs of the cell. Tissues with higher activity and energy turnover are able to increase the

A



B

	cds of the first *IDH* exon	cds of the first *TRAP* exon	noncoding region between *IDH* and *TRAP*	both first exons and intergenic region
human vs. rat %Identity / Gaps	77.8 / 0	91 / 0	70.1 / 38	73.8 / 38
human vs.mouse %Identity / Gaps	77.8 / 0	88.1 / 0	71.4 / 31	74.3 / 31
mouse vs. rat %Identity / Gaps	95.1 / 0	97 / 0	92.6 / 10	93.2 / 10

FIG. 3. Comparison of the human, rat, and mouse promoter regions of *TRAPδ* and *IDHγ* genes. Alignment of the intergenic regions (**A**). The alignment includes the first exons of human *IDHγ* and *TRAPδ* and the homologous regions of the rat and mouse cDNAs. The PCR-amplified fragments of rat and mouse are indicated in boldface. Dots indicate nucleotides identical with the consensus sequence. Differing positions are shown as letters or as dashes in the case of deletions. Transcribed regions are underlined. The ATG start codons are shaded gray. The boxed consensus sequence shows the region that is not transcribed in any of the three species (position 292–298). The other boxed region with arrows (position 325–357) indicates the 33 nt containing a predicted cAMP response element (CRE) and two predicted binding sites for the activating transcription factor (ATF) in each species. Pairwise evaluation of the alignment indicating the percentages of identity and the number of gaps introduced by the alignment program MAP (**B**).

expression of the *IDHγ* gene. Our Northern blot data provide experimental evidence for this regulation of expression.

The shared CpG island of *IDHγ* and *TRAPδ* can be considered typical in terms of G+C content and size and in its property of harboring both first exons. The observed molar G+C contents of 61% in mouse, 62% in rat, and 68% in human are in agreement with the reported, significantly lower, G+C content of mouse CpG islands compared to those of human (Antequera and Bird, 1993). A striking feature of this particular CpG island is that it is practically devoid of Sp1 sites, which are common in other CpG islands.

The usefulness of the EST databases in interpreting genomic sequences is also noteworthy. Our results reaffirm the notion that genomic sequencing is complemented and that sequence analysis, gene identification, and annotation are facilitated by comprehensive EST databases. In the case of the *TRAPδ* and *IDHγ* genes, EST sequences in the databases were a powerful and time- and labor-saving tool for the identification of coding regions and exon–intron boundaries on genomic DNA and for the rapid determination of the cDNA sequences in human and mouse. Of 4100 known human transcripts, 72% are matched by ESTs (Adams *et al.*, 1995). Although the representation of genes in the EST database is assumed to be lower than 72%, probably about half of the human genes are covered. EST projects in different species make "phylum-hopping" among organisms possible (Tugendreich *et al.*, 1994).

Our results on the comparative analysis of two closely linked loci in three organisms demonstrate this. Another characteristic of the existing EST databases is the overrepresentation of housekeeping genes. We have exploited this property to detect relatively rare events, like alternative splicing and alternative polyadenylation. In the case of *IDH*γ, 3 clones of 42 EST database entries showed alternative splicing, resulting in two new isoforms of the IDHγ protein. We confirmed these events by RT-PCR.

We have demonstrated the conserved compact head to head arrangement for two genes with unrelated function in three phyla. Although the biological significance of this phenomenon remains to be elucidated, it is an interesting object for studying the synteny of genes in other species farther down on the evolutionary tree. It would be particularly challenging to include species like *Drosophila melanogaster* that do not have DNA methylation as a regulatory option. In combination with functional studies and probing of promoter activities, this region, handy in size as it is, will be helpful in the elucidation of the regulatory functions of CpG islands and will open up avenues to access hitherto unknown regulatory elements using a comparative approach.

ACKNOWLEDGMENT

This work was supported by a research grant from the Deutsche Forschungsgemeinschaft (PL 173/2-1).

REFERENCES

Adams, M. D., Kerlavage, A. R., Fleischmann, R. D., Fuldner, R. A., Bult, C. J., Lee, N. H., Kirkness, E. F., Weinstock, K. G., Gocayne, J. D., White, O., *et al.* (1995). Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature* **377:** 3–174.

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* **215:** 403–410.

Alun, T., and Skolnick, M. H. (1994). A probabilistic model for detecting coding regions in DNA sequences. *IMA J. Math. Appl. Med. Biol.* **11:** 149–160.

Antequera, F., and Bird, A. (1993). Number of CpG islands and genes in human and mouse. *Proc. Natl. Acad. Sci. USA* **90:** 11995–11999.

Ausubel, M. F., Brent, R., Kingston, R. E., Moore, D. D., Smith, J. A., Seidman, J. G., and Struhl, K. (1988). "Current protocols in Molecular Biology," p. 4.8, Greene Publ. and Wiley-Interscience, New York.

Benbrook, D. M., and Jones, N. C. (1994). Different binding specificities and transactivation variant CRE's by CREB complexes. *Nucleic Acids Res.* **22:** 1463–1469.

Bonfield, J. K., Smith, K., and Staden, R. (1995). A new DNA sequence assembly program. *Nucleic Acids Res.* **23:** 4992–4999.

Branden, C., and Tooze, J. (1991). "Introduction to Protein Structure," pp. 141–159, Garland, New York.

De Sario, A., Geigl, E. M., Palmieri, G., D'Urso, M., and Bernardi, G. (1996). A compositional map of human chromosome band Xq28. *Proc. Natl. Acad. Sci. USA* **93:** 1298–1302.

Devereux, J., Haeberli, P., and Smithies, O. (1984). A comprehensive set of sequence analysis programs for the VAX. *Nucleic Acids Res.* **12:** 387–395.

Dynan, W. S., Saffer, J. D., Lee, W. S., and Tjian, R. (1985). Transcription factor Sp1 recognizes promoter sequences from the monkey genome that are simian virus 40 promoter. *Proc. Natl. Acad. Sci. USA* **82:** 4915–4919.

Dynan, W. S., and Tjian, R. (1983). The promoter-specific transcription factor Sp1 binds to upstream sequences in the SV40 early promoter. *Cell* **35:** 79–87.

Faranda, S., Frattini, A., Zucchi, I., Patrosso, C., Milanesi, L., Montagna, C., and Vezzoni, P. (1996). Characterization and fine localization of two new genes in Xq28 using the genomic sequence/EST database screening approach. *Genomics* **34:** 323–327.

Griffin, G. G., and Griffin, A. M. (1993). Cosmid sequencing. *In* "DNA Sequencing Protocols" (J. M. Walker, Ed.), Vol. 23, pp. 149–168, Humana Press, Totowa, NJ.

Hartmann, E., Gorlich, D., Kostka, S., Otto, A., Kraft, R., Knespel, S., Burger, E., Rapoport, T. A., and Prehn, S. (1993). A tetrameric complex of membrane proteins in the endoplasmic reticulum. *Eur. J. Biochem.* **214:** 375–381.

Holthuis, J. C., van Riel, M. C., and Martens, G. J. (1995). Translocon-associated protein TRAP delta and a novel TRAP-like protein are coordinately expressed with pro-opiomelanocortin in Xenopus intermediate pituitary. *Biochem. J.* **312:** 205–213.

Huang, X. (1994). On global sequence alignment. *Comput. Appl. Biosci.* **10:** 227–235.

Jurka, J., Klonowski, P., Dagman, V., and Pelton, P. (1996). CENSOR—A program for identification and elimination of repetitive elements from DNA sequences. *Comput. Chem.* **20:** 119–122.

Lehrach, H., Drmanac, R., Hoheisel, J., Larin, Z., Lennon, G., Monaco, A. P., Nizetic, D., Zehetner, G., and Poustka, A. (1990). "Genetic and Physical Mapping" (K. E. T. Davies, and S. M., Eds.), Vol. 1, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Lennon, G., Auffray, C., Polymeropoulos, M., and Soares, M. B. (1996). The I.M.A.G.E. Consortium: An integrated molecular analysis of genomes and their expression. *Genomics* **33:** 151–152.

Nichols, B. J., Hall, L., Perry, A. C., and Denton, R. M. (1993). Molecular cloning and deduced amino acid sequences of the gamma-subunits of rat and monkey NAD(+)-isocitrate dehydrogenases. *Biochem. J.* **295:** 347–350.

Pearson, W. R., and Lipman, D. J. (1988). Improved tools for biological sequence comparison. *Proc. Natl. Acad. Sci. USA* **85:** 2444–2448.

Peitsch, M. C., Peitsch, M. C., Wells, T. N., Stampf, D. R., and Sussman, J. L. (1996). ProMod and Swiss-Model: Internet-based tools for automated comparative protein modelling. *Biochem. Soc. Trans.* **24:** 274–279.

Prestridge, D. S., and Stormo, G. (1993). SIGNAL SCAN 3.0: New database and program features. *Comput. Appl. Biosci.* **9:** 113–115.

Rogner, U. C., Kioschis, P., Wilke, K., Gong, W., Pick, E., Dietrich, A., Zechner, U., Hameister, H., Pragliola, A., Herman, G. E., *et al.* (1994). A YAC clone map spanning 7.5 megabases of human chromosome band Xq28. *Hum. Mol. Genet.* **3:** 2137–2146.

Sehgal, A., Patil, N., and Chao, M. (1988). A constitutive promoter directs expression of the nerve growth factor receptor gene. *Mol. Cell. Biol.* **8:** 3160–3167.

Shinya, E., and Shimada, T. (1994). Identification of two initiator elements in the bidirectional promoter of the human dihydrofolate reductase and mismatch repair protein 1 genes. *Nucleic Acids Res.* **22:** 2143–2149.

Tugendreich, S., Bassett, D. E., Jr., McKusick, V. A., Boguski, M. S., and Hieter, P. (1994). Genes conserved in yeast and humans. *Hum. Mol. Genet.* **3:** 1509–1517.

Überbacher, E. C., and Mural, R. J. (1991). Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proc. Natl. Acad. Sci. USA* **88:** 11261–11265.

Wright, K. L., White, L. C., Kelly, A., Beck, S., Trowsdale, J., and Ting, J. P. (1995). Coordinate regulation of the human TAP1 and LMP2 genes from a shared bidirectional promoter. *J. Exp. Med.* **181:** 1459–1471.