Vol. 16, No. 3

Human Hepatocyte Nuclear Factor 4 Isoforms Are Encoded by Distinct and Differentially Expressed Genes

THORSTEN DREWES, SABINE SENKEL, BEATRIX HOLEWA, AND GERHART U. RYFFEL*

Institut für Zellbiologie (Tumorforschung), Universitätsklinikum Essen, D-45122 Essen, Germany

Hepatocyte nuclear factor 4 (HNF4) was first identified as a DNA binding activity in rat liver nuclear extracts. Protein purification had then led to the cDNA cloning of rat HNF4, which was found to be an orphan member of the nuclear receptor superfamily. Binding sites for this factor were identified in many tissue-specifically expressed genes, and the protein was found to be essential for early embryonic development in the mouse. We have now isolated cDNAs encoding the human homolog of the rat and mouse HNF4 splice variant HNF4α2, as well as a previously unknown splice variant of this protein, which we called HNF4α4. More importantly, we also cloned a novel HNF4 subtype (HNF4γ) derived from a different gene and showed that the genes encoding HNF4α and HNF4γ are located on human chromosomes 20 and 8, respectively. Northern (RNA) blot analysis revealed that HNF4γ is expressed in the kidney, pancreas, small intestine, testis, and colon but not in the liver, while HNF4α RNA was found in all of these tissues. By cotransfection experiments in C2 and HeLa cells, we showed that HNF4γ is significantly less active than HNF4α2 and that the novel HNF4α splice variant HNF4α4 has no detectable transactivation potential. Therefore, the differential expression of distinct HNF4 proteins may play a key role in the differential transcriptional regulation of HNF4-dependent genes.

While intensive research during the last decade has established that tissue-specific gene expression is regulated to a large extent at the level of transcription, it has also become clear that most if not all of the tissue-specific transcription factors involved in this regulation are not restricted to a single tissue. Furthermore, the molecular cloning of these factors has shown that they are usually encoded by gene families with a multitude of subtypes and splice variants, and it is believed that it is the intra- and interfamily interplay of the cell-type-specific subset of these transcription factors that determines the identity of a certain tissue.

Analysis of the regulatory regions of liver-specifically expressed genes, for example, has led to the identification of three gene families of transcription factors (i.e., the C/EBP, hepatocyte nuclear factor 1 [HNF1], and HNF3 families) and of another factor (HNF4) for which up to now splice variants but no subtypes derived from different genes had been identified (for reviews, see references 13, 19, and 25). All of these transcription factors originally considered to be liver specific are now known to be expressed also in other organs, but their combined expression is a unique feature of the liver (24).

The factor HNF4 was originally identified in rat liver nuclear extracts as a protein binding to a DNA element of the transthyretin promoter (5). Protein purification and cDNA cloning revealed that HNF4 is an orphan member of the nuclear receptor superfamily with a zinc finger DNA binding domain and a putative ligand binding domain (20). Binding sites for HNF4 have been found in the regulatory regions of many genes expressed in the liver and kidney, including the gene for the transcription factor HNF1α (reviewed in reference 18). HNF4 mRNA was detected in the liver, kidney, intestine, and pancreas (14, 20), and homozygous deletion of the gene encoding HNF4 demonstrated that HNF4 plays an important role in early embryogenesis (3).

Studying the role of tissue-specific transcription factors in tumorigenesis, we found that loss of the transcription factor HNF1α (LFB1 in reference 4) is a consistent feature of human kidney tumors (4). Because expression of the gene encoding HNF1α is controlled by HNF4 (12, 23, 26), we wanted to isolate cDNA probes for human HNF4 to analyze HNF4 transcription factors in these tumor samples (17). Since the isolation of cDNAs coding for HNF4 transcription factors from *Xenopus laevis* had revealed the existence of two HNF4 genes (HNF4α and HNF4β) in this species (10), we screened two human kidney cDNA libraries with a *Xenopus* HNF4 cDNA to identify the HNF4 isoforms expressed in human kidney. In total, we isolated 10 independent cDNA clones coding for three different HNF4 isoforms which are derived from two distinct and differentially expressed genes.

MATERIALS AND METHODS

Isolation and analysis of cDNA clones. Two human kidney cDNA libraries (HL1123a and HL3001a; Clontech) were screened with a cDNA fragment encoding the first 367 amino acids of the *Xenopus laevis* HNF4β protein (10) at 50°C in a hybridization buffer containing 0.75 M NaCl, 0.075 M sodium citrate, 1% blocking reagent (Boehringer), 0.1% (wt/vol) *N*-lauroylsarcosine, and 0.02% (wt/vol) sodium dodecyl sulfate (SDS). The DNA of the positive clones was subcloned into pBluescript II (Stratagene) and sequenced on both strands, using an ALF automated sequencer (Pharmacia).

Chromosomal assignment of the human HNF4α and HNF4γ genes. Fifty nanograms of genomic DNA isolated from a panel of rodent-human hybrid cell lines (generously provided by H.-J. Lüdecke and B. Horsthemke) was subjected to PCR, using the oligonucleotide primer pairs 4A-E (5′-CCTGTTGCAGG AGATGCTGCTG-3′)–4A-I (5′-CAGCACTTAGAACAGTGACTGGC-3′) and 4G-E (5′-ATACTTTTAGGTCCCATGTCAACA-3′)–4G-I (5′-CATGCAAA GCAAGGACTACTTGTATACTC-3′) to amplify a 260-bp fragment and a 318-bp fragment of the genes encoding HNF4α and HNF4γ, respectively. Primers 4A-E and 4C-E correspond to nucleotides (nt) 1297 to 1318 in HSHNF4A4 (amino acids 398 to 405) and nt 2847 to 2874 in HSHNF4GA (amino acids 720 to 727), respectively; primers 4A-I and 4C-I were deduced from intron sequence (data not shown). The resulting DNA fragments were separated on a 6% polyacrylamide gel and stained with ethidium bromide.

Plasmid constructions. For expression of the different HNF4 isoforms in eukaryotic cells, cDNA constructs containing the complete open reading frames were cloned into the *Hin*dIII and *Not*I sites of Rc/CMV (InVitrogen), using linker sequences from the pBluescript polylinker. The expression vectors for HNF4α2 (RcHNF4α2) and HNF4α4 (RcHNF4α4) contain the same 5′ and 3′ untranslated regions and differ only in the additional 90 bp included in the

* Corresponding author. Mailing address: Institut für Zellbiologie, Universitätsklinikum Essen, Hufelandstr. 55, D-45122 Essen, Germany. Phone: 49-(0)201-723-3110. Fax: 49-(0)201-723-5905.
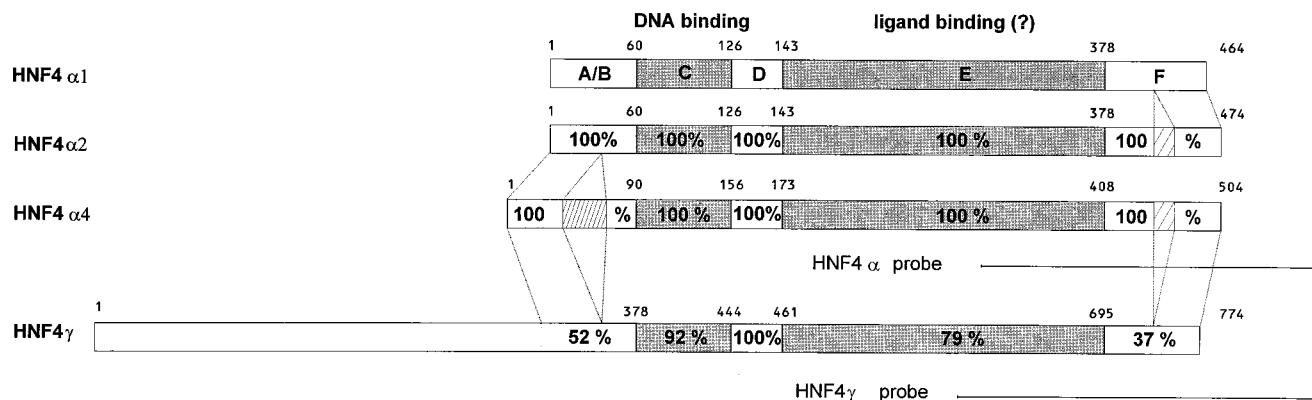
FIG. 1. Schematic comparison of the different HNF4 isoforms. Small numbers refer to the amino acids bordering the homology domains A/B, C, D, E, and F of the members of the nuclear receptor superfamily. Numbers within boxes represent percentages of amino acid identity between the corresponding domains of HNF4α1 (corresponding to HNF4B in reference 2), HNF4α2, HNF4α4, and HNF4γ. The additional 30- and 10-amino-acid sequences in HNF4α4 and HNF4α2, respectively, are represented by hatched boxes. Locations of the probes used for the detection of HNF4α and HNF4γ RNA in the Northern blots (Fig. 4) are indicated.

HNF4α4 splice variant. For the construction of RcHNF4γ, the 770-bp HindIII-StuI fragment of the HNF4γ cDNA containing 689 bp of 5′ untranslated region and the coding sequence for the first 22 amino acids was substituted by the 64-bp HindIII-StuI fragment of a PCR product generated by amplification of nt 684 to 1451 of the HNF4γ cDNA with the oligonucleotide primers 5′-TCCGGGAAG-GTGCAAGCTTGGAAGATG-3′ and 5′-AAGAACCACTGCAGAAGATAG GGCTCGAGC-3′. Besides the complete HNF4γ reading frame of 2,322 bp, the resulting expression vector RcHNF4γ thus contains only 6 and 236 bp of the 5′ and 3′ untranslated sequences of the HNF4γ cDNA, respectively. The reporter construct HNF1pro-luc (generously provided by H. Nakhei) contains the promoter region (nt −1536 to −133) of the rat HNF1α gene (23) inserted into the pXP1 vector (15). The reporter constructs H4-tk-luc and tk-luc (identical to pT81-luc in reference 15) have been described previously (7).

Northern (RNA) blot analysis. The 1,003-bp XbaI-EcoRI fragment of HNF4γ (nt 2245 to 3248 in HSHNF4GA) and the 1,211-bp PstI-XbaI fragment of HNF4α (nt 1033 to 2244 in HSHNF4A4) were labeled with [α-^{32}P]dCTP by using the RediPrime system (Amersham) and consecutively hybridized to multiple-tissue Northern blots (Clonetech no. 7759-1, lot no. 56495, and Clonetech no. 7760-1, lot no. 54244) according to the manufacturer's instructions. The blots contained 2 μg of poly(A)$^+$ RNA per lane, and the quality of the RNA was controlled by the manufacturer by hybridization of a representative blot for each lot of RNA with a human β-actin control probe. The blots were washed at high stringency (0.015 M NaCl, 0.0015 M sodium citrate, 0.1% SDS; 68°C) and exposed to X-ray film.

Cell culture and transfections. C2 hepatoma cells (6) and HeLa cells were cultured at 37°C in Dulbecco's modified Eagle's medium (Biochrome) supplemented with 10% heat-inactivated fetal calf serum, 2 mM L-glutamine, and antibiotics. C2 cells were transfected by the calcium phosphate precipitation method as described previously (16); HeLa cells were transfected with Lipofectin (Gibco BRL), using 2 μg of DNA and 6 μl of Lipofectin per 3.3-cm well. Twenty hours after transfection, the cells were harvested and analyzed for luciferase activity by using the Promega luciferase assay system and a Lumat LB 9501 luminometer (Berthold). The resulting luciferase activities were corrected for the protein concentration determined in the same extracts with the Bio-Rad protein assay system. In all transfections, the total amount of DNA was adjusted to 2 μg by the addition of Rc/CMV expression vector coding for no transcription factor, and all transfections were done with at least two independent plasmid preparations.

Nucleotide sequence accession numbers. The nucleotide sequences described here have been submitted to the EMBL nucleotide sequence database under accession numbers Z49825 (HSHNF4A4 = HNF4α4) and Z49826 (HSHNF4GA = HNF4γ).

RESULTS

Different isoforms of HNF4 transcription factors are expressed in the human kidney. Since the isolation of cDNAs coding for HNF4 transcription factors from X. laevis had revealed the existence of two distinct HNF4 genes in this species (HNF4α and HNF4β [10]), we wondered whether HNF4 transcription factors might be encoded by a gene family also in humans. Therefore, we first screened two human kidney cDNA libraries at low stringency, using a probe derived from a Xenopus HNF4β cDNA. The clones isolated in this screen were then used to rescreen the same libraries at high stringency. In total, we isolated 10 independent cDNAs coding for human HNF4 proteins. Restriction mapping and nucleotide sequence comparison of the cDNAs revealed that they could be divided into two groups, each derived from a different gene. Since the proteins encoded by the two groups of cDNAs are much more closely related to each other than to any other member of the nuclear receptor superfamily (see Discussion), we concluded that they both are HNF4 proteins and called them, in analogy to the nomenclature for the HNF1 and HNF3 families of transcription factors, HNF4α and HNF4γ. As discussed below, HNF4γ is clearly not the mammalian homolog of Xenopus HNF4β.

Sequence analysis showed that the combined sequence for the first group of four cDNAs is nearly identical to the previously published sequence for a human HNF4 cDNA (HNF4A in reference 2). In comparison with this sequence, we found only three nucleotide exchanges and an additional 70 and 759 nt at the 5′ and 3′ ends, respectively. In accordance with Sladek's nomenclature for different splice variants of HNF4 (18), the protein encoded by our cDNAs will be called HNF4α2, since it contains a 10-amino-acid insertion in the carboxy-terminal F domain in comparison with the protein HNF4α1 (HNF4B in reference 2) (Fig. 1). Furthermore, one of our cDNA clones contained an additional 90 nt coding for another 30 amino acids in the amino-terminal A/B domain, which show no sequence homologies to any protein sequence in the EMBL databases. Since the name HNF4α3 is reserved for another splice variant for which no full coding sequence has been cloned (18), we called this novel protein HNF4α4 (Fig. 1).

In comparison with the previously published human HNF4 sequence (2), the HNF4α proteins described here contain nine additional amino acids at the amino terminus (see Discussion). This elongation is the result of the translational initiation at an in-frame initiation codon which is located 27 nt upstream of the previously proposed translation initiation site and which had not been included in the previously published shorter cDNA sequence (2).

The combined sequence of the second group of six overlapping cDNAs (3,248 bp) contains an open reading frame of 2,322 nt flanked by stop codons and encodes a novel protein of 774 amino acids (HNF4γ), as well as 689 and 236 nt of 5′ and 3′ untranslated regions, respectively. As shown by the schematic comparison in Fig. 1 and the amino acid sequence com-
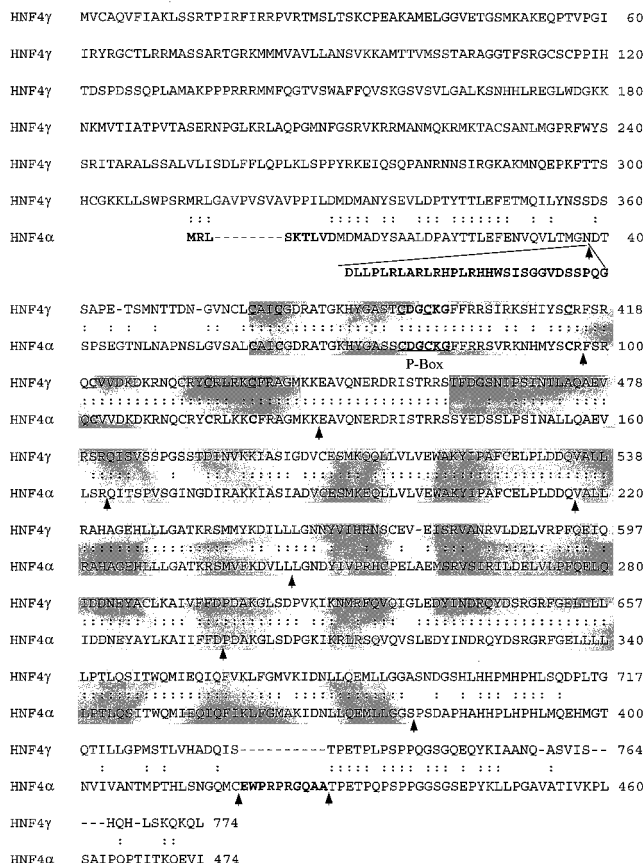
HNF4γ   MVCAQVFIAKLSSRTPIRFIRRPVRTMSLTSKCPEAKAMELGGVETGSMKAKEQPTVPGI 60

HNF4γ   IRYRGCTLRRMASSARTGRKMMMVAVLLANSVKKAMTTVMSSTARAGGTFSRGCSCPPIH 120

HNF4γ   TDSPDSSQPLAMAKPPPRRRMMFPQGTVSWAFFQVSKGSVSVLGALKSNHHLREGLWDGKK 180

HNF4γ   NKMVTIATPVTASERNPGLKRLAQPGMNFGSRVKRRMANMQKRMKTACSANLMGPRFWYS 240

HNF4γ   SRITARALSSALVLISDLFFLQPLKLSPPYRKEIQSQPANRNNSIRGKAKMNQEPKFTTS 300

HNF4γ   HCGKKLLSWPSRMRLGAVPVSVAVPPILDMDMANYSEVLDPTYTTLEFETMQILYNSSDS 360

HNF4α   MRL-------SKTLVDMDMADYSAALDPAYTTLEFENVQVLTMGNDT 40

DLLPLRLARLRHPLRHHWSISGGVDSSPQG

HNF4γ   SAPE-TSMNTTDN-GVNCLCAICGDRATGKHYGASTCDGCKGFFRRSIRKSHIYSCRFSR 418

HNF4α   SPSEGTNLNAPNSLGVSALCAICGDRATGKHYGASSCDGCKGFFRRSVRKNHMYSCRFSR 100

P-Box

HNF4γ   QCVVDKDKRNQCRYCRLRKCFRAGMKKEAVQNERDRISTRRSTFDGSNIPSINILLAQAEV 478

HNF4α   QCVVDKDKRNQCRYCRLKKCFRAGMKKEAVQNERDRISTRRSSYEDSSLPSINALLQAEV 160

HNF4γ   RSEQISVSSPGSSTDINVKKIASIGDVCESMKQQLLVLVEWAKYIPAFCELPLDDQVALL 538

HNF4α   LSRQITSPVSGINGDIRAKKIASIADVGESMKEQLLVLVEWAKYIPAFCELPLDDQVALL 220

HNF4γ   RAHAGEHLLLGATKRSMMYKDILLLGNNYVIHHNSCEV-EISRVANRVLDELVRPFQEIQ 597

HNF4α   RAHAGEHLLLGATKRSMVFKDVLLLGNDYIVPRHCPELAEMSRVSIRILDELVLPFQELQ 280

HNF4γ   IDDNEYACLKAIVFFDPDAKGLSDPVKIKNMRFQVQIGLEDYINDRQYDSRGRFGELILL 657

HNF4α   IDDNEYAYLKAIIFFDPDAKGLSDPGKIKRLRSQVQVSLEDYINDRQYDSRGRFGELLLL 340

HNF4γ   LPTLQSITWQMIEQIQEVKLFGMVKIDNLLQEMLLGGASNDGSHLHHPMHPHLSQDPLTG 717

HNF4α   LPTLQSITWQMIEQIQFIKLFGMAKIDNLLQEMLLGGSPSDAPHAHHPLHPHLMQEHMGT 400

HNF4γ   QTILLGPMSTLVHADQIS----------TPETPLPSPPQGSGQEQYKIAANQ-ASVIS-- 764

HNF4α   NVIVANTMPTHLSNGQMCEWPRPRGQAATPETPQPSPPGGSGSEPYKLLPGAVATIVKPL 460

HNF4γ   ---HQH-LSKQKQL 774

HNF4α   SAIPOPTITKOEVI 474

FIG. 2. Amino acid sequence comparison between human HNF4α2 and HNF4γ. The 9 additional amino acids at the amino terminus of HNF4α and the 30 and 10 additional amino acids inserted in HNF4α4 and HNF4α2, respectively (hatched in Fig. 1), are shown in boldface. The sequences of the DNA binding domain C and of the putative ligand binding domain E are boxed. The cysteine residues of the zinc fingers of the DNA binding domain and the P box are underlined in the HNF4γ and HNF4α sequences, respectively; the arrowheads mark the exon-intron boundaries determined for the mouse HNF4 gene by Taraviras et al. (22).

parison in Fig. 2, the main differences between HNF4γ and HNF4α are located in the amino-terminal A/B domain, which is more than 300 amino acids longer in HNF4γ than in HNF4α, and in the carboxy-terminal F domain, which shows only 37% amino acid identity between HNF4α and HNF4γ. In contrast, the DNA binding domain and parts of the putative ligand binding domain are almost identical. While these domains are also homologous to the corresponding domains of other members of the nuclear receptor superfamily (data not shown), the amino terminus of HNF4γ (amino acids 1 to 312) shows no sequence homologies with any protein in the EMBL databases.

The genes encoding HNF4α and HNF4γ are located on human chromosomes 20 and 8, respectively. For the chromosomal assignment of the human genes encoding HNF4α and HNF4γ, we performed PCR analysis of genomic DNA isolated from a panel of hybrid cell lines which each contained one specific human chromosome in addition to a mouse or hamster genome. As shown in Fig. 3, an amplification product specific for the HNF4α gene was obtained only with total human DNA (Fig. 3, lane Hu) and with DNA isolated from a mouse cell line containing human chromosome 20, while a fragment of the HNF4γ gene could be amplified only from total human DNA (lane Hu) and from the genomic DNA of a hamster-human
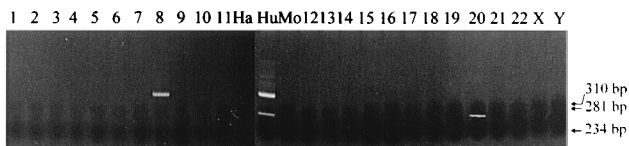

FIG. 3. Mapping of the chromosomal localization of the human HNF4α and HNF4γ genes. The human chromosome contained in the hybrid cell line serving as a source for the genomic DNA used in each PCR is indicated above the corresponding lane. Hu, human genomic DNA; Ha, hamster genomic DNA; Mo, mouse genomic DNA. The fragment lengths of a molecular weight marker are given at the right.

chromosome 8 hybrid cell line. Control reactions with genomic mouse (lane Mo) or hamster (lane Ha) DNA yielded no specific amplification products. From these results, we concluded that the genes encoding HNF4α and HNF4γ are located on human chromosomes 20 and 8, respectively.

HNF4α and HNF4γ are differentially expressed. To compare the expression patterns of the two different genes encoding HNF4 subtypes, we hybridized two human multiple-tissue Northern blots subsequently with probes specific for HNF4α and HNF4γ (Fig. 4; for the positions of the probes, see Fig. 1). HNF4α RNA was detected in RNA from the pancreas, kidney, liver, colon, small intestine, and testis, with signal intensities varying from very strong in the liver and small intestine to barely detectable in the testis. Thus, the tissue distribution of human HNF4α is comparable to that of the corresponding rat (14, 20), mouse (27), and *Xenopus* (9) proteins. HNF4γ RNA was detected in about equal amounts in RNA from the pancreas, kidney, small intestine, and testis but not in liver RNA and only very weakly in RNA from the colon. No discrete signals of HNF4α or HNF4γ RNA could be detected in RNA from skeletal muscle, lung, placenta, brain, heart, peripheral blood, ovary, prostate, thymus, and spleen (Fig. 4). Comparing the signal intensities of different exposures of the Northern
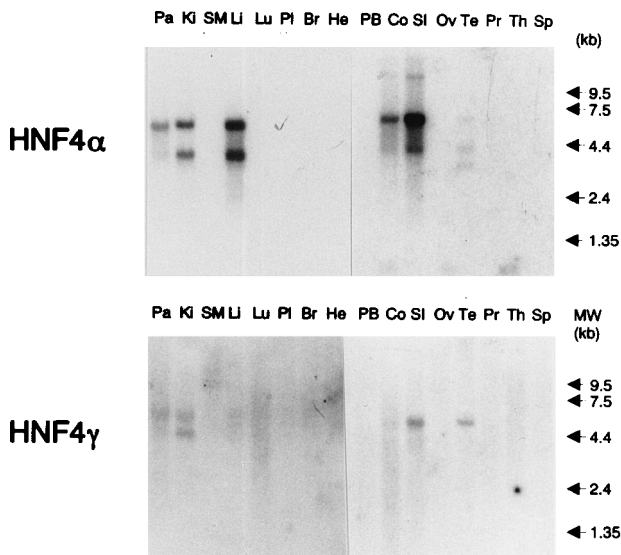

FIG. 4. Northern blot analysis of the tissue distribution of human mRNA for HNF4α and HNF4γ. The positions of RNA molecular weight markers are given at the right. Pa, pancreas; Ki, kidney; SM, skeletal muscle; Li, liver; Lu, lung; Pl, placenta; Br, brain; He, heart; PB, peripheral blood leukocytes; Co, colon; SI, small intestine; Ov, ovary; Te, testis; Pr, prostate; Th, thymus; Sp, spleen. Exposure times for the blot were 3 days (upper left), 6 days (upper right), 20 days (lower left), and 25 days (lower right).

blots, we furthermore estimated that HNF4γ RNA is more than 10-fold less abundant than HNF4α RNA in the human pancreas, kidney, and small intestine, whereas in the testis, comparable amounts of RNA of both isoforms are found. With both hybridization probes, RNA species of different sizes, corresponding to transcripts of >10, 6.0, and 4.0 kb for HNF4α and 5.5 and 4.1 kb for HNF4γ, were detected in some tissues.

HNF4α2, HNF4α4, and HNF4γ are sequence-specific transcriptional activators with different activation potentials. To investigate the transcription activation potential of human HNF4α and HNF4γ proteins, we performed transient transfection experiments. In a first series of experiments, we transfected increasing amounts of expression vectors for these factors into dedifferentiated C2 rat hepatoma cells lacking HNF4. As HNF4-dependent reporter constructs, we used H4-tk-luc, which possesses four HNF4 binding sites in front of the thymidine kinase promoter (7), and, as a more natural target, HNF1pro-luc, which contains a fragment of the rat HNF1α promoter that includes an HNF4 binding site (23). Figure 5 summarizes experiments in which the human HNF4α2 activated H4-tk-luc (Fig. 5A) and HNF1pro-luc (Fig. 5B) up to about 10- and 6-fold, respectively. This stimulation is comparable to that achieved with the rat HNF4α1 under the same conditions (reference 7 and data not shown). In comparison, the transactivation potentials reached with HNF4γ and especially with HNF4α4 were lower even under saturating conditions. In control transfections, a reporter construct lacking HNF4 binding sites (tk-luc) was not transactivated by any of the isoforms (Fig. 5C), demonstrating that the transcription activation potential of the human HNF4 proteins is strictly dependent on the presence of HNF4 binding sites. To check whether the functional differences between the isoforms are dependent on the cell system, we repeated the transfection experiments in HeLa cells. Again, we first performed titration experiments with increasing amounts of expression vector (data not shown) and subsequently used saturating amounts in all experiments. As shown in Fig. 6A and B, HNF4α2 activates both H4-tk-luc and HNF1pro-luc about sevenfold in HeLa cells, while the novel splice variant HNF4α4 is inactive on both reporters and the isoform HNF4γ shows a significant transactivation only on the HNF1 promoter construct (Fig. 6B). Again, a control reporter construct lacking HNF4 binding sites is not activated at all by any of the transcription factors (Fig. 6C), demonstrating that the transcriptional activation by HNF4 transcription factors is DNA sequence specific. To investigate possible interactions between the isoforms, we performed cotransfection experiments. In a first series of experiments, in which we used limiting amounts of expression vector, we did not observe any synergistic effect between the different HNF4 isoforms (Fig. 7A). In a second series of experiments, we used excess amounts of expression vector for HNF4α4 or HNF4γ over HNF4α2 (Fig. 7B). While an excess of HNF4α4 was able to reduce the activity of HNF4α2 about twofold, HNF4γ reduced the activity completely to the level reached with HNF4γ alone. Thus, the activities of HNF4-dependent genes can be modulated by differential expression of the various HNF4 isoforms.

DISCUSSION

HNF4 transcription factors are encoded by a gene family in humans. Using a *Xenopus* HNF4 cDNA as a probe, we isolated human cDNAs for three different isoforms of HNF4 encoded by two different genes. The sequence of one of these proteins (HNF4α2) is identical to the previously published human HNF4 protein sequence (2) except for one conservative amino
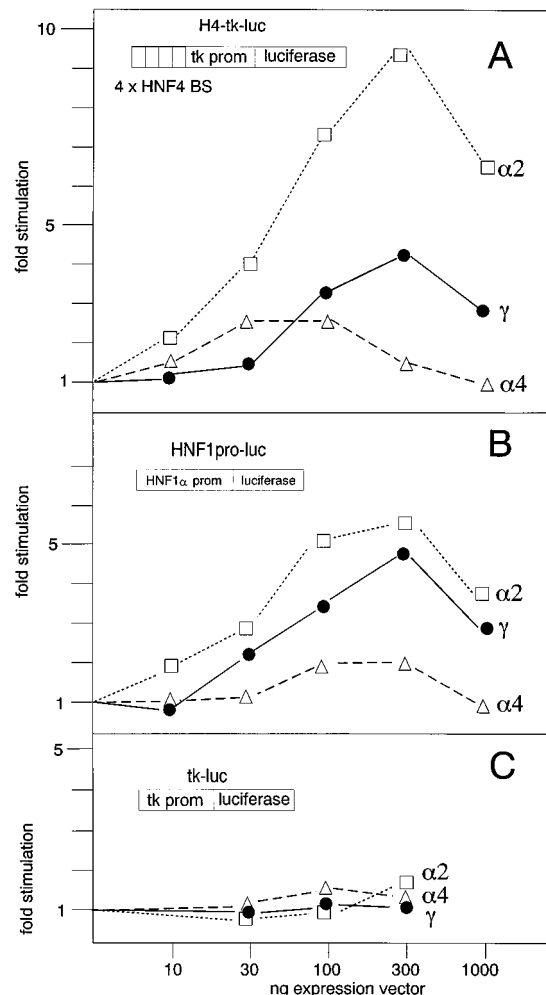

FIG. 5. Transactivation potential of the HNF4 isoforms in C2 cells. The reporter constructs H4-tk-luc, HNF1pro-luc, and tk-luc, containing four, one, and no HNF4 binding sites (HNF4BS), respectively, were cotransfected with increasing amounts of expression vectors for HNF4α2, HNF4α4, and HNF4γ. For calculation of the fold activation, luciferase activities were divided by the mean luciferase activity of the unstimulated reporter construct within each set of experiments. The values for 10 and 1,000 ng of expression vector represent the means of two experiments; all others represent the means of four experiments. tk, thymidine kinase; prom, promoter.

acid exchange (Gly-440 in HNF4α2 [this report] to Ala-431 in HNF4A [2]) in the F domain and nine additional amino acids at the amino terminus. This longer amino terminus is generated by translational initiation at an upstream AUG start codon that is not covered by cDNA previously isolated by Chartier et al. (2). We believe that this amino-terminal elongation with its alternative initiation codon is the authentic amino terminus of all of the HNF4α proteins, since a highly homologous amino acid sequence is also encoded by the rat (20), mouse (8), and *Xenopus* (9) HNF4α cDNAs (Fig. 8), and even the human HNF4γ (this report) and *Xenopus* HNF4β (10) proteins contain related sequences (Fig. 1 and data not shown).

In addition, we isolated a cDNA coding for an HNF4α protein containing 30 additional amino acids inserted in the amino-terminal A/B domain, resulting in the isoform HNF4α4 (Fig. 1). We believe that HNF4α4 is a splice variant of HNF4α, because the 90-nt sequence encoding these 30 amino acids is
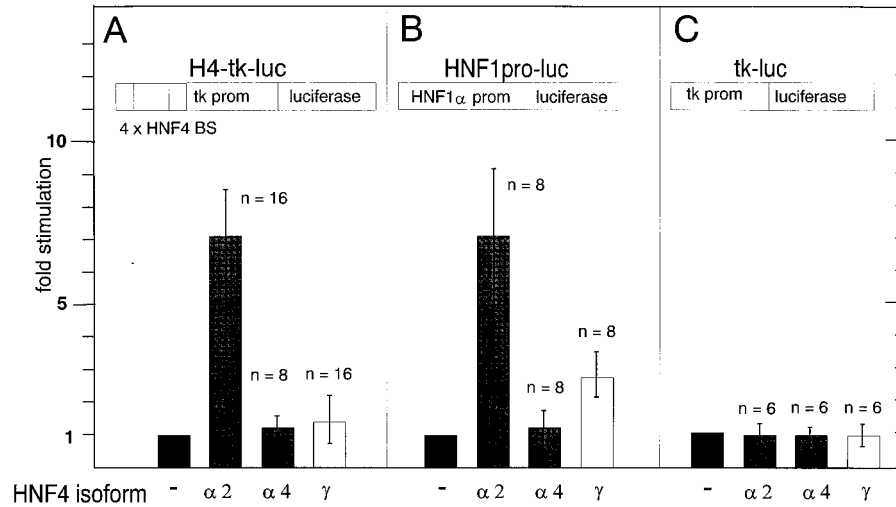
FIG. 6. Comparison of the transactivation potentials of the HNF4 isoforms in HeLa cells. The reporter constructs H4-tk-luc, HNF1pro-luc, and tk-luc were cotransfected with saturating amounts (200 ng) of expression vectors for HNF4α2, HNF4α4, and HNF4γ. For calculation of the fold activation, luciferase activities were divided by the mean luciferase activity of the unstimulated reporter construct within each set of experiments. The error bars represent the standard deviation. n, number of experiments; tk, thymidine kinase; prom, promoter.

inserted into the HNF4α cDNA sequence exactly at the boundary between exons 1 and 2 (Fig. 2) that has been identified for the mouse gene by Taraviras et al. (22). We assume that the additional 90 nt were not introduced into the mRNA by an alternative usage of splice donor or acceptor sites as is the case for the additional 10 amino acids in the carboxy terminus of HNF4α2 (HNF4A in reference 2) but rather represent an additional exon, because the sequence of neither the 5′ nor the 3′ end of the first intron of the mouse HNF4 gene (22) can be translated into a protein sequence homologous to the 30 additional amino acids. By PCR analysis, we showed that the gene encoding HNF4α is located on human chromosome 20, as had already been assumed by Avraham et al. (1), who mapped the gene for the homologous mouse HNF4 protein to mouse chromosome 2 in a region syntenic to human chromosome 20.

The third human HNF4 isoform that we identified (HNF4γ) is clearly distinct from HNF4α and is encoded by a gene located on human chromosome 8. In comparison with HNF4α and with HNF4β, a novel HNF4 subtype identified in *X. laevis* (10), it contains an amino-terminal elongation of 312 amino acids (Fig. 1 and 2 and data not shown) with no sequence homologies to any known protein. Nevertheless, such a long amino-terminal A/B domain of 377 amino acids is not unusual for a member of the nuclear receptor superfamily, since, for example, the corresponding domains of the different forms of the progesterone receptor are 380 and 555 amino acids long, respectively (11). Except for the longer amino terminus,
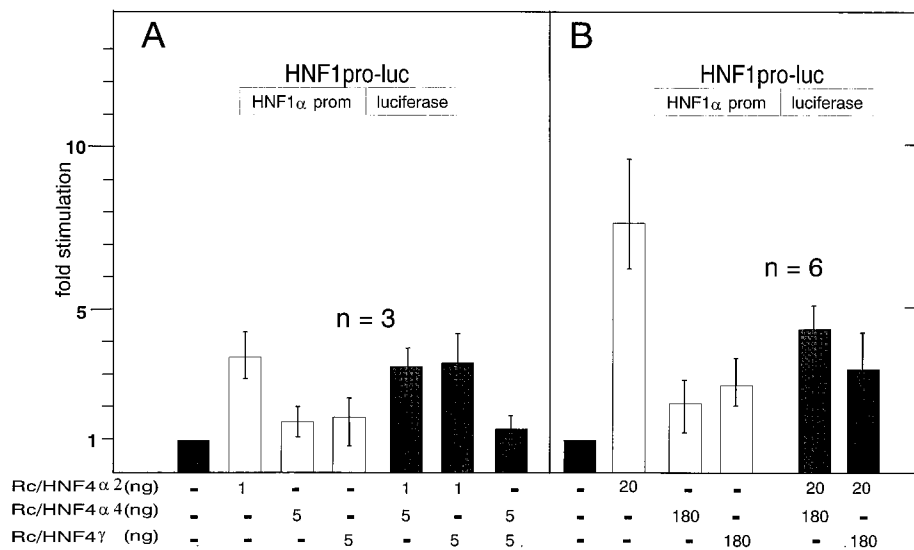

FIG. 7. Functional interactions between the different HNF4 isoforms. HeLa cells were transfected with 1.8 μg of the reporter construct HNF1pro-luc and different amounts of expression vector for the various HNF4 isoforms as indicated. For calculation of the fold activation, luciferase activities were divided by the mean luciferase activity of the unstimulated reporter construct within each set of experiments. The error bars represent the range of experimental data. n, number of experiments; prom, promoter.

```
M R L S K T L V D     Human HNF4α

M R L S K T L A D     rat HNF4

M R L S K T L A G     mouse HNF4

M R L S K A L I D     Xenopus HNF4α
```

FIG. 8. Sequence comparison of the proposed amino-terminal elongation of human HNF4α with the corresponding amino acid sequences deduced from the rat (20), mouse (27), and *Xenopus* (9) HNF4 cDNAs. Identical amino acids are in boldface.

HNF4γ and HNF4α are highly homologous, especially in the DNA binding domain C and in parts of the putative ligand binding domain E (Fig. 1 and 2). The only insertion in HNF4α2 in comparison with HNF4γ consists of the 10 amino acids 419 to 428 (Fig. 2), which determine the difference between the HNF4 splice variants HNF4α1 and HNF4α2 (Fig. 1). We therefore speculate that an exon-intron boundary may be located at this position in the HNF4γ gene as well.

HNF4 transcription factors constitute a novel subfamily of the nuclear receptor superfamily. Comparing the amino acid sequences of the DNA binding domain of the human HNF4α protein with the sequences of other proteins, we found 100% identity to the corresponding domains of the HNF4α proteins of rats (20), mice (8), and *X. laevis* (9) and 90% amino acid identity to the DNA binding domain of *Drosophila* HNF4 (DHNF4 [28]). For the DNA binding domain of human HNF4γ, these comparisons yielded 95 and 89% identity to HNF4α and DHNF4, respectively. In contrast, the DNA binding domains of all other members of the nuclear receptor superfamily showed no more than 62% amino acid identity to HNF4α or HNF4γ (data not shown). Furthermore, HNF4α, *Xenopus* HNF4β (10), human HNF4γ, and DHNF4 (27) are the only proteins containing the sequence CDGCKG in the P-box region in the first zinc finger (Fig. 2). Therefore, we conclude that HNF4α, HNF4β, HNF4γ, and DHNF4 are more closely related to each other than to any other nuclear receptor and constitute a subfamily within the nuclear receptor superfamily. Furthermore, we deduce from the comparison of human HNF4α and HNF4γ with the *Xenopus* and *Drosophila* proteins that the different genes encoding HNF4 transcription factors in humans have developed after the phylogenetic separation of invertebrates and vertebrates, because the amino acid sequence of the putative ligand binding domain of DHNF4 is about equally homologous to HNF4α and HNF4γ sequences (66 and 64% amino acid identity, respectively), but before the separation of amphibians and mammals, because human HNF4α is much more closely related to *Xenopus* HNF4α (9) than to HNF4γ (data not shown). Thus, it is most likely that more than one gene encoding HNF4 transcription factors exists also in other mammals.

What are the in vivo roles of the different HNF4 transcription factors? Our studies have shown that in comparison with HNF4α, HNF4γ has an overlapping but not identical expression pattern (Fig. 4) and a significantly lower transcription activation potential (Fig. 5 and 6). It is therefore not surprising that HNF4γ (which most likely exists also in mice; see above) cannot substitute for HNF4α in homozygous HNF4α knockout mice which die because of serious gastrulation defects before embryonic day 10.5 (3). In this respect, the HNF4 family is different from other subfamilies of the nuclear receptor superfamily, e.g., the retinoic acid receptors, for which gene knockout experiments have shown a high degree of functional redundancy in embryogenesis as well as in the adult (reviewed in

reference 21). In contrast, we assume that HNF4γ may play a modulatory role in the differential transcriptional regulation of HNF4-dependent genes in HNF4γ-expressing tissues (e.g., kidney or pancreas) in comparison with tissues which express only HNF4α (e.g., liver). One example for such a differential regulation might be the L-pyruvate kinase gene, which shows a different footprint on the HNF4 binding site in its promoter region with pancreas and liver extracts (14). Moreover, we cannot exclude the possibility that HNF4γ binds to DNA elements that are not recognized by HNF4α, and we also envisage that HNF4γ and HNF4α may cooperate differently with other transcription factors.

REFERENCES

1. **Avraham, K. B., V. R. Prezioso, W. S. Chen, E. Lai, F. M. Sladek, W. Zhong, J. E. Darnell, N. A. Jenkins, and N. G. Copeland.** 1992. Murine chromosomal location of four hepatocyte-enriched transcription factors: HNF-3 alpha, HNF-3 beta, HNF-3 gamma, and HNF-4. Genomics **13:**264–268.
2. **Chartier, F. L., J. P. Bossu, V. Laudet, J. C. Fruchart, and B. Laine.** 1994. Cloning and sequencing of cDNAs encoding the human hepatocyte nuclear factor 4 indicate the presence of two isoforms in human liver. Gene **147:**269–272.
3. **Chen, W. S., K. Manova, D. C. Weinstein, S. A. Duncan, A. S. Plump, V. R. Prezioso, R. F. Bachvarova, and J. E. Darnell.** 1994. Disruption of the HNF-4 gene, expressed in visceral endoderm, leads to cell death in embryonic ectoderm and impaired gastrulation of mouse embryos. Genes Dev. **8:**2466–2477.
4. **Clairmont, A., T. Ebert, H. Weber, C. Zoidl, P. Eickelmann, W. A. Schulz, H. Sies, and G. U. Ryffel.** 1994. Lowered amounts of the tissue-specific transcription factor LFB1 (HNF1) correlate with decreased levels of glutathione S-transferase alpha messenger RNA in human renal cell carcinoma. Cancer Res. **54:**1319–1323.
5. **Costa, R. H., D. R. Grayson, and J. E. Darnell, Jr.** 1989. Multiple hepatocyte-enriched nuclear factors function in the regulation of transthyretin and α1-antitrypsin genes. Mol. Cell. Biol. **9:**1415–1425.
6. **Deschatrette, J., E. E. Moore, M. Dubois, and M. C. Weiss.** 1980. Dedifferentiated variants of a rat hepatoma: reversion analysis. Cell **19:**1043–1051.
7. **Drewes, T., A. Clairmont, L. Klein-Hitpaβ, and G. U. Ryffel.** 1994. Estrogen-inducible derivatives of hepatocyte nuclear factor-4, hepatocyte nuclear factor-3 and liver factor B1 are differently affected by pure and partial antiestrogens. Eur. J. Biochem. **225:**441–448.
8. **Hata, S., T. Inoue, K. Kosuga, T. Nakashima, T. Tsukamoto, and T. Osumi.** 1995. Identification of two splice isoforms of mRNA for mouse hepatocyte nuclear factor 4 (HNF-4). Biochim. Biophys. Acta **1260:**55–61.
9. **Holewa, B., E. Pogge von Strandmann, D. Zapp, P. Lorenz, and G. U. Ryffel.** Mech. Dev., in press.
10. **Holewa, B., D. Zapp, T. Drewes, L. Klein-Hitpass, and G. U. Ryffel.** Unpublished data.
11. **Kastner, P., A. Krust, B. Turcotte, U. Stropp, L. Tora, H. Gronemeyer, and P. Chambon.** 1990. Two distinct estrogen-regulated promoters generate transcripts encoding the two functionally different human progesterone receptor forms A and B. EMBO J. **9:**1603–1614.
12. **Kuo, C. J., P. B. Conley, L. Chen, F. M. Sladek, J. E. Darnell, Jr., and G. R. Crabtree.** 1992. A transcriptional hierarchy involved in mammalian cell-type specification. Nature (London) **355:**457–461.
13. **Lai, E., and J. E. Darnell.** 1991. Transcriptional control in hepatocytes: a window on development. Trends Biochem. Sci. **16:**427–430.
14. **Miquerol, L., S. Lopez, N. Cartier, M. Tulliez, M. Raymondjean, and A. Kahn.** 1994. Expression of the L-type pyruvate kinase gene and the hepatocyte nuclear factor 4 transcription factor in exocrine and endocrine pancreas. J. Biol. Chem. **269:**8944–8951.
15. **Nordeen, S. K.** 1988. Luciferase reporter gene vectors for analysis of promoters and enhancers. BioTechniques **6:**454–457.
16. **Schorpp, M., W. Kugler, U. Wagner, and G. U. Ryffel.** 1988. Hepatocyte-specific promoter element HP1 of the Xenopus albumin gene interacts with transcriptional factors of mammalian hepatocytes. J. Mol. Biol. **202:**307–320.
17. **Sel, S., T. Ebert, G. U. Ryffel, and T. Drewes.** Submitted for publication.

18. **Sladek, F. M.** 1994. Hepatocyte nuclear factor 4 (HNF-4), p. 207–230. *In* F. Tronche and M. Yaniv (ed.), Liver specific gene expression. R. G. Landes Co., Austin, Tex.
19. **Sladek, F. M., and J. E. Darnell.** 1992. Mechanisms of liver-specific gene expression. Curr. Opin. Genet. Dev. **2:**256–259.
20. **Sladek, F. M., W. M. Zhong, E. Lai, and J. E. Darnell.** 1990. Liver-enriched transcription factor HNF-4 is a novel member of the steroid hormone receptor superfamily. Genes Dev. **4:**2353–2365.
21. **Sucov, H. M., and R. M. Evans.** 1995. Retinoic acid and retinoic acid receptors in development. Mol. Neurobiol. **10:**169–184.
22. **Taraviras, S., A. P. Monaghan, G. Schütz, and G. Kelsey.** 1994. Characterization of the mouse HNF-4 gene and its expression during embryogenesis. Mech. Dev. **48:**67–79.
23. **Tian, J. M., and U. Schibler.** 1991. Tissue-specific expression of the gene encoding hepatocyte nuclear factor 1 may involve hepatocyte nuclear factor 4. Genes Dev. **5:**2225–2234.
24. **Tronche, F., and M. Yaniv.** 1992. HNF1, a homeoprotein member of the hepatic transcription regulatory network. Bioessays **14:**579–587.
25. **Xanthopoulos, K. G., and J. Mirkovitch.** 1993. Gene regulation in rodent hepatocytes during development, differentiation and disease. Eur. J. Biochem. **216:**353–360.
26. **Zapp, D., S. Bartkowski, B. Holewa, C. Zoidl, L. Klein-Hitpaβ, and G. U. Ryffel.** 1993. Elements and factors involved in tissue-specific and embryonic expression of the liver transcription factor LFB1 in *Xenopus laevis*. Mol. Cell. Biol. **13:**6416–6426.
27. **Zhong, W., J. Mirkovitch, and J. E. Darnell, Jr.** 1994. Tissue-specific regulation of mouse hepatocyte nuclear factor 4 expression. Mol. Cell. Biol. **14:**7276–7284.
28. **Zhong, W. M., F. M. Sladek, and J. E. Darnell.** 1993. The expression pattern of a Drosophila homolog to the mouse transcription factor HNF-4 suggests a determinative role in gut formation. EMBO J. **12:**537–544.