The EMBO Journal vol.10 no.6 pp.1445 – 1457, 1991

vHNF1 is a homeoprotein that activates transcription and forms heterodimers with HNF1

Javier Rey-Campos, Tanguy Chouard, Moshe Yaniv and Silvia Cereghini[1]

Unité des Virus Oncogènes, UA 1149 du CNRS, Département des Biotechnologies, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cedex 15, France
[1]Present address: Unité de Génétique de la Différentiation, Département de Biologie Moléculaire, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris Cedex 15, France

Communicated by M.Yaniv

vHNF1 and HNF1 are two nuclear proteins that bind to an essential element in the promoter proximal sequences of albumin and of many other liver-specific genes. HNF1 predominates in hepatocytes but is absent in dedifferentiated hepatoma cells. These cells contain vHNF1 but fail to express most of the liver traits. In the present work we have isolated cDNA clones for vHNF1 and found that it is a homeoprotein homologous to HNF1 in regions important for DNA binding. Unexpectedly, vHNF1 transactivated the albumin promoter in transfection experiments. Like the HNF1 mRNA, the vHNF1 message was found in kidney, liver and intestine although in different proportions. The fact that vHNF1 and HNF1 readily form heterodimers *in vitro* and the biochemical characterization of vHNF1/HNF1 heterodimers in nuclear extracts of kidney, liver and several cell lines, strongly argue that such heterodimers exist *in vivo*. Our results raise the possibility that heterodimerization between homeoproteins could be a common phenomenon in higher eukaryotes, which may have implications in the regulatory network sustained between these factors.

Key words: albumin expression/heterodimerization of homeoproteins/HNF1/transcription factors/vHNF1

Introduction

Development, cell fate determination and cell differentiation are complex phenomena that ultimately depend on switching on and off the expression of particular sets of genes. The mechanisms used for such controlled gene expression act predominantly at the transcriptional level. To a great extent, the expression of a given gene depends on the existence in the cell of the pertinent transcription factors (Maniatis *et al.*, 1987; Wasylyk, 1988). In many cases these factors not only directly contact the regulatory sequences of the genes but they may also interact with other regulatory factors which will eventually fine tune the final activation state of the gene (Benezra *et al.*, 1990; Keleher *et al.*, 1988; Kelleher *et al.*, 1990; reviewed by Ptashne, 1988, and Lewin, 1990). Transcription factors can be grouped into families, the members of which share particular structural features that are important for their functions. The homeoproteins are a family of proteins that share a common 60 amino acid domain, the homeodomain, that has been conserved throughout evolution from yeast to man (reviewed by Affolter *et al.*, 1990; Scott *et al.*, 1989; Holland and Hogan, 1988; Gehring, 1987). It has been shown in recent years that the homeoproteins act as transcription factors which control the expression of genes by binding to specific regulatory sequences (Levine and Hoey, 1988). The homeodomain mediates this binding. The three-dimensional structure of the archetypal homeodomain consists of three α helices folded in a compact shape (Qian *et al.*, 1989; Otting *et al.*, 1990) with characteristics reminiscent of prokaryotic repressors.

Hepatic Nuclear Factor 1 (HNF1) is a homeodomain-containing transcriptional regulator that is essential for the liver-specific expression of albumin and many other hepatic genes (Frain *et al.*, 1989; Baumhueter *et al.*, 1990; Chouard *et al.*, 1990; Bach *et al.*, 1990; and references therein). It recognizes a pseudopalindromic sequence (consensus g/aGTTAATNATTAACc/a) that is present in the promoter regions of many liver-specific genes. HNF1 is the most distant relative of the homeoproteins so far identified. It has a rather diverged homeodomain that contains extra sequences upstream of the putative helix III (Finney, 1990; Nicosia *et al.*, 1990; Chouard *et al.*, 1990). HNF1 binds to its target sequence as a dimer, which might be related to the dyad symmetry of its recognition sequence. The dimerization domain of HNF1 is located in the N-terminus of the molecule (Nicosia *et al.*, 1990; Chouard *et al.*, 1990) and is predicted to have an α helical structure that could mediate formation of a coiled coil structure between the monomers.

Although HNF1 promotes the hepatocyte-specific expression of many liver genes, its distribution *in vivo* does not correlate uniquely with the hepatic phenotype (Baumhueter *et al.*, 1990; Blumenfeld,M., Maury,M., Chouard,T., Yaniv,M. and Condamine,H., submitted). Thus, the expression of HNF1 is not restricted only to liver: kidney and intestine also have significant levels of HNF1 message and protein. In a more simple model with hepatoma cell lines, HNF1 follows precisely the differentiated liver phenotype: HNF1 is expressed in fully differentiated hepatoma cell lines, such as H4II and Fao or HepG2, but it is absent in the dedifferentiated variants H5 and C2, as well as in extinguished somatic cell hybrids, which fail to express most of the liver genes (Cereghini *et al.*, 1990). In turn, these HNF1-negative cells express a distinct factor that binds to DNA with the same sequence specificity. This factor has been designated vHNF1 (Baumhueter *et al.*, 1988) or vAPF (Cereghini *et al.*, 1988) (hereafter referred to as vHNF1) and was thought to be incapable of activating transcription from the albumin promoter (Cereghini *et al.*, 1988; Tronche *et al.*, 1989).

In the present work, we describe the isolation of cDNA clones for vHNF1 and show that it shares a high degree of homology with HNF1 in regions important for specific DNA binding. Unexpectedly, vHNF1 behaved as a transcriptional

© Oxford University Press

activator of the albumin promoter in transient co-transfection experiments. Moreover, the tissue distribution in the adult qualitatively resembles that of HNF1 albeit with different levels of expression. A clue for untangling the potential functional redundancy between these two factors comes from the finding that HNF1 and vHNF1 form heterodimers both *in vitro* and *in vivo*. As has been shown to happen between the yeast mating type factors α2 and a1 (Goutte and Johnson, 1988; Dranginis, 1990) and was suggested for the *b* polypeptides of *Ustilago maydis* (Schulz *et al.*, 1990), heterodimerization between homeoproteins may provide novel mechanisms for controlling homeoprotein action.

Results

Isolation of HNF1-related cDNA clones from H5 cells

HNF1 was characterized as a nuclear factor involved in the transcription of a group of liver-specific genes. It is absent in dedifferentiated derivatives of hepatoma cells and in extinguished somatic cell hybrids. Instead, these cells contain a distinct protein (vAPF or vHNF1) with lower molecular weight that recognizes the same sequences as HNF1 (Cereghini *et al.*, 1988; Baumhueter *et al.*, 1988). It was postulated that vHNF1 could either be a modified form of HNF1 or alternatively it could be encoded by another gene. Since Northern blot analysis (Cereghini *et al.*, 1990) and RNase protection experiments (Baumhueter *et al.*, 1990) failed to detect HNF1 mRNA in cells containing vHNF1, it became plausible that vHNF1 was encoded by a distinct gene. When hybridization was repeated with a probe covering the homeobox of HNF1 (a 511 bp *Nco*I fragment from 334 to 845 with the numbering starting at the ATG) under low stringency conditions, a band of ~2.9 kb was observed in the RNA samples from C2 and H5 dedifferentiated cells (Figure 1). We then used the same probe to screen, under low stringency conditions, 8×10^5 plaques from an unamplified λgt10 cDNA library prepared from H5 cells. Fifty positive plaques were identified in the primary screening. Ten of the strongest ones were further purified and confirmed as real positive plaques. Their inserts ranged from 0.9 to 3 kb in length. Preliminary sequence analysis performed with specific primers derived from the homeobox of HNF1 showed strong sequence similarity to HNF1 in all of these clones. The full sequence of one of them, named tentatively vHNF1-A, was determined by using serial deletions with exonuclease III. Figure 2 shows the sequence of 2328 bp of the cDNA insert as well as the deduced amino acid sequence of the longest open reading frame. This starts at nucleotide 31 and extends along 1671 nucleotides to position 1701, followed by an in-frame termination codon, TGA. Other clones with longer 5′ untranslated regions showed the existence of a TAA termination codon (position −72) in frame with the first ATG, suggesting that this ATG is the real initiation codon. Thus, this cDNA can code for a polypeptide of 557 amino acids with an estimated molecular weight of 62 kd. The 3′ non-coding sequence contains several putative polyadenylation signals.

Figure 3 shows the alignment of the amino acid sequences of clone *vHNF1-A* with that of *HNF1*. Three strictly conserved segments stand out of the overall homology displayed between both sequences. The first segment extends from amino acids 1 to 32 and corresponds to the α helical

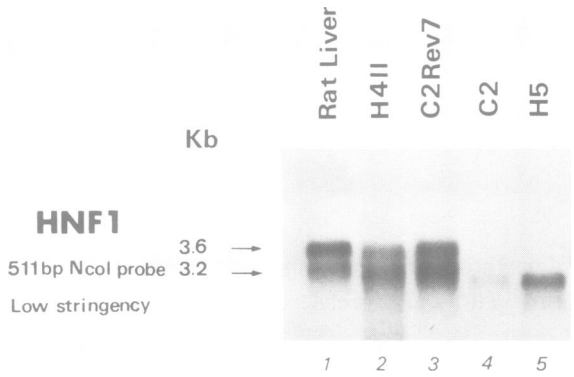

Fig. 1. Low stringency hybridization of RNA blots from liver and hepatoma cells with the homeobox of HNF1. Poly(A)$^+$ RNA from rat liver and differentiated hepatoma cells (H4II and C2Rev7) or dedifferentiated variants (H5 and C2) was separated on denaturing agarose gel and, after transfer to nylon membranes, hybridized under low stringency conditions with an *Nco*I fragment containing the homeobox of HNF1. The two HNF1 mRNA species of 3.6 kb and 3.2 kb present in liver and differentiated cells (see also Figure 6) differ in the length of their 3′ untranslated region due to utilization of distinct polyadenylation sites (Chouard *et al.* (1990)).

region shown to be indispensable for HNF1 dimerization (Nicosia *et al.*, 1990; Chouard *et al.*, 1990). The second concerns residues 89−180 of vHNF1-A and is equivalent to the B domain of HNF1 (Nicosia *et al.*, 1990). The third conserved region spans 90 amino acids (229−318 in vHNF1-A and 197−286 in HNF1) and contains the atypical homeodomain; only nine out of these 90 residues are different, six of which are conservative changes. Secondary structure predictions and alignment with conserved residues in homeodomains suggest that, like HNF1, vHNF1-A also contains, upstream of the helix III of the homeodomain, an extra segment of 18−21 amino acids distinctive for the HNF1 homeodomain among all the homeoproteins (Finney, 1990; Nicosia *et al.*, 1990; Chouard *et al.*, 1990). The conservation of this extra segment suggests that it might have some important role in the action of both factors, which has made their homeodomains evolve to the atypical structure they may have today (Finney, 1990; Nicosia *et al.*, 1990).

The sequence homology between vHNF1-A and HNF1 declines towards the C-terminal part of the sequences, even though, as in HNF1, this region is rich in serine and threonine (26% of the residues) (Frain *et al.*, 1989; Chouard *et al.*, 1990). They contain several short segments of homology although the alignment requires several gaps. Finally, the glycine/proline- rich stretch just downstream of the homeodomain of HNF1 (amino acids 288−308 in HNF1) is absent in vHNF1-A. This region has been proposed to function as a potential hinge structure in the HNF1 molecule (Chouard *et al.*, 1990) and to be essential for transcriptional activation by HNF1 (Nicosia *et al.*, 1990).

A major difference in the sequence of the N-terminal part of vHNF1-A with respect to HNF1 is an extra 26 amino acid long segment that is absent in HNF1 (Figure 3). This sequence extends from residue 183 to 208 of vHNF1-A (underlined in Figure 2). In fact, we have found a second set of clones that do not contain this extra segment. The prototype of these clones was called *vHNF1-B*. This clone has longer 5′ and 3′ untranslated sequences and consequently it uses a different polyadenylation site beyond that used in *vHNF1-A*. Except for this and the absence of the 78 nt

```
                                M  V  S  K  L  T  S  L  Q  Q  E  L  L  S  A  L  L  S  S  G  V  T  K  E  V  L  I  Q  A  L    30
CACAACCCCTTCTTTTTCCGTTCTTGGAAAATGGTGTCCAAGCTCACGTCTCTCCAGCAAGAACTCCTGAGTGCCCTGCTGAGCTCCGGAGTCACCAAGGAAGTGCTGATCCAGGCCTTG   120

 E  E  L  L  P  S  P  N  F  G  V  K  L  E  T  L  P  L  S  P  G  S  G  A  D  L  D  T  K  P  V  F  H  T  L  T  N  G  H  A    70
GAGGAATTACTGCCGTCCCCGAATTTCGGGGTGAAGCTGGAGACACTGCCCCTGTCCCCCGGGAGCGGGGCGGATCTCGACACCAAGCCGGTTTTCCATACTCTCACCAATGGCCACGCC   240

 K  G  R  L  S  G  D  E  G  S  E  D  G  D  D  Y  D  T  P  P  I  L  K  E  L  Q  A  L  N  T  E  E  A  A  E  Q  R  A  E  V   110
AAGGGCCGCTTGTCTGGGGACGAGGGCTCAGAGGACGGCGACGACTATGACACTCCTCCCATCCTCAAAGAGCTCCAGGCGCTCAACACCGAGGAGGCCGCGGAGCAGCGGGCCGAGGTG   360

 D  R  M  L  S  E  D  P  W  R  A  A  K  M  I  K  G  Y  M  Q  Q  H  N  I  P  Q  R  E  V  V  D  V  T  G  L  N  Q  S  H  L   150
GACCGGATGCTCAGTGAGGACCCATGGAGGGCCGCCAAAATGATCAAGGGATACATGCAACAGCACAACATCCCCCAGAGGGAGGTGGTTGATGTCACAGGCCTGAACCAATCCCACCTT   480

 S  Q  H  L  N  K  G  T  P  M  K  T  Q  K  R  A  A  L  Y  T  W  Y  V  R  K  Q  R  E  I  L  R  Q  F  N  Q  T  V  Q  S  S   190
TCTCAACACCTCAACAAGGGCACCCCCATGAAGACCCAGAAGCGAGCTGCCCTGTACACCTGGTACGTCAGAAAGCAACGGGAGATCCTCCGACAGTTCAACCAGACAGTCCAGAGCTCT   600

 G  N  M  T  D  K  S  S  Q  D  Q  L  L  F  L  F  P  E  F  S  Q  Q  N  Q  G  P  G  Q  S  E  D  A  C  S  E  P  T  N  K  K   230
GGAAACATGACAGACAAAAGCAGTCAGGATCAGCTGCTATTTCTCTTTCCAGAGTTCAGTCAACAGAACCAGGGCCCTGGGCAGTCGGAGGACGCCTGCTCTGAGCCCACCAACAAGAAG   720

 M  R  R  N  R  F  K  W  G  P  A  S  Q  Q  I  L  Y  Q  A  Y  D  R  Q  K  N  P  S  K  E  E  R  E  A  L  V  E  E  C  N  R   270
ATGCGCCGCAACCGGTTCAAATGGGGGCCCGCATCCCAGCAAATTTTGTACCAGGCCTACGACCGGCAAAAGAATCCCAGCAAGGAAGAGAGGGAGGCCTTAGTGGAGGAGTGTAACAGG   840

 A  E  C  L  Q  R  G  V  S  P  S  K  A  H  G  L  G  S  N  L  V  T  E  V  R  V  Y  N  W  F  A  N  R  R  K  E  E  A  F  R   310
GCAGAATGTTTGCAACGAGGGGTCTCCCCCTCCAAAGCCCATGGCCTAGGCTCCAACTTGGTCACGGAGGTCCGTGTCTACAACTGGTTTGCAAACCGCCGGAAGGAAGAGGCATTCAGA   960

 Q  K  L  A  M  D  A  Y  S  S  N  Q  T  H  N  L  N  P  L  L  T  H  G  S  P  H  H  Q  P  S  S  S  P  P  N  K  L  S  G  V   350
CAGAAGCTGGCCATGGATGCCTATAGCTCCAACCAGACGCACAACCTGAACCCCCTGCTCACCCATGGCTCCCCTCACCACCAGCCAAGCTCCTCTCCACCCAACAAGTTGTCAGGAGTG  1080

 R  Y  S  Q  P  G  N  N  E  V  T  S  S  S  T  I  S  H  H  G  N  S  A  M  V  T  S  Q  S  V  L  Q  Q  V  S  P  A  S  L  D   390
CGCTACAGCCAACCGGGAAACAATGAGGTCACTTCCTCTTCGACAATCAGTCACCATGGTAACAGTGCCATGGTGACCAGCCAGTCAGTTTTACAACAAGTCTCCCCGGCCAGCCTGGAC  1200

 P  G  H  S  L  L  S  P  D  S  K  M  I  S  V  S  G  G  G  L  P  P  V  S  T  L  T  N  I  H  S  L  S  H  H  N  P  Q  Q  S   430
CCAGGCCACAGTCTCCTCTCACCTGACAGTAAAATGATCTCGGTGTCTGGAGGAGGACTGCCCCCGGTCAGCACCTTGACGAATATCCACAGCCTCTCCCACCACAATCCCCAGCAATCT  1320

 Q  N  L  I  M  T  P  L  S  G  V  M  A  I  A  Q  S  L  N  T  S  Q  A  Q  G  V  P  V  I  N  S  V  A  S  S  L  A  A  L  Q   470
CAAAACCTCATCATGACTCCCCTGTCCGGAGTCATGGCCATTGCACAGAGCCTCAACACCTCCCAAGCCCAGGGTGTCCCAGTCATCAACAGTGTGGCTAGCAGCCTGGCAGCCCTACAG  1440

 P  V  Q  F  S  Q  Q  L  H  S  P  H  Q  Q  P  L  M  Q  Q  S  P  G  S  H  M  A  Q  Q  P  F  M  A  A  V  T  Q  L  Q  N  S   510
CCCGTCCAGTTCTCTCAACAGCTGCACAGCCCTCACCAGCAGCCCCTCATGCAGCAGAGCCCAGGCAGTCACATGGCCCAGCAGCCCTTCATGGCTGCGGTGACTCAGCTACAGAACTCC  1560

 H  M  Y  A  H  K  Q  E  P  P  Q  Y  S  H  T  S  R  F  P  S  A  M  V  V  T  D  T  S  S  I  N  T  L  T  S  M  S  S  S  K   550
CACATGTATGCACATAAGCAGGAGCCCCCTCAGTATTCCCACACCTCCCGGTTCCCATCTGCAATGGTGGTCACAGATACCAGTAGCATCAATACCCTCACCAGCATGTCTTCCAGTAAA  1680

 Q  C  P  L  Q  A  W  *
CAGTGTCCACTGCAAGCCTGGTGATACCCACATAACCACTCACTTTGTGCACAGCAACGAGGACCCCATTTTCCACACCATCAGTTTCTGGGCATCTGTCATAGAGAAGTCCAGCGACCT  1800

GACAGGCACCTGGGAAGCCAGGAGAGATCCCTGCTTACCTGACATCTGCCGGGGACCTCCGACAAGCCCAAGCCACTCTCAGGAGGTGCAGCCCGAAGCCCAGCTTCTCTTCTTTGCAGT  1920
ATTGTCGTAATGCCTCTCCCAAGATGCCAAGTGCCCTTGTTTCTCCCAGAGGTGGCCGGTGCAAAATGGTGCAACAGGAATCGAGAAGCCCATGGAGTTTCCACTGCAATCCGTCATTGA  2040
ACAAACTGATGCAAAAACTTGAATCTGTTACTGAAATAAGATGAGGACCGAGGAGGGTGTGTGCTACTGAACTGAGCCAGACACTGTAAATAACTGACAGACCCCCCTCTCCCATCCCAG  2160
TGTCTCAAGATTTCTTCTAAAGAAGTAAATTTGTCAATGGGTGTAAACTATCAACTACTGTATTAAGTGCAATTTTTCCTTCCACAGATATCTACCCCTGCCCGATATAATACTAAAGTG  2280
TCTATCACTTTTCTTGTAAAGGTAGAGTTAAATTCCAAAAAAAAAAAA...2328
```

Fig. 2. Nucleotide and amino acid sequence of vHNF1-A cDNA. Nucleotide sequence of vHNF1-A and deduced amino acid sequence. Amino acids are shown in single letter code above the nucleotide sequence. Numbers indicate the last nucleotide, or amino acid, in each line. The 78 nucleotide sequence underlined corresponds to the insertion found in the vHNF1-A cDNA clone with respect to the vHNF1-B clone. A star indicates the stop codon. The accession number in the EMBL database for the sequence shown is X56546.

insertion, the nucleotide sequence of *vHNF1-B* is identical to that of *vHNF1-A*. These observations suggest that *vHNF1-A* and *vHNF1-B* originate by alternative splicing of a single gene. Consistent with this hypothesis, the nucleotide sequence of both ends of this insertion match consensus donor and acceptor splicing sites. Since *vHNF1-B* and *vHNF1-A* behaved similarly in all the experiments done, the actual role of this segment remains to be established. We only describe here the results obtained with the *vHNF1-A* clone.

The following experiments were designed to verify whether *vHNF1-A* is indeed a genuine cDNA clone coding for the DNA binding activity present in the dedifferentiated cell H5 and previously designated as vHNF1.

vHNF1-A protein binds to the proximal element of the rat albumin promoter

To test the DNA binding properties of the vHNF1-A protein we inserted the intact coding region of the *vHNF1-A* cDNA into a pGEM1-derived T7 expression vector (see Materials and methods). We then performed gel retardation assays with the labelled albumin PE probe and the *in vitro* translated vHNF1-A protein. As can be seen in Figure 4A, vHNF1-A protein binds to PE specifically as confirmed by the competition experiments: binding was fully inhibited by an excess of unlabelled wild type PE oligonucleotide (PE), it was only slightly competed by PE mutants mut3, mut4 and DS12 and finally, mutant DS34 was unable to compete at all. This behaviour closely ressembles that of both HNF1

Fig. 3. Sequence homology between HNF1 and vHNF1-A. **A**. Alignment of the amino acid sequences of rat HNF1 and vHNF1-A. A Needelman and Wunch algorithm was used to align both sequences using the amino acid equivalence matrix described by Feng *et al.* (1985). Gaps in both sequences were allowed to optimize the alignment and are indicated by dashes. Only identities are shown. **B**. Scheme of homologous regions between HNF1 and vHNF1-A and vHNF1-B. Regions I, II and III are highly conserved and correspond to the dimerization domain, the B domain and the homeodomain respectively (Nicosia *et al.*, 1990; Chouard *et al.*, 1990). Region IV, which includes the putative hinge region of HNF1, has no homologue in vHNF1. Region V contains ADI and ADII transactivation domains of HNF1 (Nicosia *et al.*, 1990).

and vHNF1 previously described (Cereghini *et al.*, 1988).

In the gel retardation assay, the *in vitro* translated vHNF1-A protein formed a complex with PE that migrated faster than the complex obtained with vHNF1 from H5 cell extracts. This different mobility may be explained if the vHNF1 molecule undergoes post-translational modifications *in vivo*. For example, it could be glycosylated as occurs with HNF1 (Lichtsteiner and Schibler, 1989). To test this hypothesis we transiently transfected C33 human epithelial cells, which express neither vHNF1 nor HNF1, with a plasmid construct containing the coding sequence of *vHNF1-A* under control of the Rous sarcoma virus (RSV) LTR. Gel retardation assays performed with nuclear extracts of these transfected cells and labelled PE, revealed a complex

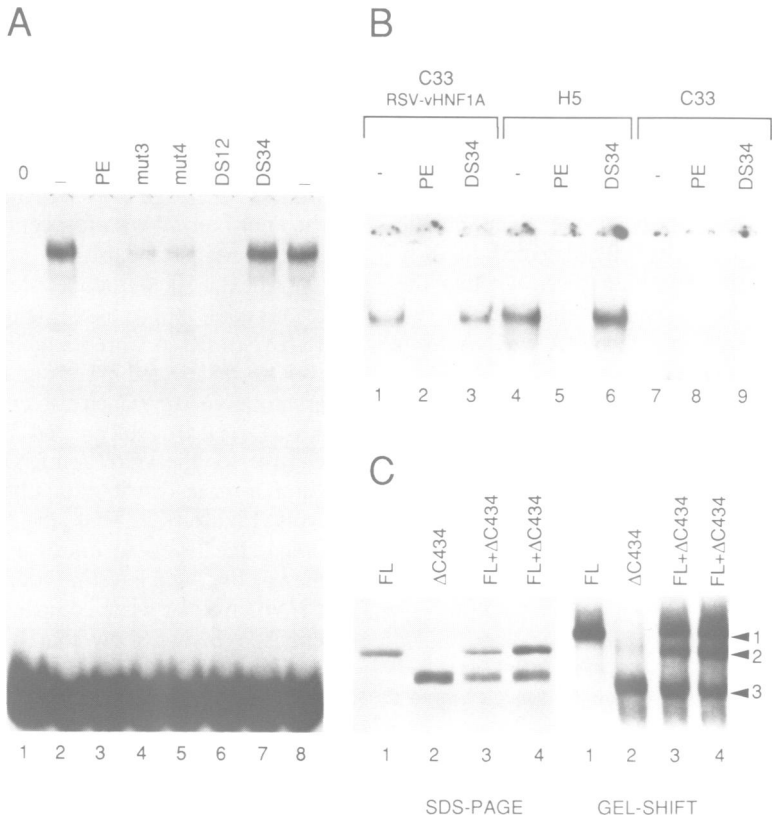

Fig. 4. Binding of vHNF1-A to the albumin proximal element. **A**. DNA binding specificity of vHNF1-A. Gel shift assay performed with *in vitro* translated vHNF1-A and labelled PE oligodeoxynucleotide in the absence (lanes 2 and 8) or the presence of a 12-fold excess of unlabelled competitor PE (lane 3), and mutant oligodeoxynucleotides mut3 (lane 4), mut4 (lane 5), and DS12 (lane 6). Mutant oligonucleotide DS34 (lane 7) was used in a 54-fold excess. Lane 1 corresponds to the control translation mixture to which no RNA was added. **B**. Gel retardation assay performed with nuclear extracts of H5 cells, C33 cells and C33 cells transfected with an expression vector containing vHNF1-A cDNA under the control of the RSV-LTR. Lanes 1, 4 and 7 were done in the absence of competitor DNA. Assays done with 12-fold excess of competitor unlabelled PE oligodeoxynucleotide or 54-fold excess of mutant DS34 are shown in lanes 2, 5 and 8 and 3, 6 and 9 respectively. **C**. vHNF1-A binds to PE as a dimer. The left part of the figure shows the SDS−PAGE analysis of *in vitro* translation mixtures programmed with *in vitro* synthesized RNA for the full length (FL, lane 1) or a truncated (ΔC434, lane 2) vHNF1-A protein. Lanes 3 and 4 correspond to *in vitro* translation mixtures programmed with different ratios of both RNAs. At the right are shown the gel retardation assays with labelled PE performed with the same *in vitro* translation mixtures. The arrows at the right indicate the position of the full length homodimer (1), the heterodimer between the full length protein and ΔC434 (2) and the truncated vHNF1-A homodimer (3).

with an electrophoretic mobility identical to the vHNF1 complex of H5 cell extracts (Figure 4B). This complex was not observed when the assay was carried out with nuclear extracts from mock transfected cells. These data strongly suggest that *vHNF1-A* can actually code for the vHNF1 activity of H5 cells and that post-translational modifications may occur *in vivo*.

It has been shown that HNF1 is present as a dimer in solution and that it binds to DNA in this form (Chouard *et al.*, 1990; Nicosia *et al.*, 1990). As the N-terminus of vHNF1-A exhibits considerable homology with the putative dimerization domain of HNF1, we wondered whether it also binds to DNA as a dimer. Figure 4C shows a gel retardation experiment done with labelled PE and *in vitro* translation mixtures directed with RNAs encoding the full size vHNF1-A protein and a truncated version ΔC434 (434 residues) of the protein that lacks 123 residues from the C-terminus. The co-translation of these RNAs yielded a mixture of two proteins of 68 and 48 kd respectively. The gel retardation assay made with this mixture produced three complexes: those corresponding to the translation products of the intact and truncated proteins, and a band with an intermediate mobility. This intermediate complex conformed well with the expected heterodimer between the truncated and the full length vHNF1-A proteins.

Anti-vHNF1-A antibodies specifically recognized vHNF1

To verify further that *vHNF1-A* represents an authentic vHNF1 cDNA clone we prepared polyclonal antibodies against the vHNF1-A protein. Rabbits were immunized with a fusion protein between glutathione-S-transferase (Smith and Johnson, 1988) and a fragment of 155 amino acids from the C-terminal part of vHNF1-A (see Materials and methods). This region of the molecule shows low homology with HNF1 and the antibodies raised specifically recognized vHNF1-A protein, and not HNF1, in immunoprecipitation experiments (not shown).

When this serum was included in the gel retardation assays with labelled PE and H5 cell nuclear extracts (Figure 5), it fully displaced the vHNF1 complex. In contrast, it did not affect the HNF1 complex obtained with liver extracts, confirming that this serum recognized the vHNF1 complex specifically. In turn, antibodies raised against a specific peptide of HNF1 did not react with the vHNF1 complex, whereas they displaced the HNF1 complex of liver extracts.

Taken together, these data strongly support our conclusion that *vHNF1-A* is a bona fide vHNF1 cDNA clone.

The tissue distribution of vHNF1 resembles that of HNF1 in the adult

vHNF1 DNA binding activity was first found in extinguished somatic cell hybrids and in dedifferentiated hepatoma cell lines, which fail to express most of the liver-specific functions, whereas it was not initially detected in fully differentiated hepatoma cells. In order to see whether the *vHNF1* transcripts followed a similar distribution, we performed Northern blot analyses of poly(A)$^+$ RNA from these cell lines, using a full length *vHNF1-A* cDNA probe. A 2.9 kb RNA band was detected in the dedifferentiated hepatoma cell lines C2 and H5 (Figure 6B) as well as in extinguished somatic cell hybrids (not shown). Unexpectedly, the same RNA was also present in H4II and C2Rev7 differentiated hepatoma cells (Figure 6B). The levels of *vHNF1* message were rather similar in all these cell lines, when normalized with respect to the *GAPDH* mRNA, used as an internal standard.

We next looked for the presence of *vHNF1* mRNA in various adult mouse tissues (Figure 6C). The highest levels of *vHNF1* mRNA were found in kidney. Significantly lower levels were detected in liver and intestine. The human colon carcinoma cell line Caco also contained low levels of *vHNF1* message. Heart, spleen and FR3T3 rat fibroblasts were negative for the *vHNF1* mRNA. *HNF1* mRNA displayed a similar tissue distribution, although with different relative levels (Figure 6C). Thus, the *HNF1* message was found at comparable levels to those of *vHNF1* mRNA, in kidney, whereas liver and intestine contained a much higher amount of *HNF1* mRNA when compared with *vHNF1*. Experiments described below show that both liver and differentiated hepatoma cells contain low amounts of vHNF1 protein. This contrasts with our apparent previous failure to detect vHNF1 in these cells (Cereghini *et al.*, 1988). In fact, vHNF1 activity was present but it was misinterpreted as degradation products of HNF1.

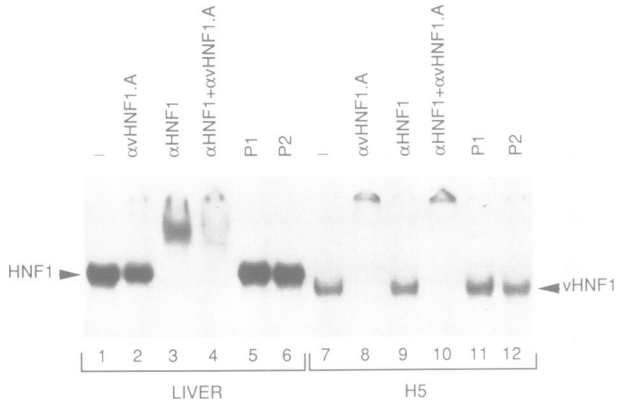

Fig. 5. Specificity of the anti-vHNF1-A antibodies. Gel retardation assay performed with liver (lanes 1–6) or H5 cell (7–12) nuclear extracts and ^{32}P-labelled PE probe with 10 μl of anti-vHNF1-A serum (lanes 2 and 8), 5 μl of anti-HNF1 serum (lanes 3 and 9), or with both sera (lanes 4 and 10). Lanes 5, 6, 11 and 12 correspond to assays done with the same amounts of the respective preimmune sera. In lanes 1 and 7 no serum was added.

vHNF1 and HNF1 can form heterodimers

Since the N-terminal part of vHNF1 shows homology with the dimerization domain of HNF1 and since both factors can co-exist within the same cell, we investigated whether they could form heterodimers and whether such heterodimers might exist *in vivo*. As a first step, we tested if vHNF1 and

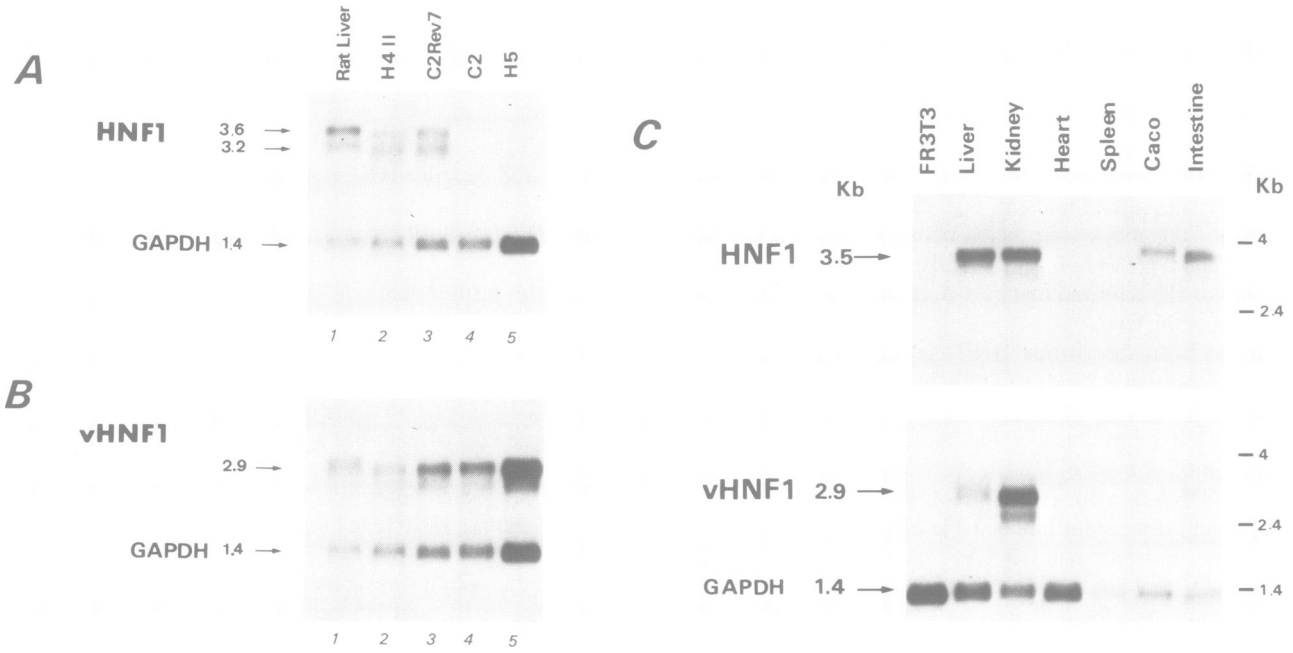

Fig. 6. Northern blot analyses with HNF1 and vHNF1-A probes. Approximately 3 μg of poly(A)$^+$ RNA from different cell lines and tissues was electrophoresed, transferred to nylon membranes and hybridized with the HNF1 cDNA (panels A and C) (Chouard *et al.*, 1990) or the vHNF1-A full length cDNA (panels B and C). As an internal standard, a *GAPDH* probe was included in the hybridization. Numbers to the left indicate the estimated sizes of the RNA bands. The high degree of homology between rat and human *HNF1* and *vHNF1* genes (Bach *et al.*, 1990; I.Bach, personal communication) allowed the use of rat probes for detecting the respective human messages from human cell.

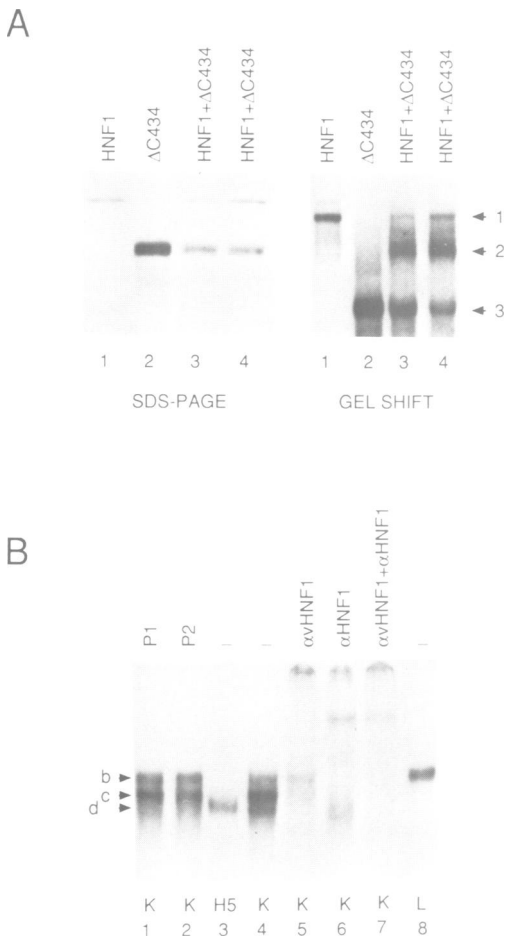

Fig. 7. Heterodimerization between HNF1 and vHNF1. **A.** Full length HNF1 and the truncated vHNF1-A, ΔC434, were translated *in vitro* separately (lanes 1 and 2) or together (lanes 3 and 4) and analysed in denaturing SDS–polyacrylamide gels. 4 μl of the same translation mixtures were used in a gel shift assay with ^{32}P-labelled PE probe. The numbers to the right indicate the three complexes obtained corresponding to the HNF1 homodimer (1), the heterodimer between HNF1 and ΔC434 (2) and the ΔC434 homodimer (3). **B.** Gel retardation assay performed with rat kidney nuclear extracts (K) and ^{32}P-labelled PE probe with 10 μl of anti-vHNF1-A serum (lane 5), 5 μl of anti-HNF1 serum (lane 6) or both sera (lane 7). Lanes 1 and 2 correspond to assays done with similar amounts of the respective preimmune sera. Assays in lane 3 (H5 cells extracts), lane 4 (kidney extracts) and lane 8 (liver extract) are without any serum added. The letters to the left indicate the position of the HNF1 homodimers (b), HNF1/vHNF1 heterodimers (c) and vHNF1 homodimers (d).

HNF1 were able to form heterodimers *in vitro* (Figure 7A). RNAs coding for HNF1 and the truncated version of vHNF1-A described above (ΔC434), were co-translated *in vitro*. SDS–PAGE of the translated products showed two distinct bands of 85 and 48 kd respectively. When incubated with radioactive PE probe, this translation mixture gave rise to three different complexes in gel retardation assays. The upper and the lower complexes coincided with the HNF1 and the truncated vHNF1-A (ΔC434) homodimers. The intermediate complex had an electrophoretic mobility compatible with that expected for an HNF1/truncated vHNF1-A (ΔC434) heterodimer. Similar results were obtained when both HNF1 and vHNF1-A full size proteins were used in these assays (Figure 8, lane 9). In order to check whether the HNF1/vHNF1-A heterodimers may form

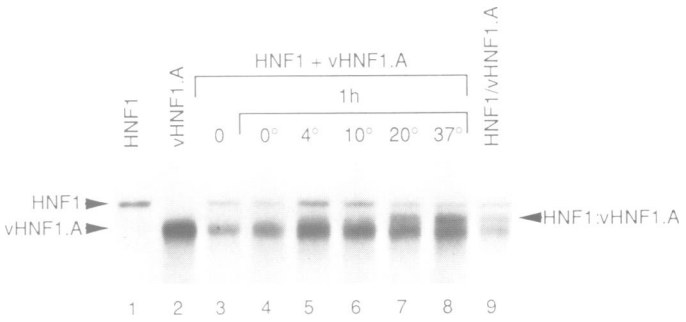

Fig. 8. Subunit exchange of HNF1 and vHNF1 homodimers. Full length HNF1 and vHNF1 were translated *in vitro* separately. 2 μl of each translation mixture were mixed and incubated for 1 h at 0, 4, 10, 20 or 37°C (lanes 4, 5, 6, 7 and 8 respectively) before using them in a gel shift assay with ^{32}P-labelled PE probe. Lanes 1 and 2 correspond to both translation mixtures assayed separately. In lane 3, both translation mixtures were mixed inmediately before performing the gel shift assay. Lane 9 corresponds to the co-translation of HNF1 and vHNF1. The complexes corresponding to the HNF1 and vHNF1 homodimers and the HNF1:vHNF1 heterodimer are indicated by arrows.

once the respective homodimers have been formed, we mixed HNF1 and vHNF1-A, which had been translated *in vitro* separately, and incubated them at different temperatures for 1 h. These mixtures were then used in a gel retardation assay with radioactive PE probe. HNF1/vHNF1-A heterodimers were only observed when the mixtures were incubated above 10°C (Figure 8). This indicates that HNF1 and vHNF1-A homodimers are relatively stable at low temperatures, and subunit exchange between both homodimers only occurs when the temperature is increased. Evidence for heterodimer formation was also obtained by transfecting C33 cells with a mixture of RSV-HNF1 and RSV-vHNF1 expression constructs. Gel retardation assays performed with nuclear extracts prepared from these cells gave three bands corresponding to both homodimers and the heterodimer (not shown).

Since high levels of both *vHNF1* and *HNF1* mRNA were detected in kidney we wondered whether vHNF1/HNF1 heterodimers might exist in this organ. To this end, we performed gel retardation assays with kidney nuclear extracts and labelled PE probe (Figure 7B). In addition to complexes with electrophoretic mobilities similar to those of HNF1 and vHNF1 homodimers, a third complex with intermediate mobility was observed. This pattern parallels that obtained with the *in vitro* translated or co-transfected vHNF1 and HNF1 and suggests that this middle complex might correspond to vHNF1/HNF1 heterodimers. To verify this possibility we tested if specific antibodies against cloned vHNF1 recognized this complex. As shown in Figure 7B, when the vHNF1-A antiserum was added to the mixture of kidney nuclear extract and radioactive PE probe, the presumed vHNF1/HNF1 heterodimer was effectively displaced. As expected, the lowest complex, presumably vHNF1 homodimers, was also displaced, whereas no change was observed in the band corresponding to the HNF1 homodimer. Reciprocally, HNF1-specific antibodies raised against a peptide within the C-terminal part of HNF1, specifically displaced the HNF1 homodimers and also the putative HNF1/vHNF1 heterodimers. When both antisera were used together, all three complexes were displaced, suggesting that no other PE binding activity was present in

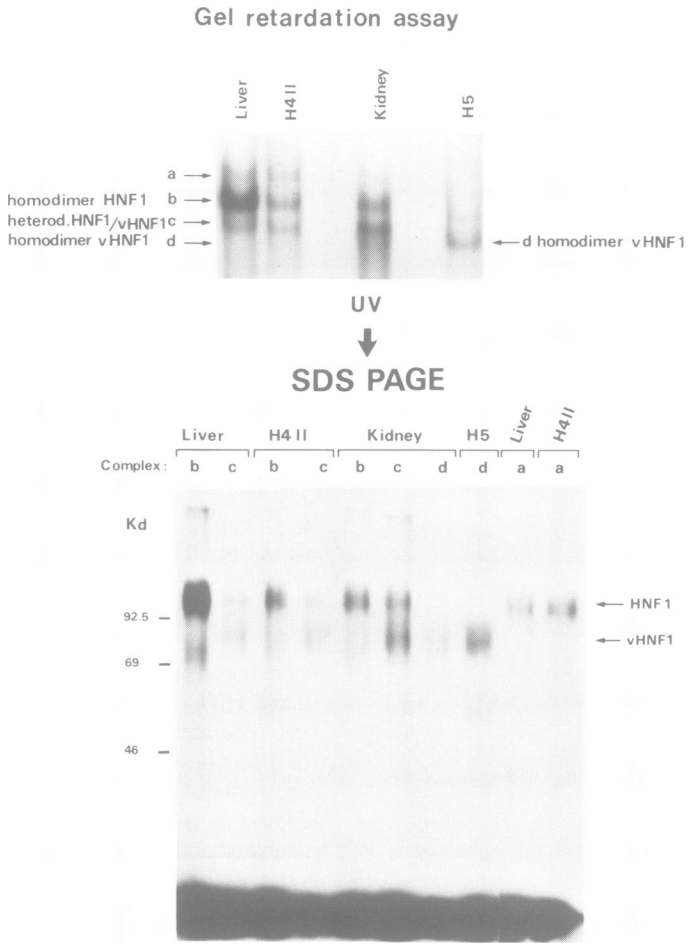

Fig. 9. UV cross-linking analysis of HNF1/vHNF1 heterodimers. The upper part of the figure shows a preparative gel shift assay performed with nuclear extracts from liver, H4II cells, kidney and H5 cells and ^{32}P-labelled, BrdU substituted PE probe. After UV irradiation of the gel (Cereghini *et al.*, 1988) the bands were excised and analysed by denaturing SDS−PAGE (lower part). The b, c and d bands correspond to the presumed HNF1 homodimers, HNF1/vHNF1 heterodimers and vHNF1 homodimers respectively. Band a is discussed in the text. The position of molecular weight markers is indicated to the left of the SDS−polyacrylamide gel.

these extracts. Preimmune sera did not displace any of the complexes. Similar experiments performed with liver and H4II cell nuclear extracts suggest that low amounts of vHNF1/HNF1 heterodimers may also exist in these cells (not shown).

Further evidence for the presence of vHNF1 and HNF1 in the intermediate complex came from UV cross-linking experiments (Figure 9). The proteins forming each of the complexes in the gel retardation assays were UV cross-linked to BrdU substituted ^{32}P-labelled PE probe by directly irradiating the retardation gel, as previously described (Cereghini *et al.*, 1988). The bands were excised and analysed on denaturing SDS−polyacrylamide gels. In accordance with previous results, this approach distinguished between the predominantly HNF1 complex present in liver and H4II cells and the vHNF1 complex present in H5 cells. Both liver and H4II cells also contained low amounts of a faster migrating band (c) that gave rise to bands of roughly equal intensities at 100 kd and 75 kd in the denaturing gel. A complex running in the same position was more abundant in kidney extracts. After UV cross-linking it also gave a roughly equal yield of the 100 kd and 75 kd bands, as expected for a heterodimer of HNF1 and vHNF1. In kidney, a rather diffuse complex migrating in a position similar to that of the H5 band, produced the 75 kd band only. Finally, the slowest migrating band in liver and H4II cells gave rise to only the 100 kd band in the denaturing gel. This complex may have a different conformation or contain an additional protein that is not in direct contact with the DNA. It was not further characterized.

vHNF1 activates transcription in vivo, mediated by the proximal element of the rat albumin promoter

Early experiments showed that the vHNF1 binding activity was present in hepatoma cell derivatives that fail to express albumin and many liver-specific genes. This led us to think that vHNF1 was unable to promote transcription from promoters containing the HNF1 site. In order to test this hypothesis we carried out transient transfection experiments with the RSV-vHNF1-A expression vector described above. We have used as a reporter gene, the CAT coding sequence driven by the rat albumin promoter (−151/+16) (Tronche *et al.*, 1989, 1990). For comparison, we also ran similar transient transfection experiments with an RSV-HNF1 expression construct described elsewhere (Tronche,F., Chouard,T., Blumenfeld,M., Griffo,G. and Yaniv,M., submitted). The human epithelial cell line C33 was used as a recipient. Surprisingly, vHNF1 expression activated the

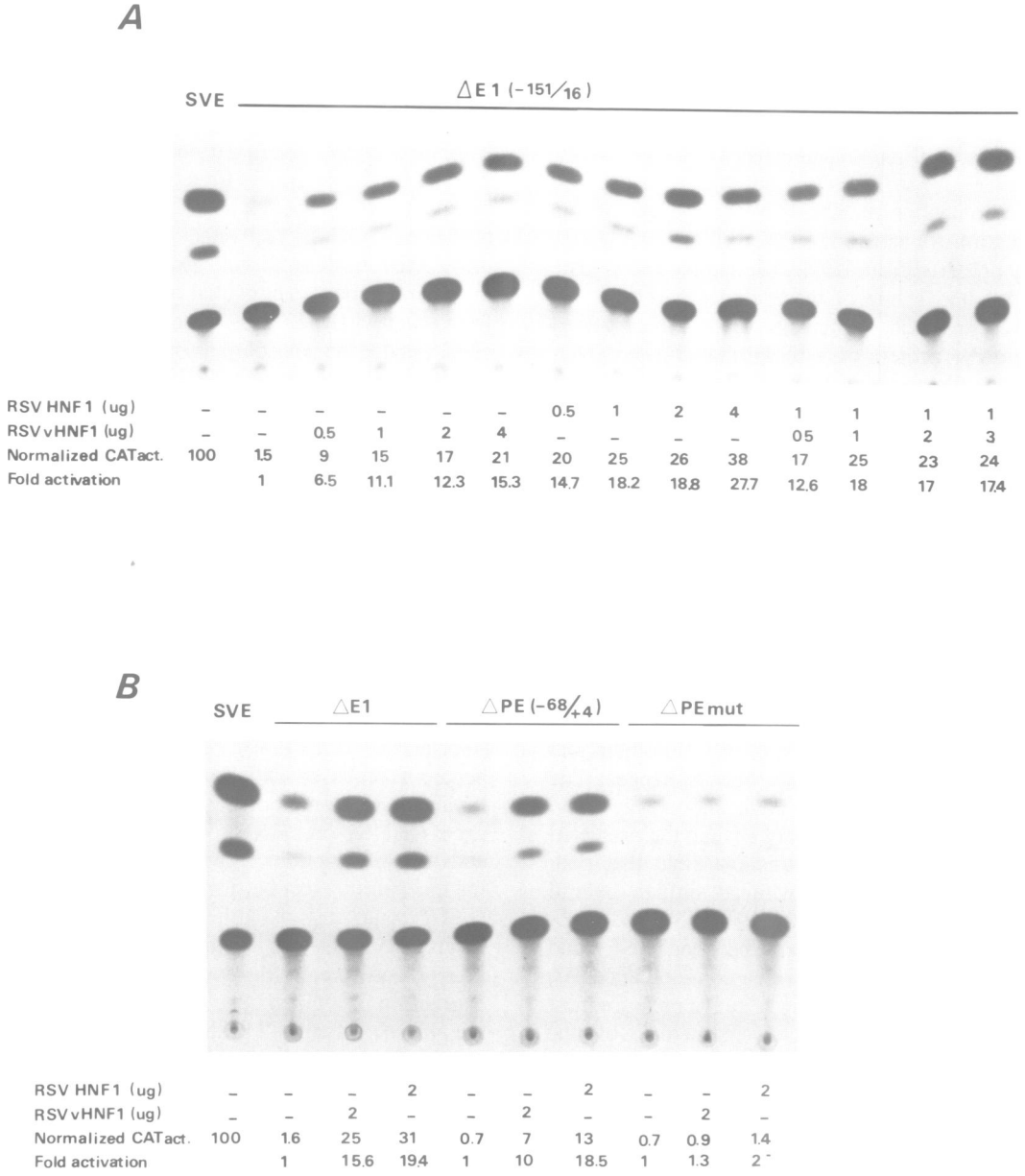

Fig. 10. Transcriptional activation by vHNF1. **A.** CAT assays performed with extracts from C33 cells transfected with different amounts of vHNF1-A and HNF1 expression constructs. CAT reporter gene was driven by the ΔE1 (−151/+16) albumin promoter. CAT activity was normalized with respect to the transfection efficiency by measuring the β-galactosidase activity obtained with a cotransfected RSV-β-galactosidase construct. It is expressed as the percentage of the normalized CAT activity obtained with the SVE promoter. The degree of activation of the albumin promoter is indicated in the bottom line. **B.** Similar CAT assay as in A but with CAT gene driven by the albumin minimal promoter, ΔPE(−68/+4), or a mutated version, ΔPEmut, which contains a mutated HNF1 site (Tronche *et al.*, 1989).

transcription of the reporter CAT gene, although at lower levels (10- to 15-fold stimulation) than those obtained when HNF1 was transfected (20- to 25-fold stimulation) (Figure 10A). A shorter version of the albumin promoter encompassing just the PE and the TATA box (−68/+4), yielded similar results. Both vHNF1 and HNF1 failed to promote transcription from an albumin promoter containing a mutated PE previously shown to be incapable of binding both factors (Cereghini *et al.*, 1988) (Figure 10B), confirming that the stimulation observed was mediated by the proximal element.

We next tested the effect on transcription of cotransfecting both RSV-vHNF1 and RSV-HNF1 expression constructs, in an attempt to assess the transactivation potential of vHNF1/HNF1 heterodimers. As shown in Figure 10A, the co-transfection of increasing amounts of the RSV-vHNF1 construct with a constant amount of RSV-HNF1, resulted in transactivation levels that did not exceed that obtained with the RSV-HNF1 construct alone. This suggests that the heterodimers between vHNF1 and HNF1 may have a lower transactivation capacity than the HNF1 homodimers.

Discussion

vHNF1 is a homeoprotein homologous to HNF1

Nuclear factors HNF1 and vHNF1 were first detected in differentiated and dedifferentiated rat hepatoma cells respectively. They interact *in vitro* with the same sequence

element, namely the proximal element of the albumin promoter or the equivalent element of other liver-specific genes (Cereghini *et al.*, 1988; Baumhueter *et al.*, 1988). This prompted us to think that both factors might share a similar DNA binding domain. Supporting this hypothesis, Northern blots of RNA from dedifferentiated cells gave a signal when hybridized under low stringency conditions with a probe containing the homeobox of HNF1. We made use of this cross hybridization for isolating cDNA clones of vHNF1 from a library prepared from H5 dedifferentiated cells. The largest open reading frame of these clones had a coding capacity of 557 amino acids. When this cDNA was expressed *in vitro* it yielded a protein of ~68 kd that showed DNA binding activity with a sequence specificity identical to that of vHNF1. Post-translational modifications could explain the smaller size of the *in vitro* translated protein compared with vHNF1, as detected in H5 cells. Supporting this hypothesis, the protein expressed in transfected cells gave rise to a complex indistinguishable from vHNF1. Further proof that our clone indeed coded for vHNF1 came from experiments with specific antibodies raised against the C-terminal portion of the cloned protein. These antibodies specifically displaced the vHNF1 complex of H5 cells in gel retardation assays.

The deduced amino acid sequence of vHNF1 shows a high degree of homology with HNF1 in regions reported to be important for DNA binding (Chouard *et al.*, 1990; Nicosia *et al.*, 1990); the dimerization domain, the B domain and the variant homeodomain show only minor changes. The sequence of vHNF1 in the homeodomain shows only a few amino acid substitutions which do not change the predicted secondary structure. Therefore, it is plausible that both factors possess homeodomains with a similar overall structure (Finney, 1990; Nicosia *et al.*,1990; Chouard *et al.*, 1990). The high degree of similarity between HNF1 and vHNF1 indicates that both come from a common ancestor, probably by gene duplication. Strong selective pressure must have acted to keep the particular structure of the homeodomain almost unchanged. The conservation also extends to the extra 18–20 amino acid loop. No specific function has been assigned to this loop, but it is dispensable for dimer formation and DNA binding. However, its conservation between the two proteins suggests that it may play an important role in the ultimate function of these two factors. The dimerization domain and the homeodomain have specific functions assigned to them: dimer formation and DNA contact. The B domain, recently shown to be crucial for the overall DNA binding activity of HNF1, has no specific function assigned. Interestingly, this domain is strictly conserved in vHNF1, which reinforces its possible importance in the molecular mechanisms of action of these two factors.

A striking difference between HNF1 and vHNF1 is the existence of a 26 amino acid long insertion between the homeodomain and the B domain. As we have shown, this insertion is not present in all vHNF1 molecules and may originate by alternative splicing. No difference was observed in DNA binding activity *in vitro* between both forms of vHNF1. The putative extra exon has been maintained through evolution, as mouse and human vHNF1 homologues display it in their sequences (S.Cereghini, unpublished data; I.Bach, manuscript in preparation). Whether there is a functional significance for the presence of this extra sequence is unknown for the moment.

vHNF1 is a transcriptional activator

vHNF1 was first found in the dedifferentiated hepatoma cell lines H5 and C2, which do not exhibit a hepatic phenotype and fail to express many liver-specific genes. In addition, when the albumin promoter is transfected into these cells, it remains silent. These data compelled us to believe that vHNF1 might be unable to activate transcription of genes bearing the PE sequence in their promoters. Unexpectedly, transfected vHNF1 activated the transcription of a co-transfected CAT reporter gene driven by the albumin promoter, in human epithelial cells. More surprisingly, we found that when H5 cells were transfected with a vHNF1 cDNA under the control of a strong promoter, transcription directed by a co-transfected albumin promoter was readily observed (not shown). One possible explanation for this apparent discrepancy could be that the levels of endogenous expression of vHNF1 in H5 cells are below the threshold required to promote transcription. Upon transfection, the amount of vHNF1 would rise to levels capable of inducing transcription from the albumin promoter. We have to recall that the PE binding activity found in H5 nuclear extracts is noticeably lower than in liver or hepatoma cell extracts. The importance of threshold levels of transcription factors in the control of gene expression has been elegantly demonstrated in the case of the *Drosophila bicoid* gene product (Struhl *et al.*, 1989) and also in the case of NF-AT in lymphocytes (Fiering *et al.*, 1990). Alternatively, dedifferentiated hepatoma cells or extinguished hybrids might contain a limited concentration of an inhibiting factor. Introduction of excess HNF1 or vHNF1 would titrate this inhibitor and activate transcription. However, whatever the explanation for this paradox can be, it is worth noting that vHNF1 and HNF1 could regulate different genes, hence the relative levels of both factors are different in different organs (i.e. kidney and liver). Moreover, our results show that the expression of a given gene does not depend exclusively on the presence in the cell of the required transcription factors.

The transactivation function of HNF1 has been broadly mapped to two domains, ADI and ADII, located in the C-terminal half of the molecule, downstream of the homeodomain (Nicosia *et al.*, 1990). These two domains are not conserved in vHNF1. Perhaps, transactivation depends on specific structural features of the proteins, more than just on sequence conservation. For example, like the transactivation domain of Pit1 (Theill *et al.*, 1989), ADI of HNF1 is particularly rich in threonine and serine. These two residues are also highly represented in the C-terminal part of vHNF1. It is puzzling however, why other short segments in the C-terminus of vHNF1 and HNF1 have been evolutionarily conserved and precisely these two domains have not. The clarification of these points requires further investigation.

vHNF1 and HNF1 form heterodimers

As Northern analyses showed, vHNF1 expression in the adult qualitatively resembles that of HNF1. Kidney, liver and intestine all express both mRNAs although at quite different levels. Kidney showed the highest amount of vHNF1 mRNA, while intestine and liver showed barely detectable levels. The amount of HNF1 mRNA was fairly similar in all three tissues.

The fact that vHNF1 shows homology with the dimerization domain of HNF1 and since both mRNAs are

co-expressed within the same cell, prompted us to investigate whether both factors could form heterodimers. *In vitro* translated vHNF1 and HNF1 proteins could in fact form such heterodimers as shown by gel retardation assays with the PE probe. By virtue of their different electrophoretic mobilities, three complexes were observed corresponding to the vHNF1 and HNF1 homodimers and, migrating in between, the vHNF1/HNF1 heterodimer. Similar evidence for heterodimer formation was obtained when both factors were expressed *in vivo* by co-transfection of C33 cells.

We next investigated whether vHNF1/HNF1 heterodimers might exist in the organism. In fact, we observed in gel retardation assays with kidney nuclear extracts, a PE binding activity that could correspond to vHNF1/HNF1 heterodimers. Its mobility was intermediate between that of HNF1 and vHNF1 homodimers which were also observed in this extract. More conclusively, specific antibodies against vHNF1 were able to inhibit the formation of this intermediate complex as well as the faster migrating vHNF1 homodimer complex. Similarly, anti-HNF1 antibodies also displaced the intermediate complex and that corresponding to HNF1 homodimers. When both sera were used together, all three complexes were displaced. Finally, UV cross-linking experiments performed with the different complexes confirmed that this intermediate band was made up of two polypeptide chains with sizes identical to that of vHNF1 and HNF1. Additional data strongly suggest that the vHNF1/HNF1 heterodimers actually exist in the cell rather than being an artefact of extract preparation. No subunit exchange between vHNF1 and HNF1 homodimers at 0°C was observed in mixing experiments, practically ruling out the possibility that heterodimerization might occur during extract preparation.

Heterodimerization has been shown to be a rather common phenomenon between members of other families of transcription factors (reviewed by Jones, 1990). A very well documented case is that of Jun and Fos, where the Jun/Fos heterodimers bind DNA more avidly and activate transcription more efficiently than Jun homodimers (reviewed by Curran and Franza, 1988). Conversely, it has been shown that the Id factor interacts with the helix−loop−helix proteins E12 and E47 and the myogenic factor MyoD (Benezra *et al.*, 1990). This interaction results in down-regulation of MyoD transactivation (*op. cit.*). In *Drosophila* a similar interaction has been suggested to occur between the *extramacrochaetae* gene product and other helix−loop−helix proteins (Ellis *et al.*, 1990; Garrell and Modolell, 1990). In all cases, the final activity of the factors is affected by the interaction. Among homeoproteins, there are two examples where direct interaction between two members has been proposed. Yeast mating type factor **a**1 may drive the change in the recognition specificity of the α2 factor by directly interacting with it (Goutte and Johnson, 1988; Dranginis, 1990). Similar change in specificity might occur in the case of the homeoprotein related *b* polypeptides of *U.maydis* (Schulz *et al.*, 1990). In this case, the combination of two different alleles of the *b* locus triggers the pathogenic development of this fungus. In both cases, the interaction would produce a qualitative change, rather than just a quantitative difference, in transactivation.

We addressed the functional consequences of vHNF1/HNF1 heterodimerization, by assaying its effect on PE dependent transcription, in transfection experiments with both factors. The results showed that, under the conditions of our assay, heterodimerization does not seem to have a marked positive or negative effect on the transactivation potential of both factors. Regardless of the apparent low effect on albumin transcription, in transient assays, heterodimerization between vHNF1 and HNF1 may have a biological function in the organism. For example, the different relative amounts of homodimers and heterodimers between HNF1 and vHNF1 present in liver and kidney, makes it plausible that different genes might be regulated by each molecular species. In addition, the heterodimer could increase the diversity of the activation domains and permit interaction with different coadaptor species that would link it to the basic transcriptional machinery (Lewin, 1990). These combinatorial effects would considerably expand the regulatory spectrum of the homeoproteins.

Finally, the co-existence of two or more homeoproteins, with identical or similar DNA binding specificities, in certain cells or regions of the developing embryo is the rule rather than the exception. Models which consider various homeoprotein combinations have been proposed (Lewis, 1978; Peifer *et al.*, 1987) but their molecular bases are not yet understood. Lymphocytes contain both the ubiquitous Oct-1 and the cell specific Oct-2 homeoproteins that bind to the same target sequence (reviewed by Schreiber *et al.*, 1989). The progress zone of the mouse limbs contains transcripts encoded by several representatives of the *Hox-4* (previously *Hox-5*) gene cluster (Dollé *et al.*, 1989) and in *Drosophila*, some cells within a given parasegment express more than one homeotic gene (e.g. *Ubx* and *abdA* in parasegment 8. Peifer *et al.*, 1987). The progressive development of somites along the anterior−posterior axis of the mouse embryo or the proximo-distal progression in the formation of the mouse limbs is accompanied by the delayed appearence in time of more posterior (or distal) homeoproteins in addition to early members of the same *Hox* cluster (Dollé *et al.*, 1989). We would like to propose that cell differentiation may progress according to a similar sequential principle. Early differentiation of the endoderm would involve vHNF1, while later specialization of some of the cells into mesenchymal cells in the liver or epithelial cells in the intestine, would require the additional presence of HNF1. This hypothesis is compatible with at least two observations: vHNF1, but not HNF1, is strongly induced upon differentiation of F9 cells into primitive endoderm *in vitro* (S.Cereghini and M.O.Ott, in preparation); and only vHNF1 is found in dedifferentiated hepatoma cells, whereas the fully differentiated phenotype is correlated with the appearance of HNF1 (Cereghini *et al.* 1990; and the present study).

Materials and methods

Cell lines and DNA probes

Cell lines and DNA transfections. Differentiated rat hepatoma cell lines H4II, Fao and C2Rev7 and dedifferentiated derivatives, H5 and C2 (Deschatrette and Weiss, 1974) were cultured and transfected as described previously (Ott *et al.*, 1984). C33 human epithelial tumour (Yee *et al.*, 1985) and Caco human colon carcinoma (Pinto *et al.*, 1983) cells were cultured in DMEM with 10% fetal calf serum and transfected by standard calcium phosphate coprecipitation procedure.

DNA probes. For gel retardation experiments a double-stranded ^{32}P-labelled albumin Proximal Element (PE) oligonucleotide was used (Cereghini *et al.*, 1988): TGTGGTTAATGATCTACAGTTA. Competitor oligonucleotides were: mut3: TGTGGTgtATGATCTACAGTTA; mut4: TGTGGTTAAT-

GAggTACAGTTA; DS12: TGTcaTTAATGATCTAttGTTA; DS34: TGTGGTgtATGAggTACAGTTA.

Isolation of cDNA clones

A cDNA library of H5 cells [oligo(dT) primed] was constructed from poly(A)$^+$ RNA by using cDNA synthesis and λgt10 cloning kits from Amersham, essentially according to the instructions of the manufacturer. The unamplified library was screened, by hybridization in low stringency conditions, with an *Nco*I fragment of 511 bp derived from the HNF1 cDNA, as a probe. This fragment extends from positions 334 to 845 starting at the ATG. Hybridization was done in 6× SSC, 1% SDS, 5% milk at 55°C overnight with 2×10^6 c.p.m./ml of probe, labelled by random priming to a specific activity of $\sim5\times10^8$ c.p.m./μg. The most stringent wash was in 1× SSC at 55°C. DNA from isolated positive clones was purified from liquid lysates and inserts were excised and subcloned in the plasmid vector Bluescript SK+ (Stratagene). Serial deletions were obtained with exonuclease III (Henikoff, 1987) using an 'Erase a Base' kit from Promega, following the manufacturer's instructions. Sequencing reactions were done with a Sequenase kit from USB, using double-stranded plasmid as template.

Computer analyses of the sequence data were carried out on a Data General MV8000 mainframe computer.

In vitro transcription and translation

vHNF1-A and vHNF1-B inserts were subcloned in a T7 polymerase transcription vector bearing the β-globin leader sequence upstream of the *Nco*I cloning site (provided by Dr R.Treisman). The 5′ end of the inserts was modified to remove the untranslated sequences using a PCR amplified fragment introducing an *Nco*I site just at the ATG initiation codon. *In vitro* transcription was carried out with a Stratagene transcription kit according to the instructions of the manufacturer. Prior to transcription, the templates were linearized with *Eco*RI, to obtain the full length (FL) protein or with *Bsp*HI (position 1331 in the *vHNF1-A* cDNA) to obtain the truncated vHNF1 protein (ΔC434). *HNF1* templates were linearized with *Eco*RI. The synthesized RNA was treated with phenol/chloroform and ethanol precipitated. *In vitro* translations of this RNA were carried out with commercial reticulocyte lysates (Amersham and Promega) and [^{35}S]methionine. The translated products were analysed in SDS−polyacrylamide denaturing gels and used directly in gel retardation assays.

Northern blot analysis

Total RNA from different tissues and cell lines was extracted according to the guanidinium thiocyanate−acid phenol procedure (Chomczynski and Sacchi, 1987). Poly(A)$^+$ RNA was isolated by oligo(dT) cellulose chromatography (Sambrook *et al.*, 1989), electrophoresed through a 1.2% agarose−2.2 M formaldehyde gel and transferred to Pall Biodyne nylon membranes. Hybridization was done at 42°C in 50% formamide, 5× SSPE, 5× Denhardt's, and random priming ^{32}P-labelled probes, for 24 h. For high stringency conditions, the membranes were washed at 60°C with 0.1× SSC, 0.25% SDS. Low stringency washes were at 40°C in 0.5× SSC, 0.25% SDS.

Gel retardation assays

Gel retardation assays and UV cross-linking experiments were performed as previously described (Cereghini *et al.*, 1988). Briefly, the reactions were done in a 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 sonicated salmon sperm DNA and 0.2 ng of ^{32}P-labelled probe. For *in vitro* translated products, 1−4 μl of the translation mixture was used. After 5 min on ice, the mixtures were loaded onto a 5% or 6% polyacrylamide gel in 0.25× TBE and run at room temperature for 2−5 h at 12 V/cm. After fixing, the gels were vacuum dried and exposed with an intensifying screen at −80°C. When ^{35}S-labelled translation products were used in the assay, two films were used to block the ^{35}S-specific radiation.

Antibody preparation

HNF1-specific antiserum. A peptide containing the 21 amino acids from F541 to T561 in the rat HNF1 sequence (FTSDTEASSEPGLHEPSSPAT, Chouard *et al.* 1990) was chemically synthesized and covalently linked to keyhole limpet haemocyanin (KLH) using glutaraldehyde (Ausubel *et al.* 1989). Various amounts of the partially linked KLH-peptide mixture were injected into a Blanc-Bouscat rabbit as follows: day 1: 400 μg with complete Freund's adjuvant into popliteal lymph nodes; day 23: 400 μg with incomplete Freund's adjuvant into the subscapular cavity; day 31: 200 μg in PBS intramuscularly; day 32: same intravenously; day 46: bleeding and serum preparation.

vHNF1-specific antiserum. vHNF1 specific antibodies were raised against the C-terminal part of the molecule (residues 403−557). This region was this part of the molecule and glutathione-S-transferase was obtained by cloning the *Sau*3AI fragment from positions 1236 to 1826 of vHNF1.A cDNA in the *Bam*HI site of the pGEX.3X bacterial expression vector (Smith and Johnson, 1988). The expression of the protein was induced by adding IPTG to 1 mM to 1 litre cultures and continuing the incubation at 37°C for 4 h. The fusion protein was purified on glutathione columms as described (Smith and Johnson, 1988). Immunization of rabbits was done by subcutaneous injection of 100 μg of purified fusion protein with Freund's adjuvant following standard procedures (Harlow and Lane, 1988). Bleedings were done regularly 10 days after each injection. For gel retardation assays, 10 μl of serum was mixed with the nuclear extract and incubated for 15 min at room temperature before adding the ^{32}P-labelled oligonucleotide probe.

Acknowledgments

We thank M.C.Weiss, J.Ham and D.Duboule for valuable comments on the manuscript; A.Doyen for technical help, O.Siffert for peptide synthesis, B.David-Watine for the gift of kidney nuclear extracts, R.Treisman for the gift of the T7 transcription vector, F.Tronche for the RSV.HNF1 expression vector, and V.Ley and F.Bourgade for help in antibody preparation. This work was supported by grants from the CNRS, INSERM, ARC, LNFCC and the FRMF. J.R.-C. was supported by fellowships from the Commission of the European Communities and the Ministerio de Educacion y Ciencia (Spain).

References

Affolter,M., Schier,A. and Gehring,W.J. (1990) *Current Opinion Cell Biol.*, **2**, 485−495.

Ausubel,F.M., Brent,R., Kingston,R.E., Moore,D.D., Seidman,J.G., Smith,J.A. and Struhl,K. (1989) *Current Protocols in Molecular Biology.* Greene Publishing Associates and Wiley Interscience, New York.

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

Baumhueter,S., Courtois,G. and Crabtree,G.R. (1988) *EMBO J.*, **7**, 2485−2493.

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.

Benezra,R., Davis,R.L., Lockshon,D., Turner,D.L. and Weintraub,H. (1990) *Cell* **61**, 49−59.

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

Cereghini,S.,Yaniv,M. and Cortese,R. (1990) *EMBO J.* **9**, 2257−2263.

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.

Curran,T. and Franza,Jr,B.R. (1988) *Cell*, **55**, 395−397.

Deschatrette,J. and Weiss,M.C. (1974) *Biochemie*, **56**, 1603−1611.

Dollé,P., Izpisua-Belmonte,J.-C., Falkenstein,H., Renucci,A. and Duboule,D. (1989) *Nature*, **342**, 767−772.

Dranginis,A.M. (1990) *Nature*, **347**, 682−685.

Ellis,H.M., Spann,D.R., Posakony,J.W. (1990) *Cell*, **61**, 28−37.

Feng,D.F., Johnson,M.S. and Doolittle,R.F. (1985) *J. Mol. Evol.*, **21**, 112−125.

Fiering,S., Northrop,J.P., Nolan,G.P., Mattila,P.S., Crabtree,G.R. and Herzenberg,L.A. (1990) *Genes Dev.*, **4**, 1823−1834.

Finney,M. (1990) *Cell*, **60**, 5−6.

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

Garrell,J. and Modolell,J. (1990) *Cell*, **61**, 39−48.

Gehring,W.J. (1987) *Science*, **236**, 1245−1251.

Goutte,C. and Johnson,A.D. (1988) *Cell*, **52**, 875−882.

Harlow,E. and Lane,D. (1988) *Antibodies: a Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Henikoff,S. (1987) *Methods Enzymol.*, **155**, 156−165.

Holland,P.W.H. and Hogan,B.L.M. (1988) *Genes Dev.*, **2**, 773−782.

Jones,N. (1990) *Cell*, **61**, 9−11

Keleher,C.A., Goutte,C. and Johnson,A.D. (1988) *Cell*, **53**, 927−936.

Kelleher,R.J.,III, Flanagan,P.M. and Kornberg,R.D. (1990) *Cell*, **61**, 1209−1215.

Levine,M. and Hoey,T. (1988) *Cell*, **55**, 537−540.

Lewin,B. (1990) *Cell*, **61**, 1161−1164.

Lewis,E.B. (1978) *Nature*, **276**, 565–570.
Lichsteiner,S. and Schibler,U. (1989) *Cell*, **57**, 1179–1187.
Maniatis,T., Goodbourn,S. and Fischer,J.A. (1987) *Science*, **236**, 1237–1244.
Nicosia,A., Monaci,P., Tomei,L., De Francesco,R., Nuzzo,M., Stunnenberg,H. and Cortese,R. (1990) *Cell*, **61**, 1225–1236.
Ott,M.O., Sperling,L., Herbomel,P., Yaniv,M. and Weiss,M. (1984) *EMBO J.*, **3**, 2505–2510.
Otting,G., Qian,Y.Q., Billeter,M., Müller,M., Affolter,M., Gehring,W.J. and Wüthrich,K. (1990) *EMBO J.*, **9**, 3085–3092.
Peifer,M., Karch,F. and Bender,W. (1987) *Genes Dev.*, **1**, 891–898.
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.
Ptashne,M. (1988) *Nature*, **335**, 683–689.
Qian,Y.Q., Billeter,M., Otting,G., Müller,M., Gehring,W.J. and Wüthrich,K. (1989) *Cell*, **59**, 573–580.
Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) *Molecular Cloning: A Laboratory Manual* (Second edition). Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.
Schreiber,E., Muller,M.M., Schaffner,W. and Matthias,P. (1989) In Renkawitz,R. (ed.), *Tissue-specific Gene Expression*. VCH Verlagsgesellschaft, Weinheim.
Schulz,B., Banuett,F., Dahl,M., Schlesinger,R., Schäfer,W., Martin,T., Herskowitz,I. and Kahmann,R. (1990) *Cell*, **60**, 295–306.
Scott,M.P., Tamkun,J.W. and Hartzell,G.W.,III (1989) *Biochim. Biophys. Acta*, 989, 25–48.
Smith,D.B. and Johnson,K.S. (1988) *Gene*, **67**, 31–40.
Struhl,G., Struhl,K. and MacDonald,P.M. (1989) *Cell*, **57**, 1259–1273.
Theill,L.E., Castrillo,J.-L., Wu,D. and Karin,M. (1989) *Nature*, **342**, 945–948.
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., Cereghini,S., Weiss,M. and Yaniv,M. (1990) *Mol. Biol. Med.*, **7**, 173–185.
Wasylyk,B. (1988) *CRC Crit. Rev. Biochem.*, **23**, 77–120.
Yee,D., Krishnan-Hewlett,I., Baker,C.C., Schegel,R. and Howley,P.M. (1985) *Am. J. Pathol.*, **119**, 361–366.

Received on December 17, 1990; revised on February 21, 1991