

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 stimulated transcription. The affinity of HNF-6 α and HNF-6 β 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 exten-

sively 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)¹-1 family, the winged helix proteins HNF-3 α , - β , and - γ , and the 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 (4) 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-3 β* 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

* This work was supported in part by grants from the Belgian State Program on Interuniversity Poles of Attraction, Prime Minister's Office, Federal Office for Scientific, Technical, and Cultural Affairs; from the Délégation Générale. Higher Education and Scientific Research of the French Community of Belgium; from the Fund for Scientific Medical Research (Belgium); from the National Fund for Scientific Research (Belgium); and from the Fonds de Développement Scientifique (Louvain University). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EBI Data Bank with accession number(s) Y14933.

§ Holds a fellowship from the Fonds pour la Formation à la Recherche dans l'Industrie et l'Agriculture (Belgium).

¶ Supported by Grants NF 3130-038786.93 and NF 3100-040843.94 from the Swiss National Science Foundation.

** Research Associate of the National Fund for Scientific Research (Belgium). To whom correspondence should be addressed: Hormone and Metabolic Research Unit, Louvain University Medical School and Christian de Duve Institute of Cellular Pathology (ICP), Avenue Hippocrate 75, B-1200 Brussels, Belgium. Tel.: 32 2 764 7583; Fax: 32 2 762 74 55; E-mail: lemaigre@horm.ucl.ac.be.

¹ The abbreviations used are: HNF-6, hepatocyte nuclear factor-6; EMSA, electrophoretic mobility shift assay; PFK-2, 6-phosphofructo-2-kinase; RT-PCR, reverse transcription-polymerase chain reaction; FCS, fetal calf serum; bp, base pair; oligo, oligonucleotide(s); TTR, transthyretin.

Caenorhabditis elegans (4).² 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 (200 ng) with primers whose 5' ends correspond to positions 210 and 1374 of rat HNF-6 cDNA (4), namely the BGL primer (5'-gaagatc-tatgTTGcagccCGCTCACCC-3') and the RKE primer (5'-CTT-CCCGTGTTCCTGCTCTTTCC-3'), followed by Southern blotting of the PCR product with a primer (78S) specific for the β isoform (5'-ATGGGTGGCAGTGTGCCTC-3').

Plasmid Constructions—pSPHNF6 α and pSPHNF6 β contain the rat HNF-6 α and - β cDNA cloned in the *Bam*HI and *Sac*I sites of pSP72 (Promega). pSPcut-hd and pSPcut-78-hd code for amino acids 292–444 and 292–470 of HNF-6 α and - β , respectively. The corresponding cDNA fragments were obtained by RT-PCR on rat liver RNA with the cut 5' (5'-CCCAAGCTTACCATTGCAATCAATACCAAGAGGTGCTC-3') and hd 3' (5'-CGGGATCCTCAGTCCAGACTCCTCTTCTCGC-3') primers and ligated in the *Hind*III and *Bam*HI sites of pSP72. pSPHNF6 α hd and pSPHNF6 β hd code for amino acids 1–384 and 1–410 of HNF-6 α and - β , respectively, and were obtained by an *Eco*NI-*Eco*RV deletion. pSPHNF-6 α cut codes for amino acids 1–291 and 371–465 of HNF-6 α and was generated by an *Xmn*I-*Eco*NI deletion. In pSPHNF6 β cut, which codes for amino acids 1–291 and 366–491 of HNF-6 β , the cut domain was deleted by a *Bln*I-*Xmn*I digestion. pSPHNF6 Δ (cut + hd) codes for amino acids 1–291 and was made by deleting the portion 3' of the *Xmn*I site. The site-directed mutants pSPHNF6F48W, pSPHNF6M50H, and pSPHNF6F48W,M50H were made using a PCR-based strategy with the following primers: 5'-CTGTGCACTGCTGTTTCATGAATGCG-3' (F48W), 5'-CAGCACTTCT-TCCACAATGCGAGAAG-3' (M50H), and 5'-CAGCACTGCTGTTCCACAAT-GCGAGAAG-3' (F48W,M50H); mutated codons are underlined. pHNF-6/HNF-3 β (6 \times)-TATA-luc and pHNF-6/TTR (6 \times)-TATA-luc contain six HNF-6-binding sites from the *hnf-3 β* promoter (–141 to –127) or from the transthyretin (TTR) promoter (–11 to –86) upstream of a TATA box and a luciferase gene. pECE-HNF6 α , which is identical to pECE-HNF6 described in Lemaigre *et al.* (4), and pECE-HNF6 β contain the cDNA of HNF-6 α and - β downstream of the SV40 early promoter in pECE72. pECE-HNF6F48W, pECE-HNF6M50H, and pECE-HNF6F48W,M50H were generated by deleting a *Bst*EII/*Xba*I fragment from pECE-HNF6 α and replacing it by the corresponding *Bst*EII/*Eco*RV fragment derived from pSPHNF6F48W, pSPHNF6M50H, and pSPHNF6F48W,M50H. pRL-138 contains the rat *pfk-2* liver promoter (–138 to +87, *i.e.* devoid of HNF-6-binding site) cloned upstream of the Renilla luciferase coding sequence. pGEM-3 was from Promega. The cDNA of *F22D3.1/ceh-38* (yk34h2) and *ceh-21* (yk90c4) was excised from λ ZAPII (Stratagene) libraries by standard procedures. The fragments containing the cut domain plus the homeodomain of *F22D3.1/ceh-38* and of *ceh-21* were isolated by PCR from the corresponding cDNA with the *F22D3.1/ceh-38* cut 5' (5'-ACCATGGCAATAGACACAAA-GATCTTTGCG-3') and *F22D3.1/ceh-38* hd 3' (5'-TCAACCGAGTCGACTTC-TACGACG-3') primers and with the *ceh-21* hd cut 5' (5'-ACCATGGCACTC-GACACTGTGACATTGCTC-3') and *ceh-21* hd 3' (5'-TCAGTCAATGCG-AAGACGACGCGC-3') primers, respectively. The *F22D3.1/ceh-38* and *ceh-21* fragments were ligated in the *Hind*III and *Bam*HI and in the *Xba*I and *Xho*I sites of pSP72 to yield pSPF22D3.1/*ceh-38* and pSPceh-21, respectively. pGEXHNF6 β , coding for a GST-HNF6 β fusion protein, was obtained by subcloning a *Clal*/*Hind*III fragment from pSPHNF6 β into the *Eco*RI and *Hind*III sites of pGEX-KG.

In Vitro Synthesis of Recombinant Proteins and Electrophoretic Mobility Shift Assay (EMSA)—To produce wild-type or mutant recombinant HNF-6 α and - β proteins, 1 μ g of the corresponding HNF-6 cDNA under the control of the SP6 promoter was combined with 25 μ l of TNT-coupled wheat germ extract system (Promega) in a 50- μ l incubation mixture according to the supplier's instructions. To produce the DNA-binding domains of Ceh-38 and of Ceh-21, pSPF22D3.1/*ceh-38* and pSPceh-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 μ l of the reaction mixture were incubated on ice for 20 min in a final volume of 20 μ l containing 10 mM Hepes (pH 7.6), 1 mM dithiothreitol, 1 mM MgCl₂, 0.5 mM EGTA, 50 mM KCl, 10% (v/v) glycerol, 4 μ g 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 \times TBE buffer and electrophoresed at 200 V. The sequence of the double-stranded oligonucleotides used as probes was as follows: oligo HNF-4, 5'-AGGATAGAAGTCAATGATCTGGGA-3' (–394 to –371 of the mouse *hnf-4* promoter); oligo HNF-3 β , 5'-AGCTTAAGGCCCG-ATATTGATTTTTTTTCTCC-3' (–150 to –118 of the rat *hnf-3 β* promoter); oligo phosphoenolpyruvate carboxykinase, 5'-CAAAGTTAGT-CAATCAAGCTTG-3' (–263 to –240 of the rat phosphoenolpyruvate carboxykinase (*pepck*) gene promoter); oligo TTR, 5'-GTCTGCTAAGT-CAATAATCAGAAT-3' (–110 to –87 of the mouse *transthyretin* promoter); oligo GRUc, 5'-AGCTTCAAACAAACAAAAATCCATAA-CTTTCA-3' (in the glucocorticoid response unit of the first intron of the rat *pfk-2* gene A); oligo PFK-2, 5'-gacGCTTTGAAATTGATTC-AAAGC-3' (–195 to –216 of the *pfk-2* L promoter); oligo CYP2C12, 5'-GCAAAATATTGATTTTTATGGTG-3' (–52 to –30 of the rat *cyp2c12* gene promoter). For affinity constant (K_D) determination, EMSAs were performed with a constant amount of recombinant protein and various concentrations of probe (1.3–20.8 nM) 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 $K_D = -1/\text{slope}$.

Cell Culture and Transfection—Human hepatoma HepG2 cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal calf serum (FCS). Cells (8 \times 10⁵ cells per 60-mm dish) were transfected overnight in medium without FCS using DOTAP, 3 μ g of pHNF-6/HNF3 β (6 \times)-TATA-luc or pHNF-6/TTR(6 \times)-TATA-luc reporter construct, 400 ng of pECE-HNF6 α , pECE-HNF6 β , pECE-HNF6F48W, pECE-HNF6M50H, or pECE-HNF6F48W,M50H, 500 ng of pRL-138 as internal control, and pGEM-3 to adjust plasmid amounts to 5 μ g. 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) to internal control activity (Renilla luciferase).

Cos-7 cells (6 \times 10⁵) were transfected for 6 h in Dulbecco's modified Eagle's medium without FCS using DOTAP and 10 μ g 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-Cl (pH 7.5), 1 mM EDTA, 150 mM NaCl. The cells were pelleted and resuspended in 50 mM Tris-Cl (pH 7.9), 500 mM KCl, 0.5 mM EDTA, 2.5 μ g/ml leupeptin, 1 mM dithiothreitol, 0.1% (v/v) Nonidet P-40, 1 mM phenylmethylsulfonyl fluoride, and 20% (v/v) glycerol. After three freeze-thaw cycles, the lysates were centrifuged and the supernatants collected. 5 μ l 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 MSE (generously provided by W. Gilbert). *C. elegans* sequence searches were performed using the blast server.³ Open reading frames of unfinished *C. elegans* cosmid sequences were analyzed using GENEFINDER within ACEDB (24) on a SUN SPARCStation5. Cosmids were examined in detail using a special dot matrix program⁴ to ascertain that no additional cut domains were present in the *C. elegans* ONECUT genes. Accession numbers are as follows: T26C11 (containing *ceh-21* also

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

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

⁴ T. R. Bürglin, manuscript in preparation.

known as *T26C11.6*, *ceh-39* also known as *T26C11.7*, and *T26C11.5*, U41017; F22D3 (containing *ceh-38* also known as *F22D3.1*), U28993; *F17A9.6*, AF016417; *dgclox*, X69017; *hCDP*, M74099, P39880; *rcdp-2*, U09229; *rhnf-6α*, X96553; *rhnf-6β*, Y14933; *mhnf-6*, U95945; *hHNF-6*, U77975; *mcux*, X75013; *mcux-2* (also known as *cutl2*), U45665; *msatb1*, U05252; *hSATB1*, Q01826, M97287; *cm18e7*, AF023470. cDNAs for *ceh-21*, *cm18e7* and *yk90c4*; cDNA for *F22D3.1/ceh-38*, *yk34h2*. *R07D10.x* is an open reading frame found on the overlapping cosmids *R07D10* and *C17H12*.

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 indicated 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(6×)-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. 1J). 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

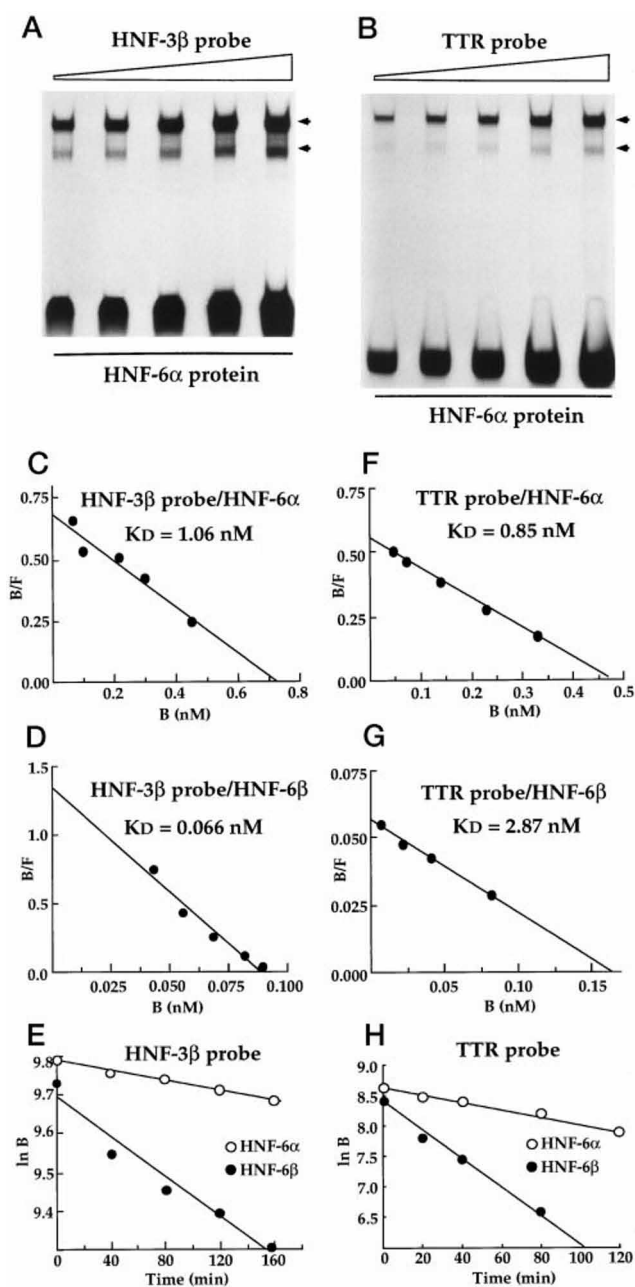


FIG. 1. The HNF-6α and HNF-6β isoforms differ in DNA-binding kinetics and site recognition. A, B, and J, EMSA with the recombinant proteins and probes indicated. Arrows point to specific complexes. C, D, F, and G, representative Scatchard plots for binding of recombinant proteins to the probes indicated. I, summary of calculated K_D , k_{on} (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).

HNF-6α (Fig. 1I). 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.

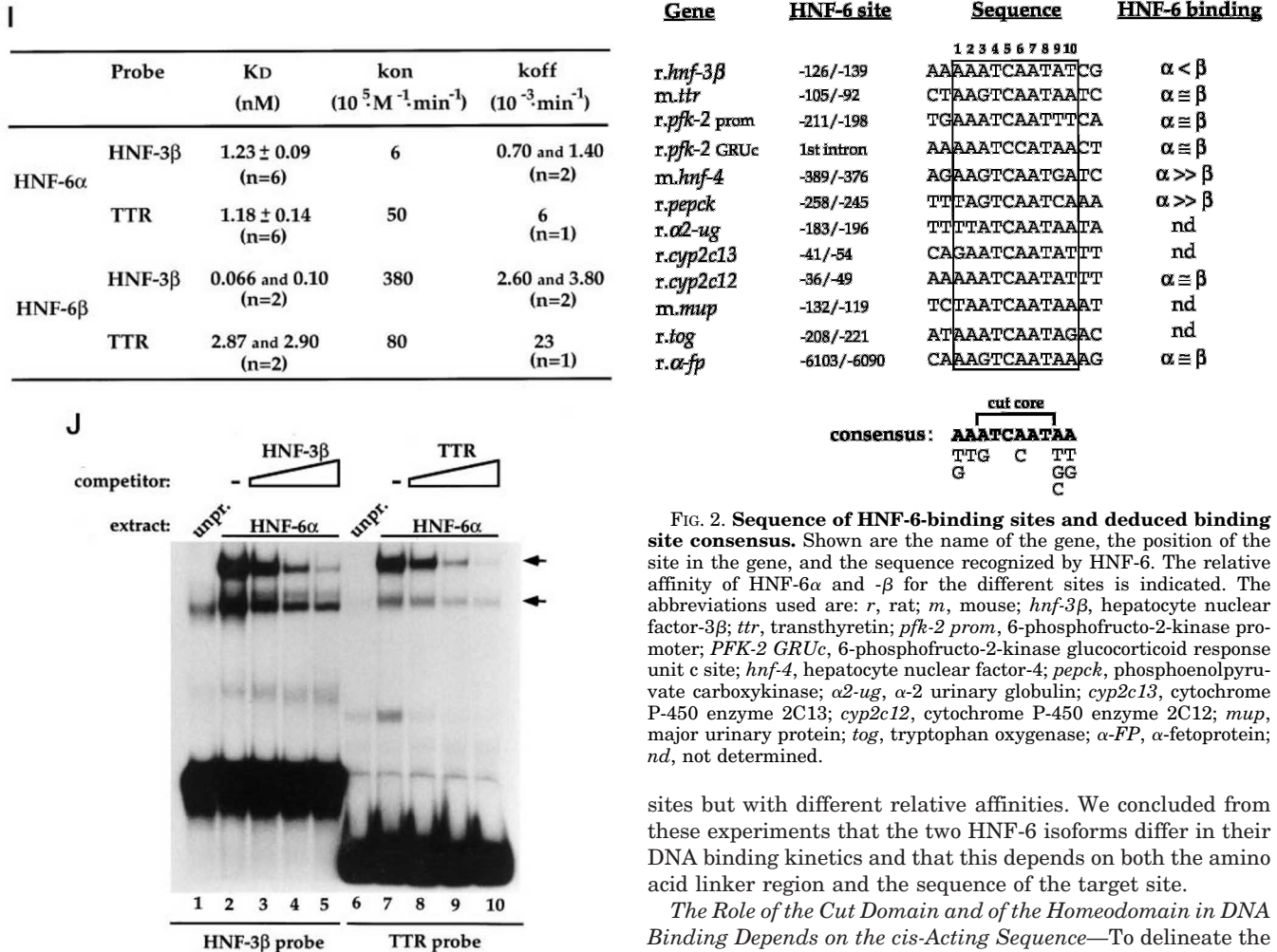


FIG. 1—continued

stants (k_{on}) from the relationship $k_{on} = k_{off}/K_D$ (Fig. 1I). As shown in Fig. 1E, the k_{off} of HNF-6β was actually 3-fold higher than that of HNF-6α, whereas its k_{on} (values calculated from the K_D in Fig. 1, C and D) was 60-fold higher than that of HNF-6α. Thus, if HNF-6α and HNF-6β were to compete at identical concentrations for the HNF-3β probe, HNF-6β would leave little chance to HNF-6α for occupying this binding site.

We had shown earlier that HNF-6 binds in the TTR promoter to a site whose sequence differs from that of the HNF-3β site (25). We therefore determined whether the differences in affinity of HNF-6α and HNF-6β reported above with the HNF-3β probe held true for the TTR probe. The data in Fig. 1, B, F, G, and I, show that this was not the case. The affinity of HNF-6β for the TTR probe was not higher but was in fact 3-fold lower than that of HNF-6α (Fig. 1I). Representative experiments are shown in Fig. 1, F and G. The off-rate constants (Fig. 1H) for dissociation from the TTR probe showed that HNF-6β dissociates four times faster than HNF-6α from this probe also (Fig. 1I). It is noteworthy that both HNF-6α and -β dissociate 10 times faster from the TTR probe than from the HNF-3β probe. Thus, HNF-6β differed from HNF-6α not only in terms of affinity for a given binding site but also in exhibiting differences in affinity for different targets. We therefore compared the binding of the two recombinant isoforms to several of the known HNF-6 sites. Fig. 2 shows an alignment of HNF-6 sites based on our observations and those of Samadani and Costa (26) and summarizes the relative affinities of HNF-6α and -β for these sites. As can be seen, the two isoforms bound to all

sites but with different relative affinities. We concluded from these experiments that the two HNF-6 isoforms differ in their DNA binding kinetics and that this depends on both the amino acid linker region and the sequence of the target site.

The Role of the Cut Domain and of the Homeodomain in DNA Binding Depends on the cis-Acting Sequence—To delineate the DNA-binding region of HNF-6, and to gain insight into the respective roles of the cut domain and of the homeodomain, we constructed deletion mutants and synthesized them *in vitro* (Fig. 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 Δ(cut + 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

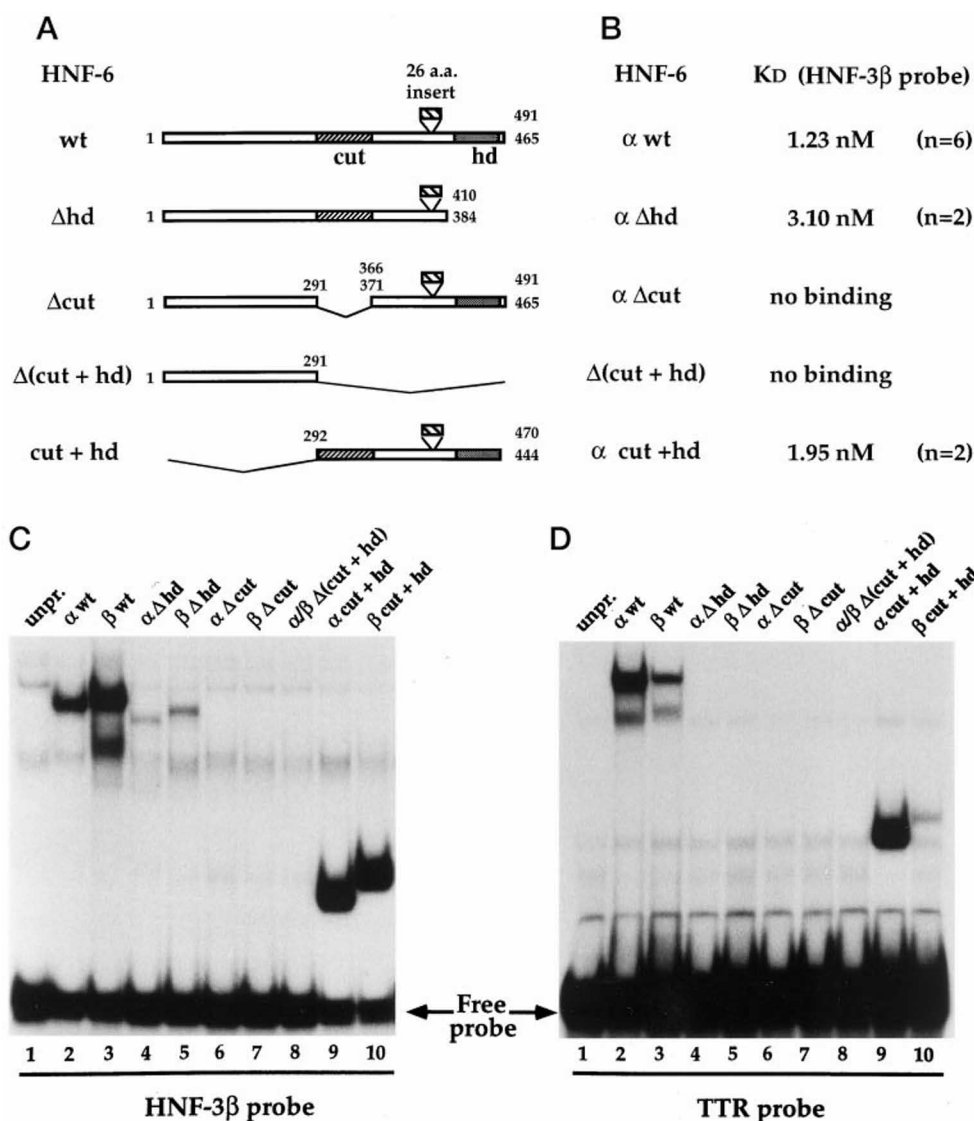


FIG. 3. Role of the cut domain and of the homeodomain in DNA binding. A, schematic representation of wild-type (*wt*) HNF-6 α and - β (with 26-amino acid insert) and of the HNF-6 deletion mutants used in EMSA. Numbers refer to amino acid positions. B, affinity of HNF-6 proteins for the HNF-3 β probe. C and D, EMSA with the HNF-3 β 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.

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-3 β 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-3 β 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-6 α and - β . The same amount of HNF-6 β yielded less complexes, relative to HNF-6 α , with the TTR probe than with the HNF-3 β probe (lanes 2 and 3 in Fig. 3D versus C). Here again, the amino-terminal half of the HNF-6 proteins was dispensable for binding (Fig. 3D, lanes 9 and 10). The poor binding of HNF-6 β cut + hd was in keeping with the low affinity of HNF-6 β for the TTR probe. Deletion of both the cut domain and the homeodomain abolished binding (Fig. 3D, lane 8). In contrast to what we had seen with the HNF-3 β probe, both the cut domain and the homeodomain were required for binding to the TTR probe since deletion of either of the two domains abolished binding (Fig. 3D, lanes 4–7). We concluded from these experiments that (i) the cut

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 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-6 α 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

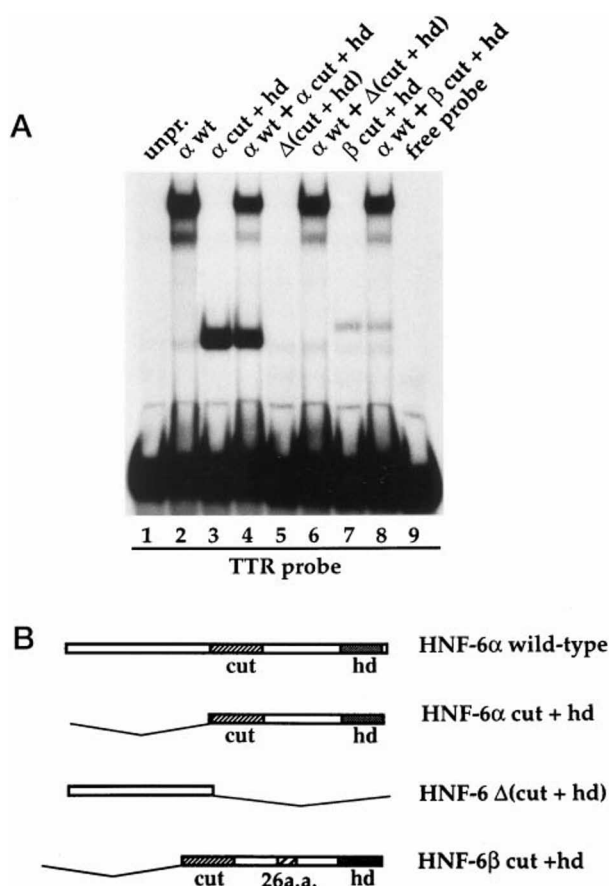


FIG. 4. **HNF-6 binds as a monomer to DNA.** A, EMSA with the TTR probe and with the HNF-6 proteins indicated above the lanes. When two proteins were combined, the synthesis of each protein was performed separately before their mixing in the binding reaction. *unpr.*, unprogrammed; *wt*, wild type. B, schematic representation of the HNF-6 proteins used in the experiments.

HNF-6 α (Fig. 4A, lane 2) and the carboxyl-terminal portion of HNF-6 that contains the cut domain and the homeodomain of the α isoform (HNF-6 α cut + hd) (Fig. 4A, lane 3) or of the β isoform (HNF-6 β cut + hd) (Fig. 4A, lane 7) produced the expected complexes. When full-length HNF-6 α was combined with these deletants, no complex indicative of dimerization was seen (Fig. 4A, lanes 4 and 8), excluding the occurrence of a dimerization interface in the DNA-binding region. There was no dimerization domain in the remaining portion of the protein either, as no complex indicative of dimerization was seen when combining full-length HNF-6 α with the amino-terminal portion of HNF-6 (HNF-6 Δ (cut + hd), which does not bind alone, see lane 5) (Fig. 4A, lane 6). Identical results were obtained by using extracts containing cotranslated proteins of different length rather than by mixing the extracts primed for synthesizing individual proteins (data not shown). In addition, when a fusion protein containing full-length HNF-6 β fused to glutathione S-transferase (GST) was incubated with full-length HNF-6 β , no intermediate complex was detected (data not shown). Results identical to those observed with the TTR probe were obtained with the HNF-3 β probe (data not shown). These results suggest that the two isoforms of HNF-6 bind to DNA as monomers and that they do not form heterodimers.

HNF-6 and *C. elegans* Genes Define a New Class of Cut Homeobox Genes—The sequence of the DNA-binding domains of HNF-6 is quasi-identical in rat, mouse, and man (Fig. 5). As basic developmental processes of mammals are very similar to those in primitive organisms, it is not surprising that DNA-binding domains of developmentally crucial transcription fac-

tors have been conserved between distantly related species. 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 homeodomain.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 homeodomain 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.

cut domain

m HNF-6
h HNF-6
c R07D10.x
c ceh-39
c ceh-21
c T26C11.5
c F22D3.1
c F17A9.6

d cut
m Cux
h CDP
r CDP2
dg Clox
m Cux2

m SATB1
h SATB1

cut domain

r HNF-6
m HNF-6
h HNF-6
c R07D10.x
c F22D3.1
c ceh-39
c ceh-21
c F17A9.6

d cut r1
m Cux r1
h CDP r1
dg Clox r1
m Cux2 r1

d cut r2
m Cux r2
h CDP r2
r CDP2 r2
dg Clox r2
m Cux2 r2

d cut r3
m Cux r3
h CDP r3
r CDP2 r3
dg Clox r3
m Cux2 r3

m SATB1 r1
h SATB1 r1
m SATB1 r2
h SATB1 r2

FIG. 5. **Alignment of the amino acid sequences of the homeodomain (*upper panel*) and the cut domain (*lower panel*) of CUT 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, M, 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*, *C. elegans*; *d*, *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*).

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 *C. elegans*

and vertebrates. This would suggest that other *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

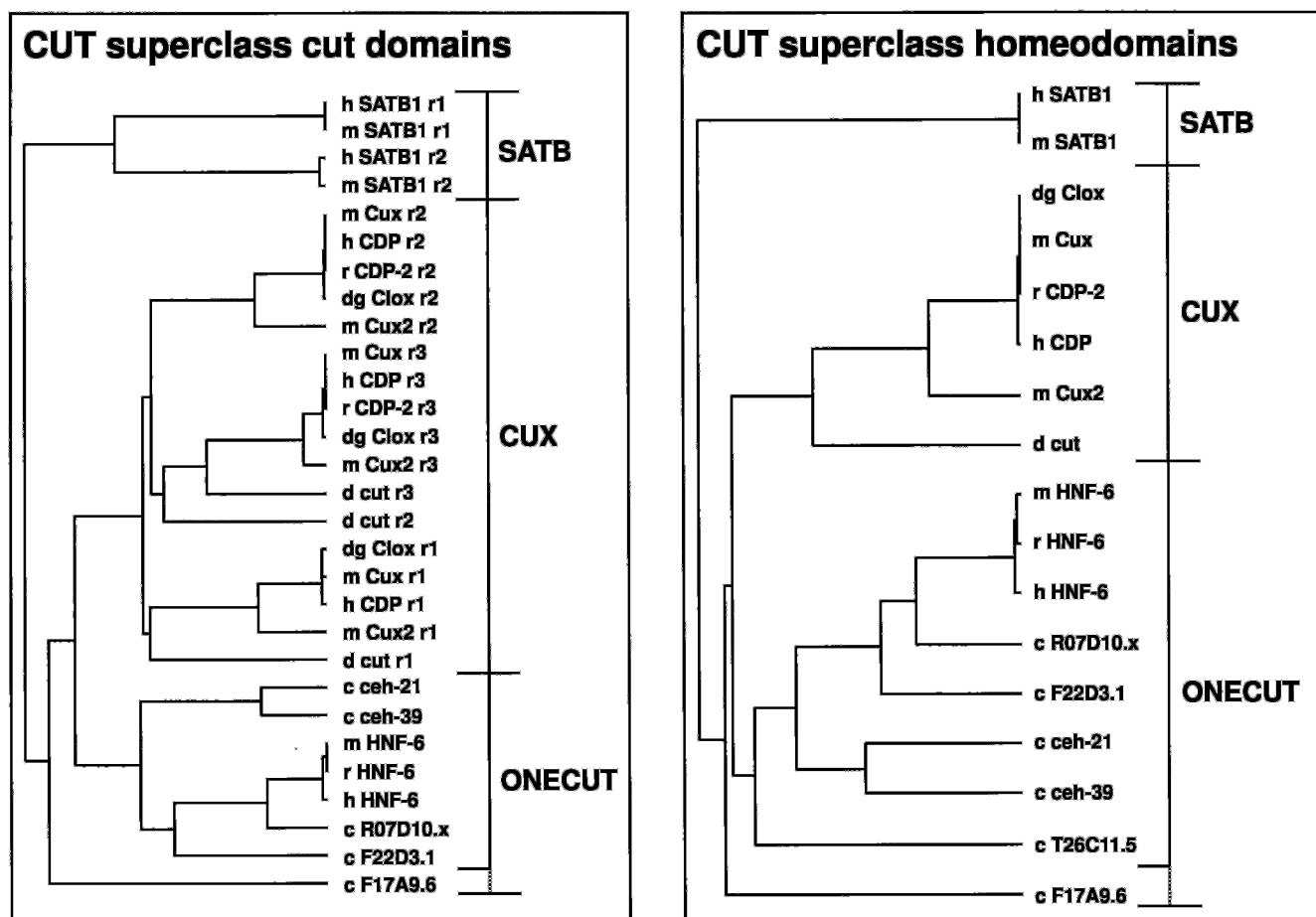


FIG. 6. Comparative trees of the cut domains and of the homeodomains of proteins of the CUT superclass. The CUT superclass is divided into the following three classes: the CUX class (three cut domains), the SATB class (two cut domains), and the newly defined ONECUT class (a single cut domain). F17A9.6 is a divergent CUT superclass gene with a single cut domain and a Tyr residue instead of Trp at position 48 of the homeodomain. It might be a highly diverged member of the ONECUT class or might represent a distinct novel class. Species codes as in Fig. 5.

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, Depending on the Target Sequence—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-6 α into a tryptophan (HNF-6 α F48W) or Met-50 into a histidine (HNF-6 α M50H) and also constructed the double mutant (HNF-6 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-3 β probe) the presence of the HNF-6 ho-

meodomain 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-6 α , the affinity of HNF-6 α F48W was unchanged, the affinity of HNF-6 α M50H was 3-fold lower, and the affinity of HNF-6 α 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-6 α or HNF-6 α 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-3 β* gene promoters. When cotransfected with the TTR site-driven construct, the three types of mutants activated transcription as efficiently as wild-type HNF-6 α (Fig. 9A). In contrast, when cotransfected with the HNF-3 β site-driven construct, the double mutant HNF-6 α F48W,M50H showed a 50% decrease in transcriptional activation compared with wild-type HNF-6 α . HNF-6 α M50H showed a small, but significant, decrease in transcriptional stimulation, whereas HNF-6 α F48W was as active as wild-type HNF-6 α . 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

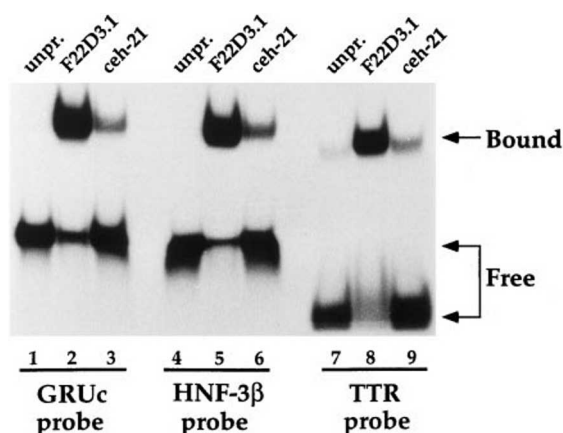
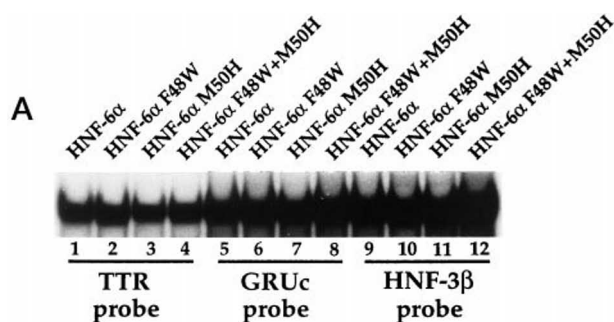


FIG. 7. *C. elegans* members of the ONECUT class recognize HNF-6-binding sites from rat genes. The probes shown were incubated in EMSA with an unprogrammed wheat germ extract (*unpr.*) or with extracts programmed for the bipartite DNA-binding domain (cut + hd) of the *C. elegans* proteins coded by the F22D3.1/*ceh-38* or *ceh-21* genes as indicated above the lanes.



B

	KD (TTR probe)
HNF-6α	1.18 nM (n=6)
HNF-6α F48W	1.10 nM (n=2)
HNF-6α M50H	3.20 nM (n=2)
HNF-6α F48W+M50H	2.20 nM (n=1)

FIG. 8. Effect of mutations F48W and/or M50H in the HNF-6 homeodomain on DNA recognition and DNA binding affinity. A, EMSA with the probes and the HNF-6 proteins indicated. B, affinity of wild-type and mutant HNF-6α for the TTR probe.

dyad did not affect activation when the homeodomain was required for HNF-6 binding. When the homeodomain was not required for binding, as was the case with the HNF-3β site, an intact F48M50 dyad was required for full activation. Since the HNF-6α F48W,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 data⁸ point to the presence of a transcription activation domain in the amino-terminal half of HNF-6. We concluded from these experiments that

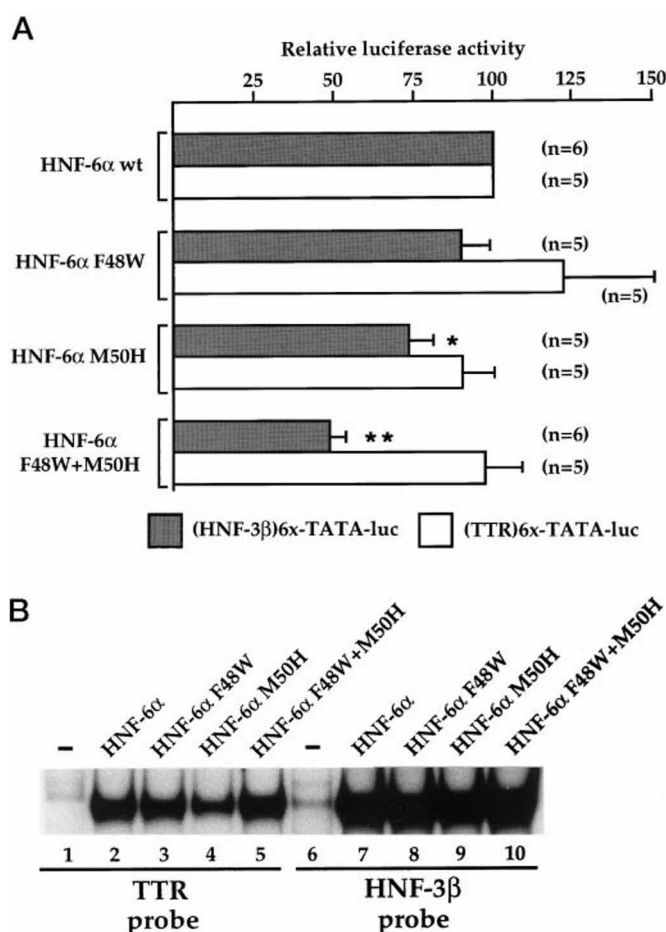


FIG. 9. Effect of mutations F48W and/or M50H in the HNF-6α homeodomain on transcriptional stimulation. A, the transfection data are expressed as percentage of stimulation relative to wild-type HNF-6α in each individual experiment. Depending on the experiment, wild-type HNF-6α stimulated the pHNF-6/HNF-3β(6×)-TATA-luc construct 15–30-fold and the pHNF-6/TTR(6×)-TATA-luc construct 4–6-fold. The fold stimulation by the HNF-6 proteins was calculated by normalizing for the fold stimulation observed with the empty expression vector (pECE72). Data are means ± S.E. *, $p < 0.05$; **, $p < 0.001$, compared with wild-type HNF-6α. B, recombinant wild-type HNF-6α and mutant proteins are expressed at similar levels after transfection. Extracts from Cos-7 cells transfected with the expression vectors indicated were prepared and used as a source of proteins in EMSA with the probes indicated.

the homeodomain of HNF-6 can be involved in both DNA binding and transcriptional activation.

DISCUSSION

We have characterized here an isoform of HNF-6 (HNF-6β) which differs from HNF-6α by a 26-amino acid insert in the linker region between the cut domain and the homeodomain, and we have shown that this isoform can activate transcription. Several homeodomain proteins contain a second type of DNA-binding domain. The role of the linker between these two domains has been extensively studied in the POU factors, in which a POU homeodomain is associated with a POU-specific domain. In this family of proteins, the use of mutant (21) and chimeric (41) proteins showed that both linker length and linker composition affect DNA binding affinity and specificity. For OCT-1, the POU homeodomain and the POU-specific domain bind to DNA cooperatively, even when expressed as separate peptides not joined by the linker (42). Therefore, the linker of the POU factors facilitates binding by maintaining a high local concentration of the two domains. To our knowledge, HNF-6 provides the first example of naturally occurring vari-

⁸ 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-6 β much better than HNF-6 α , whereas others that bind HNF-6 α might not bind HNF-6 β . A kinetic analysis of HNF-6 binding to the HNF-3 β site showed that this difference results from a strong effect of linker length on the docking (k_{on} of HNF-6 β 60-fold higher than that of HNF-6 α) without much influence on the stability of the complex (k_{off} of HNF-6 β 3-fold higher than that of HNF-6 α). Since these two isoforms of HNF-6 originate from the same gene,⁵ 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-6 α and HNF-6 β 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-6 α stimulated the *hnf-3 β* promoter in transfection experiments, addition of increasing amounts of HNF-6 β 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 Δ hd and HNF-6 Δ cut in EMSA and double hybrid experiments failed to show protein-protein contacts between the cut domain and the homeodomain of HNF-6.⁸ 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 data⁸ 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 amino-terminal arm of the Msx-1 homeodomain and the general transcription factor TBP leads to transcriptional repression (46). Also, transcriptional repression by the HOXD8 or HOXA7 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-3 β (6 \times)-TATA and HNF-6/TTR(6 \times)-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., Weinberger, C., Ray, M., and Darnell, J. E. (1981) *Cell* **23**, 731–739
2. Cereghini, S. (1996) *FASEB J.* **10**, 267–282

3. Sladek, F. M., and Darnell, J. E. (1992) *Curr. Opin. Genet. & Dev.* **2**, 256–259
4. Lemaigre, F. P., Durviaux, S. M., Truong, O., Lannoy, V. J., Hsuan, J. J., and Rousseau, G. G. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 9460–9464
5. Landry, C., Clotman, F., Hioki, T., Oda, H., Picard, J. J., Lemaigre, F. P., and Rousseau, G. G. (1997) *Dev. Biol.* **15**, 247–257
6. Lahuna, O., Fernandez, L., Karlsson, H., Maiter, D., Lemaigre, F. P., Rousseau, G. G., Gustafsson, J.-A., and Mode, A. (1997) *Proc. Natl. Acad. Sci. U. S. A.* **94**, 12309–12313
7. Blochliger, K., Bodmer, R., Jack, J., Jan, L. Y., and Jan, Y. N. (1988) *Nature* **333**, 629–634
8. Andrés, V., Nadal-Ginard, B., and Mahdavi, V. (1992) *Development* **116**, 321–334
9. Neufeld, E. J., Skalik, D. G., Lievens, P. M.-J., and Orkin, S. H. (1992) *Nat. Genet.* **1**, 50–55
10. Quaggin, S. E., Vanden Heuvel, G. B., Golden, K., Bodmer, R., and Igarashi, P. (1996) *J. Biol. Chem.* **271**, 22624–22634
11. Valarché, I., Tissier-Seta, J.-P., Hirsch, M.-R., Martinez, S., Goridis, C., and Brunet, J.-P. (1993) *Development* **119**, 881–896
12. Yoon, S. O., and Chikaraishi, D. M. (1994) *J. Biol. Chem.* **269**, 18453–18462
13. Andrés, V., Chiara, M. D., and Mahdavi, V. (1994) *Genes Dev.* **8**, 245–257
14. Aufiero, B., Neufeld, E. J., and Orkin, S. H. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91**, 7757–7761
15. Coqueret, O., Bérubé, G., and Nepveu, A. (1996) *J. Biol. Chem.* **271**, 24862–24868
16. Harada, R., Dufort, D., Denis-Larose, C., and Nepveu, A. (1994) *J. Biol. Chem.* **269**, 2062–2067
17. Harada, R., Bérubé, G., Tamplin, O. J., Denis-Larose, C., and Nepveu, A. (1995) *Mol. Cell. Biol.* **15**, 129–140
18. Ades, S. E., and Sauer, R. T. (1994) *Biochemistry* **33**, 9187–9194
19. Hanes, S. D., and Brent, R. (1989) *Cell* **57**, 1275–1283
20. Gehring, W. J., Qian, Y. Q., Billeter, M., Furukubo-Tokunaga, K., Schier, A. F., Resendez-Perez, D., Affolter, M., Otting, G., and Wüthrich, K. (1994) *Cell* **78**, 211–223
21. van Leeuwen, H. C., Strating, M. J., Rensen, M., de Laat, W., and van der Vliet, P. C. (1997) *EMBO J.* **16**, 2043–2053
22. Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) *J. Mol. Biol.* **215**, 403–410
23. Devereux, J., Haerberli, P., and Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395
24. Durbin, R., and Thierry-Mieg, J. (1991) *C. Elegans Database*, Code and data available from anonymous ftp servers lirmm.lirmm.fr, cele.mrc-lmb.cam.ac.uk, and.ncbi.nlm.nih.gov
25. Lemaigre, F. P., Durviaux, S. M., and Rousseau, G. G. (1993) *J. Biol. Chem.* **268**, 19896–19905
26. Samadani, U., and Costa, R. H. (1996) *Mol. Cell. Biol.* **16**, 6273–6284
27. Wilson, D. S., Guenther, B., Desplan, C., and Kuriyan, J. (1995) *Cell* **82**, 709–719
28. Hope, I. A., and Struhl, K. (1987) *EMBO J.* **6**, 2781–2784
29. Waterston, R., Martin, C., Craxton, M., Huynh, C., Coulson, A., Hillier, L., Durbin, R., Green, P., Shownkeen, R., Halloran, N., Metzstein, M., Hawkins, T., Wilson, R., Berks, M., Du, Z., Thomas, K., Thierry-Mieg, J., and Sulston, J. (1992) *Nat. Genet.* **1**, 114–123
30. Wilson, R., Ainscough, R., Anderson, K., Baynes, C., Berks, M., Bonfield, J., Burton, J., Connell, M., Copsey, T., Cooper, J., Coulson, A., Craxton, M., Dear, S. Du, Z., Durbin, R., Favello, A., Fraser, A., Fulton, L., Gardner, A., Green, P., Hawkins, T., Hillier, L., Jier, M., Johnston, L., Jones, M., Kershaw, J., Kirsten, J., Laisster, N., Latreille, P., Lightning, J., Lloyd, C., Mortimore, B., O'Callaghan, M., Parsons, J., Percy, C., Rifken, L., Roopra, A., Saunders, D., Shownkeen, R., Sims, M., Smaldon, N., Smith, A., Smith, M., Sonhammer, E., Staden, R., Sulston, J., Thierry-Mieg, J., Thomas, K., Vaudin, M., Vaughan, K., Waterston, R., Watson, A., Weinstock, L., Wilkinson-Sproat, J., and Wholdman, P. (1994) *Nature* **368**, 32–38
31. Dickinson, L. A., Kohwi, J. T., and Kohwi-Shigematsu, T. (1992) *Cell* **70**, 631–645
32. Bürglin, T. R. (1994) in *Guidebook to the Homeobox Genes* (Duboule D., ed) pp. 25–71, Oxford University Press, Oxford
33. Bürglin, T. R. (1995) in *Biodiversity and Evolution* (Arai, R., Kato, M., and Doi, Y., eds) pp. 291–336, The National Science Museum Foundation, Tokyo
34. Holland, P. W. H., Garcia-Fernández, J., Williams, N. A., and Sidow, A. (1994) *Development* (suppl.) 125–133
35. Pendleton, J. W., Nagai, B. K., Murtha, M. T., and Ruddle, F. H. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 6300–6304
36. Ruddle, F. H., Bentley, K. L., Murtha, M. T., and Risch, N. (1994) *Development* (suppl.) 155–161
37. Cassata, G., Kagoshima, H., Prétôt, R. F., Aspöck, G., Niklaus, G., and Bürglin, T. R. (1998) *Gene (Amst.)*, in press
38. Schnabel, C. A., and Abate-Shen, C. (1996) *Mol. Cell. Biol.* **16**, 2678–2688
39. Zappavigna, V., Sartori, D., and Mavilio, F. (1994) *Genes Dev.* **8**, 732–744
40. Zhang, H., Catron, K. M., and Abate-Shen, C. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 1764–1769
41. Aurora, R., and Herr, W. (1992) *Mol. Cell. Biol.* **12**, 455–467
42. Klemm, J. D., and Pabo, C. O. (1996) *Genes Dev.* **10**, 27–36
43. Dickinson, L. A., Dickinson, C. D., and Kohwi-Shigematsu, T. (1997) *J. Biol. Chem.* **272**, 11463–11470
44. Wilson, D. S., Sheng, G., Jun, S., and Desplan, C. (1996) *Proc. Natl. Acad. Sci. U. S. A.* **93**, 6886–6891
45. Stepchenko, A. G., Luchina, N. N., and Pankratova, E. V. (1997) *Nucleic Acids Res.* **25**, 2847–2853
46. Catron, K. M., Zhang, H., Marshall, S. C., Inostroza, J. A., Wilson, J. M., and Abate, C. (1995) *Mol. Cell. Biol.* **15**, 861–871