The EMBO Journal vol.12 no.11 pp.4229–4242, 1993

# More potent transcriptional activators or a transdominant inhibitor of the HNF1 homeoprotein family are generated by alternative RNA processing

**Ingolf Bach[1] and Moshe Yaniv[2]**

Unité des Virus Oncogènes, UA 1644 du CNRS, Département des Biotechnologies, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cedex 15, France
[1]Present address: Eukaryotic Regulatory Biology Program, School of Medicine, University of California at San Diego, La Jolla, CA 92093-0648, USA
[2]Corresponding author

Communicated by M.Yaniv

**We report the isolation of cDNAs from human liver encoding several isoforms of the hepatocyte nuclear factor homeoproteins HNF1 and vHNF1 generated by the differential use of polyadenylation sites and by alternative splicing. In the novel isoforms intron sequences that are excised in the previously described forms are translated in the same frame as exon sequences until the first termination codon is encountered. Hence, the newly found isoforms all contain different C-terminal domains. For HNF1 it has been shown that its C-terminal region is responsible for the activation of transcription. In transient transfection assays the two novel HNF1 isoforms, HNF1-B and -C, transactivate 5-fold better than the previously described HNF1 protein (HNF1-A). The newly isolated isoform of vHNF1, designated vHNF1-C, is unable to transactivate and behaves as a transdominant repressor when cotransfected with HNF1-A, -B or -C. All of the different isoforms of HNF1 and vHNF1 can form homo- and heterodimers and their mRNAs are differentially expressed in fetal and adult human liver, kidney and intestine, suggesting distinct roles during development. Our studies show that the transactivation domain of the members of the HNF1 homeoprotein family is organized in modules which can be exchanged to generate either more potent transcriptional activators or a transdominant repressor.**
*Key words:* alternative splicing/heterodimerization/HNF1 homeoprotein family/liver/polyadenylation/transcriptional activator/inhibitor

## Introduction

The transcriptional activator proteins HNF1 (hepatocyte nuclear factor 1, also called HNF1$\alpha$ or LF-B1) and vHNF1 (variant HNF1, also called HNF1$\beta$ or LF-B3) belong to the same homeoprotein family that is important for the liver-specific expression of a variety of genes including albumin and $\beta$-fibrinogen (for review see Mendel and Crabtree, 1991; De Simone and Cortese, 1991; Tronche and Yaniv, 1992). Analysis of the tissue distribution of HNF1 and vHNF1 mRNA by Northern blotting, RNase protection and *in situ* hybridization has revealed that their transcripts are found not only in liver but also in kidney, intestine, stomach and pancreas (Baumhueter *et al.*, 1990; Kuo *et al.*, 1990; Bach *et al.*, 1991; Blumenfeld *et al.*, 1991; De Simone *et al.*, 1991; Mendel *et al.*, 1991a; Ott *et al.*, 1991; Rey-Campos *et al.*, 1991; Emens *et al.*, 1992; Lazzaro *et al.*, 1992).

HNF1 and vHNF1 proteins bind to DNA either as homodimers (Chouard *et al.*, 1990; Nicosia *et al.*, 1990) or as heterodimers (Bach *et al.*, 1991; De Simone *et al.*, 1991; Mendel *et al.*, 1991a; Rey-Campos *et al.*, 1991). Both proteins are highly conserved in their dimerization domains, B-domains and homeodomains. The dimerization domain was mapped to a short segment in the N-terminal portion of the proteins. This domain is bound by DCoH (dimerization cofactor of HNF1), a small protein of 11 kDa which reduces the exchange rate of HNF1 molecules in an HNF1 dimer (Mendel *et al.*, 1991b). A region designated the B-domain (Nicosia *et al.*, 1990), which displays weak homologies to the POU-A domain and which is necessary for high specificity binding of HNF1 (Tomei *et al.*, 1992), is located between the dimerization and the DNA binding domains. HNF1 and vHNF1 are thought to contact DNA via an extra-large homeodomain, containing a loop of 21 additional amino acids between the predicted helices 2 and 3 (Frain *et al.*, 1989; Bach *et al.*, 1990; Baumhueter *et al.*, 1990; Chouard *et al.*, 1990; Finney, 1990). Thus, all the sequences necessary for dimerization and DNA binding are located in the N-terminal halves of the proteins. For HNF1 it was shown that transcriptional activation is conferred by its C-terminal half. In particular, the C-terminal 150 amino acids are essential for transactivation and include a region rich in serine and threonine residues (Kuo *et al.*, 1990; Nicosia *et al.*, 1990; Raney *et al.*, 1991; Sourdive *et al.*, 1993).

Several mechanisms have been described to generate from a single gene distinct proteins with modified function. Alternative splicing and the differential use of initiation codons have been shown to modify the properties of transcription factors (for review see Foulkes and Sassone-Corsi, 1992).

In this work we report the isolation of cDNAs encoding novel isoforms of the human HNF1 and vHNF1 proteins which differ in their C-termini from the previously described human HNF1 and vHNF1 proteins (Bach *et al.*, 1990, 1991). Comparison of HNF1 cDNA sequences with the genomic structure of the HNF1 gene (Bach *et al.*, 1992), shows that HNF1-A and HNF1-C (as well as vHNF1-A/B and vHNF1-C) mRNAs are generated by the differential use of polyadenylation sites and that the HNF1-B message is an alternatively spliced product of a common HNF1-B/C pre-mRNA. In transient transfection assays, the novel HNF-1 forms are stronger transactivators than the initially described human HNF1-A, while the novel form of vHNF1 is a transdominant repressor. The concentrations of mRNAs specific for HNF1-A and HNF1-C or for vHNF1-A/B and vHNF1-C vary in tissues and differ between fetus and adult in some tissues, suggesting that the isoforms might control gene expression during organ development.

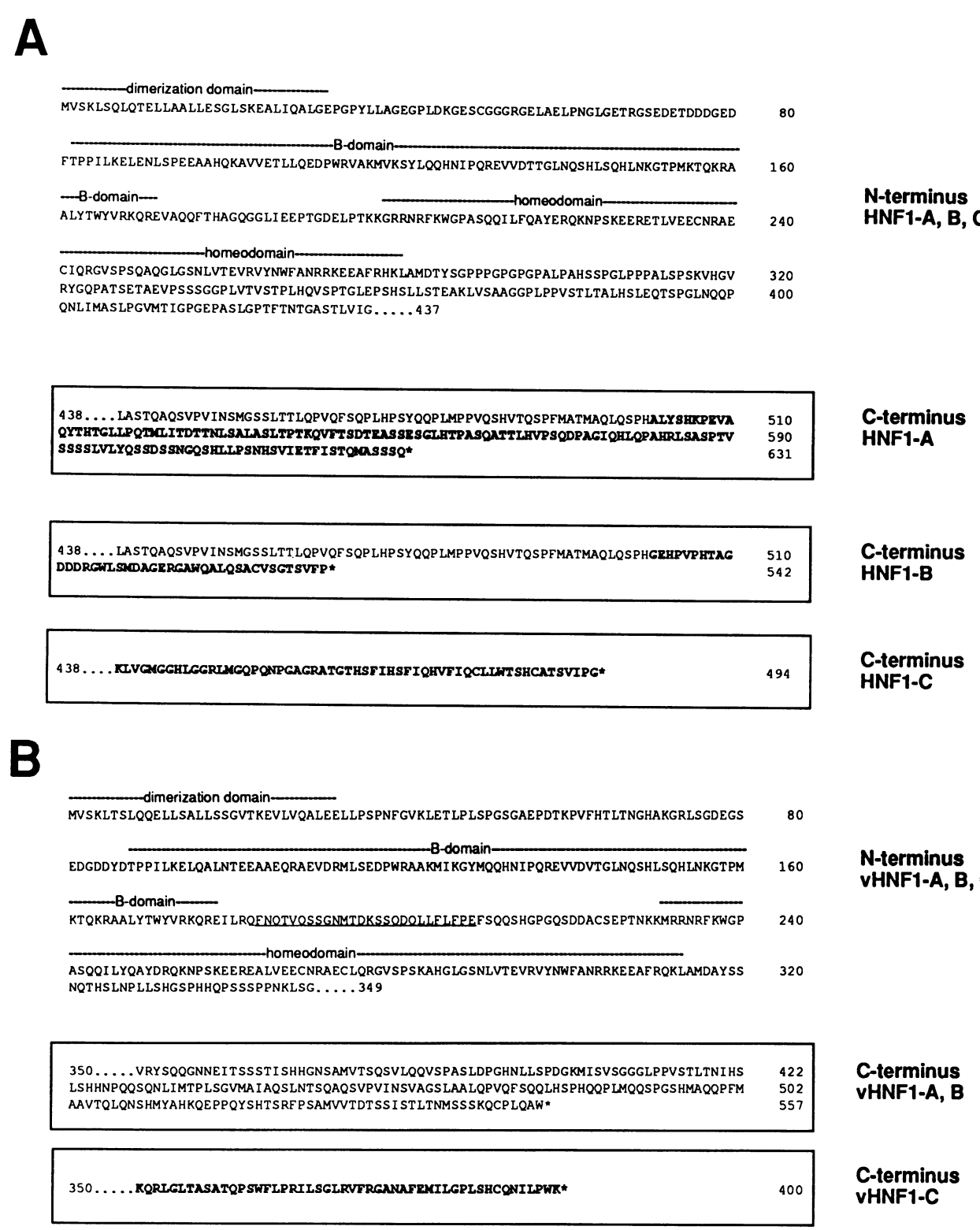


**Fig. 1.** Amino acid sequences of the different HNF1 and vHNF1 isoforms. The dimerization domains, B-domains and homeodomains are indicated. The amino acids specific to each form are shown in bold letters and asterisks mark the end of the proteins. (**A**) Amino acid sequences of the HNF1 isoforms A, B and C. (**B**) The vHNF1-A, vHNF1-B and vHNF1-C protein sequences. The 26 amino acids of the extra exon present in vHNF1-A and -C but missing in vHNF1-B are underlined.

## Results

### *Multiple transcripts from the HNF1 and vHNF1 genes*

A human liver cDNA library in λgt10 was screened with the 511 bp *Nco*I fragment containing the rat HNF1 homeodomain. Out of 16 positive clones the majority encoded the human HNF1 protein which we identified previously (Bach *et al.*, 1990). This form of HNF1 is highly homologous to the one encoded by the rat and mouse clones isolated in several laboratories. However, two positive clones, HCL13 and 24, were chosen for further analysis

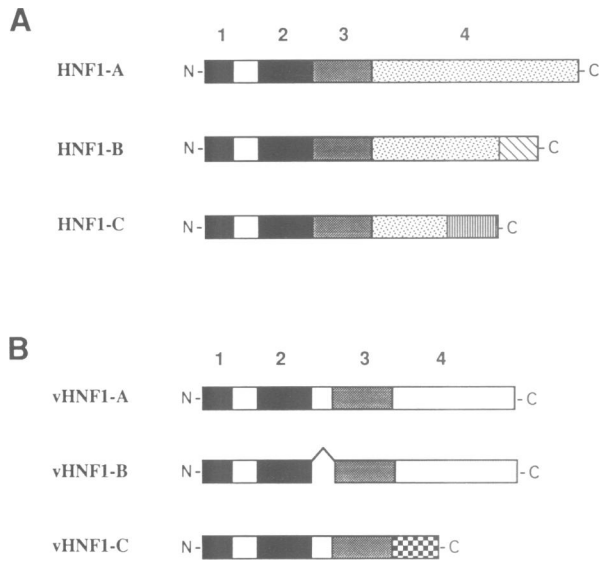


**Fig. 2.** Schematic diagram of the different isoforms of HNF1 and vHNF1. (**A**) HNF1-A, -B and -C proteins and (**B**) vHNF1-A, -B and -C proteins. Functional important protein regions and the specific C-terminal regions of HNF1-B and -C and vHNF1-C are indicated. 1, dimerization domains; 2, B-domains; 3, homeodomains; 4, transactivation domains.

since their sequences diverged from known HNF1 sequences in their 3′ regions. The same human liver λgt10 cDNA library was rescreened with a vHNF1-specific 517 bp *Pst*I−*Acc*I fragment. Fourteen positive clones were obtained, and two (HV2 and HV8) that diverged from the canonical vHNF1 cDNA sequence (Bach *et al.*, 1991) were chosen for further study. As a first step, all clones were sequenced.

Clone HCL13 is 2329 bp long, starting 197 bp upstream of the ATG initiator codon and terminating in a poly(A) tail downstream of a putative AACAGA polyadenylation signal. Nucleotide sequences were deposited in the EMBL sequence data bank and are not shown here. The sequence up to position 1697 is identical to that of the human HNF1 cDNA (Bach *et al.*, 1990). No significant homologies with other known sequences could be found downstream of nucleotide 1697. Translation of the cDNA sequence reveals an open reading frame of 1626 bp, coding for a 542 amino acid polypeptide chain. In this novel isoform, the most C-terminal 131 residues of the previously described HNF1 protein are replaced by 42 unrelated amino acids (Figures 1A and 2A). The novel sequence contains six negatively charged amino acids. The protein encoded by this clone was designated HNF1-B, while the previously described human HNF1 protein (Bach *et al.*, 1990) was renamed HNF1-A.

Clone HCL24 is 3002 bp in length and ends with a poly(A) tail. It contains all the sequences of clone HCL13 (except the first 64 nucleotides in the 5′ non-coding region) but has in addition, an insert of 736 bp at position 1441 (Figure 3). This insertion displayed no homologies to sequences of HNF1-A or -B. The insert contains an in-frame stop codon at position 1615 replacing the last 194 amino acids of the HNF1-A protein with 57 amino acids; it thus shares the N-terminal 437 amino acids with HNF1-A. This shortest HNF1 form was designated HNF1-C (Figures 1A and 2A). The HNF1-C-specific C-terminal tail is positively charged and rather hydrophobic. *In vitro* translation of the three cDNA species gave rise to polypeptide chains with the expected molecular weight, confirming the assignment of the open reading frames (data not shown).

The vHNF1-related clone HV8 contains 3436 bp upstream of a poly(A) tail. A putative CAGAAA polyadenylation signal is located 27 bases upstream. This clone starts at nucleotide 742 on the human vHNF1 clone HV17 (Bach *et al.*, 1991) in a region that corresponds to the beginning of the 26 amino acid extra exon which is found in vHNF1-A but not in vHNF1-B (Rey-Campos *et al.*, 1991). Thus, this clone lacks the coding sequences for the N-terminal 183 amino acids of the human vHNF1 protein (Bach *et al.*, 1991). The independently isolated clone HV2 is 2692 nucleotides in length and is fully contained in clone HV8. It starts at position 879 on the human vHNF1 sequence of clone HV17, just upstream of the homeodomain. Since HV2 and HV8 are partial clones, we refer, for the sake of simplicity, in the nucleotide and amino acid sequence numberings to the full-length clone HV17 and its encoded amino acid sequence (Bach *et al.*, 1991), unless otherwise indicated.

Clone HV8 shares 498 identical nucleotides (positions 742−1240) with the human vHNF1 sequence, encompassing the complete homeodomain. However, downstream from position 1240 the sequence of clone HV8 does not show any significant sequence homology to clone HV17. An in-frame TAA stop codon is located 152 bp further downstream of nucleotide 1240. Thus, the last 257 amino acids corresponding to the entire C-terminal half of human vHNF1 are replaced with 50 vHNF1-C specific residues.

Since clones HV2 and HV8 encoded only partial vHNF1-C proteins we first checked if a full-length vHNF1-C transcript containing the N-terminal vHNF1-A sequences exists. We performed reverse transcription−polymerase chain reaction (RT-PCR) on RNA derived from HepG2 cells, using as primers oligonucleotide PVH5, including the ATG start codon of vHNF1-A, and oligonucleotide 26, overlapping the TAA stop codon of vHNF1-C. One single band was amplified, migrating in the position expected for a vHNF1 cDNA encoding the dimerization and DNA binding domains of vHNF1-A and the novel C-terminal part (not shown). These results show that HepG2 cells contain mRNA that corresponds to the hypothetical full-length vHNF1-C mRNA.

A full-length vHNF1-C (VHCFL) was constructed by ligating the *Hin*dIII−*Apa*I fragment of clone HV17 (vHNF1-A) containing the N-terminal sequences of vHNF1 into the HV2 *Hin*dIII- and *Apa*I-cut plasmid. The N-terminal sequences of vHNF1-A were used for the full-length vHNF1-C construction, since clone HV8 contains the additional 26 amino acid exon of vHNF1-A which is lacking in vHNF1-B. The amplification product obtained from the VHCFL construct with the primers used in the RT-PCR experiment described above gave a product identical to that obtained from HepG2 RNA (not shown)

The amino acid stretches specific for HNF1-B, HNF1-C and vHNF1-C were compared with seqeunces in the Los Alamos gene bank. No significant sequence homologies were found. An amphipathic α-helix is predicted in the region specific for vHNF1-C from position 375 to 386, whereas the 42 or 57 amino acid sequences specific for HNF1-B and -C respectively did not show any characteristic α-helix or β-sheet structure. A characteristic feature in the HNF1-C-specific protein sequence is that it contains three repeats of an HS(V)FI motif.

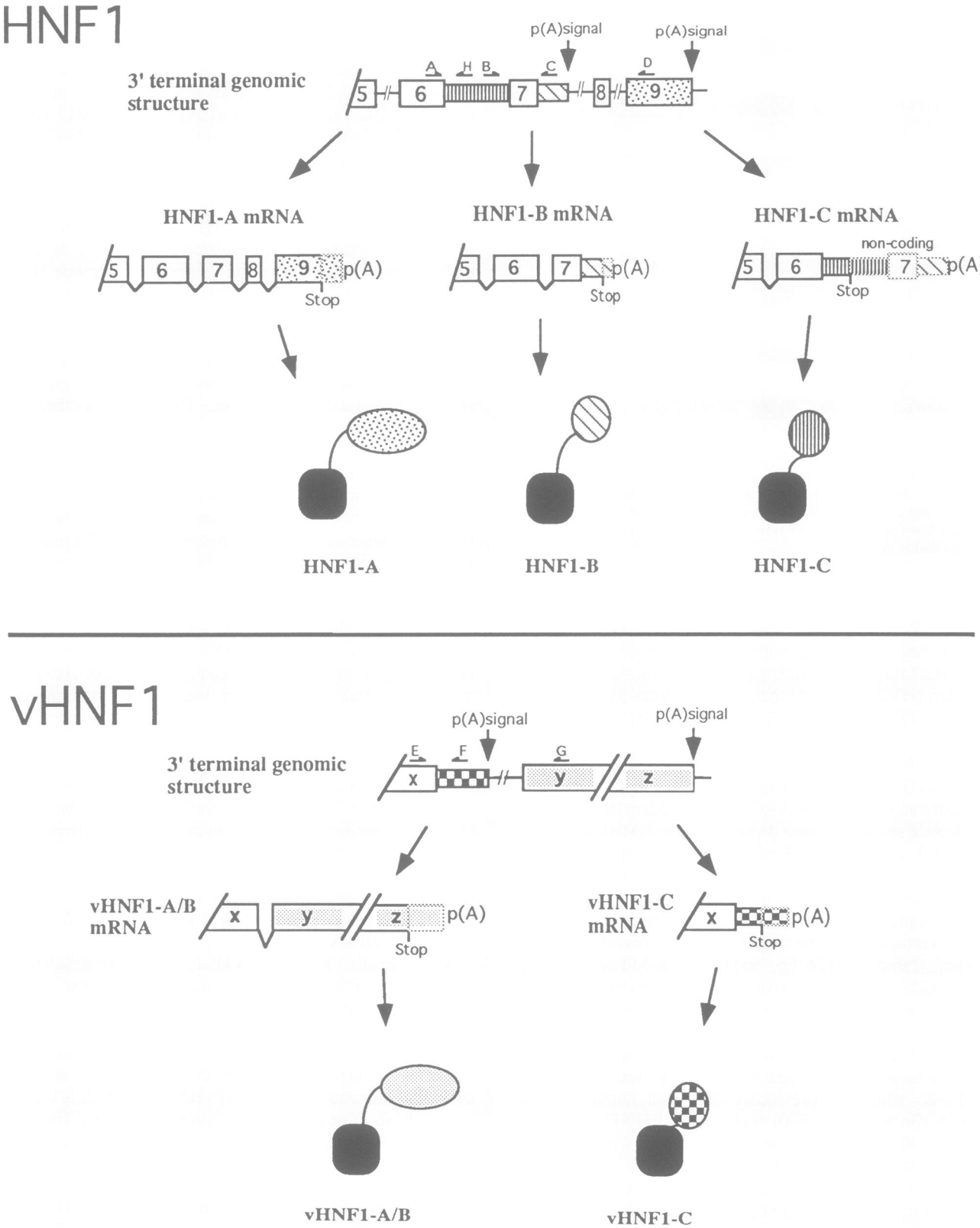


**Fig. 3.** Comparison of the respective genomic, cDNA and protein structures of the HNF1 and vHNF1 isoforms in their C-terminal regions. The exons in the HNF1 gene are in large boxes and numbered according to the rat and mouse gene structure. Since the upstream vHNF1 gene structure is not known we have used letters (x, y and z) to indicate exons. The introns which are partially translated in the novel isoforms of HNF1 and vHNF1 are indicated by small boxes, the splicing events occurring and stop codons are shown. Locations of polyadenylation signals used in the cDNAs are indicated. Non-coding regions in the cDNAs are drawn in dotted boxes. Primers used for PCRs on genomic DNA or for the RT-PCR analysis are drawn above the genomic structure. The N-terminal halves of the respective proteins containing all domains necessary for DNA binding are drawn in black and the specific C-termini are indicated.

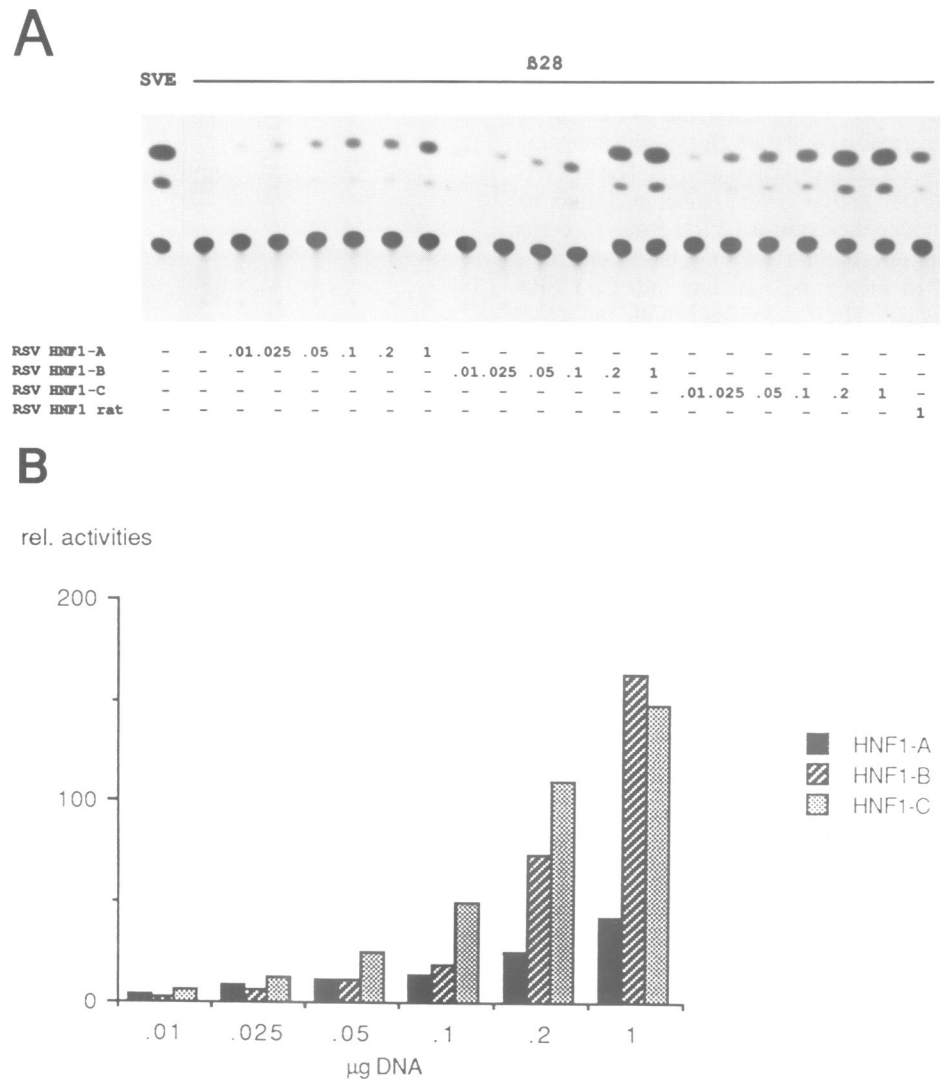


**Fig. 4.** HNF1-B and -C are more potent transcriptional activators than HNF1-A. (**A**) CAT assays of C33 cells transfected with increasing amounts of HNF1-A, -B or -C (in $\mu$g) together with the ($\beta$28)3 promoter linked to the CAT gene. The rat HNF1 construction corresponding to human HNF1-A was used as a control. (**B**) Diagram of the relative activation levels after $\beta$-gal normalization representing the mean of four independent transfection experiments. Transfections were adjusted to constant DNA with RSV-luciferase as carrier.

### *Differential selection of polyadenylation sites and alternative splicing are responsible for the generation of the HNF1 and vHNF1 isoforms*

The perfect homology in the 5′ parts of the nucleotide sequences of the three HNF1 cDNAs strongly argues that they originate from the same gene corroborating with the detection of a unique HNF1 gene by Southern blot hybridization (J.Rey-Campos and M.Pontoglio, unpublished results). This observation suggests that the different cDNA forms are generated by alternative RNA processing. The divergence in the sequences preceding the poly(A) tail between the HNF1-A and the B/C isoforms as well as between vHNF1-A/B and vHNF1-C shows that different polyadenylation signals are used for each class of mRNA. We have recently shown that the rat and mouse HNF1 genes have identical intron−exon structures (Bach *et al.*, 1992). Thus, it is reasonable to suppose that the human gene will have a very similar structure. Comparison of the cDNA sequences of all HNF1 clones with the genomic structure (Figure 3) shows that the regions where the HNF1-B and HNF1-C cDNAs diverge from HNF1-A sequences coincide with the exon−intron junction after the 7th or 6th exon, respectively. This suggests that the novel 3′ sequences of the HNF1-B/C cDNAs arise either from a polyadenylation event in the 7th intron or by splicing exon 7 to a new exon with a novel polyadenylation site. To try to distinguish between these hypotheses we performed PCRs on human genomic DNA with primers (A/H and A/C for HNF1; E/F for vHNF1) indicated in Figure 3 that should generate identical fragments in the cDNA and genomic DNA if the novel sequences derive from contiguous introns. This was indeed the case for both HNF1-C and vHNF1-C. In PCRs on human genomic DNA with primers A/D or E/G no amplified band was detected (data not shown). Taken together these results show that the use of alternative polyadenylation sites present in introns generates forms HNF1-B and vHNF1-C. In addition, non-excision of the last intron from the transcript giving rise to HNF1-B generates the HNF1-C form (Figure 3).

### *HNF1-B and -C are more potent transcriptional activators than HNF1-A*

Since most of the sequences of the HNF1 protein which were shown to be essential for transcriptional activation were

deleted and replaced with shorter segments in the HNF1-B and -C isoforms (Raney *et al.*, 1991; Sourdive *et al.*, 1993), we wondered what the functional significance of these differences was and whether the novel forms would still be able to activate transcription. We investigated this question by carrying out transient cotransfection assays. The cDNAs encoding HNF1-A, -B and -C were subcloned into an expression vector under the control of the Rous sarcoma virus long terminal repeat (RSV LTR). These constructs were cotransfected in increasing amounts into C33 cells which lack endogenous HNF1, together with the CAT reporter gene under the control of a chimeric $\beta$28 promoter containing three HNF1 binding sites and the TATA box of the $\beta$-fibrinogen gene (Courtois *et al.*, 1987) or the rat albumin promoter construct alb-151/+4 (Tronche *et al.*, 1989, 1990). An example of a typical experiment using the ($\beta$28)3-CAT construct is shown in Figure 4A. Figure 4B summarizes the normalized activities calculated for each form under the conditions mentioned above. These experiments were repeated four times with two different DNA preparations. HNF1-B and -C displayed a transcriptional activity which was ~5-fold higher than that of HNF1-A, when 1 $\mu$g of each was transfected. Cotransfection with the rat HNF1 expression vector resulted in a CAT activity similar to that of HNF1-A, the corresponding human isoform. Very similar activation levels (4.5-fold) were obtained when the alb-151/+4 rat albumin promoter construct was used as a reporter plasmid (data not shown). When two HNF1 forms were cotransfected, intermediate activities were obtained, depending on which forms were tested (not shown). Addition of a plasmid expressing the dimerization factor DCoH (Mendel *et al.*, 1991b) to the transfection mixture increased only marginally the values obtained without this plasmid, showing that dimerization was not a limiting step under our experimental conditions (not shown).

Even though all three HNF1 forms share the same N-terminal half, which was shown to be necessary and sufficient for dimerization and DNA binding, it was still possible that the new C-terminal sequences might affect dimerization specificity and DNA binding. To investigate whether these forms can heterodimerize, the proteins were expressed by *in vitro* transcription/translation and the resulting products were subsequently tested alone or in pairwise combinations in electrophoretic mobility shift assays. A double-stranded oligonucleotide, PE56, containing the proximal element of the rat albumin promoter was used as probe. In addition to the bands corresponding to the homodimers, intermediate bands in positions typical for heterodimers could be seen at least for the HNF1-A/B and HNF1-A/C combinations (data not shown). Since HNF1-B and -C dimers migrate at a very similar position in this type of experiment (see also Figure 6A) a heterodimer between these two proteins could not be clearly resolved from the corresponding homodimers. These results demonstrate that the novel HNF1 isoforms dimerize and bind specific DNA similarly to HNF1-A.

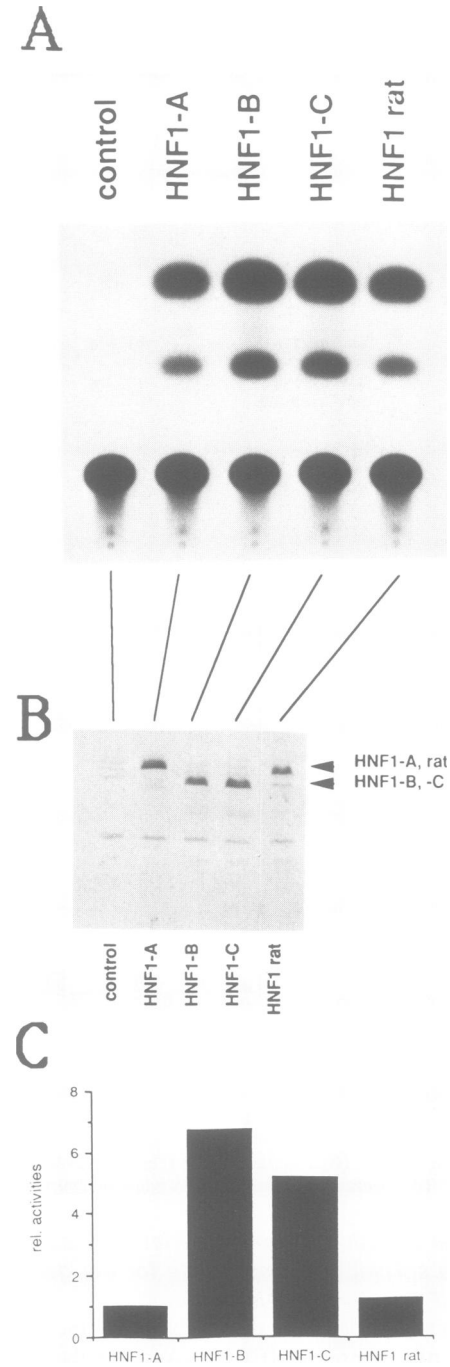


**Fig. 5.** The three HNF1 isoforms are present in equal amounts in transfected C33 cells. (**A**) CAT assays of mock, HNF1-A, HNF1-B, HNF1-C and rat HNF1 transfected C33 cells. (**B**) Quantitative Western blots on total cell extracts of the same transfected cells. (**C**) The CAT activities normalized against the amount of HNF1 protein as measured in panel B.

### *The amino acid sequences specific for HNF1-B and -C are responsible for their higher transcriptional activities*

Several different hypotheses could explain why HNF1-B and -C transactivate more efficiently than the A isoform: the B and C isoforms could be more stable intracellularly or more efficiently synthesized and transported to the nucleus, or they could have higher affinities for their binding site, or their specific C-terminal sequences could be more potent transactivating domains.

To investigate whether the higher transcriptional activity might be due to an increased concentration of HNF1-B and -C proteins within the cells, we normalized the CAT activities to the amount of the different HNF1 proteins that

had accumulated in the transfected cells. This was done by splitting the transfected cells into three aliquots: one was used for the CAT assay, another for Western blotting analysis of total cellular HNF1 protein and the third for the preparation of nuclear protein extracts. The polyclonal antibody rHNt-283, directed against the N-terminal part of rat HNF1 (Sourdive *et al.*, 1993) was used in all Western blot experiments as primary antibody. This antibody cross-reacts with human HNF1, thus recognizing all three isoforms, which contain identical N-terminal halves, with equal efficiencies. A commercial $^{35}$S-labelled anti-rabbit antibody was employed as secondary antibody. The amount of radioactivity in specific bands of both the CAT assay (Figure 5A) and the Western blot from the same transfected cells (Figure 5B) was quantified with a PhosphorImager and the relative activities of each form were calculated (Figure 5C). As a control the rat HNF1 clone corresponding to HNF1-A was included in these experiments. As can be seen in panel C of Figure 5 the different transcriptional activities of the HNF1 isoforms do not reflect differences in the amount of protein accumulated in the cells.

The subcellular distribution of the different HNF1 isoforms within the transfected cells was analysed immunocytochemically to exclude the possibility that HNF1-A is less active because it is transported to the nucleus less efficiently than HNF1-B and -C. Immunostaining of C33 cells transfected with the different HNF1 expression vectors showed that HNF1-A, -B and -C were localized equally well in the nucleus. Furthermore, in quantitative Western blot assays of cytoplasmic and nuclear extracts of the same transfected cells, no difference in the nuclear localization of the three proteins could be detected, confirming the results obtained by immunolocalization (results not shown).

To compare the affinities of the proteins encoded by the three clones for an HNF1 binding site, the proximal element of the rat albumin promoter, we performed gel retardation assays with nuclear extracts of an aliquot of the transfected cells mentioned above. As illustrated in Figure 6A the gel retardations were done with two different amounts of nuclear extracts. The radioactivity contained in the specific retarded bands was quantified using a PhosphorImager and normalized against the amount of radioactivity in specific bands of the quantitative Western blot of the same extracts. The results, summarized in Figure 6B, show that the yield of the retarded bands was similar in all four extracts. This result strongly argues that the affinities of the three HNF1 forms for their DNA target are very similar if not identical.

These results, taken together with the finding by C-terminal deletion analysis that the last 170 amino acids are responsible for 85% of the transactivation potential conferred by HNF1-A (Raney *et al.*, 1991; Sourdive *et al.*, 1993), demonstrate that the residues specific for HNF1-B and -C confer their high transcriptional activity.

### *HNF1-C is differentially expressed in human tissues*

The isoforms of HNF1 all vary in their transactivation potentials. In order to examine whether the events leading to the generation of these forms are differentially regulated in different tissues or during development, we analysed the distribution of the corresponding mRNAs in certain fetal and adult human tissues and cell lines by a semi-quantitative RT-PCR based method. In these experiments we did not attempt to determine the absolute amounts of each mRNA, but

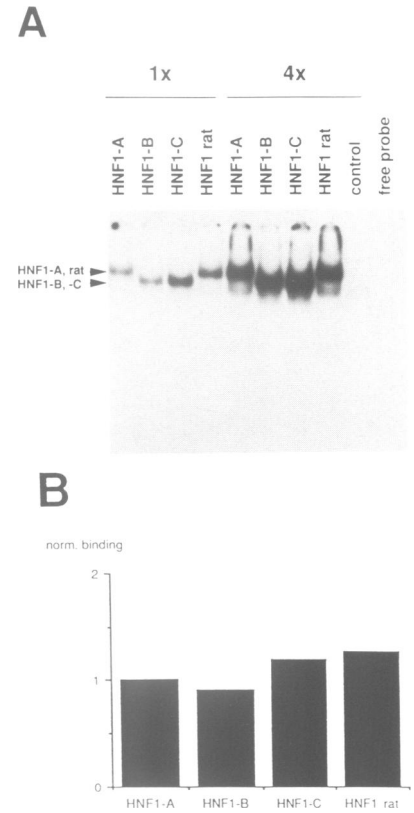


**Fig. 6.** The three human isoforms and rat HNF1 display similar affinities for the proximal element of the albumin promoter. **(A)** Gel retardation of 1× and 4× the volume of nuclear extracts of HNF1-A, HNF1-B, HNF1-C and rat HNF1 transfected C33 cells.
**(B)** Diagram of the amount of radioactivity in retarded bands after normalization against the amount of protein measured on a quantitative Western blot of the same nuclear extracts.

instead compared the ratios of the amounts of mRNA encoding the different isoforms in a given organ or cell line. Briefly, total RNA was prepared from several fetal and adult human tissue samples or cell lines. The RT-PCR analysis was performed as described in Materials and methods. Specific segments of the different HNF1 isoforms were coamplified with a set of four primers (A, B, C and D) resulting in specific bands for HNF1-A, -B and -C. To visualize the HNF1-C mRNA better, each cDNA sample was also amplified for 18 cycles with a single pair of primers (B and C), amplifying exclusively the HNF1-C message. PCR control reactions were carried out with this set of primers on plasmids HCL13, HCL16 and HCL24, to ensure that all three segments were amplified with equal efficiencies (not shown). PCR products were fractionated in polyacrylamide gels and electrotransferred on to nylon membranes. DNA fragments were detected by hybridization with a $^{32}$P-labelled probe and quantified. The exponential phase of amplification was verified for all PCR samples by comparing the radioactivities of the bands after different numbers of cycles. The accuracy of the RT-PCR method was verified by RNase protection experiments on total RNA derived from HepG2 cells. A very similar ratio of mRNAs encoding HNF1-A and HNF1-C was detected by both approaches. However, the RNase protection method was not sensitive enough to detect the low amounts of specific HNF1-B mRNAs (data not shown).

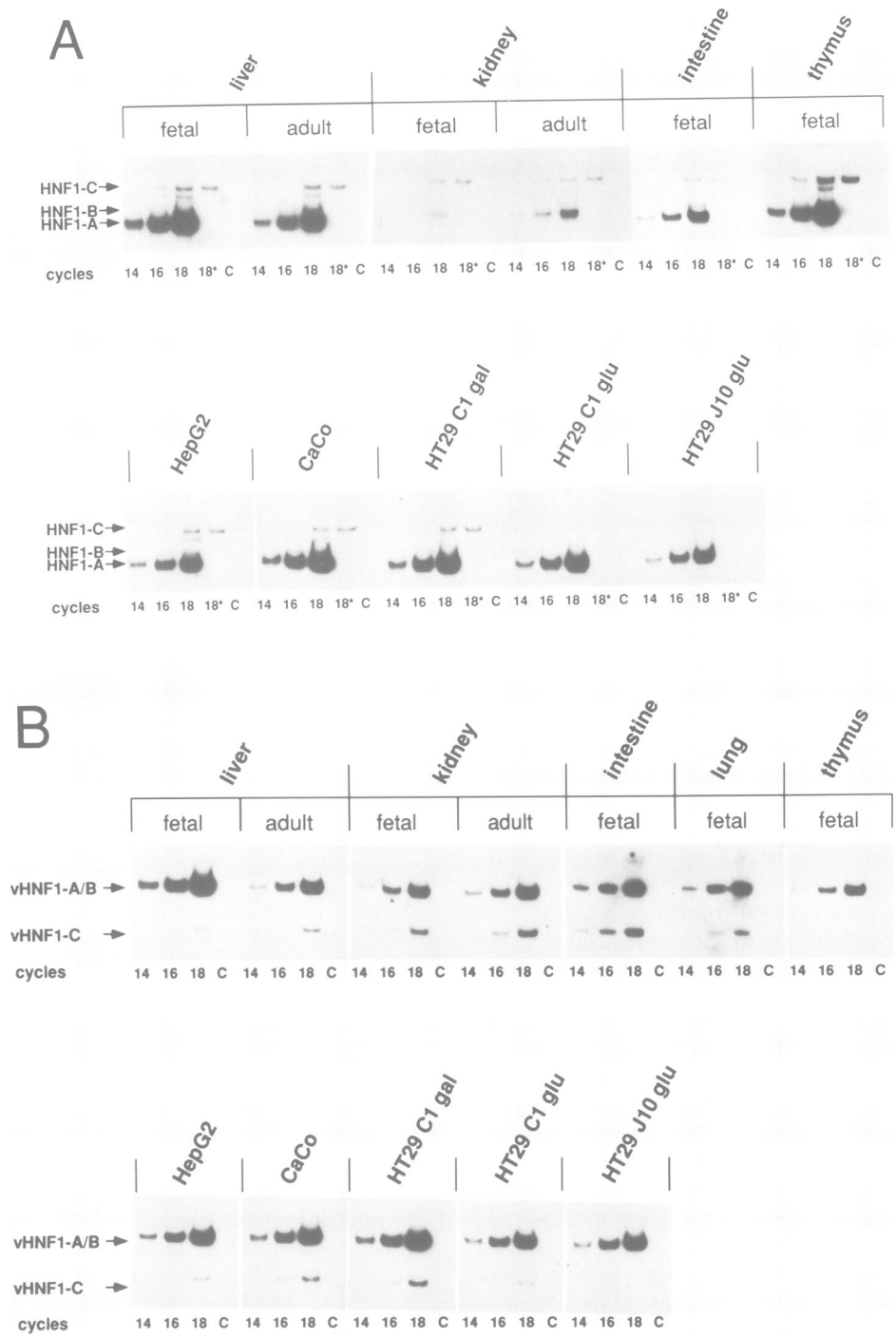


**Fig. 7.** Ratios of RT-PCR amplified fragments specific for the different human HNF1 or vHNF1 isoforms in fetal and adult tissues and cell lines. The RNA sources and the number of cycles are indicated. RNA of the respective tissue or cell line without adding the reverse transcriptase was used as a control sample (C) and was amplified for 18 cycles. (**A**) RT-PCR amplified DNA fragments of HNF1-A, -B and -C mRNA with primers A, B, C and D indicated in Figure 3. To improve the visualization of the HNF1-C-specific fragment an 18 cycle reaction was performed on each cDNA sample exclusively with primers B and C (18*). (**B**) RT-PCR analysis of vHNF1-A/B and vHNF1-C mRNA with primers E, F and G, indicated in Figure 3. These primers do not distinguish between vHNF1-A and -B isoforms.

As can be seen in Figure 7A and in Table I, mRNAs of the different HNF1 isoforms are expressed in human liver, kidney, intestine and thymus. As control, the RNAs of each tissue or cell line sample were amplified in the absence of the reverse transcriptase. No band was obtained, showing that our RNA preparations were not contaminated with genomic DNA. The HNF1-A form was predominant in all tissues analysed, so its mRNA levels were taken as reference and set to 100% in Table I. The HNF1-C mRNA levels are expressed as percentage of HNF1-A mRNA. In human liver, HNF1-C transcripts are clearly detectable and increase from 3.8% in fetal liver to 7.1% in adult liver. In the differentiated human hepatoma cell line HepG2 the fraction of specific HNF1-C mRNA was 9.4%. In fetal kidney the level of HNF1-C mRNA is almost as high (86%) as that of HNF1-A. This high level drops in the adult kidney to 27%. In thymus, HNF1-C mRNA expression of 10.9% was detected. We also tested various states of the human colon adenocarcinoma cell

**Table I.** Occurrence of mRNAs encoding the different HNF1 and vHNF1 isoforms in several fetal or adult human tissues and in human cell lines

| Tissue/cell line | HNF1-A | HNF1-C | vHNF1-A/B | vHNF1-C |
|---|---|---|---|---|
| Fetal liver | 100 | 3.8 | 100 | 4.7 |
| Adult liver | 100 | 7.1 | 100 | 15.1 |
| Fetal kidney | 100 | 86 | 100 | 43.5 |
| Adult kidney | 100 | 27 | 100 | 40 |
| Fetal intestine | 100 | 9.7 | 100 | 28.8 |
| Fetal thymus | 100 | 10.9 | 100 | 4.7 |
| Fetal lung | – | – | 100 | 8.3 |
| Fetal brain | – | – | – | – |
| HepG2 | 100 | 9.4 | 100 | 4 |
| Caco-2 | 100 | 7.4 | 100 | 26.3 |
| HT29-18-C1/gal | 100 | 7.7 | 100 | 10.9 |
| HT29-18-C1/glu | 100 | 8.1 | 100 | 8.9 |
| HT29-18-J10/glu | 100 | 4.2 | 100 | 6.8 |
| HaCaT | – | – | – | – |
| MCF-7 | – | – | – | – |
| SW13 | – | – | – | – |

Fragments of the different mRNAs were amplified by the RT-PCR procedure, fractionated and quantified as described in Materials and methods. The primers used are described in the text and indicated in Figure 3. No attempt was made to quantify the concentration of the minor form HNF1-B. HNF1-A or vHNF1-A/B levels were set at 100% and the amounts of HNF1-C or vHNF1-C mRNA detected are expressed in relation to the HNF1-A or vHNF1-A/B mRNA.

line HT29 (which can be induced to differentiate into enterocytes by the removal of glucose from the culture medium) (Pinto *et al.*, 1982; Huet *et al.*, 1987), the more differentiated human Caco-2 colon carcinoma cell line (Pinto *et al.*, 1983) and fetal intestine for the expression of the HNF1 isoforms. HNF1-C mRNA expression increased from 4.2% in the least differentiated human colon cancer cells, HT29-18-$J_{10}$ cells grown in glucose, to ~8% in HT29-$C_1$/glu cells. Further differentiation does not result in higher HNF1-C RNA levels.

HNF1-B-specific bands could be detected in all tissues or cell lines where the HNF1-C isoform was present (see also Figure 7A), but its mRNA levels were 3- to 20-fold lower and therefore difficult to quantify precisely. In Table I only the radioactivity in the specific bands for HNF1-A and -C isoforms is given. In cells related to hepatocytes and in thymus a band in addition to the expected ones can be seen, migrating between the bands specific for HNF1-B and -C (Figure 7A). Thus, we cannot rule out the possibility that other isoforms of HNF1 exist.

No HNF1-specific bands were detected in fetal lung and brain or in a human keratinocyte cell line, HaCaT, a human breast cancer derived cell line, MCF-7, or a human adrenal cortex derived cell line, SW13, indicating either that no mRNA is synthesized in these tissues or that the RNA levels are so low that 18 cycles of PCR are not sufficient to visualize an amplified cDNA (data not shown).

To verify that the mRNAs encoding the HNF1 isoforms are properly transported into the cytoplasm for translation we performed RT-PCR experiments on cytoplasmic and polysomal RNA from HepG2 cells as described above. A very similar ratio of HNF1-A:HNF1-B:HNF1-C was measured in these experiments, indicating that these mRNAs are functional (data not shown).

### *vHNF1-C inhibits transcriptional activation by the other HNF1 homeoprotein family members and is expressed in certain human tissues*

vHNF1-C contains almost none of the amino acid sequences located in the C-terminal halves of the vHNF1-A and -B isoforms (Figures 1B and 2B). When tested in transient transfection assays, vHNF1-C did not display any trans-activational activity, whereas human and rat vHNF1-A were able to transactivate. However, when cotransfected with the transcriptionally most potent forms of HNF1, i.e. HNF1-B and -C, vHNF1-C transdominantly inhibited their activation function (Figure 8A and B). The activities dropped 5-fold when equal amounts of HNF1-B or -C and vHNF1-C expression vectors were cotransfected. When a 10-fold excess of vHNF1-C was cotransfected, the activity dropped down to background levels. The amounts of protein produced were monitored in gel shift assays with nuclear extracts from the same transfected cells. vHNF1-C dependent DNA binding did not exceed that of HNF1 when equimolar amounts of each expression vector were cotransfected (not shown). The same inhibitory effect was obtained when human HNF1-A or vHNF1-A expression vectors were cotransfected with vHNF1-C (not shown).

It has recently been shown that in all tissues where vHNF1-A transcripts can be detected, vHNF1-B mRNA is also expressed, although 3- to 5-fold less abundantly (Cereghini *et al.*, 1992; Ringeisen *et al.*, 1993). We therefore analysed the ratio between vHNF1-A/B and vHNF1-C mRNAs in certain fetal and adult human tissues and cell lines by an RT-PCR based method. cDNAs were coamplified with a set of three primers (E, F and G), giving rise to two bands, one specific for vHNF1-A/B and one for vHNF1-C. PCR control reactions were carried out with this set of primers on plasmids HV17 and VHCFL, to ensure that both segments were amplified with equal efficiencies (not shown).

As can be seen in Figure 7B and Table I, mRNAs for the different forms of vHNF1 are expressed at least in human fetal and adult liver and kidney and in fetal intestine, thymus and lung. The last three tissues were not examined in the adult state. No bands were obtained when RNAs were amplified without adding reverse transcriptase. The vHNF1-A/B form was predominant in all tissues analysed, so its mRNA levels were taken as 100% in Table I. vHNF1-C levels are expressed as the percentage of

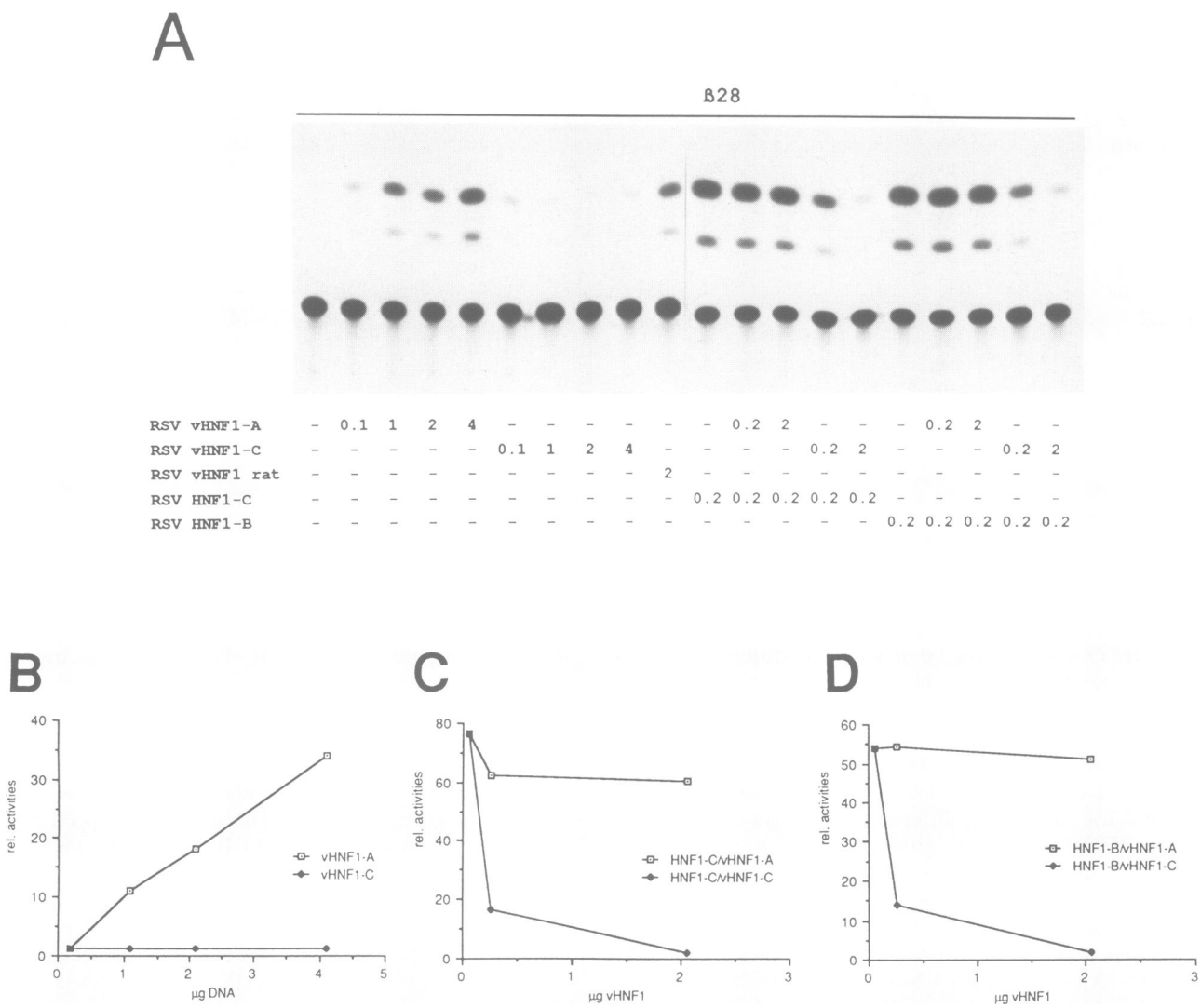


**Fig. 8.** vHNF1-C has lost its transactivational activity and inhibits the transcriptional activation conferred by HNF1-C and HNF1-B. (**A**) CAT assays of C33 cells cotransfected with increasing amounts of vHNF1-A and vHNF1-C in the absence or presence of low amounts of HNF1-C or HNF1-B. (**B**) Diagram of the relative activation values of vHNF1-A and vHNF1-C after β-gal normalization. (**C**) Diagram of the relative activation values of HNF1-C, cotransfected with increasing amounts of vHNF1-A or vHNF1-C after β-gal normalization. (**D**) Diagram of the β-gal-normalized relative activation values of HNF1-B, cotransfected with increasing amounts of vHNF1-A or vHNF1-C.

vHNF1-A/B levels. Comparable vHNF1-C levels of 43 and 40% were detected in fetal and adult kidney, respectively. In the least differentiated HT29-18 cells, $J_{10}$/glu, vHNF1-C mRNAs were found at a level of 6.8%, increasing to 10.9% in the galactose differentiated HT29-18-$C_1$ cells. Caco-2 cells, which are considered as even more differentiated, gave a value of 26%. In human fetal intestine a relative value of 29% of vHNF1-C mRNAs was obtained. Thus, the increase of vHNF1-C mRNA seems to be associated with intestinal differentiation. In fetal brain, HaCaT, MCF-7 and SW13 cells no specific vHNF1 bands were detected (data not shown).

The expression of the vHNF1 isoforms A and C was also determined from cytoplasmic and polysomal RNA preparations of HepG2 cells and both were found to be present in a ratio similar to that in total RNA (data not shown) showing that these mRNAs are functional.

## Discussion

### *Isoforms from the HNF1 and vHNF1 genes with altered transcriptional activities*

We have shown that the mRNAs encoding the novel HNF1-B, HNF1-C and vHNF1-C isoforms are generated by the differential use of polyadenylation sites and by alternative splicing. All isoforms differ from the previously known HNF1 and vHNF1 proteins in their C-terminal regions. These regions are thought to interact with other proteins to induce the transcription of the responsive genes. HNF1-B and -C proteins are stronger transcriptional activators than HNF1-A, whereas vHNF1-C is a trans-dominant inhibitor of the other HNF1 and vHNF1 isoforms. This is true at least for the chimeric (β28)3 promoter (Figures 4 and 8) where no upstream promoter elements other than the HNF1 binding sites are present (Courtois *et al.*, 1987),

or the albumin promoter construct alb-151/+4 (data not shown) containing 155 bp of the native rat albumin promoter (Tronche *et al.*, 1989, 1990).

For HNF1-B and -C, the two most plausible scenarios imaginable to explain their higher transcriptional activation potential are (i) that HNF1-B and/or -C interact with the same mediator protein as HNF1-A (but the interaction occurs on a different domain of this mediator protein), leading somehow to a stronger transcriptional activation or (ii) that HNF1-B and/or -C interact with a different mediator protein which can induce higher transcriptional activation. With the methods and constructs applied, we cannot distinguish between these two possibilities. Further investigations are necessary to clarify this point.

Equimolar amounts of transfected vHNF1-C decrease the transactivation conferred by the HNF1 isoforms 5-fold (Figure 8). In addition, vHNF1-C expression in stably transfected HepG2 cells almost completely blocks transcription of the endogenous albumin gene (I.Bach, unpublished results). Again, the molecular mechanism of this repression is not known. Assuming random heterodimer formation between vHNF1-C and all forms of HNF1, this would result in a 1:2:1 distribution of homodimers:heterodimers:homodimers when equimolar amounts of proteins are made. Quantitative consideration of our data suggests that the vHNF1-C−HNF1 heterodimers are inactive in such a mixture. These results suggest either that the presence of a single transactivation domain in the heterodimer is not sufficient for transcriptional activation or that the unique C-terminal segment of vHNF1-C interferes with the function of HNF1 transactivation domains.

Examination of C-terminal amino acid sequence changes between HNF1-A and the B and C isoforms does not give a clear molecular explanation for the differences in their transcriptional activities. The last 131 amino acids of HNF1-A were shown to be responsible for ~80% of its transcriptional activity. This segment is acidic and rich in serine, threonine and glutamine residues (Sourdive *et al.*, 1993), which are typical of activation domains. This region is replaced with an acidic segment in HNF1-B and with a short positively charged and moderately hydrophobic region in the C isoform.

Since HNF1 binding sites exist in a wide variety of different promoters with different contexts of upstream promoter elements it would be interesting to test the different isoforms of HNF1 and vHNF1 in the natural contexts of those promoters. The results of these experiments might give some clues about the mechanisms of the activation or inhibition of transcription by the different HNF1 and vHNF1 isoforms.

Several examples have been described in which both a transcriptional activator and inhibitor molecule can be generated from a single gene either by the use of internal initiation codons or by splicing out parts of or the entire transactivation domain (Koenig *et al.*, 1989; Descombes and Schibler, 1991; Dobrzanski *et al.*, 1991; Foulkes *et al.*, 1991; Roman *et al.*, 1991; Yen *et al.*, 1991; see Foulkes and Sassone-Corsi, 1992 for review). Removing or adding a sequence encoding an active part of a molecule is a relatively easy way for a cell to alter protein activity. However, in the example of the HNF1 isoforms a functionally active part of the protein is replaced with an even more active set of amino acids. This replacement mechanism is more complex because for one protein several active domains must be encoded and, according to the cell type and/or developmental state, brought into the right context. To our knowledge the isoforms B and C from the HNF1 gene are the first example of an transcription factor in which the active part of the transactivation domain is exchanged. In addition, it has not yet been observed that either a more potent transcriptional activator (HNF1-C) or a transdominant inhibitor (vHNF1-C) is generated by this mechanism.

### *Biological roles for the different isoforms of HNF1 and vHNF1?*

All isoforms of HNF1 and vHNF1 differ functionally in their transactivation potentials. In a search of a possible physiological relevance for the existence of the different forms we first examined whether the ratio of the levels of mRNAs for the different isoforms is constant or variable between different sites of their expression by a semi-quantitative RT-PCR based approach. This method was the most suitable for the analysis of the very limited amounts of human material available.

As illustrated in Figure 7A and B and Table I, mRNAs for all of the different forms of HNF1 and vHNF1 are synthesized at least in human liver, kidney, intestine and thymus. In fetal lung, vHNF1-A/B and vHNF1-C mRNA could be detected, but not HNF1 transcripts, which is consistent with the previous detection of vHNF1 but not HNF1 mRNAs in this tissue (De Simone *et al.*, 1991; Mendel *et al.*, 1991a; Cereghini *et al.*, 1992). The ratio between mRNAs encoding the different isoforms HNF1-A:C and vHNF1-A/B:C varies ~20- and 10-fold, respectively, depending on the developmental state of the cells and the tissue type (see Table I). Taken together these results suggest that the different isoforms of HNF1 and vHNF1 might play a role in the regulation of gene expression during liver, kidney and intestinal development.

Some target genes of HNF1 or vHNF1, e.g. the albumin gene (Tilghman and Belayew, 1982), are known to be expressed in a more liver-specific manner in the adult organism. However, other target genes of the HNF1 homeoprotein family, like the genes encoding $\alpha_1$-antitrypsin (Tripodi *et al.*, 1991), aminopeptidase N (Olsen *et al.*, 1991) and $\alpha$-fetoprotein (Tyner *et al.*, 1990), are known to be expressed also in tissues other than liver, e.g. intestine and/or kidney and, in the case of the insulin I gene, in the pancreas (Emens *et al.*, 1992). Differential levels of the HNF1 and vHNF1 isoforms in these organs might help to explain such a differential gene expression pattern. It should be emphasized that since these newly found isoforms are more potent, even low expression levels or small differences in their levels of expression during development might alter specific gene expression significantly and hence might be of physiological relevance. For example, in adult liver HNF1-C mRNA is detected 'only' at 7% of the level of the HNF1-A mRNA. In such a tissue most of the HNF1-C protein would be heterodimerized with HNF1-A. Taking into account that HNF1-C homodimers can transactivate 5-fold more strongly and that the cotransfection of HNF1-A and -C expression vectors results in intermediate activation levels (I.Bach, unpublished results), suggest that the transcriptional

activity conferred by HNF1-C in these cells is more than one-third that of HNF1-A. The occurrence of the HNF1/vHNF1 isoforms adds a level of complexity to the tissue-specific combination of transcription factors resulting in a network of regulatory transcription factors made responsible for specific gene expression in each organ [Xanthopoulos *et al.*, 1991; for reviews see Struhl (1991) and Tronche and Yaniv (1992)]. This augmentation in complexity is also demonstrated by the fact that in a cell producing all HNF1/vHNF1 isoforms, 21 different HNF1/vHNF1 homo- or heterodimers could be formed, provided that there was no selectivity in dimer formation.

Our results show that all different isoforms of HNF1 and vHNF1 are expressed in liver, kidney and intestine, but also in HepG2, Caco-2 and HT-29 cell lines, suggesting that they might be coexpressed in the same cells. However, it might well be that some isoforms are expressed differentially, by the cells of a certain region of an organ or only a spread subset of cells within this organ. Since our analysis of the tissue distributions was only done on RNA originating from part of each organ, we cannot at present exclude this last possibility.

### *Mechanism of alternative RNA processing*

The next question concerns the mechanism underlying the alternative processing of mRNA and its regulation. In the case of the immunoglobulin $\mu$ heavy chain gene it has been shown that two different mRNAs are transcribed, the membrane-associated ($\mu_m$) and the secreted ($\mu_s$) forms of the immunoglobulin heavy chain. The occurrence of each form is regulated during the course of pre-B cell maturation. $\mu_m$-encoding mRNAs are formed in early-stage B-lymphocytes whereas in mature plasma cells $\mu_s$ transcript formation is highly favoured. It has been shown that this switch occurs mainly at the level of differential use of polyadenylation sites (Early *et al.*, 1980; Peterson and Perry, 1986). Although differential polyadenylation usage has been shown in mRNAs from other genes (e.g. the *Drosophila doublesex* gene; Burtis and Baker, 1989) the situation with the HNF1 and vHNF1 genes is more similar to the immunoglobulin $\mu$ heavy chain gene.

Putative polyadenylation signals, AGUAAA and AUUAAA, preceding the poly(A) sequences have been found in the human HNF1-A and vHNF1-A transcripts, respectively (Bach *et al.*, 1990, 1991). Similarly to the AAUAAA consensus polyadenylation sequence, both sequences confer efficient polyadenylation activity *in vitro* (Sheets *et al.*, 1990). The putative polyadenylation signals present in the HNF1-B/C and vHNF1-C mRNAs (AACAGA and CAGAAA, respectively) are both more divergent from the consensus sequence. Comparison of these signals with different sequences which were tested *in vitro* (Sheets *et al.*, 1990; reviewed by Wickens, 1990) suggests that they confer only weak polyadenylation activity. Taken together with our results on the structure of the human HNF1 gene (Figure 3), this observation provides a model to explain our data. Two different pre-mRNAs are synthesized from a single gene by the use of two alternative polyadenylation sites. The different polyadenylation activities of the polyadenylation signals might reflect their involvement in the regulation of pre-mRNA synthesis. Regulation of the different pre-mRNA abundances would be principally achieved by competition between cleavage/polyadenylation, elongation and RNA splicing. The alternative splicing of these two mRNAs would extend even further the combinatorial set-up of the resulting proteins. Intron sequences that are excised in one class of transcripts are part of the translated sequences in the other class. The translation of the latter transcripts proceeds into these intronic sequences and terminates at the first stop codon encountered. This model is supported by results obtained from two different genes (HNF1 and vHNF1). It is consistent with the observations that transcripts of all isoforms are detectable in all tissues where the HNF1 or vHNF1 genes are transcribed and that the mRNA abundancies of the HNF1-B, HNF1-C and vHNF1-C mRNAs are always lower than that of the respective A isoforms in all tissues and cell lines analysed. The fact that the 5′ splice sites used alternatively in the HNF1 and vHNF1 genes correspond to 'normally' conserved donor sites favours the possibility that regulation of RNA processing occurs at the level of differential use of polyadenylation sites.

Since the polyadenylation sites of HNF1-B/C and vHNF1-C are quite divergent from the AAUAAA consensus, they are predicted to be very weak. The observation that the mRNA levels of these isoforms vary in different cell types and during organ development suggests that the polyadenylation and alternative splicing events may be differentially regulated. Several mechanisms which regulate alternative RNA processing have been proposed, including the existence of proteins that bind to specific RNA sequences and that are able to regulate specific use of polyadenylation signals and/or RNA splicing (McLaughlan *et al.*, 1989; Hedley and Maniatis, 1991; for review see Maniatis, 1991).

In conclusion, our results show that differential RNA processing of transcripts coding for transcription factors in different cell types and during development adds another level of complexity to the repertoire of transcription factors.

## Materials and methods

### *Screening the cDNA library and analysis of positive clones*

$1.2 \times 10^6$ plaques of a rat genomic library in λgt10 (generous gift of D.Lamy) were screened with the 511 bp *Nco*I fragment of the rat HNF1 cDNA (Chouard *et al.*, 1990) or the human vHNF1 cDNA-derived 517 bp *Pst*I–*Acc*I fragment (Bach *et al.*, 1991). The probes were radioactively labelled by random priming with a Multiprime DNA labelling system kit (Amersham) according to the manufacturer's instructions. The screenings were carried out as previously described (Bach *et al.*, 1990, 1991).

To prepare the DNA of positive phage isolates, the liquid culture DNA preparation method according to Sambrook *et al.* (1989) was applied. Inserts of positive clones were isolated and subcloned in the Bluescribe plasmid. The HNF1 and vHNF1 encoding regions of the inserts were sequenced directly from double-stranded plasmid DNA using the Sequenase DNA sequencing kit (USB) and successive oligonucleotides as primers.

### *Plasmid constructs*

HNF1-A, -B and -C expression vectors were constructed on the basis of the rat HNF1 expression vector under the control of the RSV LTR. The HNF1-A cDNA of clone HCL20 (Bach *et al.*, 1990), cut with *Xho*I and *Apa*LI and the *Sac*I–*Sca*I fragments of HNF1-B and -C (clones HCL13 and HCL24, respectively) were inserted into the RSV rat HNF1 vector after removal of most of the rat sequences by cleavage with *Sac*I and *Eco*NI. The *Apa*LI and *Eco*NI sites were made flush-ended by Klenow treatment according to Sambrook *et al.* (1989). The first 16 N-terminal amino acids of HNF1-A and the first 10 of HNF1-B and -C are still encoded by rat cDNA sequences. However, since the 37 N-terminal residues are identical in the rat and human HNF1 proteins (Bach *et al.*, 1990), the proteins encoded by the HNF1-A, -B and -C expression vectors are the authentic human proteins.

A full-length vHNF1-C in Bluescribe (VHCFL) was constructed by ligating the *Hin*dIII−*Apa*I fragment of clone HV17 (vHNF1-A) containing the N-terminal sequences of vHNF1-A into HV2 cut with *Hin*dIII and *Apa*I. The expression vectors for vHNF1-A and -C were constructed on the basis on the rat vHNF1-A expression vector under the control of the RSV LTR (Rey-Campos *et al.*, 1991). *Sac*I−*Dra*I and *Sac*I−*Rsr*II fragments of clones HV17 and VHCFL, respectively, were ligated in the RSVvHNF1 (rat) vector with most of the vHNF1 coding region deleted by *Sac*I and *Bal*I digestion. Before this ligation, the *Rsr*II end of the VHCFL fragment was made blunt-ended by Klenow treatment. Thus, the first 17 N-terminal amino acids of the vHNF1-A and -C expression vectors are encoded by the corresponding rat cDNA sequences but this does not affect the protein sequence, since the human and rat vHNF1 protein sequences are identical throughout this region (Bach *et al.*, 1991; Rey-Campos *et al.*, 1991). The correct sequence was verified in all these constructs by DNA sequencing.

### *In vitro transcription and translation*

The clones HCL16 (HNF1-A), HCL13 (HNF1-B), HCL24 (HNF1-C), HV17 (vHNF1-A) and VHCFL (vHNF1-C), subcloned in Bluescribe and all encoding full-length proteins, were used directly for *in vitro* transcription with a Stratagene transcription kit. HCL16 and HCL24 transcripts were initiated from the T7 promoter of a *Bam*HI and *Sca*I linearized plasmid, respectively, whereas the HCL13, HV17 and VHCFL RNAs were obtained from *Sca*I, *Dra*I and *Rsr*II linearized plasmids, respectively from the T3 promoter. *In vitro* translations were performed with Promega rabbit reticulocyte lysates and [$^{35}$S]methionine.

### *Cell lines and transient transfections*

The C33 human cervical carcinoma cell line (Yee *et al.*, 1985), the Caco-2 human colon carcinoma (Pinto *et al.*, 1983), the human breast cancer cell line MCF-7 (Soule *et al.*, 1973), the human keratinocyte cell line HaCaT (Boucamp *et al.*, 1988) and the human adrenocortical carcinoma cell line SW13 (Leibovitz *et al.*, 1973) were cultured in Dulbecco's modified Eagle's medium (DMEM) supplemented with 7% fetal calf serum. Transfections were done by the standard calcium phosphate coprecipitation procedure. For all transfections the amount of plasmid DNA containing an RSV promoter was adjusted using an RSV−luciferase plasmid.

Extracts of transfected C33 cells ($2 \times 10^7$ cells) were prepared from two 10 cm Petri dishes. Before plating these cells a coverslip was added to each plate for the immunocytochemical studies. After harvesting the transfected cells, they were split into three aliquots, two of which were used to prepare the total, cytoplasmic and nuclear cell extracts. The third aliquot was used for the CAT assays which were quantified on a PhosphorImager (Molecular Dynamics) after autoradiography.

### *Extract preparation from transfected cells*

For total cell extracts, an aliquot of transfected cells was directly added to 2 × SDS sample buffer before sonication and subsequent electrophoresis on a 10% SDS−polyacrylamide gel. Nuclear extracts from transfected cells were prepared from another aliquot essentially as previously described (Gorski *et al.*, 1986; Cereghini *et al.*, 1988) with some variations. Briefly, the pellet of transfected cells was resuspended in 500 μl of HNB solution (0.5 M sucrose, 15 mM Tris−HCl pH 7.5, 60 mM KCl, 0.25 mM EDTA, 0.125 mM EGTA, 0.5 mM spermidine, 0.15 mM spermine, 1 mM DTT, 0.5 mM PMSF, 5 μg/ml bastatin, 5 μg/ml aprotinin, 5 μg/ml pepstatin, 5 μg/ml leupeptin). 250 μl of 1% NP40 in HNB were added and the nuclei were pelleted by centrifugation at 6000 r.p.m. for 3 min. The supernatant was added to 2 × SDS sample buffer and kept as cytoplasmic extract. The nuclei were resuspended in 100 μl of NEB1 solution (20 mM HEPES pH 7.9, 25% glycerol, 1.5 mM $MgCl_2$, 0.14 M NaCl, 0.25 mM EDTA, 0.5 mM DTT, 0.5 mM PMSF, 5 μg/ml bastatin, 5 μg/ml aprotinin, 5 μg/ml pepstatin, 5 μg/ml leupeptin). 100 μl of 0.7 M NaCl in NEB1 buffer were added and incubated under slow agitation at 4°C for 30 min followed by 50 min centrifugation at 10 000 *g*. The supernatant was frozen in liquid nitrogen.

### *Western blots*

For Western blots, proteins from total, cytoplasmic and nuclear extracts of transfected cells were separated on a 10% SDS−polyacrylamide gel and blotted on to 0.2 μm nitrocellulose sheets (Bio-Rad) essentially as previously described (Towbin *et al.*, 1979). The nitrocellulose filters were incubated in 5% dried milk in PBS for 30 min and subsequently washed twice for 15 min in PBS supplemented with 0.05% Tween 20. The filters were then incubated with a polyclonal anti-HNF1 antibody, rHNt-283, in PBS/Tween 20 and 0.2% BSA for at least 1 h. After three washes in PBS/Tween 20 the membranes were hybridized to the $^{35}$S-labelled anti-rabbit antibody (Amersham) for 1 h. The filters were washed again three times in PBS/Tween 20 and air dried. After autoradiography the radioactivity of the specific bands normalized against background bands was quantified on a PhosphorImager instrument using ImageQuant software (Molecular Dynamics).

### *Immunodetection in transfected cells*

Cells were grown and transfected on 18 mm glass coverslips as described above. For the immunofluoresence studies, the cells on the coverslip were rinsed in PBS, fixed for 20 min in 95% ethanol and 5% acetic acid at −20°C, rinsed again in PBS and incubated with the primary anti-HNF1 antibody, rHNt-283, in PBST+10% FCS. The coverslips were rinsed three times in PBST and incubated with a Texas Red coupled donkey anti-rabbit antibody (Amersham) in PBST+10% FCS. The cells were rinsed first in PBS and then in distilled water, mounted in anti-fade medium (Citiflour Ltd, London) and viewed on a Zeiss Axiophot epifluorescence microscope. Kodak 2415 Technical Pan film was used for photographing results.

### *Gel retardations*

The gel retardation assays were performed with the *in vitro* expressed proteins or with nuclear extracts of tranfected cells essentially as described in Cereghini *et al.* (1988). The reactions were done in a total volume of 14 μl containing 10 mM HEPES, 4 mM $MgCl_2$, 0.1 mM EDTA, 1.5 mM spermidine, 15% glycerol, 1.5 μg poly(dI−dC), 1 μg of sonicated salmon sperm DNA and 0.2 ng of $^{32}$P-labelled probe. A double-stranded oligonucleotide, PE56, containing the sequence of the albumin proximal element (Cereghini *et al.*, 1988) was used as probe. The gel was exposed overnight and the radioactivity of the retarded bands was quantified on a PhosphorImager.

### *RT-PCR analysis*

Total RNA was isolated by the acid guanidinium thiocyanate−phenol−chloroform extraction method essentially as described by Chomczynski and Sacchi (1987). Cytoplasmic and polysomal RNA were prepared as described by Ausubel *et al.* (1987) and Berger and Kimmel (1987) respectively. The RT-PCR experiments were performed essentially as described in Montarras *et al.* (1989). Briefly, 3 μg of total RNA were reverse transcribed in a total volume of 60 μl. The conditions of the reverse transcription and PCR were essentially as described by Montarras *et al.* (1989). Each PCR was performed on 5 μl of the reverse transcription cDNA mixture in a total volume of 100 μl. Aliquots of 30 μl were taken at cycle 14, 16 or 18 (each cycle consisted of 30 s at 94°C, 30 s at 60°C and 1 min at 72°C) in an automated thermal cycler. The mixture of oligonucleotides A, B, C and D was used as primer set for the amplification of specific HNF1-A, -B and -C segments. To visualize better the specific HNF1-C band, a PCR was performed with only primers B and C. For specific vHNF1-A/B and -C amplifications, the primer set consisted of oligonucleotides E, F and G. 20 μl aliquots of the PCRs were electrophoresed on a 6% native polyacrylamide gel and blotted on to a Hybond N membrane (Amersham) according to the manufacturer's instructions. The probe for the different HNF1 isoforms was a fragment obtained by PCR on plasmid HCL16 with primers PHABC 1336 and PHABC 1689, encompassing an internal sequence common to all three amplified segments of HNF1-A, -B and -C and hence hybridizing equally well to all three HNF1 isoforms. Specific vHNF1 bands were visualized by a probe obtained by PCR with primers ΔVHC1 and ΔVHC2 on plasmid HV17 encompassing an internal sequence common to both amplified vHNF1-A/B and -C segments. The probes were $^{32}$P-labelled by random priming with an Amersham Megaprime DNA labelling kit and hybridized according to the manufacturer's instructions. The filters were exposed for various times and the radioactivity of the bands was quantified with a PhosphorImager instrument.

### *Oligonucleotides*

A: 5′-CCAACACAGGTGCCTCCACCCTGGT-3′
B: 5′-TTCTCCAAACCACGGGCTCTGGGA-3′
C: 5′-CCCTGCATCCATTGACAGCCAACC-3′
D: 5′-CCGTGTGGGTGTACTGGGCCACCT-3′
PHABC 1336: 5′-TGGCCTCCACGCAGGCACAGAGTGT-3′
PHABC 1689: 5′-CTCTGCAGCTGAGCCATGGTGGCCA-3′
E: 5′-ACGGCCTGGGCTCCAACTTGGTCACT-3′
F: 5′-AACCCTTAAACCAGATAAGATCCGT-3′
G: 5′-TTCAACCTCCTCCTGAGACTGAGATCAT-3′
H: 5′-TCTATCTTATGTACTGGGCTTCTGCAT-3′
ΔVHC1: 5′-AGGTCCGTGTCTACAACTGGTTT-3′
ΔVHC2: 5′-AAAATGGCCATCAACCTGACAGCTTGTTTGGAGGA-3′
PVH5: 5′-CTTCTTTTTCCGTCCTTGGAAAATG-3′
26: 5′-ATTAGAATTCAGAGCCACACATTATTTCCAGGG-3′

29: 5′-TATAGGATCCTTGGAGGAGTTGCTGCCATCCCC-3′
30: 5′-TTTTGAATTCGGTGTGTCATAGTCGTCGCCGTCCT-3′
PE56: 5′-TGTGGTTAATGATCTACAGTTA-3′ (double-stranded).

## Acknowledgements

We wish to thank D.Lamy for the gift of the human liver cDNA library, G.Butler-Browne from the Clinique Baudelocque for the human fetal tissue samples, C.Antignac from hôpital Necker for the gift of human kidney RNA, J.-P.Hardelin for the gift of human brain RNA and genomic DNA and G.Crabtree for the (β28)3-CAT construct. We are indebted to K.Friess for help in sequence analysis, T.Chouard for the rHNt-283 antibody, C.Pfarr for helpful advice in immunofluorescence and Western blot experiments, A.Scherf for help in antibody preparation, J.Ham and F.Tronche for critically reading the manuscript and S.Cereghini for valuable discussions. This work was supported by grants from the CEE BAP (0117F), INSERM, ARC, LNFCC and FRMF to M.Y. I.B. was a Boehringer Ingelheim Fonds fellow and then supported by the Académie des Sciences; Société de Secours des Amis des Sciences.

## References

Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (1987) *Current Protocols in Molecular Biology*. Greene Publishing Associates and Wiley-Interscience, New York.

Bach,I., Galcheva-Gargova,Z., Mattei,M.-G., Simon-Chazottes,D., Guénet,J.-L., Cereghini,S. and Yaniv,M. (1990) *Genomics*, **8**, 155–164.

Bach,I., Mattei,M.-G., Cereghini,S. and Yaniv,M. (1991) *Nucleic Acids Res.*, **19**, 3553–3559.

Bach,I., Pontoglio,M. and Yaniv,M. (1992) *Nucleic Acids Res.*, **20**, 4199–4204.

Baumhueter,S., Mendel,D.B., Conley,P.B., Kuo,C.J., Turk,C., Graves,M.K., Edwards,C.A., Courtois,G. and Crabtree,G.R. (1990) *Genes Dev.*, **4**, 372–379.

Berger,S.L. and Kimmel,A.R. (1987) *Methods Enzymol.*, **152**.

Blumenfeld,M., Maury,M., Chouard,T., Yaniv,M. and Condamine,H. (1991) *Development*, **113**, 589–599.

Boucamp,P., Petrussevska,R.T., Breitkreuz,D., Hornung,J., Markham,A. and Fusenig,N.E. (1988) *J. Cell Biol.*, **106**, 761–771.

Burtis,K.C. and Baker,B.S. (1989) *Cell*, **56**, 997–1010.

Cereghini,S., Blumenfeld,M. and Yaniv,M. (1988) *Genes Dev.*, **2**, 957–974.

Cereghini,S., Ott,M.-O., Power,S. and Maury,M. (1992) *Development*, **116**, 783–797.

Chomczynski,P. and Sacchi,N. (1987) *Anal. Biochem.*, **162**, 156–159.

Chouard,T., Blumenfeld,M., Bach,I., Vandekerckhove,J., Cereghini,S. and Yaniv,M. (1990) *Nucleic Acids Res.*, **18**, 5853–5863.

Courtois,G., Morgan,J.G., Campbell,L.A., Fourel,G. and Crabtree,G.R. (1987) *Science*, **238**, 688–692.

Descombes,P and Schibler,U. (1991) *Cell*, **67**, 569–579.

De Simone,V. and Cortese,R. (1991) *Curr. Opin. Cell Biol.*, **3**, 960–965.

De Simone,V., De Magistris,L., Lazzaro,D., Gerstner,J., Monaci,P., Nicosia,A. and Cortese,R. (1991) *EMBO J.*, **10**, 1435–1443.

Dobrzanski,P., Noguchi,T., Kovary,K., Rizzo,C.L., Lazo,P.S. and Bravo,R. (1991) *Mol. Cell. Biol.*, **11**, 5470–5478.

Early,P., Rogers,J., Davis,M., Calame,K., Bond,M., Wall,R. and Hood,L. (1980) *Cell*, **20**, 313–319.

Emens,L.A., Landers,D.W. and Moss,L.G. (1992) *Proc. Natl Acad. Sci. USA*, **89**, 7300–7304.

Finney,M. (1990) *Cell*, **60**, 5–6.

Foulkes,N.S. and Sassone-Corsi,P. (1992) *Cell*, **68**, 411–414.

Foulkes,N.S., Borelli,E. and Sassone-Corsi,P. (1991) *Cell*, **64**, 739–749.

Foulkes,N.S., Mellström,B., Benusiglio,E. and Sassone-Corsi,P. (1992) *Nature*, **355**, 80–84.

Frain,M., Swart,G., Monaci,P., Nicosia,A., Stämpfli,S., Frank,R. and Cortese,R. (1989) *Cell*, **59**, 145–157.

Gorski,K., Carneiro,M. and Schibler,U. (1986) *Cell*, **47**, 767–776.

Hedley,M.L. and Maniatis,T. (1991) *Cell*, **65**, 579–586.

Huet,C., Sahuquillo-Merino,C., Coudrier,E. and Louvard,D. (1987) *J. Cell Biol.*, **44**, 345–357.

Koenig,R.J., Lazar,M.A., Hodin,R.A., Brent,G.A., Larsen,P.R., Chin,W.W. and Moore,D.D. (1989) *Nature*, **337**, 659–661.

Kuo,C.J., Conley,P.B., Hsieh,C.-L., Francke,U. and Crabtree,G.R. (1990) *Proc. Natl Acad. Sci. USA*, **87**, 9838–9842.

Lazzaro,D., De Simone,V., DeMagistris,L., Lehtonen,E. and Cortese,R. (1992) *Development*, **114**, 469–479.

Leibovitz,A., McCombs,W.B., Johnston,D., McCoy,C.E. and Stinson,J.C. (1973) *J. Natl Cancer Inst.*, **51**, 691–697.

McLaughlan,J., Simpson,S. and Clements,J.B. (1989) *Cell*, **59**, 1093–1105.

Maniatis,T. (1991) *Science*, **251**, 33–34.

Mendel,D.B. and Crabtree,G.R. (1991) *J. Biol. Chem.*, **266**, 677–680.

Mendel,D.B., Hansen,L.P., Graves,M.K., Conley,P.B. and Crabtree,G.R. (1991a) *Genes Dev.*, **5**, 1042–1056.

Mendel,D.B., Khavari,P.A., Conley,P.B., Graves,M.K., Hansen,L.P., Admon,A. and Crabtree,G.R. (1991b) *Science*, **245**, 1762–1767.

Montarras,D., Pinset,C., Chelly,J., Kahn,A. and Gros,F. (1989) *EMBO J.*, **8**, 2203–2207.

Nicosia,A., Monaci,P., Tomei,L., DeFrancesco,R., Nuzzo,M., Stunnenberg,H. and Cortese,R. (1990) *Cell*, **61**, 1225–1236.

Olson,J., Lausten,L., Kärnström,U., Sjöström,H. and Noren,O. (1991) *J. Biol. Chem.*, **27**, 18089–18096.

Ott,M.-O., Rey-Campos,J., Cereghini,S. and Yaniv,M. (1991) *Mech. Dev.*, **36**, 47–58.

Peterson,M.L. and Perry,R.P. (1986) *Proc. Natl Acad. Sci. USA*, **83**, 8883–8887.

Pinto,M., Appay,M.D., Simon-Assmann,P., Chevalier,G., Dracopoli,N., Fogh,J. and Zweibaum,A. (1982) *Biol. Cell.*, **44**, 193–196.

Pinto,M., Robine-Leon,S., Appay,M.D., Kedinger,M., Triadou,N., Dussaulx,E., Lacroix,B., Simon-Assmann,P., Haffen,K., Fogh,J. and Zweibaum,A. (1983) *Biol. Cell.*, **47**, 323–330.

Raney,A.K., Easton,A.J., Milich,D.R. and McLachlan,A. (1991) *J. Virol.*, **65**, 5774–5781.

Rey-Campos,J., Chouard,T., Yaniv,M. and Cereghini,S. (1991) *EMBO J.*, **10**, 1445–1457.

Ringeisen,F., Rey-Campos,J. and Yaniv,M. (1993) *J. Biol. Chem.*, in press.

Roman,C., Cohn,L. and Calame,K. (1991) *Science*, **254**, 94–97.

Sambrook,J, Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual*. 2nd edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sheets,M.D., Ogg,S.C. and Wickens,M.P. (1990) *Nucleic Acids Res.*, **18**, 5799–5805.

Soule,H.D., Vasquez,J., Long,A., Albert,S. and Brennan,M.J. (1973) *J. Natl Cancer Inst.*, **51**, 1409–1413.

Sourdive,D.J.D., Chouard,T. and Yaniv,M. (1993) *Compt. Rend. Acad. Sci. Paris*, **316**, 385–394.

Struhl,K. (1991) *Neuron*, **7**, 177–181.

Tilghman,S.M. and Belayew,A. (1982) *Proc. Natl Acad. Sci. USA*, **79**, 5254–5257.

Tomei,L., Cortese,R. and DeFrancesco,R. (1992) *EMBO J.*, **11**, 4119–4129.

Towbin,H., Staehelin,T. and Gordon,J. (1979) *Proc. Natl Acad. Sci. USA*, **76**, 4350–4354.

Tripodi,M., Abott,C., Vivian,N., Cortese,R. and Lovell-Badge,R. (1991) *EMBO J.*, **10**, 3177–3182.

Tronche,F. and Yaniv,M. (1992) *Bioessays*, **14**, 579–587.

Tronche,F., Rollier,A., Bach,I., Weiss,M. and Yaniv,M. (1989) *Mol. Cell. Biol.*, **9**, 4759–4766.

Tronche,F., Rollier,A., Herbomel,P., Bach,I., Weiss,M. and Yaniv,M. (1990) *Mol. Biol. Med.*, **7**, 173–185.

Tyner,A.L., Godbout,R., Compton,R.S. and Tilghman,S.M. (1990) *J. Cell Biol.*, **110**, 915–927.

Wickens,M. (1990) *Trends Biochem. Sci.*, **15**, 277–281.

Xanthopoulos,K.G., Prezioso,V.R., Chen,W.S., Sladek,F.M., Cortese,R. and Darnell,J.E. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 3807–3811.

Yee,D., Krishnan-Hewlett,I., Baker,C.C., Schlegel,R. and Howley,P.M. (1985) *Am. J. Pathol.*, **119**, 361–366.

Yen,J., Wisdom,R.M., Tratner,I. and Verma,I.M. (1991) *Proc. Natl Acad. Sci. USA*, **88**, 5077–5081.

*Received on May 4, 1993; revised on July 27, 1993*

## Note added in proof

The sequence data reported here have been deposited in the EMBL/GenBank nucleotide sequence databases under the accession numbers X71346 (HNF1-B), X71347 (HNF1-C) and X71348 (vHNF1-C).