Isoforms of Hepatocyte Nuclear Factor-6 Differ in DNA-binding Properties, Contain a Bifunctional Homeodomain, and Define the New ONECUT Class of Homeodomain Proteins*

(Received for publication, February 23, 1998, and in revised form, March 23, 1998)

Vincent J. Lannoy§§, Thomas R. Bürglin¶¶, Guy G. Rousseau‡, and Frédéric P. Lemaigre***

From the ¶¶Hormone and Metabolic Research Unit, Louvain University Medical School and Christian de Duve Institute of Cellular Pathology (ICP), B-1200 Brussels, Belgium, and the ‡Department of Cell Biology, Biozentrum, University of Basel, CH-4056 Basel, Switzerland

Hepatocyte nuclear factor-6 (HNF-6) contains a single cut domain and a homeodomain characterized by a phenylalanine at position 48 and a methionine at position 50. We describe here two isoforms of HNF-6 which differ by the linker that separates these domains. Both isoforms stimulate transcription. The affinity of HNF-6α and HNF-6β for DNA differed, depending on the target sequence. Binding of HNF-6 to DNA involved the cut domain and the homeodomain, but the latter was not required for binding to a subset of sites. Mutations of the F48M50 dyad that did not affect DNA binding reduced the transcriptional stimulation of constructs that do not require the homeodomain for DNA binding, but did not affect the stimulation of constructs that do require the homeodomain. Comparative trees of mammalian, Drosophila, and Caenorhabditis elegans proteins showed that HNF-6 defines a new class, which we call ONECUT, of homeodomain proteins. C. elegans proteins of this class bound to HNF-6 DNA targets. Thus, depending on their sequence, these targets determine for HNF-6 at least two modes of DNA binding, which hinge on the homeodomain and on the linker that separates it from the cut domain, and two modes of transcriptional stimulation, which hinge on the homeodomain.

The phenotype of multicellular organisms is determined in part by the cell type-specific expression of genes. Since the initial observation that expression of liver-specific genes is controlled at the level of transcription (1), several liver-enriched transcription factors have been identified and extensively studied. These factors contain DNA-binding domains that have been conserved throughout evolution. Based on the structure of their functional domains, the liver-enriched transcription factors were classified into five families (2, 3). These include the CCAAT/enhancer binding proteins and the proline acid-rich factors, which both contain a leucine zipper, the homeodomain proteins of the hepatocyte nuclear factor (HNF)1-4 family, the winged helix proteins HNF-4 family, the proline zinc finger orphan receptor HNF-4 family. These factors are expressed not only in liver but also in a restricted number of other tissues, where they control gene transcription both in the adult and during development (2).

We have recently cloned a new type of liver-enriched transcription factor that we called HNF-6. This factor, which was originally characterized as a transcriptional activator of the liver promoter of the 6-phosphofructo-2-kinase (pfk-2) gene, is expressed in liver, brain, spleen, pancreas, and testis. Moreover, the developmental pattern of expression of HNF-6 in the mouse and the demonstration that HNF-6 can control transcription of the hnf-3β and hnf-4 genes (5) suggest a role of HNF-6 in several developmental programs. In adult rat liver, HNF-6 mediates sex-dependent effects of growth hormone (6). HNF-6 contains a cut domain and a homeodomain. Cut homeodomain proteins were originally described as the products of the Drosophila cut gene (7) and of its mammalian homologs, the mclox genes (8–12). These proteins contain three cut domains upstream of a homeodomain. Such cut repeats, which are well conserved sequences of 66–88 amino acids, function as DNA-binding domains, alone or in combination with the homeodomain (13–17).

In contrast, HNF-6 contains a single cut domain. Moreover, its homeodomain differs by the nature of its residues 48 and 50 from the 400-odd homeodomains described so far (4). Residue 48 of homeodomains, which is part of their hydrophobic core, is a phenylalanine in HNF-6, whereas it is invariably a tryptophan in the other homeodomain proteins, including the cut homeodomain proteins. Residue 50, which is located in the DNA recognition helix and is a key determinant of sequence-specific DNA binding (18, 19; for a review on homeodomain-DNA interactions, see Ref. 20), is a methionine in HNF-6, an amino acid never found at this position in other homeodomain proteins. These characteristics of the cut domain and of the homeodomain of HNF-6 have been evolutionarily conserved since we found them in the ceh-21 and F22D3.1/ceh-38 genes of...
Caenorhabditis elegans (4). In proteins that contain two DNA-binding domains, the distance between these domains may be crucial for their proper function (21). Our preliminary data suggested the existence of an isoform of HNF-6 characterized by an additional sequence of 26 amino acids in the linker region between the cut domain and the homeodomain (4). All these features of HNF-6 raised important questions that we have addressed in the present work, namely (i) does the putative isoform of HNF-6 differ from the bona fide HNF-6 only by the linker region? (ii) how does the linker sequence affect the DNA-binding properties of HNF-6? (iii) are the single cut domain and the peculiar homeodomain of HNF-6 involved in DNA binding and transcriptional stimulation? (iv) could HNF-6 dimerize on DNA? and (v) are the DNA-binding properties conserved among the cut homeodomain proteins related to HNF-6?

**EXPERIMENTAL PROCEDURES**

**Reverse Transcription (RT)-PCR—**To obtain the β isoform of HNF-6 we performed RT-PCR (Gene Amp kit from Perkin-Elmer) on rat liver RNA. The putative HNF-6 primer whose 5′ end corresponds to position 5′-GTACTGTCGACATTGCTC-3′ and 1374 of rat HNF-6 cDNA (4), namely the BGL primer (5′-ggagactattggtgacgctgcggctc-3′) and the RRE primer (5′-CTTGGCTTGTGGCTTCTCC-3′), followed by Southern blotting of the PCR product with a primer (78S) specific for the β isoform (5′-ATGGTGGCGACTGTTGCTC-3′).

**Plasmid Constructions—**pSPHNF6F6 and pSPHNF6J6 carry the rat HNF-6a and β cDNA cloned in the BamHI and XcoI sites of pSP72 (Promega), pSPcl8-hd and pSPcl8-hd/β for amino acids 1–291 and 371–410 of HNF-6α and β, respectively. The corresponding cDNA fragments were obtained by RT-PCR on rat liver RNA with the cut 5′-CTTTGAC-3′ and 1374 of rat HNF-6 cDNA (4), namely the BGL primer (5′-ggagactattggtgacgctgcggctc-3′) and the RRE primer (5′-CTTGGCTTGTGGCTTCTCC-3′), followed by Southern blotting of the PCR product with a primer (78S) specific for the β isoform (5′-ATGGTGGCGACTGTTGCTC-3′).

**DNA Binding and Transactivation Properties of HNF-6**

In Vitro Synthesis of Recombinant Proteins and Electrophoretic Mobility Shift Assay (EMSA)—To produce wild-type or mutant recombinant HNF-6α and β proteins, 1 μg of the corresponding HNF-6 cDNA under the control of the SP6 promoter was combined with 25 μl of TNT-coupled wheat germ extract system (Promega) in a 50-μl incubation reaction according to the manufacturer’s instructions. The incubation reaction was used to transfect the DNA-binding domains of Ceh-38 and of Ceh-21, psPF22D3.1/ceh-38 and pSpcEcl-21 were incubated with the wheat germ extract to yield translated products of 169 (Ceh-38) and of 157 (Ceh-21) amino acids. For EMSAs, 5 μl of the reaction mixture were incubated on ice for 20 min in a final volume of 20 μl containing 10 mM Hepes (pH 7.6), 1 mM dithiothreitol, 1 mM MgCl₂, 0.5 mM EDTA, 5 μl 25% glycerol, 4 μg of poly(dI-dC), and the 32P-labeled probe (20,000 cpm, except for Scatchard analysis, see below). These solutions were loaded on an 8% acrylamide gel (acrylamide/bisacrylamide ratio was 29:1) in 0.25× TBE buffer and electrophoresed at 200 V. The sequence of the double-stranded oligonucleotides used as probes was as follows: oligo HNF-4, 5′-AGGATAGAAATCAGATCAGTCGA-3′ (384 to −371 of the mouse hnf-4 promoter); oligo HNF-3β, 5′-AGCTTAAACCGCCTATATTGATTTTCTTC-3′ (150 to −118 of the rat hnf-3 promoter); oligo phosphoepolyryvptarboxikinase, 5′-CAAGATTGTTG-CAATCAACACTTGTTG-3′ (263 to −240 of the rat phosphoepolyryvptarboxikinase (pepck) gene promoter); oligo TTR, 5′-GGTCTGCTAAGT-CAATATCAATGTAAT-3′ (110 to −87 of the mouse transthyretin promoter); oligo GRU, 5′-ACGCCCTAACACACACCAAAAGGAATCTCAAA-CTTTCTACCC-3′ (in the glucocorticoid response unit of the first intron of the rat pgh-2 gene A); oligo PFK-2, 5′-gatgcCTTGTGATGATGATGAT-3′ (−195 to −216 of the pgh-2 L promoter); oligo CYP2C12, 5′-GCAAAATATGAGATTTTATGGTG-3′ (−53 to −30 of the rat cyp2c12 gene promoter. For affinity constant (Kᵦ) determination, EMSAs were performed with a constant amount of recombinant protein and various concentrations of probe (1.3–20.8 mU) of decreasing specific radioactivity to avoid overlabeling of the gels. Incubations lasted 1 h on ice to reach equilibrium. Bound and free probe were quantified with an Instant Imager (Packard). The bound/free ratio was calculated, and the values were plotted against the concentration of bound probe (nm) to obtain Scatchard plots in which the Kᵦ = −B/Δslope. Cell Culture and Transfection—Human heparins HepG2 cells were grown in Dulbecco’s modified Eagle’s medium supplemented with 10% fetal calf serum (FCS). Cells (8 × 10⁶ cells per 60-mm dish) were transfected overnight in medium without FCS using DOTAP, 3 μg of pSPHNF6F6β (6×)•TATA-luc or pSPHNF6F6β (6×)•TATA-luc reporter construct, 400 ng of pCE-HEFNF6α, pCE-HEFNF6β, pCE-HEFNF6αF48W, pCE-HEFNF6αM50H, pCE-HEFNF6αF48W,M50H, or pECE-HNF6αF48W, F48W, M50H, 500 ng of pRL-138 and pECE-HNF6αF48W expression control, and pDNA 1:1 total plasmid DNA. The cells were washed with phosphate-buffered saline and further incubated for 24 h in Dulbecco’s modified Eagle’s medium plus 10% FCS before measuring luciferase activities with the Dual-Luciferase kit (Promega). Luciferase activities were measured with a Lumac Luminometer and expressed as the ratio of reporter activity (firefly luciferase divided by internal control) activity (Renilla luciferase).

Cells (7 cells × 10⁶) were transfected for 6 h in Dulbecco’s modified Eagle’s medium without FCS using DOTAP and 10 μg of expression vector. Forty-eight h after transfection, the cells were washed with phosphate-buffered saline and harvested in 1 ml of 40 mM Tris-CI (pH 7.5), 1 mM EDTA, 150 mM NaCl. The cells were pelleted and resuspended in 50 mM Tris-CI (pH 7.9), 50 mM KCl, 0.5 mM EDTA, 2.5 μg/ml leupeptin, 1 μM dithiothreitol, 0.1% (v/v) Nonidet P-40, 1 μM phenylmethylsulfonyl fluoride, and 20% (v/v) glycerol. After three freeze-thaw cycles, the lysates were centrifuged and the supernatants collected. 5 μl of supernatant was used in EMSA as described above.

**Sequence Analysis—**Data base searches were performed using BLAST at the National Center for Biotechnology Information (22). Sequences were extracted using FETCH from the GCG package (23). Further packages used from GCG were GAP for pairwise comparisons and PILEUP for multiple sequence comparisons and for classification of sequences. Multiple sequence alignments were performed using MEGA (generously provided by W. Gilbert). C. elegans sequence searches were performed using the blast server available at http://www.sanger.ac.uk/DataSearch/. T. R. Bürglin, manuscript in preparation.

2 T. R. Bürglin and G. Ruskun, unpublished data.

3 Blast server available at http://www.sanger.ac.uk/DataSearch/.

4 T. R. Bürglin, manuscript in preparation.
DNA Binding and Transactivation Properties of HNF-6

RESULTS

There Are Two Isoforms of HNF-6—Besides the fully coding cDNA originally described for the 465-residue rat HNF-6 (HNF-6α), we had found a partial cDNA that was identical to the 3′-half of HNF-6α cDNA, except for a 78-bp insert within the spacer that separates the cut domain from the homeodomain (4). To extend this partial sequence, we performed an RT-PCR reaction on rat liver RNA with a PCR primer (oligo RKE) located 3′ of the 78-bp insert and a PCR primer (oligo BGL) located 5′ of the HNF-6α ATG initiator codon. Southern blotting of the PCR products with an oligonucleotide (oligo 78S) specific for the 78-bp insert showed a 1130-bp band that corresponded to the missing portion of the cDNA (data not shown).

Data base scanning with the sequence of this fragment revealed that it was identical to the corresponding HNF-6α cDNA fragment, with the exception of the 78-bp insert. This defined an isoform of HNF-6 (HNF-6β, 491 residues) that differs from HNF-6α (465 residues) only by the presence of an additional sequence of 26 amino acids in the linker between the cut domain and the homeodomain. Cloning of the rat HNF-6 gene sequence of 26 amino acids in the linker between the cut domain and the homeodomain. This suggested that the two isoforms originate from the same gene by alternative splicing.5

HNF-6α stimulates the transcription of several genes. To test the transcriptional activity of HNF-6β, we cotransfected HepG2 cells with an HNF-6α (pECE-HNF6α) or HNF-6β (pECE-HNF6β) expression vector and a luciferase reporter construct driven by a TATA box and six copies of a HNF-6 site derived from the hnf-3β (pHNF-6/HNF-3β×TATA-luc) promoter. HNF-6α and HNF-6β had similar (15–30-fold) stimulatory effects (data not shown).

HNF-6α and HNF-6β Differ in DNA-binding Properties—HNF-6β differs from HNF-6α by the length of the linker region between the cut domain and the homeodomain. This suggested that the two HNF-6 isoforms may have distinct DNA-binding properties. We therefore compared the affinity of HNF-6α and -β for the hnf-3β promoter site (HNF3β probe). To do so, the corresponding HNF-6 cDNAs were transcribed and translated in a wheat germ extract. The integrity of the recombinant proteins was controlled by SDS-polyacrylamide gel electrophoresis after incorporation of [14C]leucine (data not shown). This showed a major band of the expected molecular weight and a band of lower molecular weight that was interpreted as a product of degradation or of initiation from an internal initiation codon. Accordingly, EMSAs with the synthesized proteins showed an upper major band and a lower minor band (Fig. 1A). These bands were competed for by the cold probe and were not seen with unprogrammed extracts (Fig. 1F). We excluded that the upper band was the result of dimerization of HNF-6 with the probe (see below). Scatchard plots were performed to determine the equilibrium dissociation constants (K_D) of HNF-6α and -β. In these experiments, various amounts of probe were used with a fixed amount of protein. The K_D determined in this manner is independent of protein concentration and of the proportion of active protein in the preparation. HNF-6β bound to the HNF-3β probe with a 12-fold higher affinity than HNF-6α (Fig. 1J). Representative experiments are shown in Fig. 1, C and D. To determine whether this was due to a faster on-rate or to a slower off-rate of HNF-6β, we measured off-rate constants (k_off) experimentally and deduced the on-rate con-

5 M. Rastegar, G. G. Rousseau, and F. P. Lemaigre, unpublished observations.

FIG. 1. The HNF-6α and HNF-6β isoforms differ in DNA-binding kinetics and site recognition. A, B, and J, EMSA with the recombinant proteins and probes indicated. Arrows point to specific complexes. C, D, F, and G, representative Scatchard plots for binding of recombinant proteins to the probes indicated. I, summary of calculated K_D (calculated from the experiments shown in E and H), and k_off. In the experiments (A–D, F, G, and I), various amounts of probe were used with a fixed amount of in vitro synthesized protein. For Scatchard plots the upper complex only (see A and B) was taken into account to calculate the concentration of bound probe. E and H, rate of dissociation of HNF-6α and HNF-6β from the probes indicated. After 20 min of incubation (zero time) a 60-fold excess of cold probe was added, and binding was determined by EMSA at the time points indicated. J, the two slowest migrating complexes were not seen with unprogrammed (unpr.) extract (lanes 1 and 6 versus 2 and 7) and were inhibited by an excess of cold probe (lanes 3–5 and 8–10). Molar excess of competitor probe was 10-fold (lanes 3 and 8), 30-fold (lanes 4 and 9), and 70-fold (lanes 5 and 10).
DNA Binding and Transactivation Properties of HNF-6

| Probe  | $K_D$ (nM) | $k_{on}$ | $k_{off}$ |
|--------|------------|---------|---------|
| HNF-6β| 1.23 ± 0.09| 6       | 0.70 and 1.40 |
| TTR    | 1.18 ± 0.14| 30     | 6       |
| HNF-6α| 0.066 and 0.10| 380 | 2.60 and 3.80 |
| TTR    | 2.87 and 2.90| 80    | 25      |

Table: DNA Binding Constants of HNF-6 isoforms

Figure 1—continued

Figure 2. Sequence of HNF-6-binding sites and deduced binding site consensus. Shown are the name of the gene, the position of the site in the gene, and the sequence recognized by HNF-6. The relative affinity of HNF-6α and -β for the different sites is indicated. The abbreviations used are: r, rat; m, mouse; hnf-3β, hepatocyte nuclear factor-3β; ttr, transthyretin; pfk-2 2 pmr, 6-phosphofructo-2-kinase promoter; ffr 2 gruc, 6-phosphofructo-2-kinase glucocorticoid response unit c site; hnf-4, hepatocyte nuclear factor-4; pepck, phosphoenolpyruvate carboxykinase; α2-ug, α-2 urinary globulin; cyplc13, cytochrome P-450 enzyme 2C13; cyplc12, cytochrome P-450 enzyme 2C12; mup, major urinary protein; log, tryptophan oxygenase; α-FP, α-fetoprotein; nd, not determined.

Figure 3A. The integrity of the recombinant proteins was controlled by SDS-polyacrylamide gel electrophoresis after incorporation of [14C]leucine (data not shown). The ability of the HNF-6 mutants to bind specifically to DNA was tested by EMSA with the HNF-3β probe (Fig. 3C). Consistent with the presence of the 26-residue insert, the HNF-6β proteins migrated slower than the HNF-6α proteins. Deletion of the amino-terminal half (amino acids 1–291) of HNF-6α and -β (HNF-6 cut + hd) did not prevent binding (Fig. 3C, lanes 9 and 10 versus lanes 2 and 3). Deletion of the homeodomain (HNF-6 Δhd) was still compatible with binding (Fig. 3C, lanes 4 and 5). Deletion of the cut domain (HNF-6 Δcut) led to the loss of DNA binding (Fig. 3C, lanes 6 and 7), indicating that the homeodomain alone was unable to bind to the HNF-3β probe. Deletion of both the cut domain and the homeodomain (HNF-6 Δcut + hd) also abolished DNA binding (Fig. 3C, lane 8). An isolated cut domain or homeodomain could not be tested as they were very unstable in the wheat germ extracts (data not shown).

Wheat germ extracts produced different quantities of recombinant HNF-6 depending on the mutant tested (hence the apparently poor binding of HNF-6 Δhd in Fig. 3C, lanes 4 and 5). We therefore determined whether the differences in amount of complexes of HNF-6 proteins with the HNF-3β probe seen by EMSA reflected true differences in affinity of the HNF-6 proteins for DNA. The $K_D$ values derived from Scatchard plots for wild-type HNF-6 and for its deletants are given in Fig. 3B. These values indicate that the carboxyl-terminal half of HNF-6 (HNF-6 cut + hd) bound the probe with an affinity similar to
that of wild-type HNF-6, thereby demonstrating that the entire DNA-binding domain is located within this portion of the protein. They also show that deletion of the homeodomain barely decreased the affinity of HNF-6 for the HNF-3β probe. The latter finding suggested that the homeodomain is not required for the binding of HNF-6 to a subset of sites that includes the HNF-6 site of the HNF-3β promoter.

To address this possibility we tested the same HNF-6 deletion mutants of HNF-6 on the TTR probe, which was shown above to bind wild-type HNF-6α and -β (with 26-amino acid insert) and of the HNF-6 deletion mutants used in EMSA. Numbers refer to amino acid positions. B, affinity of HNF-6 proteins for the HNF-3β probe. C and D, EMSA with the HNF-3β probe or with the TTR probe as indicated below the lanes with unprogrammed (unpr.) wheat germ extracts or with wheat germ extracts programmed for the HNF-6 proteins indicated above the lanes.

HNF-6 Binds as a Monomer to DNA—The homeodomain of the Paired class of homeodomain proteins serves as a dimerization interface (27), and the Clox cut domain and the Clox homeodomain are known to interact with each other (13). We therefore investigated the possibility that HNF-6α and -β bind to DNA as dimers and perhaps form heterodimers. To address this point, we used the assay described by Hope and Struhl (28), in which a short and a long version of the protein studied are mixed and tested by EMSA. If dimerization occurs, one observes a complex corresponding to the long homodimer, a faster migrating complex corresponding to the short homodimer, and a complex of intermediate mobility corresponding to the heterodimer. Full-length and deleted forms of HNF-6 were synthesized in vitro and tested in this way with the TTR probe. When incubated alone with DNA, the full-length domain and the homeodomain constitute the DNA-binding region of HNF-6; (ii) the homeodomain of HNF-6 is unable to bind on its own; (iii) the homeodomain is nevertheless required for HNF-6 binding to a subset of sites that includes the HNF-6 site of the TTR promoter; and (iv) the homeodomain is not required for HNF-6 binding to a subset of sites that includes the HNF-3β promoter.
HNF-6 proteins used in the experiments. Data base searches with HNF-6 revealed that several C. elegans open reading frames are most similar to HNF-6. Of these, ceh-21 (cm18e7) was first found as an expressed sequence tag by the genome project (29), which upon sequencing was found to have a highly divergent homedomain.6 Currently, the genome sequencing project (30) has finished more than 75% of the C. elegans genome, and we have undertaken a systematic search of these sequences for cut-like genes. Fig. 5 shows the sequence alignments of the cut domains and homeodomains of all cut-related genes found to date. This list includes six C. elegans genes, as well as a MAR/SAR DNA-binding protein, SATB1, which was initially not recognized as containing cut domains and a homeodomain (31).

We then compared all cut domains and homeodomains using Pileup (23). The comparative tree based on the homeodomain sequences shows clearly three distinct groups (Fig. 6). One group is formed by several C. elegans genes and the mammalian hnf-6 genes; another group is formed by the fly and vertebrate cut/cux genes; and the last group is formed by the SATB genes. The tree of the cut domains gives the same overall picture. The homeodomain sequences are very divergent. For example, the homeodomain of rat HNF-6 is only 28% identical to that of Drosophila Cut. Thus, we modify the existing classification scheme (32, 33) to elevate the cut class to the CUT superclass containing the three following classes: SATB, CUX (cut, CDP, Cux), and ONECUT. The SATB is a highly divergent class, containing two cut domains which have arisen by duplication and an atypical homeodomain that appears to have two extra amino acids between helix 1 and helix 2. Still, the precise alignment of the first third of the SATB homeodomains is somewhat difficult, because of the high sequence divergence. The CUX class genes have three cut domains, which have also arisen by duplication from a single ancestral domain. In vertebrates, at least two families exist, Cux and Cux2, consistent with the fact that other homeobox genes have also been duplicated in vertebrate evolution (e.g. Refs. 34–36). HNF-6 and its C. elegans relatives share a single cut domain and a homeodomain; thus, we name this new class the ONECUT class. An unusual feature of the ONECUT and SATB class genes is that the absolutely conserved Trp residue of position 48 of the homedomain is replaced with a Phe. Given the important role of this residue for the formation of the hydrophobic core, we suggest that the ONECUT and SATB class genes might have a differently organized hydrophobic core where the angle between helix 1 and helix 3 might be slightly changed. Another feature of the ONECUT class genes is the presence of a Met residue at position 50, which is critical for DNA binding, suggesting that the recognized sequence and affinity could differ from other homeodomains.

Three of the C. elegans genes (i.e. ceh-21, ceh-39, and T26C11.5) are located next to each other in tandem on cosmid T26C11. This suggests that these three genes arose through duplication, which is supported by the fact that ceh-21 and ceh-39 are most similar to each other (Fig. 6). T26C11.5 is highly divergent and does not contain a cut domain; nevertheless, because it is located next to ceh-21 and ceh-39, it is most likely derived from them by a duplication event. C. elegans F17A9.6 has also only one cut domain and a highly diverged homeodomain, in which the Trp residue is replaced by a Tyr residue. It could represent a highly diverged ONECUT member as well or could be the founding member for a second different ONECUT class. R07D10.x,7 from an unfinished cosmid, is the

---

6 T. R. Bürglin and G. Ruvkun, unpublished observations.
7 L. Hillier, personal communication.
one most similar to the HNF-6 genes and most probably represents the direct orthologue. The sequence identity was 89% in the cut domain and 73% in the homeodomain, a typical value for orthologous homeodomain sequences between \textit{C. elegans} and vertebrates. This would suggest that other \textit{C. elegans} ONECUT class genes might be members of new ONECUT families and yet be found in flies and vertebrates. Expression analysis using F22D3.1/ceh-38-GFP reporter constructs

```plaintext
FIG. 5. Alignment of the amino acid sequences of the homeodomain (upper panel) and the cut domain (lower panel) of CUX superclass homeobox genes. Dots represent identities to rat HNF-6. Shaded regions indicate more highly conserved positions. Conserved or similar residues that are present in ONECUT class genes as well as in at least one of the other classes, have been marked (similar residues: hydrophobic, I, L, V, F, Y; and W; basic, K and R; acidic, D and E; OH, S and T). Dashes indicate gaps introduced to maintain alignment. The abbreviations used are: r, rat; m, mouse; h, human; c, \textit{C. elegans}; d, \textit{Drosophila}; dg, dog. Upper panel, the homeodomain (delineated by the box) sequences are shown for the three classes: the upper part shows the ONECUT class, the middle shows the CUX class, and the bottom shows the SATB class. Position 48 is highly unusual in the ONECUT and SATB class genes, as the absolutely conserved Trp is replaced by a Phe or Tyr. Furthermore, to maintain alignment with the highly conserved hydrophobic residues of helix 1, two amino acids need to be looped out in the SATB genes between helix 1 and helix 2. Note also some sequence conservation downstream of the homeodomain. Lower panel, the cut domain (delineated by the box) sequences are arranged as in A, except that the three and two, respectively, repeats of the CUX and SATB class are grouped separately (r1, r2, and r3).
```
showed that F22D3.1/ceh-38 is expressed rather ubiquitously, suggesting that this gene might have a more general role in transcription (37).

The amino acid sequence similarity of the DNA-binding domains of the members of the ONECUT class suggested that they may have similar sequence-specific DNA recognition properties. We tested this hypothesis by synthesizing in vitro the bipartite DNA-binding region, i.e., the cut domain plus the homeodomain of the C. elegans Ceh-38 and Ceh-21 proteins. According to the comparative trees (Fig. 6), these two proteins are more distantly related to HNF-6 than R07D10.x in the ONECUT class. The in vitro synthesized peptides were used in EMSAs with probes known to bind HNF-6. Fig. 7 shows that these two C. elegans peptides bound to the three mammalian probes tested. This demonstrates the relatedness not only in sequence but also in DNA-binding properties of Ceh-38, Ceh-21, and HNF-6.

Mutations of Phe-48 and Met-50 Affect the Transcriptional Activation Properties of HNF-6—The conserved F48M50 dyad in the homeodomain of the ONECUT class suggested that these amino acids confer specific properties to the proteins of this class. We therefore mutated Phe-48 of HNF-6a into a tryptophan (HNF-6a F48W) or Met-50 into a histidine (HNF-6a M50H) and also constructed the double mutant (HNF-6a F48W,M50H) to obtain the dyad that is typical of the CUX-type homeodomain (see Fig. 5). These mutants were tested with probes that require (TTR probe) or not (HNF-3β probe) the presence of the HNF-6 homeodomain for binding. As shown in Fig. 8A, these mutations still allowed binding of HNF-6 to the two probes. Similar data were obtained with the GRUC probe (Fig. 8A). Scatchard plots were constructed with the TTR probe. These demonstrated that, compared with wild-type HNF-6a, the affinity of HNF-6a F48W was unchanged, the affinity of HNF-6a M50H was 3-fold lower, and the affinity of HNF-6a F48W,M50H was 2-fold lower (Fig. 8B).

Since homeodomains are known to be involved in protein-protein contacts (38–40), we investigated the possibility that Phe-48 and Met-50 of HNF-6 influence transcriptional activity. We therefore cotransfected wild-type HNF-6a or HNF-6a whose dyad was mutated as above with reporter constructs containing the luciferase gene controlled by a TATA box and by HNF-6 sites derived from the TTR or hnf-3β gene promoters. When cotransfected with the TTR site-driven construct, the three types of mutants activated transcription as efficiently as wild-type HNF-6a (Fig. 9A). In contrast, when cotransfected with the HNF-3β site-driven construct, the double mutant HNF-6a F48W,M50H showed a 50% decrease in transcriptional activation compared with wild-type HNF-6a. HNF-6a M50H showed a small, but significant, decrease in transcriptional stimulation, whereas HNF-6a F48W was as active as wild-type HNF-6a. We verified that the various forms of recombinant HNF-6 proteins were expressed and bound to DNA at similar levels after transfection by performing EMSA with nuclear extracts prepared from cells transfected with the corresponding plasmids (Fig. 9B). Thus, mutation of the F48M50

![Fig. 6. Comparative trees of the cut domains and of the homeodomains of proteins of the CUT superclass.](image-url)
dyad did not affect activation when the homeodomain was not required for binding, as was the case with the HNF-3β site, an intact F48W,M50H dyad was required for full activation. Since the HNF-6αF48W,M50H mutant bound as well as wild-type HNF-6α to the HNF-3β probe (Fig. 9B), these transfection data suggest that the homeodomain of HNF-6 can behave as a transcription activation domain. The residual stimulation of the hnf-3β promoter-derived construct by the double mutant (Fig. 9A) could depend on regions of HNF-6 located outside the homeodomain. Indeed, our unpublished data* point to the presence of a transcription activation domain in the amino-terminal half of HNF-6. We concluded from these experiments that

*V. J. Lannoy, G. G. Rousseau, and F. P. Lemaigre, unpublished observations.
ants of the linker length. In HNF-6, this linker influenced the affinity in a different way for different targets. Thus, some genes will bind HNF-6β much better than HNF-6α, whereas others that bind HNF-6α might not bind HNF-6β. A kinetic analysis of HNF-6 binding to the HNF-3β site showed that this difference results from a strong effect of linker length on the docking ($k_{on}$ of HNF-6β 60-fold higher than that of HNF-6α) without much influence on the stability of the complex ($k_{off}$ of HNF-6β 3-fold higher than that of HNF-6α). Since these two isoforms of HNF-6 originate from the same gene, alternative splicing leads to an increase in the DNA binding repertoire of HNF-6. This mechanism has not been described so far for cut homeodomain proteins.

As HNF-6α and HNF-6β can be expressed in the same tissues (4), we have explored the possibility that these isoforms influence each other’s activity when binding to the same target. This was ruled out. First, there was no evidence for formation of heterodimers. Second, when HNF-6α stimulated the hnf-3β promoter in transfection experiments, addition of increasing amounts of HNF-6β failed to amplify or inhibit this effect (data not shown).

Another question that we addressed in this work is the role of the cut domain and of the homeodomain of HNF-6. Others have studied the role of the three cut domains of human CDP and of its canine homolog Clox (13, 14, 16, 17). They showed that GST fusion proteins containing either one of these cut repeats can bind DNA, with a broad but overlapping sequence specificity characterized by the presence of a 5′-ATCGAT-3′ motif. This motif resembles the 5′-ATCAAT-3′ sequence found in the HNF-6 binding consensus. The CDP/Clox homeodomain alone also binds DNA when fused to GST, and it does so with a sequence specificity that overlaps that of the cut domains (13). Furthermore, the CDP/Clox cut repeats and homeodomain synergize in DNA binding (16), probably as a consequence of interactions between the cut domain and the homeodomain (13). Interestingly, in target selection experiments by PCR a cut domain alone and a cut domain associated with the homeodomain yield the same consensus, which contains the cut core 5′-ATCGAT-3′ (17). The DNA-binding properties of HNF-6 bear similarities with those of CDP/Clox but also display major differences. In HNF-6, the cut domain was essential for binding. Contrary to what is seen with CDP/Clox, the HNF-6 homeodomain could not bind DNA on its own. Still, the presence of a HNF-6 homeodomain was an absolute requirement for binding to a subset of targets. Thus, there are at least two ways in which HNF-6 binds to DNA, depending on the target sequence. The requirement of a homeodomain, which in itself is unable to bind DNA, for high affinity DNA binding with another type of domain was recently observed with the SATB-1 protein, a member of the CUT superclass of homeodomain proteins. The atypical homeodomain of SATB-1 promotes DNA recognition by its matrix attachment region binding domain (43). Coincubation of HNF-6 5β and HNF-6 Δcut in EMSA and double hybrid experiments failed to show protein-protein contacts between the cut domain and the homeodomain of HNF-6. In any case, our data suggest that the linker, the cut domain, and the homeodomain are all used combinatorially to determine the binding affinity, the binding kinetics, and the sequence specificity of HNF-6.

Sequence comparisons between the cut domains and the homeodomains of cut homeodomain proteins led us to update the classification of these proteins and to define the ONECUT class and the CUT superclass. The ONECUT class is characterized by the presence of a single cut domain and by a homeodomain that contains the F48M50 dyad. The corresponding amino acids in the CUX and SATB classes are W48H50 and F48Q50, respectively. Amino acid 48 in homeodomains is part of the hydrophobic core. Given the chemical similarity between phenylalanine and tryptophan, it is not surprising that the F48W mutation in HNF-6 was devoid of effect on DNA binding. As to amino acid 50, it plays a crucial role in the DNA recognition properties of homeodomains by interacting with nucleotides 3′ of the canonical TAAT consensus (Ref. 44; for a review on homeodomain-DNA interactions, see Ref. 20). Mutation of this amino acid can modify the sequence specificity as in, for example, the case in Bicoid (19) or alter the discrimination among probes, for example, as in POU factors (45) or else leave the DNA-binding properties unaffected as the Q50A mutation in the engrailed homeodomain (18). We found here that the homeodomain of HNF-6 is required for binding to probes such as the TTR probe. As binding of the HNF-6 homeodomain alone was undetectable and crystallographic data are lacking, we cannot conclude whether the homeodomain itself contacts the TTR probe or whether it influences the binding of the cut domain. Our unpublished data indicate that the TAAT sequence in the TTR probe is not crucial for HNF-6 binding. Therefore, the HNF-6 homeodomain is unlikely to interact with DNA in a way similar to that of homeoproteins known to bind optimally to TAAT-containing sequences. Wilson et al. (44) have drawn an amino acid consensus for homeodomains that interact with TAAT-containing sequences. Interestingly, our alignment (not shown) of this consensus with the entire homeodomain of ONECUT proteins showed that the latter diverges significantly from the consensus only in the third helix, which is the DNA-recognition helix in homeodomain crystals. Our experiments indicated that mutation of the HNF-6 F48M50 dyad into W48H50, the dyad found in CUX proteins, is conservative in terms of DNA binding. However, the same mutation did decrease transcriptional stimulation of a target to which HNF-6 bound without need for the homeodomain. In contrast, this mutation did not affect activation of a construct containing a target strictly requiring the homeodomain for HNF-6 binding. Thus, when the homeodomain is not involved in DNA binding, it might be available for protein-protein contacts involved in transcriptional activation. Some homeodomains of other homeoproteins are involved in transcriptional repression. For instance, the interaction between the aminoterminal arm of the Max-1 homeodomain and the general transcription factor TBP leads to transcriptional repression (46). Also, transcriptional repression by the HOXD8 or HOX7 homeodomain requires its first or third helix, respectively (38, 39). Our study of HNF-6 now provides evidence that the homeodomain can stimulate transcription via amino acids that are specific to a class of homeodomain proteins. Indeed, C. elegans members of the ONECUT class are able to bind to HNF-6 sites and, like HNF-6, they possess an F48M50 dyad. It is therefore tempting to speculate that the C. elegans proteins will regulate transcription in a way similar to HNF-6. Also, it is likely that the evolutionary conservation of the dyad in the ONECUT class relies on similarities in the mode of transcriptional activation and DNA binding.

Acknowledgments—We thank L. Hillier, Genome Sequencing Center, Washington University, St. Louis, for communication of DNA sequence data prior to publication. We thank G. Ruvkun in whose lab part of the C. elegans project was initiated; Y. Kohara (National Institute of Genetics, Japan) and A. Coulson (Sanger Center, UK) for providing C. elegans clones; R. Costa for the HNF-6/HNF-3β(6×)-TATA and HNF-6/TTR(6×)-TATA reporter constructs; L. Bertrand and J. P. Herveg for assistance in computer work; J.-L. Danan and H. Nacer-Cherif for communication of data; and S. Neou for technical help.

REFERENCES

1. Derman, E., Krauter, K., Walling, L., Weinerberger, C., Ray, M., and Darnell, J. E. (1981) Cell 23, 731–739
2. Cereghini, S. (1996) PASEB J. 10, 267–282
