Hepatocyte nuclear factor 3α belongs to a gene family in mammals that is homologous to the *Drosophila* homeotic gene *fork head*

Eseng Lai,[1] Vincent R. Prezioso,[2] Wufan Tao,[1] William S. Chen,[2] and James E. Darnell, Jr.[2,3]

[1]Division of Endocrinology and Program of Cell Biology and Genetics, Memorial Sloan-Kettering Cancer Center, New York, New York 10021 USA; [2]Laboratory of Molecular Cell Biology, Rockefeller University, New York, New York 10021 USA

By analysis of cDNA clones that cross-hybridized with a portion of the cDNA encoding the recently described rat protein hepatocyte nuclear factor 3α (HNF-3α, previously called HNF-3A), we now describe two additional members, HNF-3β and HNF-3γ, of this gene family. A 110-amino-acid region in the DNA-binding domain of this family is not only very highly conserved in rodents (HNF-3α, -3β, and -3γ are identical in 93 of 110 amino acids in this region) but also in *Drosophila* where the homeotic gene *fork head* has 88 of the 93 residues that are identical in the three rat genes. The HNF-3 family in rodents is expressed in cells that derive from the lining of the primitive gut; some of the embryonic *Drosophila* cells in which *fork head* is expressed also give rise to gut and salivary glands. Thus, it appears that this gene family, the DNA-binding portion of which is unlike that of any previously recognized DNA-binding proteins, may contribute to differentiation of cells in internal organs in both vertebrates and invertebrates.

Using the techniques of molecular genetics to identify required DNA-binding sites for gene function and then protein purification by oligonucleotide affinity chromatography of proteins cognate with these sites, several laboratories including our own have identified transcription factors that direct expression of genes specifically in hepatocytes (Johnson et al. 1987; Frain et al. 1989; Baumhueter et al. 1990; Lai et al. 1990; Sladek et al. 1990). Hepatocyte-specific gene expression of such characteristic products as albumin begins quite early (~11 or 12 days of gestation) in the mouse embryo (Tilghman and Belayew 1982), indicating the probable early activity of transcription factors that are also active in the adult liver. Furthermore, some of these factors, for example, hepatocyte nuclear factors 1 and 3 (HNF-1 and HNF-3) appear to be present in yolk sac cells (Costa et al. 1990), which, like the hepatocyte, are derived from gut endoderm. Because the genes for some of the liver-enriched transcription factors mentioned above are transcriptionally controlled themselves (Xanthopoulos et al. 1989, and in prep.), a study of their regulation should provide access to genes that are important in early developmental cascades.

Hepatocyte nuclear factor 3α (HNF-3α, previously called HNF-3A), one of the factors purified from liver and whose cDNA we cloned recently, is transcriptionally active in liver but not in a number of other tissues. Liver extracts, however, interact with a labeled oligonucleotide containing an HNF-3 site to produce several bands in a gel mobility–retardation assay (Costa et al. 1989; Lai et al. 1990), and several mRNAs were detected in Northern blots with the cloned sequence of the HNF-3α cDNA (Lai et al. 1990). These results most likely indicate either a single transcriptional product that is differentially processed or the transcription of several related but distinct genes.

Characterization of additional HNF-3-like cDNA from rodent liver has now revealed that HNF-3α belongs to a gene family with a tightly conserved ~110-amino-acid sequence within the region shown earlier to be required for DNA binding. This conserved DNA-binding domain is unlike that of any previously recognized family of transcription factors. A recently reported sequence comparison further heightens the interest in the HNF-3 family as consisting of important contributing members to early development (Weigel and Jäckle 1990). The protein encoded by the *Drosophila* gene *fork head*, mutants of which fail to make proper terminal embryonic tissues (the proctodeum and stomodeum, which contribute cells to the foregut and hindgut), has an extensive amino acid sequence homology to the HNF-3A sequence described. The greatest sequence conservation is in the DNA-bind-

[3]Corresponding author.

ing domain of all the HNF-3 family members. The homeo box proteins that participate in Drosophila embryonic larval segmentation are segmentally expressed; and, presumably, similarly functioning homologs in mammals (Hox genes) are also segmentally expressed (Akam 1989; Graham et al. 1989). By analogy, it seems likely that the HNF-3 family and fork head (as well as any other Drosophila genes like fork head) may function in constructing internal organs from the embryonic gut. Thus, there may be an evolutionarily conserved genetic pathway operating in the differention of cells of internal organs of both vertebrates and invertebrates.

In addition to delineating three distinct members of the HNF-3 family, we also show that (1) the mRNAs for these proteins in rats are distributed differently in different organs that derive from the primitive gut, and (2) HNF-3 proteins are positive activating proteins. The study of the HNF-3 family whose members are differentially expressed in different cell types promises to be of considerable value in opening new insights into early mammalian development.

Results

Isolation of cDNA clones for HNF-3β and HNF-3γ

Liver extracts exhibit multiple DNA-binding activities for a labeled double-stranded deoxyoligonucleotide containing an HNF-3 site, and Northern analysis of liver mRNA with labeled DNA of the original HNF-3A cDNA clone shows multiple mRNAs (Lai et al. 1990). Therefore, we suspected that there might be several related genes that encode transcription factors sharing binding specificity for HNF-3 sites in DNA. When we selected and sequenced a murine cDNA that cross-hybridized with an HNF-3A probe, we isolated a partial cDNA clone, suggesting that more than one HNF-3-like gene existed. This clone, M2, was highly similar in one region of the two sequences but diverged markedly outside of this region. We used a probe derived from the divergent sequence to identify the rat counterpart of the mouse M2 cDNA and obtained a clone, termed K2, which was >90% identical to the entire mouse M2 sequence and distinctly different from the original HNF-3A (rat) gene in most of its sequence. The K2 rat cDNA contained a single large open reading frame (ORF) that would encode a protein of 458 amino acids (Fig. 1). The selection of an initiator methionine was suggested by an adjacent sequence that conformed well to the consensus for eukaryotic initiation (Kozak 1986) and the presence of a termination codon immediately upstream. Comparison of the amino acid sequence with that of HNF-3A indicated that these two proteins shared a highly homologous central domain but differed in most of the coding region.

To obtain additional members of this family, we rescreened 500,000 recombinants of the rat liver cDNA library, using a probe to the conserved sequence, and isolated 20 positive clones. We then used probes derived from nonhomologous sequences in the HNF-3A and the K2 cDNAs to identify 7 clones of the 20 that did not

hybridize to either one and therefore represented potential additional family members. Six of these clones were shown by restriction mapping and partial sequencing to be overlapping fragments of a single cDNA that was different from the two cDNA clones isolated previously. A cDNA containing the complete coding sequence of this third family member was reconstructed from two overlapping clones, L11 and L35. The joint L11/L35 sequence contains an ORF of 322 amino acids or ~140 amino acids shorter than the proteins encoded by the first two cDNAs (Fig. 2). The position of the translation initiation site in this third sequence was verified by use of a series of 5' deletions that were transcribed in vitro with SP6 or T7 RNA polymerase and translated in reticulocyte lysates (Fig. 3B). Deletion of 311 bp of the 5' end of the cDNA did not alter the size of the translated protein (d7). An additional deletion of 84 nucleotides (d8) resulted in a shortened translation product (presumably initiating at the ATG at position 537–539), confirming that the major translation initiation site must be one of the two tandem ATGs at position 312–314 and 315–317 (Fig. 2).

We renamed these three related HNF-3 cDNAs, HNF-3α (the original HNF-3A), HNF-3β (K2), and HNF-3γ (L11/L35). The nucleic acid sequence comparison of HNF-3α, HNF-3β, and HNF-3γ suggests that these cDNAs are products of three separate but related genes, rather than the products of alternative splicing. For example, in the regions of greatest homology, the amino acid sequence is more highly conserved than the nucleic acid sequence. Genomic Southern analyses using probes specific to each cDNA showed hybridization to a different pattern of genomic DNA fragments, thus supporting the existence of three separate and likely single-copy genes. Autoradiographs from Southern analysis of rat genomic DNA for HNF-3α and HNF-3β are shown in Figure 4.

HNF-3β and HNF-3γ, like HNF-3α, encode sequence-specific DNA-binding proteins

RNA transcripts from the HNF-3α, HNF-3β, and HNF-3γ clones were produced and translated in reticulocyte lysates, and the products were analyzed on SDS-polyacrylamide gels (Fig. 3A). The monomer molecular masses of HNF-3α, HNF-3β, and HNF-3γ, respectively, are 50, 47, and 42 kD. The product of HNF-3α has been shown previously to bind specifically to the proximal HNF-3 site in the transthyretin (TTR) promoter (Lai et al. 1990). We now show that the products of both HNF-3β and HNF-3γ also bind specifically to an oligonucleotide representing the original HNF-3-binding site (Fig. 3C). For comparison we show the complexes formed with crude liver extract. Three of these complexes have been referred to previously as A, B, and C in decreasing size (or increasing gel mobility).

Because we purified and sequenced fragments of the HNF-3A protein, we can definitively state that it is identical to the product of the HNF-3α gene. For the products of HNF-3β and HNF-3γ, indirect evidence indicates that they are identical or closely related to the HNF-3B and

```
TCTCCCGGGTATTGGCTGCAGCTAAGCGGGGCTTCCTGGGCCGACTGAGGTGGGTAGCCAGAAAAAGGCCTGAGGTGGCTGACAACCAGGGCGGCCAGACAACGCGAGTCCTGCGCGCCTTCTGAGGCCTC    131

CCTGGGACTTAACTGTAACGGGGAGGGGCCTCTGGAGCAGCGGCCAGCGAGTTAAAGT ATG CTG GGA GCC GTG AAG ATG GAA GGG CAC GAG CCA TCC GAC TGG AGC AGC TAC    243
                                                           M   L   G   A   V   K   M   E   G   H   E   P   S   D   W   S   S   Y     18

TAC GCG GAG CCT GAG GGC TAC TCT TCC GTG AGC AAC ATG AAC GCC AGC CTG GGG ATG AAT GGC ATG AAC ACT TAC ATG AGC ATG TCC GCG GCT GCA ATG    342
 Y   A   E   P   E   G   Y   S   S   V   S   N   M   N   A   S   L   G   M   N   G   M   N   T   Y   M   S   M   S   A   A   A   M     51

GGC AGT GGT TCC GGC AAC ATG AGC GCA GGC TCC ATG AAC ATG TCA TCC TAT GTG GGC GCT GGA ATG AGC CCG TCG CTG GCT GGC ATG TCC CCG GGC GCG    441
 G   S   G   S   G   N   M   S   A   G   S   M   N   M   S   S   Y   V   G   A   G   M   S   P   S   L   A   G   M   S   P   G   A     84

GGC GCC ATG GCG GGC ATG AGC GGC TCA GCT GGG GCG GCC GGC GTG GCG GGC ATG GGA CCG CAC CTG AGT CCG AGT CTG AGC CCA CTC GGG GGA CAG GCG    540
 G   A   M   A   G   M   S   G   S   A   G   A   A   G   V   A   G   M   G   P   H   L   S   P   S   L   S   P   L   G   G   Q   A    117

GCC GGG GCT ATG GGT GGC CTT GCT CCC TAC GCC AAT ATG AAC TCC ATG AGT CCT ATG TAC GGG CAG GCG GGC CTG AGC CGC GCT CGG GAC CCC AAG ACG    639
 A   G   A   M   G   G   L   A   P   Y   A   N   M   N   S   M   S   P   M   Y   G   Q   A   G   L   S   R   A   R   D   P   K   T    150

TAC CGG CGC AGC TAC ACT CAC GCC AAG CCT CCC TAC TCG TAC ATC TCG CTC ATC ACC ATG GCC ATC CAG CAG AGC CCC AAC AAG ATG CTG ACG CTG AGC    738
 Y   R   R   S   Y   T   H   A   K   P   P   Y   S   Y   I   S   L   I   T   M   A   I   Q   Q   S   P   N   K   M   L   T   L   S    183

GAG ATC TAT CAG TGG ATC ATG GAC CTC TTC CCT TTC TAC CGG CAG AAC CAG CAG CGC TGG CAG AAC TCC ATC CGT CAT TCT CTC TCC TTC AAC GAC TGC    837
 E   I   Y   Q   W   I   M   D   L   F   P   F   Y   R   Q   N   Q   Q   R   W   Q   N   S   I   R   H   S   L   S   F   N   D   C    216

TTT CTC AAG GTG CCC CGC GCG CCA GAC AAG CCT GGG GGC TCC TTC TGG ACC CTG CAC CCT GAC TCT GGC AAC ATG TTC GAG AAC GGT TGC TAC CTG CGC    936
 F   L   K   V   P   R   A   P   D   K   P   G   G   S   F   W   T   L   H   P   D   S   G   N   M   F   E   N   G   C   Y   L   R    249

CGC CAG AAG CGC TTC AAG TGT GAG AAG CAA CTG GCG TTG AAG GAA GCA GCG GGT GCG GGC AGT GGC GGA GGC AAG AAG ACC GCT CCT GGG ACA CAG GCT    1035
 R   Q   K   R   F   K   C   E   K   Q   L   A   L   K   E   A   A   G   A   G   S   G   G   G   K   K   T   A   P   G   T   Q   A    282

TCT CAG GTT CAG CTC GGG GAG GCC GCA GGC TCG GCC TCT GAG ACT CCG GCG GGC ACC GAG TCC CCC CAT TCC AGC GCT TCT CCG TGT CAG GAG CAC AAG    1134
 S   Q   V   Q   L   G   E   A   A   G   S   A   S   E   T   P   A   G   T   E   S   P   H   S   S   A   S   P   C   Q   E   H   K    315

CGA GGT GGC CTG AGC GAG CTG AAG GGA ACA CCT GCC TCT GCG CTG AGT CCT CCG GAG CCG GCG CCC TCG CCT GGG CAG CAG CAG CAG GCT GCA GCC CAC    1233
 R   G   G   L   S   E   L   K   G   T   P   A   S   A   L   S   P   P   E   P   A   P   S   P   G   Q   Q   Q   Q   A   A   A   H    348

CTG CTG GGC CCA CCT CAC CAT CCT GGC CTG CCA CCA GAG GCC CAC CTG AAG CCC GAG CAC CAT TAC GCC TTC AAC CAC CCC TTC TCT ATC AAC AAC CTC    1332
 L   L   G   P   P   H   H   P   G   L   P   P   E   A   H   L   K   P   E   H   H   Y   A   F   N   H   P   F   S   I   N   N   L    381

ATG TCC TCC GAG CAG CAA CAT CAT CAC AGC CAC CAC CAC CAT CAG CCC CAC AAA ATG GAC CTC AAG ACC TAC GAA CAG GTC ATG CAC TAC CCT GGG GGC    1431
 M   S   S   E   Q   Q   H   H   H   S   H   H   H   H   Q   P   H   K   M   D   L   K   T   Y   E   Q   V   M   H   Y   P   G   G    414

TAC GGT TCC CCC ATG CCA GGC AGC TTG GCC ATG GGC CCA GTC ACG AAC AAA GCC GGC CTG GAT GCC TCG CCC CTG GCT GCA GAC ACT TCC TAC TAC CAG    1530
 Y   G   S   P   M   P   G   S   L   A   M   G   P   V   T   N   K   A   G   L   D   A   S   P   L   A   A   D   T   S   Y   Y   Q    447

GGA GTG TAC TCC AGG CCT ATT ATG AAC TCG TCC TAA GAAGATGGCTTTCAGGCCCTGCTAGCTCTGGCCACTGGGGAGAAGGACAAGGGAAATGACAGGCTGAGTGGAAACATTGGGGG    1649
 G   V   Y   S   R   P   I   M   N   S   S   -                                                                                          458

AACTTTGAGGAAAAAGTAGCCACCACACTTAAGGCCCCAAAGGAGCAGTTTCACCTGTCTGTGTCCCTAAATAGCTGGGCCACACTGATCTGTCATTCTAAATAGGGAAGGGAATGGAAATATATATGTAT    1780
ACACATAAACTTGTTTTAATGGAGGACCTTTGGTTCCCACTATGTAGACTACTGCTTCTCAAGGCACCTGCAGATTTTGATTTTTGTTCTTGATTCTCTCTCTATTGCTGTCGTTGACAGAGAAGTCTGAC    1911
TTAAAAAAAAAACAAACTAAACAAAAAAAAAAACTTTTGATGAGTGACTTGAGTGTAAAACCATGTAGTTTTAACAGAAAACCAGAGGGTTGTATTGATGTTTAAAAGAGAAAAAAAAATAATGATGTAAG    2042
AGTCTGGTATAAATGACCAGGAGAAAGGACAAAAATGCATCCATTCTGGACATGGTGAAATCCAGGTCTCGGGTCTGATTTAATTTATGGTTTCTGCGTGCTTTATTTATGGCTTATAAATGTGTGTTCTG    2173
GCTAGAATGCCAGAGTTCCACAAATCTATATTAAAGTATTATTGC                                                                                          2217
```

Figure 1. Nucleic acid sequence of HNF-3β. The cDNA clone (K2) contains the complete protein-coding sequence but is missing the 3' end of the mRNA. The protein sequence is shown below, beginning at the initiator methionine at base 190, which is preceded by an in-frame termination codon.

HNF-3C proteins, respectively. Definitive identification of the proteins that form each of the gel mobility-shift complexes would require their purification and direct sequencing. The gel mobility-shift complex formed with the HNF-3β-translated product migrates identically to that of HNF-3B. The gel mobility-shift complex formed with the HNF-3γ product migrates slightly faster than the complex labeled HNF-3C. However, we obtained two antisera that aid in the identification of the proteins in the liver extracts. The first antiserum, which was raised in response to a synthetic peptide corresponding to the amino-terminal 15 amino acids of HNF-3α, reacts with the protein products of the HNF-3α and HNF-3β genes (anti-α). HNF-3β shares 12 of 15 amino acids at the amino terminus with HNF-3α. The second antibody was raised to a bacterial fusion product containing 77 amino acids of the HNF-3γ sequence and is specific to the HNF-3γ protein (anti-γ). The activity of anti-α is shown in Figure 5A. This antiserum binds to the HNF-3α produced by translation of this mRNA in reticulocyte ly-

sates (lane 2), resulting in a much slower mobility complex, labeled T, which presumably represents a complex of probe, DNA-binding protein, and antibody. The anti-α serum also reacts weakly with the HNF-3β product, shifting a small part of that complex to a slower mobility (lane 4).

The anti-γ antiserum reacts with the HNF3γ-translated product (Fig. 5A, lane 6). This antiserum appears to disrupt the binding of the protein to the DNA rather than to change the mobility of the gel-shift complex. The corresponding controls for the antibody reactions with preimmune serum are shown in Figure 5, lanes 1 and 3.

With crude liver extracts, the anti-α serum shifts the A complex and a small part of the B complex (Fig. 5B, lane 3). The anti-γ serum disrupts complex formation of the C protein without affecting the complexes formed with A and B (Fig. 5B, lane 4). The specificity of the anti-γ antiserum is shown more clearly using partially purified HNF-3C; anti-α serum does not affect binding, whereas

```
                                                               TGTCGAGTGACTCCTCCTCTTCAGAAGAGGGCAATCTAGGGATAGTGAAGCT      52
GAAACACAGTGGAACAAGTCCTTGAACCACAAGCTCATCGAACAGTTTTCTCGTTCCCCTTTGAGATATCAAGGAAGCTGTCACAGTCACTGTCTCATTCAGATGGTATAGATGCAGGCAGATTTTCTCAG      183
GGGGTCTCGGTGTAGATCTGGGACGGACCAGCACCATATACTTCGAGTTTCCATTGAGCAAGGAAGCAAAAGGCAGGAAGGCAAGAAGAGCTGAAAAAAGGCACAGCTGAGCCTCTCTGTCCTTCTTCATG      314
```

```
ATG GAA GGA CAG GAG GAC AGG ACG ACA GTA AGA CAA GAG GAC AGG GCT TAT GCT GTC CTC TTC CTT AAC ACT GCC TCC CCT CTG CCT ACA GGA CCC CTG      413
 M   E   G   Q   E   D   R   T   T   V   R   Q   E   D   R   A   Y   A   V   L   F   L   N   T   A   S   P   L   P   T   G   P   L        33

GCA CCC CCA GCC CCC ACC GCT CCC CTG GGG CCC ACC TTC CCA GGC CTG GGC GCA GGC AGC GGC ACC GGA GGC AGT GCT TCC GGG TAT GGC GCC CCA GGG      512
 A   P   P   A   P   T   A   P   L   G   P   T   F   P   G   L   G   A   G   S   G   T   G   G   S   A   S   G   Y   G   A   P   G        66

CCC GGG CTT GTA CAC GGA AAG GAA ATG GCC AAG GGG TAC CGG CGG CCA CTG GCT CAT GCC AAA CCA CCT TAT TCC TAC ATC TCT CTC ATC ACC ATG GCT      611
 P   G   L   V   H   G   K   E   M   A   K   G   Y   R   R   P   L   A   H   A   K   P   P   Y   S   Y   I   S   L   I   T   M   A        99

ATC CAG CAG GCA CCA GGC AAG ATG CTG ACC CTG AGT GAA ATC TAC CAA TGG ATC ATG GAC CTC TTC CCG TAC TAC CGG GAG AAC CAG CAA CGT TGG CAG      710
 I   Q   Q   A   P   G   K   M   L   T   L   S   E   I   Y   Q   W   I   M   D   L   F   P   Y   Y   R   E   N   Q   Q   R   W   Q       132

AAC TCC ATC CGG CAC TCG CTG TCC TTC AAT GAC TGC TTC GTC AAG GTG GCA CGC TCC CCA GAC AAA CCG GGC AAA GGC TCC TAC TGG GCC TTG CAT CCC      809
 N   S   I   R   H   S   L   S   F   N   D   C   F   V   K   V   A   R   S   P   D   K   P   G   K   G   S   Y   W   A   L   H   P       165

AGC TCT GGG AAC ATG TTT GAG AAC GGC TGC TAT CTC CGC CGT CAG AAG CGC TTC AAG CTG GAG GAG AAG GCA AAG AAA GGA AAC AGT GCC ACA TCC GCC      908
 S   S   G   N   M   F   E   N   G   C   Y   L   R   R   Q   K   R   F   K   L   E   E   K   A   K   K   G   N   S   A   T   S   A       198

ACC AGG AAT GGA ACA GTG GGG TCA GCC ACC TCT GCC ACC ACT ACA GCT GCC ACT GCA GTC ACT TCC CCG GCT CAG CCC CAG CCT ACG CCG CCG TCT GAG      1007
 T   R   N   G   T   V   G   S   A   T   S   A   T   T   T   A   A   T   A   V   T   S   P   A   Q   P   Q   P   T   P   P   S   E       231

CCC GAG GCC CAG AGT GGG GAA GAT GTG GGG GGT CTG GAC TGC GCC TCG CCT CCT TCA TCC GCA CCC TAT TTC ACT GGC CTG GAG CTC CCA GGG GAA CTG      1106
 P   E   A   Q   S   G   E   D   V   G   G   L   D   C   A   S   P   P   S   S   A   P   Y   F   T   G   L   E   L   P   G   E   L       264

AAG TTG GAT GCG CCC TAC AAC TTC AAC CAC CCT TTC TCT ATC AAC AAC CTG ATG TCA GAA CAG ACA TCA ACA CCT TCC AAA CTA GAT GTG GGG TTT GGG      1205
 K   L   D   A   P   Y   N   F   N   H   P   F   S   I   N   N   L   M   S   E   Q   T   S   T   P   S   K   L   D   V   G   F   G       297

GGC TAC GGT GCT GAG AGT GGG GAA CCT GGG GTC TAC TAC CAG AGC CTC TAT TCC CGC TCT CTG CTC AAT GCA TCC TAG CGGCGCAACTGGGAGAAATGCGGTGAT      1310
 G   Y   G   A   E   S   G   E   P   G   V   Y   Y   Q   S   L   Y   S   R   S   L   L   N   A   S   -                                    322
```

```
GGGGGTTTGCTGTGACAGATGACCGGTTCTTCCGCCCTGATCTTTCTGGTTACACTCTGCTTATCCAGTTAATTAACATCGTATTTGGTCTATTACTGTGATATGACCCATTGGCTACTGTGGTAACTGCA      1441
TGGACTCTTCGGTGGGCCTAGGGTTGGGGTATTGTGAAGGCAGATGCATTTGGGAGTGTTACGAAGGTGGCCATGTCGGACATACCGTGAAGGCAATTAGACTGGTGTACTATAAAAGCTGCATGTTAAGT      1572
GAGTGATCCACTGGGTGCCTGATGGCCGCGATGTCGGAGGACATATTGTTTGGCCCTTTGGATGCT      1638
```

Figure 2. Nucleic acid sequence of HNF-3γ. The L11/35 cDNA contains the complete protein-coding sequence but is missing the 3′ end of the mRNA. The protein sequence is shown below, beginning with two potential initiator methionines at bases 312 and 315. The second conforms much better to the consensus translation initiation sequence (Kozak 1986).

anti-γ serum completely disrupts DNA binding (Fig. 5B, lanes 5–8). These data indicate that the HNF-3γ product and HNF-3C share epitopes in the region of the protein that is distinct from HNF-3α or HNF-3β. The molecular mass on SDS gels of HNF-3C purified from rat liver extracts was determined previously to be 43 kD (Lai et al. 1990). Because the translated product of HNF-3γ is distinctly smaller on SDS–polyacrylamide gels, we conclude that the two proteins result from differential modifications of the same gene product or are very closely related but not identical in sequence.

The major conserved region in HNF-3 proteins is in the DNA-binding domain

A comparison of the amino acid sequences for the HNF-3 proteins revealed three regions (labeled I, II, and III) that are highly conserved in all three proteins (Fig. 6). We demonstrated previously that the region between amino acids 124 and 288 (numbering refers to HNF-3α) is required for DNA binding. Within this sequence is found region I, a 110-amino-acid stretch that is remarkably highly conserved in all three proteins. Ninety-three of the 110 amino acids are identical in the three proteins, whereas only one position is not shared between at least two of the three proteins. Nearly all of the differences represent conservative changes. Two other short, but well conserved, regions, II and III, are found in the carboxyl end of the proteins. As reported earlier (Lai et al. 1990), the amino acids in these two regions are not es-

sential for DNA binding. The HNF-3α and HNF-3β proteins are similar throughout the amino-terminal half of the protein and are particularly rich in methionine in this region (Lai et al. 1990; Fig. 1). Most of the rest of the HNF-3α and HNF-3β proteins have recognizable sequence similarity (~30% identity) but are considerably divergent. The HNF-3γ protein is ~140 amino acids shorter than HNF-3α and HNF-3β, does not have a large number of methionines in its amino terminus, and is highly divergent outside of the three conserved regions.

After we completed the sequence of HNF-3A we conducted computer searches for matches with other proteins and failed to discover any. Later, when both the sequences of HNF-3A and the Drosophila gene fork head had been entered in the available data bases, Weigel and Jäckle (1990) found a remarkable amino acid identity in these two sequences within the domain shown to be important for DNA binding of HNF-3A (HNF-3α). We compared the fork head sequence to each of the HNF-3 family members and found that there is greatest similarity to HNF-3β, where 100 of 110 amino acids are identical in the binding domain (Fig. 6). The homology among all four proteins extends to regions II and III, as well as to the binding domain, region I.

HNF-3 proteins have different affinities for different binding sites

The promoter of the TTR gene contains two HNF-3-binding sites, both of which appear necessary for max-
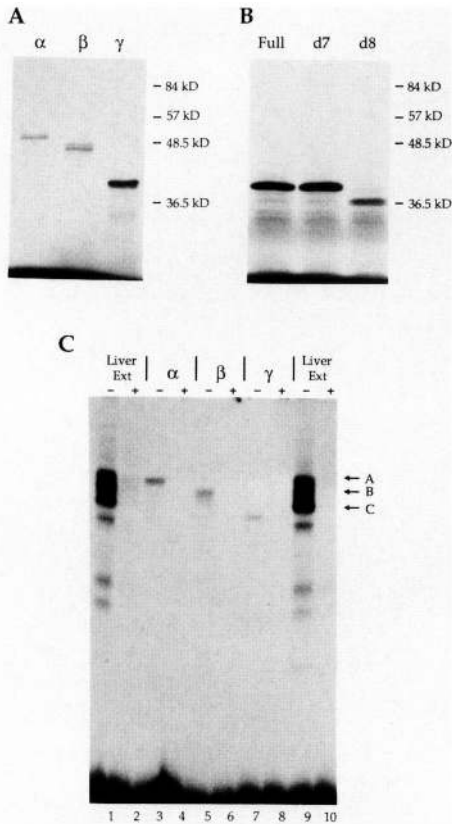
Figure 3. The HNF-3β and HNF-3γ cDNAs encode sequence-specific DNA-binding proteins. (A) Autoradiograph of the translated products of the HNF-3α, HNF-3β, and HNF-3γ cDNA analyzed by SDS-PAGE on a 9% acrylamide gel. (B) Confirmation of site of translation initiation in the HNF-3γ cDNA. Autoradiograph of translation products of full-length template and templates shortened by 5' deletions of L11/35, d7, and d8. (C) Gel mobility-shift assay of the translated products compared to liver nuclear extract. Each of the translated proteins binds to the HNF-3 oligonucleotide (−111 to −85 of the TTR promoter) in the presence of a nonspecific oligonucleotide (−175 to −151 of the TTR promoter, indicated by the (−) in lanes 1, 3, 5, 7, and 9) and is competed specifically by unlabeled HNF-3 oligonucleotide [indicated by the + in lanes 2, 4, 6, 8 and 10]. The three complexes, A, B, and C, observed with liver extract are indicated. Two micrograms of crude liver nuclear extract or 1 μl of reticulocyte lysate, containing the translation products of HNF-3α, HNF-3β, or HNF-3γ, was assayed.

imal activity of transfected DNA constructs and both of which are occupied in vivo by protein, presumably one of the HNF-3 family members, as detected by a new in vivo footprint method (Mirkovitch and Darnell 1991). The total set of binding proteins present in liver cell extracts, however, exhibits different affinities for the two DNA sites. A strong site is located between −96 and −106, and a weak site is located between −131 and −140 nucleotides upstream of the cap site (Costa et al. 1989). In an effort to gain insight into which proteins might actually function in vivo, we compared the relative affinities of each of the three HNF-3 proteins for each binding site

in vitro (Fig. 7). Using equal amounts of in vitro-translated protein, determined by [35S]methionine incorporation, we performed gel mobility-shift assays with labeled probes containing either the strong site, oligo S (−111 to −85) or the weak site, oligo W (−151 to −130). The amount of oligo S probe bound to HNF-3α or to HNF-3γ was greater than the amount bound to HNF-3β (Fig. 7, lanes 1–6). In contrast, with the oligo W probe, the amount complexed with HNF-3α or HNF-3γ was less than that bound to HNF-3β (Fig. 7, lanes 7–12). Thus, HNF-3α and HNF-3γ have a higher affinity for the strong site than does HNF-3β, whereas HNF-3β has a higher affinity for the weak site than either HNF-3α or HNF-3γ. This difference was most striking between HNF-3α and HNF-3β. While HNF-3α has a much higher affinity for the strong site compared to the weak site, HNF-3β has similar affinities for the two sites.

To illustrate these differences more specifically, competition experiments with both oligonucleotides were carried out. Using labeled oligo S probe, competition with 10-fold molar excess of unlabeled oligo S reduced binding by ~90% for HNF-3α (Fig. 8A, lane 2). The same amount of competitor only reduced the binding of HNF-3β to labeled probe by ~50% (Fig. 8A, lane 8). HNF-3α complex formation was not significantly reduced in the presence of a 40-fold molar excess of oligo W and a 200-fold excess of oligo W only decreased the HNF-3α complex by ~50% (Fig. 8A, lanes 5 and 6). In contrast, HNF-3β complex formation was reduced at least 75% with a 40-fold of excess oligo W and abolished by 200-fold excess competitor (Fig. 8A, lanes 11 and 12).

Competition studies with weak-site oligo W as the probe are consistent with the results using the strong-site oligo S as probe. A labeled oligo W probe in the presence of a 10-fold excess of unlabeled oligo S effec-
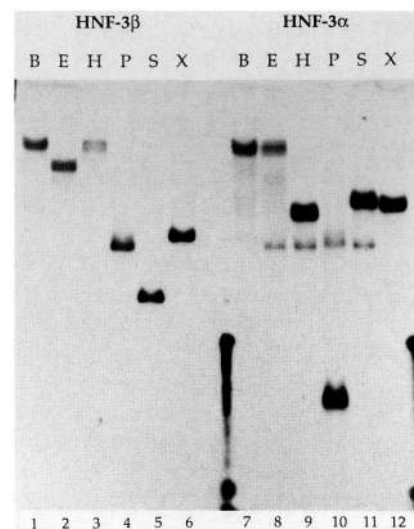
Figure 4. Southern hybridization of rat genomic DNA with specific probes to HNF-3α and HNF-3β. Digestion of genomic DNA was performed with restriction endonucleases BamHI (B), EcoRI (E), HindIII (H), PstI (P), SstI (S), and XhoI (X).
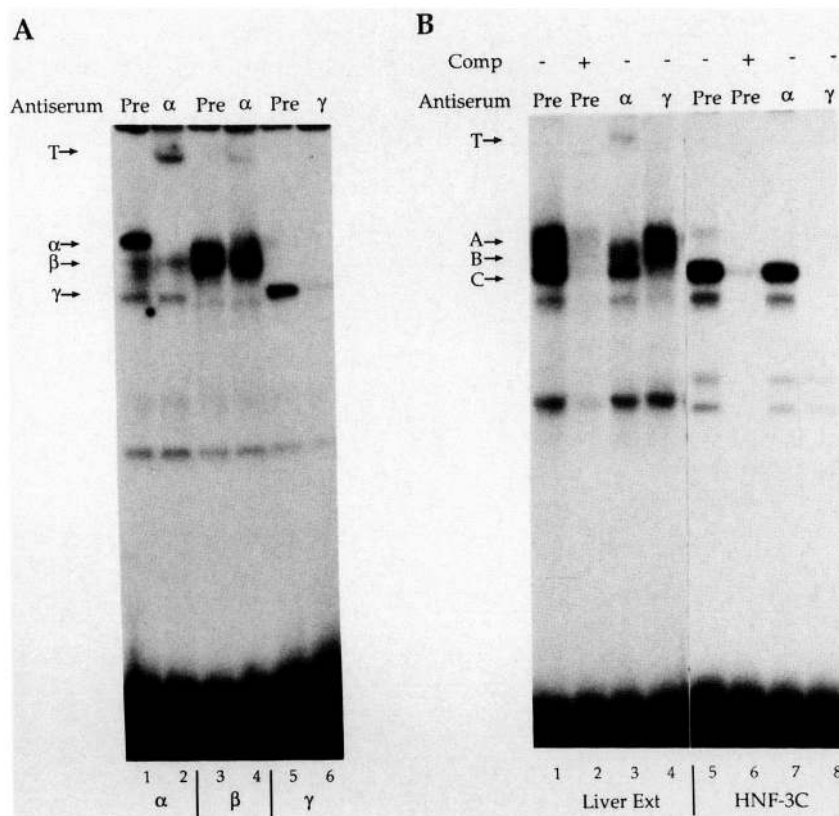
Figure 5. The HNF-3γ protein is closely related to the HNF-3C protein. (A) Anti-α serum reacts with the HNF-3α protein (lane 2) and weakly with the HNF-3β protein (lane 4). Anti-γ serum disrupts binding to DNA of the HNF-3γ protein (lane 6). Preimmune or immune serum (1 μl) was added to the standard gel mobility-shift reactions containing reticulocyte lysate-translated protein. (B) Anti-α serum specifically reacts with the A and B (weakly) complexes formed with liver nuclear extract. The C complex is not affected (lanes 3 and 7). The anti-γ serum specifically disrupts the C complex formed with liver extracts (lane 4) and with purified HNF-3C (lane 8). The reactions were performed in the presence of a nonspecific competitor (−) or a specific competitor (+) oligonucleotide (see legend to Fig. 3).

tively abolishes complex formation with HNF-3α but not with HNF-3β (Fig. 8B, lanes 5 and 11). Furthermore, a 200-fold oligo W excess is able to compete effectively for binding to HNF-3β but not to HNF-3α (Fig. 8B, lanes 9 and 3). Thus, it is clear that HNF-3α has a higher affinity for oligo S than HNF-3β, and the reverse is true for oligo W. Since both of those sites are occupied in mouse liver it seems likely that HNF-3α is bound to the −111 to −85 site (oligo S), whereas HNF-3β is bound upstream at −151 to −130. Because the DNA-binding properties of HNF-3α and HNF-3γ are more similar to each other than to those of HNF-3β, it is possible that HNF-3α and HNF-3γ are interchangeable.

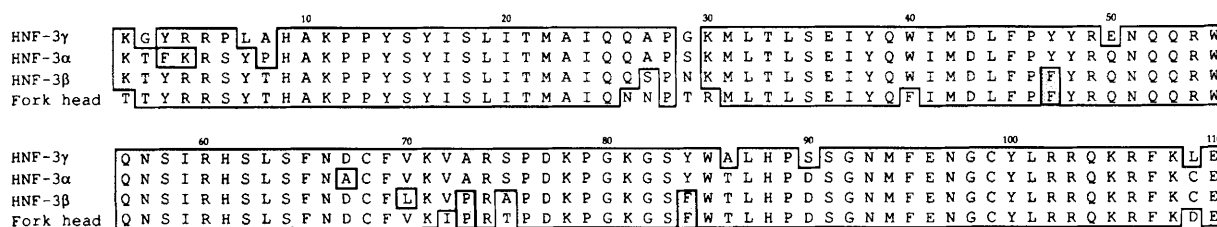### HNF-3 family members are transcriptional activators

Because early studies showed that mutation of the HNF-3 strong site markedly reduced transcription at the TTR promoter (Costa et al. 1989), it was expected that this site would be bound by transcriptional activators. With the identification of a family of proteins, however, we wanted to test which of the proteins were positive activators of transcription. This was accomplished by cotransfection of expression vectors containing the cDNA for each of the HNF-3 family members and a reporter construct with the TTR promoter (−202 to +9) containing the two HNF-3-binding sites described above fused to the chloramphenicol acetyltransferase (CAT) gene. Figure 9 shows that the cotransfection into HepG2 cells of either HNF-3α or HNF-3β stimulated the expres-

sion of the reporter gene compared to the control expression construct, which in this experiment contained a partial HNF-3β cDNA in the antisense orientation. Similar results were seen with the HNF-3γ expression construct (data not shown). In addition, the expression of a reporter gene with multiple HNF-3 sites ligated to a heterologous promoter was stimulated by cotransfection with HNF-3 expression constructs, whereas the reporter construct lacking the multiple HNF-3 sites did not respond (not shown). Taken together, the data on cotransfection of plasmids producing HNF-3 and bearing HNF-3 binding sites clearly indicate that each of the three proteins, HNF-3α, HNF-3β, and HNF-3γ, can act as positive activators of transcription.
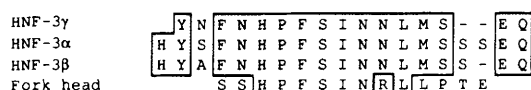
HNF-3 family members are differentially expressed

In our initial report of the cloning of an HNF-3 cDNA, the Northern analysis had shown two prominent mRNA bands at 2.0 and 2.2 kb, as well as a minor band at 3.4 kb (Lai et al. 1990). In those experiments, we used a probe that we now know spanned the homologous region of the three genes HNF-3α, HNF-3β, and HNF-3γ. Thus, we presumably detected all three of the mRNAs that we now know to exist. Using probes that are specific to each HNF-3 cDNA, we repeated the Northern analyses and found that the HNF-3α transcript is 3.4 kb, the HNF-3β transcript is 2.2 kb, and the HNF-3γ transcript is 2.0 kb (Fig. 10A). These studies also show that in the liver HNF-3β and HNF-3γ mRNAs were both more abundant
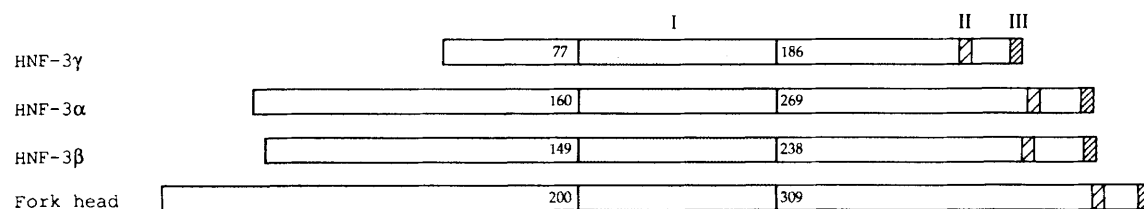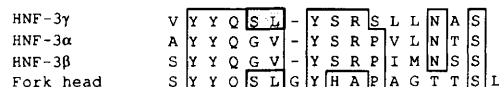
I

```
                    10              20                    30            40              50
HNF-3γ    K G Y R R P L A H A K P P Y S Y I S L I T M A I Q Q A P G K M L T L S E I Y Q W I M D L F P Y Y R E N Q Q R W
HNF-3α    K T F K R S Y P H A K P P Y S Y I S L I T M A I Q Q A P S K M L T L S E I Y Q W I M D L F P Y Y R Q N Q Q R W
HNF-3β    K T Y R R S Y T H A K P P Y S Y I S L I T M A I Q Q S P N K M L T L S E I Y Q W I M D L F P F Y R Q N Q Q R W
Fork head T T Y R R S Y T H A K P P Y S Y I S L I T M A I Q N N P T R M L T L S E I Y Q F I M D L F P F Y R Q N Q Q R W

                    60              70              80              90            100                   110
HNF-3γ    Q N S I R H S L S F N D C F V K V A R S P D K P G K G S Y W A L H P S S G N M F E N G C Y L R R Q K R F K L E
HNF-3α    Q N S I R H S L S F N A C F V K V A R S P D K P G K G S Y W T L H P D S G N M F E N G C Y L R R Q K R F K C E
HNF-3β    Q N S I R H S L S F N D C F L K V P R A P D K P G K G S F W T L H P D S G N M F E N G C Y L R R Q K R F K C E
Fork head Q N S I R H S L S F N D C F V K I P R T P D K P G K G S F W T L H P D S G N M F E N G C Y L R R Q K R F K D E
```

II

```
HNF-3γ      Y N F N H P F S I N N L M S - - E Q
HNF-3α    H Y S F N H P F S I N N L M S S S E Q
HNF-3β    H Y A F N H P F S I N N L M S S - E Q
Fork head       S S H P F S I N R L L P T E
```

III

```
HNF-3γ    V Y Y Q S L - Y S R S L L N A S
HNF-3α    A Y Y Q G V - Y S R P V L N T S
HNF-3β    S Y Y Q G V - Y S R P I M N S S
Fork head S Y Y Q S L G Y H A P A G T T S L
```

```
                               I                    II  III
HNF-3γ          |---------77|--------186|----------------Ø--Ø
HNF-3α        |----------160|--------269|------------------Ø--Ø
HNF-3β        |----------149|--------238|------------------Ø--Ø
Fork head     |----------200|--------309|-------------------------Ø--Ø
```

Figure 6. Homology of amino acid sequence among HNF-3α, HNF-3β, HNF-3γ, and the *Drosophila* nuclear protein *fork head*. Sequence comparison of the DNA-binding domain (*I*) and two short regions at the carboxyl end of the proteins (*II* and *III*). Shaded areas show sequence identity between at least two of the proteins. (*Bottom*) A schematic diagram of the proteins, showing the location of the homologous regions. The positions of residues at both ends of region *I* are indicated.

than HNF-3α mRNA. The stronger gel mobility-shift activity of HNF-3A (HNF-3α) with crude liver extracts (Fig. 5A; Lai et al. 1990) probably reflects the higher affinity of HNF-3α for the binding site (Figs. 7 and 8) rather than a larger amount of HNF-3α protein.

Using gene-specific probes, we confirmed by Northern analysis that all three mRNAs were expressed in liver (Fig. 10A) but were absent in brain, spleen (data not shown), and kidney (Fig. 10B). There was a small but definite amount of expression in the small intestine of all three mRNAs (Fig. 10A). Longer exposures of the blots of Figure 10A showed definite signals for HNF-3α and HNF-3β mRNAs.

A more extensive tissue survey revealed additional information about differential expression. HNF-3β was expressed in the lung at high levels, approximately twofold greater than in liver (Fig. 10A,B). In contrast, HNF-3γ mRNA was absent in the lung but present in the testis (Fig. 10B). The HNF-3γ-like mRNA in the testis is distinctly larger than that found in the liver. These differences were not due to differences in the amount of intact mRNA, as shown by hybridization with a probe to the GAPDH mRNA, which is present in all tissues (Fig. 10B). The distribution of HNF-3α was identical to that of HNF-3β for all of the tissues examined to date.

We have not been able to score the transcription rate of each of the family members independently because the transcription rate is low and the cDNA-derived probes necessary to separate signals from the HNF-3α, HNF-3β, and HNF-3γ genes are too short. Only when genomic clones are available will we be able to determine whether different tissue distributions reflect proportionally different transcription rates. However, no transcription signal was obtained in kidney or brain, whereas a clear signal was obtained in liver and lung using the full-length probe for HNF-3α (Lai et al. 1990; K.G. Xanthopolous et al., in prep.). Therefore, the absence of any of the HNF-3 mRNAs in a tissue likely indicates no transcription of that gene in that tissue.

Discussion

We originally identified HNF-3α as a protein that was important in the transcriptional activation of two genes, TTR and α_1-antitrypsin, which are expressed mainly in hepatocytes (Costa et al. 1989). Our current studies show that HNF-3α is only one member of a gene family that encodes at least three transcription factors in rats. Sequence comparison of the three proteins reveals a highly conserved 110-amino-acid domain, which lies within the region defined previously as neccessary for DNA binding, and all three proteins bind the same DNA sequence,
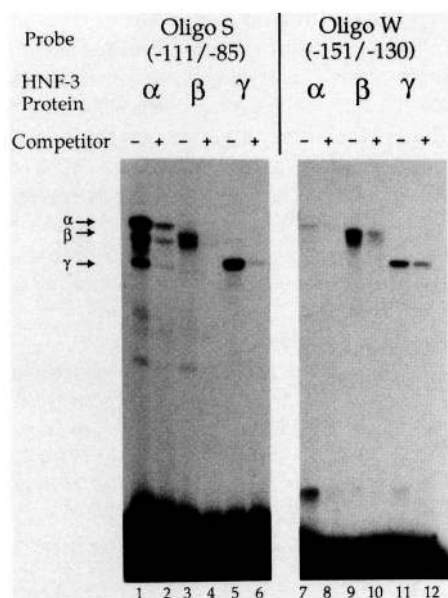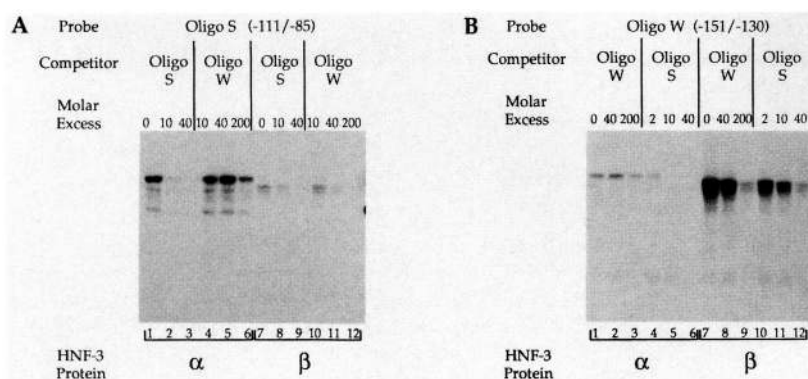
Figure 7. Relative affinity of HNF-3 proteins for two binding sites in the TTR promoter. Gel mobility-shift assays with translated proteins using the strong-site oligo S (lanes 1–6) or the weak-site oligo W (lanes 7–12) as labeled probe. The reactions were performed in the presence of a nonspecific competitor or a specific competitor (oligo S for lanes 2, 4, and 6; oligo W for lanes 8, 10, and 12). Approximately equal amounts of translated HNF-3 protein were used in all assays as estimated from the amount of [³⁵S]methionine incorporated. This experiment is representative of several using products from three separate translation reactions for each HNF-3 protein.

homeotic gene *fork head* product. Since all three rat proteins and fork head conserved the entire 110 amino acids of the DNA-binding domain, it seems likely that all of these amino acids may be required for DNA binding. For example, the DNA-binding domain in homeo box genes is conserved in vertebrates and invertebrates over a 60-amino-acid region that is required and sufficient for DNA binding (Gehring 1987). Furthermore, in the Pou homeo gene family, which has members identified in *Drosophila, Caenorhabditis elegans,* and humans, an ~50-amino-acid conserved domain (the Pou region) has greater identity than the homeo box region and is required for DNA binding (Herr et al. 1988). The other carboxy-terminal conserved regions of the HNF-3 family, which are not required for DNA binding, may conceivably be related to activating functions and either activation or repression for the *fork head* gene.

The DNA-binding domain does not resemble that of any previously identified protein and thus represents a novel DNA-binding motif. At present, we can only speculate on the nature of the HNF-3α binding. The proteins give no evidence of dimerization. For example, cotranslation of full-length and truncated versions of HNF-3α or mixtures of HNF-3α, HNF-3β, or HNF-3γ proteins does not yield complexes of intermediate mobility compared to the mobility of each complex alone. Within the 110-amino-acid conserved region, there is a cluster of basic amino acids near the carboxyl end (RRQKREK). Such regions are known to be required for DNA binding of helix–loop–helix and leucine zipper proteins and are thought to represent DNA contact points, whereas the helical portions of these proteins direct dimerization (Kouzarides and Ziff 1988; Murre et al. 1989; Turner and Tjian 1989). There is only one fairly long region (from Y/F at position 47 to V at position 72) within the 110-amino-acid conserved region of the HNF-3 family that has characteristics associated with helix formation. Except for this region, prolines and glycines, which are considered helix breakers, are liberally scattered through the remainder of region I. Such an amino acid distribution probably does not rule out helix formation but also does not allow speculation of its existence. The actual structure of the HNF-3 class of proteins is perhaps one that

albeit with different affinities. This conservation in the DNA-binding region is striking compared to the variation in the remainder of the coding region of these proteins. For example, HNF-3γ is ~140 amino acids shorter than HNF-3α, and HNF-3β and has only two short stretches of conserved sequence (II and III in Fig. 6) in addition to the DNA-binding domain. The strong conservation of the DNA-binding and the two carboxy-terminal regions (II and III) in the three rat genes is mirrored by conservation of the same regions in the *Drosophila*

Figure 8. Competition of HNF-3α and HNF-3β binding to two binding sites in the TTR promoter. (A) Binding to the strong-site oligo S of translated HNF-3α (lanes 1–6) or HNF-3β (lanes 7–12) in the presence of varying amounts of unlabeled competitor oligo S (lanes 2, 3, 8, and 9) or oligo W (lanes 4–6 and 10–12). (B) Binding to the weak-site oligo W of translated HNF-3α (lanes 1–6) or HNF-3β (lanes 7–12) in the presence of varying amounts of unlabeled competitor oligo W (lanes 2, 3, 8, and 9) or oligo S (lanes 4–6 and 10–12). Approximately equal amounts of translated HNF-3 protein were used in all assays as estimated from the amount of [³⁵S]methionine incorporated.
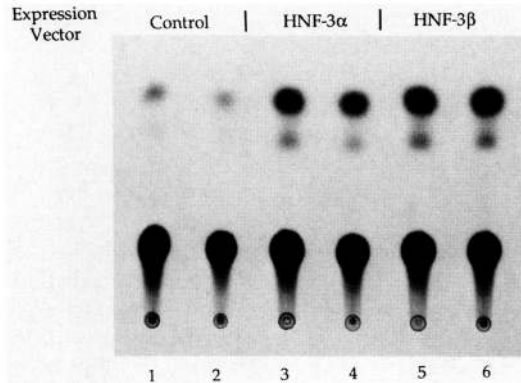
Figure 9. Transcriptional activation by cotransfection in HepG2 cells. Autoradiograph of TLC plate measuring activity of CAT produced from a reporter construct TTR–CAT when cotransfected with a control expression plasmid (lanes 1 and 2) or a plasmid expressing HNF-3α (lanes 3 and 4) or HNF-3β (lanes 5 and 6).

allows the protein to bind to DNA as a monomer, thus presenting several unique challenges to the crystallographer.

What can we say of the broader biologic role of the HNF-3 family of proteins? The identity of the DNA-binding region of HNF-3a, HNF-3β, and HNF-3γ with that of the homeotic gene *fork head* product suggests that they comprise a third group of mammalian and invertebrate genes that encode proteins having both similar DNA-binding regions and at least somewhat similar determining roles in early development. One group, the homeo box genes that control the proper segment formation in the early *Drosophila* embryo, have a counterpart in vertebrates in Hox genes that have the homeo box sequence (Akam 1989). Because of the segmental expression of Hox genes, these genes are expected to participate in proper segmental division of vertebrate embryos (Graham et al. 1989). A second group encodes the helix–loop–helix proteins, one of which, myoD, has a demonstrated role in muscle cell differentiation when overexpressed in certain susceptible cells and is a normal product during myogenesis. Another helix–loop–helix protein, E47, is thought to play a role in lymphocyte differentiation. *Drosophila* homologs of the helix–loop–helix genes include *daughterless* (Caudy et al. 1988) and *achaete-scute* (Cabrera et al. 1987; Villares and Cabrera 1987); *daughterless* affects sex determination and the formation of sensory precursor cells, whereas *achaete-scute* operates in cuticle formation. Thus, this second group of genes may be devoted mainly to cell specialization.

We have shown here that the HNF-3 family in rodents is expressed in cell types (liver, lung, and intestine) that derive from the embryonic gut tube by the outpouching of endodermal cells, which on correct contact with mesenchyme, form structures such as salivary glands, lungs, pancreas, liver, and small and large intestine. It hardly seems accidental that the *Drosophila* protein homolog of the HNF-3 family, *fork head*, is required to correctly

form the terminal embryonic structures of the larval embryo that contribute to such tissues as the salivary glands, foregut, and hindgut (Hartenstein et al. 1985; Jürgens and Weigel 1988; Weigel et al. 1989). It is logical to propose that this family of genes has maintained a role from early in evolution in forming gut-related organs. We came to the study of HNF-3 by its apparent regulatory action on the TTR gene and α_1-antitrypsin gene in adult liver. Whether one of the presently characterized HNF-3 proteins or some other member of this family has a role to play in earlier steps of hepatocyte determination now seems an important question.

Furthermore, we can ask what the function may be of different proteins of similar but not identical binding affinities such as the HNF-3 family in adult rodent tissues. First, it is obvious that different HNF-3 family members have widely different affinities for different binding sites on DNA. Thus, different HNF-3 proteins may be required to act on various sites in different genes or even in the same gene. From our analysis of cell distribution, it is clear that different tissues may vary in amounts of the different HNF-3 proteins, with HNF-3γ being absent from lung, for example. All cells deriving from endoderm may share some HNF-3-like proteins, but some family members could be more restricted in expression and serve more local functions than others. Further tissue analyses and genomic searches for more HNF-3 family members are needed to illuminate these questions.

Finally, we must search more widely in adults and embryos for members of this family to attempt to explore the hypothesis about the restricted cell lineage in which this gene family is active. For example, fork head is present in brain cells in *Drosophila*, and we did not
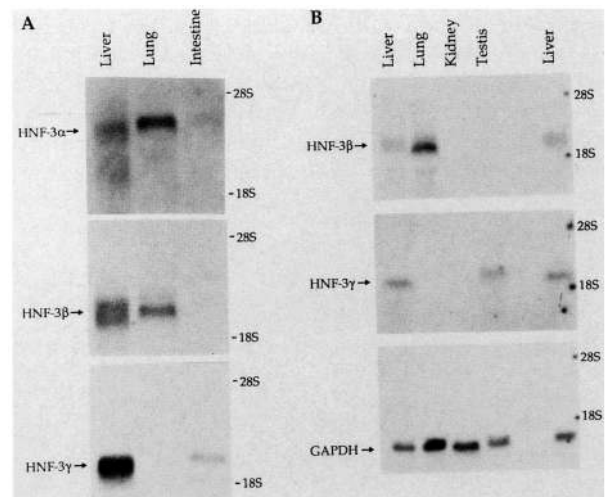
Figure 10. Northern analysis of mRNA levels for each HNF-3 gene. (A) Specific probes distinguish the major mRNA species for HNF-3α, HNF-3β, and HNF-3γ in the liver. The blot for HNF-3α was exposed for 72 hr; the blots for HNF-3β and HNF-3γ were exposed for 24 hr each. (B) HNF-3β and HNF-3γ mRNA expression in several tissues.

observe HNF-3 mRNAs in total brain tissue. However, we have not yet examined individual cell types with in situ hybridization or with specific antisera. We already know that we only detected CAAT enhancer binding protein (C/EBP), another transcription protein present in liver, by in situ hybridization in brain cells (Kuo et al. 1990). Thus, a search for region-specific expression of HNF-3 in the brain is strongly indicated.

Materials and methods

Isolation of HNF-3β and HNF-3γ cDNA

A random primer-labeled probe was prepared from the H3 cDNA (HNF-3α) and used to screen a mouse liver cDNA library (generously provided by Dr. K.E. Paulson). Two clones were isolated that proved to be homologous to the rat cDNA. The longer clone, M2, was sequenced and found to encode a protein of 329 amino acids that was missing the amino-terminal end. A highly homologous region of ~100 amino acids was followed by a divergent sequence. A 270-bp fragment from this region was used to prepare a probe for screening a rat liver cDNA library (Stratagene). One clone was identified, K2 (HNF-3β), which was homologous to the M2 clone.

The HNF-3γ cDNA was obtained by screening the rat liver cDNA library with a random primer-labeled fragment of HNF-3α. Twenty positive clones were isolated and excised as plasmids (λZAP protocol, Stratagene). Southern hybridization to the plasmid DNA digested with EcoRI was performed with probes specific to HNF-3α and HNF-3β cDNA. Seven clones did not hybridize, and six were found by restriction mapping and partial sequence determination to be overlapping fragments of a distinct cDNA. Two of these clones were ligated at a shared NcoI site to create L11/35, which encodes the full-length HNF-3γ protein sequence. Sequence was performed by the dideoxy chain-termination method on double-stranded templates. The K2 and L11/35 sequences were determined completely on both strands with overlapping subcloned and exonuclease-shortened templates, as well as with ITP-containing sequencing mixes (U.S. Biochemical Sequenase Kit) to eliminate compression artifacts.

In vitro translation and gel mobility-shift assays

Linearized plasmids were used as templates for transcription with T7 or T3 RNA polymerase (Melton et al. 1984). In vitro translation was performed by use of nuclease-treated rabbit reticulocyte lysate (Promega) according to the manufacturer's protocol. [^{35}S]Methionine (translation grade) was purchased from New England Nuclear. The reticulocyte lysate containing the translated protein was used directly in gel mobility-shift assays as described previously (Kovesdi et al. 1986) with end-labeled double-stranded oligonucleotide probes. For the comparison of DNA-binding affinities, the relative amount of translated protein was estimated by SDS-PAGE and autoradiography of the dried gel. The relative amount of label incorporated was quantitated by densitometry and then adjusted for the number of methionine residues in each HNF-3 protein being translated to arrive at an estimate of the relative amount of protein.

Preparation of antisera to HNF-3 proteins

Anti-α was obtained by immunization of two rabbits with a synthetic peptide MLGTVKMEGHESNDWC, corresponding to the 15 amino-terminal residues of HNF-3α. The terminal cys-

teine residue was added to facilitate coupling to keyhole limpet hemocyanin. Injections of 100 μg of coupled peptide with Freund's adjuvant were made subcutaneously at 4-week intervals. High-titer antiserum was obtained after three injections. Anti-γ serum was prepared by immunization of two rabbits with a fusion peptide produced in Escherichia coli. The XmnI–SstI fragment of the HNF-3γ cDNA was inserted in the T7 polymerase expression plasmid pET-3a (Rosenberg et al. 1987), which yielded an insoluble fusion peptide containing 77 amino acids of the HNF-3γ protein that was distinct from HNF-3α or HNF-3β. Bacterial cells were sonicated in 50 mM Tris-HCl (pH 7.5), 2 mM EDTA, 0.5 mM phenylmethylsulfonyl fluoride, and 0.1 M NaCl. Insoluble protein was recovered by centrifugation at 12,000g for 10 min. The pellet was resuspended in sonication buffer plus 1% SDS and centrifuged for 10 min. SDS-PAGE analysis of the supernatant showed that it was comprised of 50% of a single polypeptide of 10 kD. This preparation (200 μg) was used for each injection.

Preparation of probes specific to HNF-3α, HNF-3β, and HNF-3γ

Oligonucleotide primers were synthesized corresponding to the coding strand of nucleotides 808–827 and the noncoding strand of nucleotides 1252–1233 from the HNF-3α sequence. These sequences are identical in the HNF-3β cDNA, except for one mismatch, and flank an ~300-nucleotide nonhomologous region. These primers were used in the polymerase chain reaction with plasmids containing HNF-3α and HNF-3β cDNA as templates to prepare fragments that are specific to each gene, except for the primer sequences. The corresponding region of the HNF-3γ cDNA could be obtained by isolating the 234-bp fragment generated by restriction enzymes XmnI and SstI. Probes were prepared by random primer labeling of these fragments after denaturation.

Transfection constructs and assays

The vectors used for the expression of HNF-3 cDNAs were described previously (Chen et al. 1987). The H3 cDNA was used for HNF-3α, the K2 cDNA was used for HNF-3β, and the L11/35 cDNA was used for HNF-3γ. The control construct contained the murine HNF-3β partial cDNA, M2, in the antisense orientation. Each of these was cloned into the 5' XbaI and 3' HindIII sites of the expression vector. The XbaI and HindIII sites of each cDNA were generated by cloning each EcoRI fragment into pBluescript KS in the appropriate orientation. The reporter construct TTR–CAT was made by purifying the mouse TTR promoter sequence −202 to +9 of the TTR minigene (Costa et al. 1986), adding HindIII linkers, and fusing to the HindIII site of the CAT gene (Lew et al. 1991) without any other promoter sequence.

Transfection of human HepG2 cells was carried out by calcium phosphate precipitation (Wigler et al. 1979), except that 5 mM sodium butyrate was used after glycerol shock. pCMV-β-gal (MacGregor and Caskey 1989) was used as an internal control for transfection efficiency. For each transfection, 1 μg of pCMV-β-gal, 2 μg of TTR–CAT, and 10 μg of each expression vector were added to 5 × 10^6 HepG2 cells in 100-mm dishes. CAT assays and β-gal normalizations were carried out as described (Lew et al. 1991).

RNA analysis

RNA was isolated from rat tissues by the acid guanidinium thiocyanate/phenol-chloroform extraction method (Chomczyn-

ski and Sacchi 1987; Puissant and Houdebine 1990). Poly(A)$^+$ RNA was selected by chromatography on oligo(dT)-cellulose. Two micrograms of poly(A)$^+$ RNA from each tissue was separated on a formaldehyde–1.2% agarose gel and transferred to Zeta-probe nylon membranes (Bio-Rad) by capillary action. Hybridization was performed in 0.5 M sodium phosphate (pH 7), 1 mM EDTA, and 7% SDS at 65°C for 18–24 hr. The high-stringency wash was performed in 0.1% SSC/0.1% SDS at 65°C for 10 min. The blot was exposed to X-ray film at −80°C with two intensifying screens for 24–72 hr.

Acknowledgments

Note added in proof

Sequence data described in this paper have been submitted to the EMBL/GenBank Data Libraries.

References

Akam, M. 1989. Hox and HOM: Homologous gene clusters in insects and vertebrates. *Cell* **57**: 347–349.

Baumhueter, S., D.B. Mendel, P.B. Conley, C.J. Kuo, C. Turk, M.K. Graves, C.A. Edwards, G. Courtois, and G.R. Crabtree. 1990. HNF-1 shares three sequence motifs with the POU domain proteins and is identical to LF-B1 and APF. *Genes & Dev.* **4**: 372–379.

Cabrera, C.V., A. Martinez-Arias, and M. Bate. 1987. The expression of three members of the achaete-scute gene complex correlates with neuroblast segregation in Drosophila. *Cell* **50**: 425–433.

Caudy, M., H. Vassin, M. Brand, R. Tuma, L.Y. Jan, and Y.N. Jan. 1988. daughterless, a Drosophila gene essential for both neurogenesis and sex determination. *Cell* **55**: 1061–1067.

Chen, W.S., C.S. Lazar, M. Poenie, R.Y. Tsien, G.N. Gill, and M.G. Rosenfeld. 1987. Requirement for intrinsic protein tyrosine kinase in the immediate and late actions of the EGF receptor. *Nature* **328**: 820–823.

Chomczynski, P. and N. Sacchi. 1987. Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal. Biochem.* **162**: 156–159.

Costa, R.H., E. Lai, and J.E. Darnell Jr. 1986. Transcriptional control of the mouse prealbumin (transthyretin) gene: Both promoter sequences and a distinct enhancer are cell specific. *Mol. Cell. Biol.* **6**: 4697–4708.

Costa, R.H., D.R. Grayson, and J.E. Darnell Jr. 1989. Multiple hepatocyte-specfic nuclear factors function in the regulation of the transthyretin and a1-antitrypsin genes. *Mol. Cell. Biol.* **9**: 1415–1425.

Costa, R.H., T. Van Dyke, C. Yan, C.F. Kuo, and J.E. Darnell Jr. 1990. Similarities in transthyretin gene expression and differences in transcription factors: Liver and yolk sac compared to choroid plexus. *Proc. Natl. Acad. Sci.* **87**: 6589–

6593.

Frain, M., G. Swart, P. Monaci, A. Nicosia, S. Stampfli, R. Frank, and R. Cortese. 1989. The liver-specific transcription factor LF-B1 contains a highly diverged homeobox DNA binding domain. *Cell* **59**: 145–157.

Gehring, W.J. 1987. Homeo boxes in the study of development. *Science* **236**: 1235–1252.

Graham, A., N. Papalopulu, and R. Krumlauf. 1989. The murine and Drosophila homeobox gene complexes have common features of organization and expression. *Cell* **57**: 367–378.

Hartenstein, V., G.M. Technau, and J.A. Campos-Orterga. 1985. Fate-mapping in wild type Drosophila melanogaster. *Wilhelm Roux's Arch. Dev. Biol.* **194**: 181–195.

Herr, W., R.A. Sturm, R.G. Clerc, L.M. Corcoran, D. Baltimore, P.A. Sharp, H.A. Ingraham, M.G. Rosenfeld, M. Finney, G. Ruvkun, and H.R. Horvitz. 1988. The POU domain: A large conserved region in the mammalian *pit-1*, *oct-1*, *oct-2*, and *Caenorhabditis elegans unc-86* gene products. *Genes & Dev.* **2**: 1513–1516.

Johnson, P.F., W.H. Landschulz, B.J. Graves, and S. McKnight. 1987. Identification of a rat liver protein that binds to the enhancer core element of three animal viruses. *Genes & Dev.* **1**: 133–146.

Jürgens, G. and D. Weigel. 1988. Terminal versus segmental development in the Drosophila embryo: The role of the homeotic gene fork head. *Wilhelm Roux's Arch. Dev. Biol.* **197**: 345–354.

Kouzarides, T. and E. Ziff. 1988. The role of the leucine zipper in the fos-jun interaction. *Nature* **336**: 646–651.

Kovesdi, I., R. Reichel, and J.R. Nevins. 1986. Identification of a cellular transcription factor involved in E1A trans-activation. *Cell* **45**: 219–228.

Kozak, M. 1986. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**: 283–292.

Kuo, C.F., K.G. Xanthopoulos, and J.E. Darnell Jr. 1990. Fetal and adult localization of C/EBP: Evidence for combinatorial action of transcription factors in cell-specific gene expression. *Development* **109**: 473–481.

Lai, E., V.R. Prezioso, E. Smith, O. Litvin, R.H. Costa, and J.E. Darnell Jr. 1990. HNF-3A, A hepatocyte-enriched transcription factor of novel structure is regulated transcriptionally. *Genes & Dev.* **4**: 1427–1436.

Lew, D.J., T. Decker, I. Strehlow, and J.E. Darnell. 1991. Overlapping elements in the guanylate-binding protein gene promoter mediate transcriptional induction by alpha and gamma interferons. *Mol. Cell. Biol.* **11**: (in press).

MacGregor, G.R. and C.T. Caskey. 1989. Constructions of plasmids that express E. coli β-galactosidase in mammalian cells. *Nucleic Acids Res.* **17**: 2365.

Melton, D.A., P.A. Krug, M.R. Rebaglicti, T. Maniatis, K. Zinn, and M.R. Green. 1984. Efficient in vitro synthesis of biologically active RNA and RNA hybridization probes from plasmids containing a bacteriophage SP6 promoter. *Nucleic Acids Res.* **12**: 7035–7056.

Mirkovitch, J. and J.E. Darnell Jr. 1991. Rapid in vivo footprinting technique identifies proteins bound to the TTR gene in liver. *Genes & Dev.* **5**: 83–93.

Murre, C., P.S. McCaw, and D. Baltimore. 1989. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. *Cell* **56**: 777–783.

Puissant, C. and L. Houdebine. 1990. An improvement of the single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *BioTechniques* **8**: 148–149.

Rosenberg, A.H., B.N. Lade, D. Chui, S. Lin, I.J. Dunn, and F.W. Studier. 1987. Vectors for selective expression of cloned cD-NAs by T7 RNA polymerase. *Gene* **56:** 125–135.

Sladek, F.M., W. Zhong, E. Lai, and J.E. Darnell Jr. 1990. Liver-enriched transcription factor HNF-4 is a novel member of the steroid hormone receptor superfamily. *Genes & Dev.* **4:** 2353–2365.

Tilghman, S.M. and A. Belayew. 1982. Transcriptional control of the murine albumin/α-fetoprotein locus during development. *Proc. Natl. Acad. Sci.* **79:** 5254–5257.

Turner, R. and R. Tjian. 1989. Leucine repeats and an adjacent DNA binding domain mediate the formation of functional cFos-cJun heterodimers. *Science* **243:** 1689–1694.

Villares, R. and C.V. Cabrera. 1987. The achaete-scute gene complex of D. melanogaster: Conserved domains in a subset of genes required for neurogenesis and their homology to myc. *Cell* **50:** 415–424.

Weigel, D. and H. Jäckle. 1990. Fork head: A new eukaryotic DNA binding motif? *Cell* **63:** 455–456.

Weigel, D., G. Jürgens, F. Kuttner, E. Seifert, and H. Jäckle. 1989. The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the Drosophila embryo. *Cell* **57:** 645–658.

Wigler, M., A. Pellicer, S. Silverstein, R. Axel, G. Urlaub, and L. Chasin. 1979. DNA-mediated transfer of adenylphosphoribosyl transferase locus into mammalian cells. *Proc. Natl. Acad. Sci.* **76:** 1373–1376.

Xanthopoulos, K.G., J. Mirkovitch, T. Decker, C.F. Kuo, and J.E. Darnell Jr. 1989. Cell-specific transcriptional control of the mouse DNA-binding protein C/EBP. *Proc. Natl. Acad. Sci.* **86:** 4111–4121.