A distal dimerization domain is essential for DNA-binding by the atypical HNF1 homeodomain

Tanguy Chouard, Marta Blumenfeld, Ingolf Bach, Joël Vandekerckhove[1], Silvia Cereghini and Moshe Yaniv*

Unité des virus oncogènes, UA 1149 du CNRS, Département de Biologie Moléculaire, Institut Pasteur, 25 rue du Dr Roux, 75724 Paris cedex 15, France and [1]Laboratorium voor Genetica, Rijksuniversiteit Gent, B 9000 Gent, Belgium

Received March 8, 1990; Revised and Accepted August 8, 1990 EMBL accession no. X54423

ABSTRACT

Hepatic Nuclear Factor 1 (HNF1, also referred to as LFB1, HP1 or APF) is a liver-specific transcription factor required for the expression of many hepatocyte specific genes. We report here the purification of this rat liver nuclear protein and the cloning of its cDNA using a PCR-derived approach. Seven independent clones reveal 3 alternative polyadenylation sites and a unique open reading frame. Both a motif homologous to the homeodomain and a distal dimerization domain are required for specific DNA binding. Sequence comparisons reveal several atypical features at key positions in the segment corresponding to helices III and IV of the *Antaennapedia* homeodomain as well as a potential 24 amino acid loop in place of the universal turn between helices II and III. Together with its property to dimerize in the presence or absence of DNA, these features place HNF1 as the prototype of a novel subclass of transcription factors distantly related to homeoproteins.

INTRODUCTION

Transcriptional analysis of the albumin promoter has led to the identification of the cis-elements (1, 2) and the corresponding transacting factors (3–5) that are required for efficient and hepatocyte-specific transcription. Mutations in the proximal element immediately upstream of the TATA motif cause the most drastic and most hepatocyte-specific effect on transcription (6, 7). The factor interacting with this site has been characterized in several laboratories (8–11) as binding to proximal control regions in the promoters of many liver-specific genes as well as to several of their distal enhancer elements (reviewed in 12, 13). The consensus sequence derived from these binding sites, g/aGTTAATNATTAACc/a, is palindromic, although it is never fully symmetrical in its various natural occurences. The HNF1 target site is able to drive tissue-specific transcription of heterologous, ubiquitous promoters when cloned in both orientations or even as single or multimerized copies upstream of a TATA motif (7, 14–16). In the *Xenopus laevis* albumin promoter, the HNF1 target site seems to be the only functional element upstream of the TATA motif (10).

In vitro studies confirm the crucial role of HNF1 in hepatocyte-specific transcription. Titration of HNF1 protein in rat liver nuclear extracts using an excess of its specific DNA target abolishes the transcription driven by liver-specific promoters containing the HNF1 site (10, 11, 14). Furthermore, highly purified rat HNF1 protein is sufficient to complement a spleen nuclear extract for the *in vitro* transcription of the mouse albumin promoter (17). We report, here, the isolation of cDNA clones coding for HNF1 and the preliminary mapping of its DNA-binding domain which reveals several atypical features.

MATERIALS AND METHODS

Purification and microsequencing of HNF1

The double stranded oligonucleotide PE56a (generated by annealing the following oligonucleotides : 5′ TCGAGTGTGG-TTAATGATCTACAGTTA-3′ and 5′-TCGATAACTGTA-GATCATTAACCACAC-3′), encompassing the HNF1 binding site of the rat albumin promoter, was phosphorylated and stepwise ligated to a Sepharose-bound oligonucleotide bearing a XhoI cohesive end (18), to generate a specific DNA-affinity chromatography medium carrying 50–75 μg DNA/ml gel. HNF1 purification was monitored by a band-shift assay using 5′-end-labeled PE56a. Liver nuclear extracts were prepared from 150 rats as described (11), except that the second sucrose cushion was omitted; proteins were resuspended in Elution Buffer (EB: 20% glycerol, 20 mM hepes pH 7.9, O.5 mM EDTA, O.5 mM DTT, O.5 mM PMSF, 12 mM $MgCl_2$, 1μg/ml pepstatin, leupeptin and aprotinin) to a conductivity lower than that of 0.15 mM KCl in EB and loaded on five 200 ml columns of heparin Ultrogel (IBF). HNF1 containing fractions were eluted with 0.35 M KCl, diluted with EB to 0.3 M KCl supplemented with poly-dIdC (15 μg/mg protein) and O.1% NP40 and loaded on PE56a-DNA-affinity columns (20 ml total volume). Columns were washed with 0.3 M KCl and HNF1 activity was eluted at 0.6 M KCl, with a yield of 20% and a purification factor of 6600 fold, relative to nuclear extracts. 30 ml of pooled fractions

* To whom correspondence should be addressed

containing 25–30 μg of 87–93 kDa HNF1 were concentrated by lyophilisation to 10 ml and dialysed against 10 mM ammonium acetate containing 0.01% SDS before further concentration to 300 μl; the protein was electrophoresed on an SDS-10% polyacrylamide gel (19) and electroblotted onto a polyvinylidene difluoride membrane (PVDF); the amidoblack stained 87–93 kd HNF1 band was cut out, digested *in situ* with trypsin and the eluted tryptic peptides were fractionated by Reversed Phase HPLC Chromatography followed by gas-phase amino acid sequence analysis as previously described (20 and references therein).

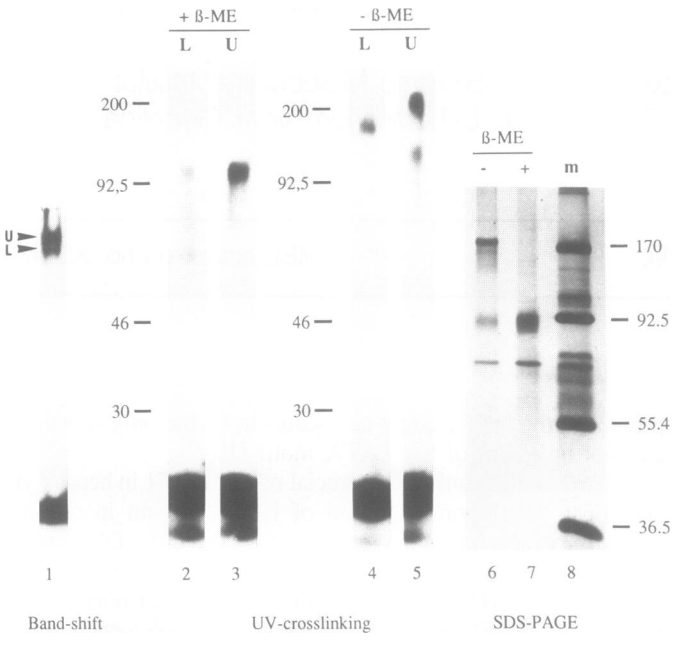

Figure 1: Analysis of affinity-purified HNF1. Dimers stabilized by S-S bonds. 10 ng of labelled PE56a oligonucleotide, in which several thymine residues had been replaced by bromodeoxyuridine residues, were incubated with 10 μl of affinity-purified fraction and complexes were resolved by preparative 5% native PAGE (lane 1); proteins were UV-crosslinked to DNA and two bands (U and L for upper and lower bands respectively) separately cut and analyzed by 10% SDS-PAGE in reducing (lanes 2, 3; 0.1% β-mercaptoethanol) or nonreducing conditions (lanes 4, 5). The estimated molecular weights of UV-crosslinked peptides should be reduced by 10–15 kDa, due to the contribution of the PE-56a oligonucleotide (11). The fraction corresponding to the peak of DNA-binding activity was analyzed by silver-stained 8% SDS-PAGE in reducing (lane 7) or nonreducing conditions (lane 6). m : markers for molecular weights (given in kDa).

Production of a nondegenerate DNA probe using polymerase chain reaction (PCR)

Enzymatic DNA amplifications with Taq DNA polymerase were performed in the buffers indicated by Saïki et al. (21) using 10 ng of oligo-dT primed cDNA from Fao cells (see below) and 1 μg each of oligonucleotides C1 or C3 and D2 or D4 (see text and Figure 2). After 50 cycles of primary amplification (one cycle = 15″ at 94°C, 1′ at 60°C, 1′ at 72°C) a 10% aliquot was futher amplified for 50 cycles and amplification products were analyzed on a 20% polyacrylamide gel. DNA of samples displaying only the expected amplified band (47 bp) in addition to the single-stranded primers was directly extracted with phenol-chloroform, phosphorylated, blunt-end ligated to a Bluescribe vector linearized with SmaI and sequenced. A 41-mer oligonucleotidic probe (CD41, see Figure 2) was derived from the sequence data.

Preparation and screening of cDNA libraries

An amplified λ-gt10 cDNA library (oligo-dT primed) from rat hepatoma Fao cells treated with cycloheximide was prepared according to standard procedures (22) and was screened with 5′-end-labeled CD41 and A15 probes (Figure 2). Hybridizations were performed in 6×saline sodium citrate (SSC), 1×Denhardt's solution, 50 μg/ml tRNA, 0.05% sodium pyrophosphate at 42°C; washes were done at 55°C in 1×SSC. Inserts of positive clones were subcloned in a Bluescribe vector and sequenced on both strands by the dideoxynucleotide method modified for double-stranded DNA (23, 24) using successive primers.

Construction of HNF1 expression vectors

Full length HNF1 cDNA was obtained as follows: the A4 clone was digested with SacI, treated with T4 DNA-polymerase and then digested with MluI to generate an MluI-SacI HNF1 fragment with a blunt SacI end (nt 200–2240, Figure 4); the 5′ part of HNF1 cDNA was synthesized by reverse transcription of rat liver

```
A          V  Y  N  W  F  A  N  R
B          A  A  L  Y  T  W  Y  V  R
CD         -  -  F  T  S  D  T  E  A  S  S  E  P  G  L  H  E  A/P S/L
C1           5'-  TTT ACT AGT GAT ACT GAA GC -3'  3'- CTT GGA CCA AAT GTA CTT CG -5'   D2
                  C   C   C   C   C   G                 C   G   G   C   G   C G
                      G                                 C   C
                      A                                 T   T
C3           5'-  TTT ACT TCT GAT ACT GAA GC -3'  3'- CTT GGA CCA GAA GTA CTT CG -5'   D4
                  C   C   C   C   C   G                 C   G   G   G   G   C G
                      G   G       G                     C   C   C
                      A   A       A                     T   T   T
CD41                  5'- TCG GAC ACG GAG GCC TCC AGT GAG CCC GGC CTG CAC GAA CC -3'
HNF-1 cDNA              ..A ... ..A ... ... ... ... ... ..T ..G ..T ..T ..G ..
A15           5'-GCATAAGCTGGCCATGGACACGTATAACGGGCCTCCACCC-3'
HNF1hn        5'-CCGCGAAGCTTGGCGGTAGAGGAGCCATGG-3'
HNF1m         5'-TCTTCCGAGCCACGCGTCTC-3'
```

Figure 2: HNF1 tryptic peptides and PCR derived probe. Sequences of tryptic peptides A, B and CD are given in single letter code, the amino acids encoded by six codons in CD are underlined. PCR primers used to determine the CD41 sequence are aligned underneath the CD peptide sequence, two of the coding strand (C1 and C3) on the left and two of the noncoding strand on the right. The nucleotides in the coding strand of the HNF1 cDNA clones which matched CD41 are replaced by dots, those which did not are indicated by their letter code. Oligonucleotide A15 probe was derived from the 5′ end of the CD13 clone coding strand; HNF1hn PCR primer, encompassing the initiation ATG, from the coding strand of the BP14 genomic clone with creation of an artificatial HindIII site; HNF1m, encompassing the unique MluI site, from the noncoding strand of the A4 clone.

total RNA (10 μg) using a primer complementary to nt 230−250 (Figure 4); the purified cDNA then served as template to generate a PCR fragment using HNF1hn and HNF1m oligonucleotides (Figure 2) and the 241 bp fragment thus obtained was digested at the HindIII and MluI sites included in the PCR primers. Both 3′ and 5′ cDNA fragments were ligated to an HindIII/HpaI digested pRSV-CAT vector (25), thus generating the pRSV-HNF1 construct. pRSV-HNF1 was digested partially with NcoI and then with BamHI to obtain a 2369 bp fragment encompassing the segment 1−2240 of the HNF1 sequence (Figure 4) followed by the 129 bp HpaI-BamHI segment from pRSV-CAT. This fragment was cloned between the NcoI and BamHI sites of a (T7 promoter-βglobin leader)-vector, derived from pGEM1 (provided by Dr. R. Treisman), thus putting the whole HNF1 coding sequence in frame with the βglobin initiator ATG (pT7βH plasmid).

HNF1 deletion vectors were constructed as follows: the 774 bp SmaI fragment from pT7βH was cloned in frame with the first ATG in an NcoI digested pT7βH vector repaired with the Klenow enzyme, thus generating pT7βH-34/291; pT7βH-Δ18/53 was obtained by XhoI digestion of pT7βH followed by repair with the Klenow enzyme and auto-ligation; pT7βH-Δ34/208 was constructed by digestion of pT7βH with ApaI, further treatment with T4 DNA polymerase and auto-ligation.

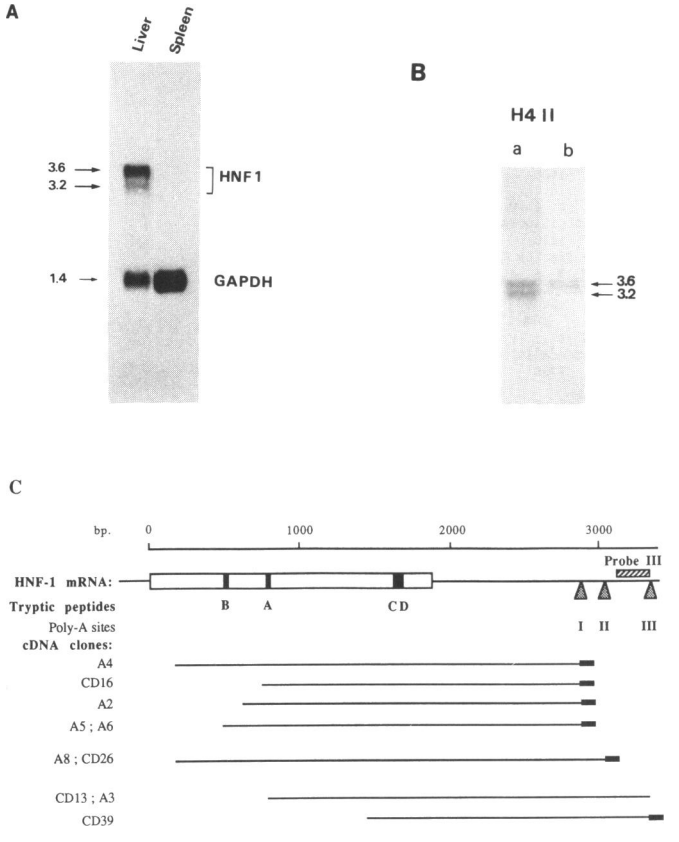

Figure 3A: Northern blot analysis of HNF1 transcripts. 3 μg of rat liver or spleen poly A+RNA were fractionated and hybridized with a mixture of full length HNF1 (clone CD26, Figure 3C) and GAPDH probes. Arrows to the left indicate the estimated size of the different bands. **B:** Northern blot analysis of alternative polyadenylation sites usage. In lane b, 12.5 μg of total RNA from H4II cells were hybridized with an HNF1 antisense RNA probe specific for transcripts of group III (see text; Probe III: nt 3158−3397, Figure 3C); after the final high stringency wash, the blot was rehybridized with a full length HNF1 cDNA probe (lane a). **C:** Schematic map of HNF1 cDNA clones presenting three polyadenylation sites. The extent of each of the seven independent HNF1 cDNA clones is indicated relative to the combined sequence of the HNF1 mRNA in front of the list of the corresponding redundant clones. The HNF1 open reading frame is indicated by a box; tryptic peptides are shown in black. PolyA tails are indicated by a black box at the end of each clone and the three polyadenylation sites included in the longest sequence are indicated by triangles. The group III specific probe (Probe III) is indicated by a hatched box.

In vitro transcription and translation

Sense RNA was synthesized *in vitro* using 10 units of T7-RNA polymerase (Stratagene transcription kit) and 1 μg of DNA templates linearized as follows: the T7βH-wt, T7βH-1/390, T7βH-1/281 and T7βH-1/264 transcripts were generated after digestion of pT7βH with BamHI, AatII, BalI and AccI respectively. The T7βH-34/291, T7βH-Δ18/53 and T7βH-Δ34/208 transcripts were obtained after digestion with BamHI of the corresponding deletion plasmids.

After 1/2 hour incubation at 37°C, phenol/chloroform extraction and ethanol precipitation, the synthesized RNA was resuspended in 20 μl of water and 2 μl was used for the *in vitro* translation with a rabbit reticulocyte lysate system (Amersham). The reaction mixture (15 μl) included 12 μl of the cell lysate and 1 μl of diluted [^{35}S]L-methionine (<1000 Ci/mmol, 1mmol/ml). The translation was carried out for 1 hour at 30°C, and quantity and quality of the products were checked by counting TCA-insoluble ^{35}S and by SDS-PAGE.

Gel retardation assays, SDS-PAGE and UV-crosslinking experiments

Gel retardation assays and UV-crosslinking experiments with liver proteins were performed as previously described (11), without any competitor DNA in the case of affinity-purified fractions.

For each *in vitro* translated protein, the volume indicated in the legend for Figure 5A was preincubated for 10 min on ice with sonicated salmon sperm DNA at a ratio of 0.5 μg/μl translation mixture and with competitor oligonucleotide where mentioned, in 20 μl of binding buffer containing 10 mM Hepes, 4mM $MgCl_2$, 0.1 mM EDTA, 4 mM spermidine, 15% glycerol. After addition of 1 ng of ^{32}P-labelled PE56 oligonucleotide probe, mixtures were further incubated for 10 min on ice and then analyzed by 5% PAGE in 0.25× TBE. The gel was dried and exposed to two sheets of X-ray film with intensifying screen at −80°C. The film closest to the gel shows both ^{35}S and ^{32}P signals and the second one only ^{32}P intensified signals. Figure 5A shows only the pattern of the second film. SDS-PAGE were done according to Laemmli (19) and, in the case of *in vitro* translated protein analysis, treated for fluorography (22) before exposure.

Northern blot analysis

Total RNA from rat liver, spleen or H4II cells was extracted using the guanidium thiocyanate-CsCl method (26). Poly A+ RNA was isolated by oligo dT cellulose chromatography (22), electrophoresed through 1.2% agarose-2.2 M formaldehyde gels and transferred to a Hybond N membrane (Amersham). RNA markers (0.24−9.5 kb; BRL) were run in parallel and visualized under UV. cDNA inserts were isolated from agarose gels, labelled by random priming to a specific activity of $3-7\times10^8$ cpm/μg for HNF1 and 1×10^7cpm/μg for GAPDH and used as probes. Hybridization was performed at 42°C in 50% formamide, 5×SSPE, 5×Denhardt with the ^{32}P-labelled probe for 24 hours. The membranes were washed at 60°C with 5×SSC, 0.25% SDS

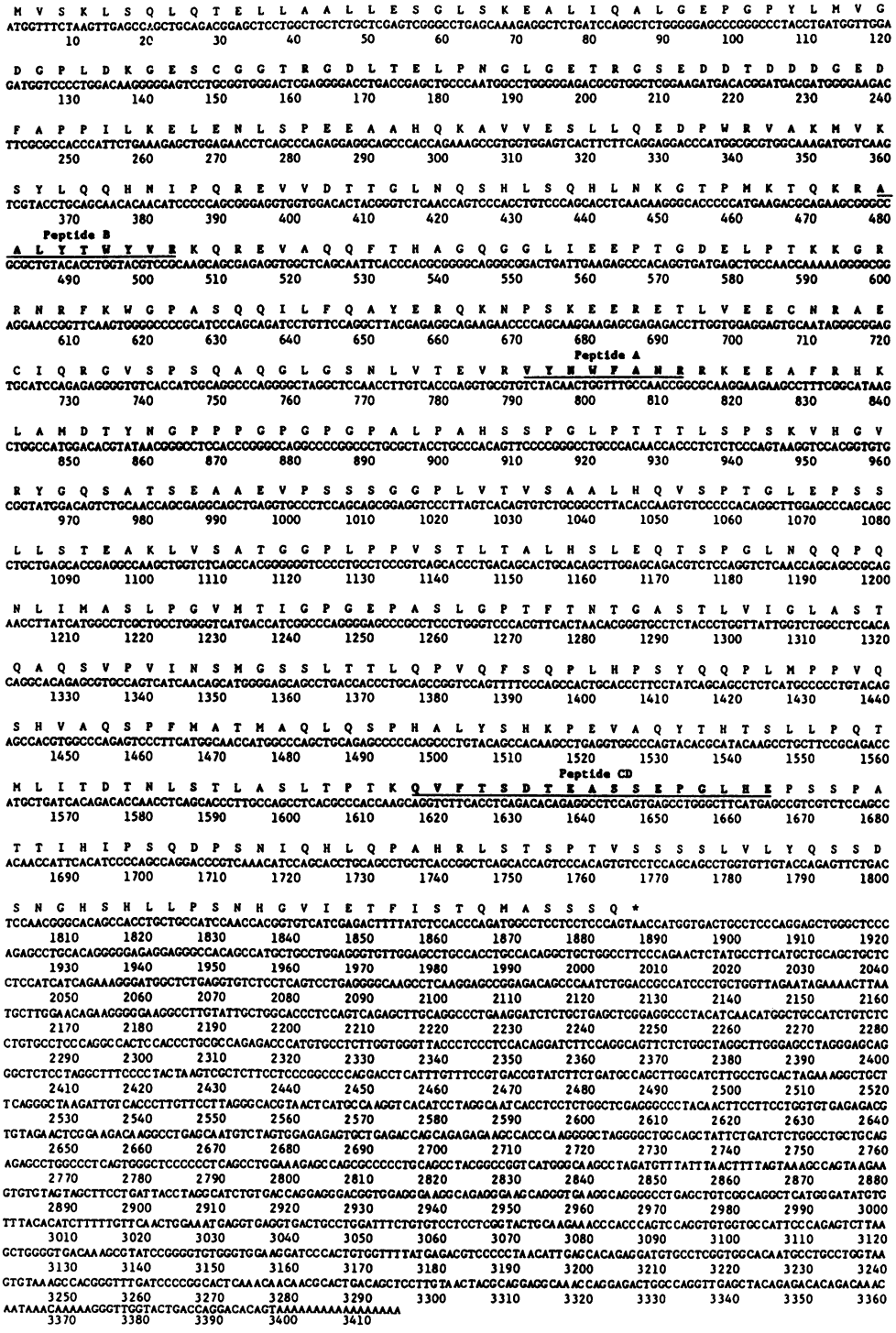

Figure 4: HNF1 cDNA sequence. Amino acids are given in single letter code above the nucleotide sequence. The numbering adopted, for both the nucleotide and the amino acid sequences, is initiated at the first codon of the HNF1 open reading frame (see text). Underline bold regions indicate tryptic peptides that were sequenced. * indicates the first nonsense codon.

and 0.1×SSC, 0.25% SDS. Alternatively, an HNF1 antisense RNA probe (nt 3158–3397) was prepared by *in vitro* transcription with T7 RNA polymerase of a BamHI digested CD13 clone (Figure 3C). Membranes were hybridized in 50% formamide, 5× SSC, 5× Denhardt at 52°C for 16 h. and washed at a final stringency of 0.1×SSC at 74°C for 1 h.

RESULTS

Native HNF1 forms dimers in the presence or absence of its DNA-target

HNF1 protein was purified from rat liver in a three step procedure: preparation of nuclear extracts, chromatography on heparin

Ultrogel and specific DNA-affinity column as described in Materials and Methods. The peak of HNF1 DNA-binding activity produced two complexes in gel retardation assay (Figure 1, lane 1: arrows U and L) and displayed a major 87−93 kDa protein, a 72 kDa polypeptide and several minor polypeptides (Figure 1, lane 7). To determine which of these polypeptides directly interacted with DNA, we performed UV-crosslinking experiments, in which ^{32}P-labelled oligonucleotide containing BrdU residues was covalently linked to the proteins in the retardation gel, before the upper and lower band were excised and run on a SDS-polyacrylamide gel. The upper complex gave rise to a roughly100 kDa band corresponding to the 87−93 kDa protein linked to the 15 kDa oligonucleotide, which is in agreement with the molucular weight previously reported for HNF1 (27, 11, 12, 17). The heterogeneity of the 87−93 kDa band (lane 7) is probably caused by the glycosylation of HNF1 (see below). The lower complex gave rise to the same 100 kDa band in addition to faster migrating species, likely to be proteolyzed chains since the yield of this lower retarded band increased during the purification procedure. The fact that the HNF1 binding site was almost palindromic suggested a hypothesis in which the upper complex would contain two intact 87−93 kDa polypeptide chains and the lower band one intact and one partially degraded chains; the profile obtained in lanes 2 and 3 is compatible with this hypothesis. Analysis of the UV-crosslinked DNA-protein complexes in nonreducing conditions gave slower migrating bands, suggesting that HNF1 might either associate with another polypeptide chain or form dimers covalently linked by S-S bonds (Figure 1, lanes 4 and 5). This interaction appeared to be specific since the same patterns were observed whether crude liver nuclear extracts or various purified fractions were used (data not shown). Further support for potential dimerization came from the observation that a major fraction of the 87−93 kDa HNF1 band ran as a 180 kDa species under nonreducing conditions (Figure 1, lane 6). In addition, this suggested that the protein might preexist as a dimer in the absence of DNA ; neither addition of specific or unrelated DNA nor irradiation with UV-light increased the yield of dimerization observed in lane 6 (not shown). These observations were further strengthened using cloned HNF1 (see below). Gel retardation assays performed in the presence of different reducing agents, indicated that S-S bond formation was not required for DNA-binding *in vitro* (data not shown), nevertheless, the factor might still need to dimerize for this purpose.

In order to obtain protein sequence data for the 87−93 kDa chain, 150 pmoles were further purified by SDS-PAGE, since no other polypeptide was visible on two-dimensional gels at the same molecular weight level (not shown), then transferred to a PVDF membrane and digested *in situ* with trypsin. The tryptic peptides were fractionated on reversed phase HPLC and submitted to amino acid sequence analysis. The three sequences obtained are listed in Figure 2. Peptides A and B were used to derive redundant oligonucleotides for direct screening of cDNA libraries from rat liver or hepatoma cell lines. After several unsuccessful attempts, we turned to a PCR-derived approach to get better probes for HNF1.

Use of PCR to generate a unique probe for HNF1

From the longest peptide (CD), four degenerate primers, two of the coding strand (C1 and C3) and two of the noncoding strand (D2 and D4), were designed to reduce the high degeneracy due to serine and leucine residues (see Figure 2). These oligonucleotides were used in an enzymatic amplification on Fao cells cDNA, followed by cloning and sequencing of the expected 47 bp PCR-amplified fragment. The nucleotides coding for the alanine and the two central serine residues of the CD peptide were thus unambiguously determined and a 41-mer nondegenerated probe was synthesized according to this sequence data (CD41 in Figure 2).

Since the C and D primers were highly degenerate (respectively 512 and 1024 mixed sequences) and PCR is a primer-limiting procedure, only about 1% of the 47 bp amplified fragments that were visible on a gel might display the exact HNF1 sequence. The remaining must contain some mismatches mainly in the ends of the fragments, since elongation by Taq-DNA-polymerase requires a better annealing at the 3′ end of the primers. The nucleotides that were actually wrong in the single PCR fragment that we sequenced as compared to the final cDNA sequence are indicated in Figure 2.

Seven independent HNF1 cDNA clones display three polyadenylation sites and a single open reading frame

With the CD41 probe, we screened 10^6 clones of a λ-gt10 cDNA library prepared from rat hepatoma Fao cells (28), treated with cycloheximide to increase the representation of potentially unstable transcripts (29). We first isolated 3 partial cDNA clones (CD13, CD16, CD39); one of them (CD13) included the two others. A 45 bp probe (A15, Figure 2) from the most 5′ region of this clone was synthesized and used to screen the same library. Seven new clones (A-clones and CD26) were isolated that were also positive with CD41. The structure of the ten clones including three pairs of identical ones is described schematically in Figure 3C.

The total length of the combined nucleotide sequence (CD26 and the 3′ end of CD39 without the oligoA tail) was 3205 bases. The cDNA clones could be ordered in three groups (I, II, III, see figure 3C) according to the position of their polyA tails relative to the combined sequence. When these clones were used to probe Northern blots with rat liver polyA RNA, a minor and a major species of 3.2 and 3.6 kb respectively were detected (figure 3A). Hence, our longest clone could not cover the entire length of the 3.6 kb mRNA and was probably missing a few hundred base pairs.

In addition, using a probe derived from the most 3′ sequence of clones of the group III, only the 3.6 kb species was detected, thus demonstrating that more than one polyadenylation site were actually used *in vivo* (Figure 3B). Examination of the sequences upstream of the polyA tails found in the three groups of clones revealed three possible alternative polyadenylation signals. The four independent clones of group I utilize the AGTAAA sequence at position 2864 (numbering of the figure 4, see below) with a polyA chain 14 to 16 nt downstream; the single clone of group II might use the AATGAG sequence at position 3027 with a polyA chain 17 nt downstream; finally, the two clones of group III utilize the AATAAA sequence at position 3361 with a polyA chain 31 nt downstream (30). The Northern blot analysis suggested that, in rat liver, the most 3′ site is used predominantly in rat liver. The situation was different in rat hepatoma cell lines (S.C., unpublished observations).

Conceptual translation of the cDNA sequence revealed a single open reading frame of 564 triplets, followed by a long nontranslated region. The open reading frame included peptides A, B and CD (underlined in figure 4) as well as two additional peptides that were not separated by the HPLC and gave short

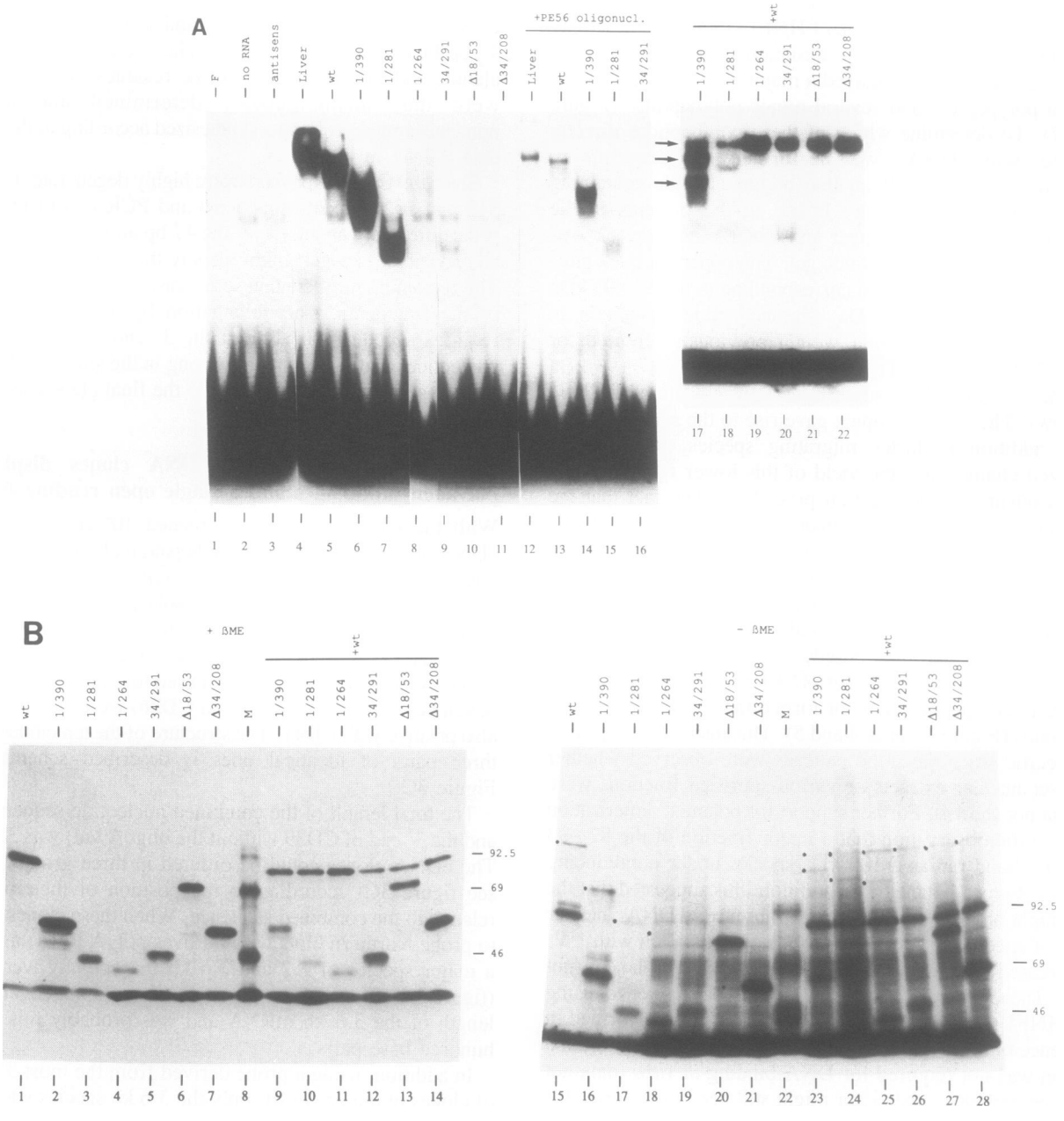

Figure 5: Functional mapping of the HNF1 protein. **A:** *In vitro* translated HNF1 analyzed by Band-shift assay. 1 ng of ^{32}P-labelled PE56a probe was incubated as described in Materials and Methods with 1 μg of rat liver nuclear proteins (lanes 4, 12) or with *in vitro* products of following constructions, in absence (lanes 1−11 and 17−22) or presence of 9 ng of unlabelled PE56a as competitor (lanes 12−16); the respective volumes of translation mixture used in binding assay and lanes in the Figure are indicated in paranthesis: no proteins (lane 1); reticulocyte lysate incubated without exogenous RNA (1 μl, lane 2); antisens HNF1 RNA (1 μl, lane 3); T7βH-wt (1 μl, lanes 5, 13); T7βH-1/390 (1 μl, lanes 6, 14); T7βH-1/281 (1 μl, lanes 7, 15); T7βH-1/264 (1 μl, lane 8); T7βH-34/291 (3 μl, lanes 9, 16); T7βH-Δ18/53 (3 μl, lane 10); T7βH-Δ34/208 (3 μl, lane 11); co-translation of approximately 1 μg of both mRNA from T7βH-wt and T7βH-1/390 (1 μl, lane 17), T7βH-1/281 (1 μl, lane 18), T7βH-1/264 (1 μl, lane 19), T7βH-34/291 (1 μl, lane 20), T7βH-Δ18/53 (1 μl, lane 21) or T7βH-Δ34/208 (1 μl, lane 22). Band-shift assay were performed as described in Materials and Methods. Note that the band visible in lane 2 is common to all lanes where *in vitro* translated proteins were used and is due to a reticulocyte DNA-binding activity. Arrows to the left of lane 17 indicate the positions, from top to bottom, of wilt type homodimers, heterodimers and T7βH1/390 homodimers. **B:** *In vitro* translated HNF1 analyzed by SDS-PAGE. 2 μl of translation mixtures were diluted with 8 μl of water, left 1 hour at room temperature, mixed with 10 μl of Laemmli's sample buffer (19) in presence (lanes 1−14) or absence (lanes 15−28) of 0.1% β-mercaptoethanol and separated on two identical SDS-8%-polyacrylamide gels that were subsequently autoradiographed for ^{35}S-met-labelled protein detection. In the translation or co-translation reactions, approximately 1 μg of mRNA from each of the following constructions were used and the respective lanes on gels are indicated in paranthesis: T7βH-wt (lanes 1, 15); T7βH-1/390 (lanes 2, 16); T7βH-1/281 (lanes 3, 17); T7βH-1/264 (lanes 4, 18); T7βH-34/291 (lanes 5, 19); T7βH-Δ18/53 (lanes 6, 20); T7βH-Δ34/208 (lanes 7, 21); T7βH-wt and T7βH-1/390 (lanes 9, 23), T7βH-1/281 (lanes 10, 24), T7βH-1/264 (lanes 11, 25), T7βH-34/291 (lanes 12, 26), T7βH-Δ18/53 (lanes 13, 27) or T7βH-Δ34/208 (lanes 14, 28). ^{14}C -labelled molecular weight markers (92.5, 69, 46 and 27 kDa; Amersham) were loaded in lanes 8 and 22. Dots on the left of bands on the autoradiogram indicate the homo- or hetero-dimers. Since an excess of intact protein was used in lanes 9−11 and 23−25, we do not detect homodimers of truncated proteins in these experiments.

mixed sequences; all were preceded by arginine or lysine residues as expected for tryptic peptides. The first methionine in this open reading frame was in the 54th position (amino acid no. 118 in Figure 4), thus generating a polypeptide of 511 amino acids with a molecular weight of 54.7 kDa, probably too short when compared with the 87−93 kDa protein that we purified. The absence of the initiation ATG in the cDNA clones was confirmed by analysis of a genomic clone (BP14) obtained by screening a rat library with a cDNA probe. The sequence of this clone partially overlapped the 5′ sequence of the cDNA clones. The first ATG codon preceded by a nonsense codon in phase was located 192 nucleotides upstream of our cDNA 5′ end (I.B., unpublished results). Using a PCR primer overlapping this putative initiation codon and a second primer encompassing a unique MluI restriction site in our cDNA sequence (respectively HNF1hn, and HNF1m in Figure 2), we amplified a 241 bp fragment from a rat liver cDNA preparation, thus making sure that there was no splicing event within the first 192 coding base pairs. This was confirmed by cloning and sequencing of the PCR fragment. This sequence, combined with that of the cDNA clones, gives rise to an open reading frame encoding a polypeptide of 628 amino acids with a calculated molecular weight of 69 kDa (see Figure 4). This sequence is virtually identical to recently published cDNA sequence for the LFB1 rat liver factor that was purified as a 45 kDa protein : the exceptions are 1 nt in the codon no. 434, coding for a valine in both sequences, and 7 nt in the noncoding region (31). A partial cDNA sequence of rat HNF1 was recently published by another group (32).

In vitro translated HNF1 requires a motif homologous to the putative recognition helix of the homeodomain for specific DNA-binding

To confirm that the cDNA we obtained coded for a protein that could indeed bind specifically to the albumin proximal element, we inserted its complete coding sequence in a pGEM1 derived vector and synthesized HNF1 by *in vitro* transcription and translation. The ^{35}S-met-labelled protein was analyzed by both a gel retardation assay and SDS polyacrylamide electrophoresis. When incubated with its specific DNA-binding site as probe (PE56a), *in vitro* translated HNF1 gave rise to a complex that was specifically displaced by the homologous oligonucleotide (see lanes 5 and 13 in Figure 5A) but not by a mutated HNF1 DNA target (DS34 (11); not shown), thus demonstrating that our cDNA clone actually encoded the activity that had been purified from rat liver. *In vitro* translated HNF1 had an apparent molecular weight of 80 kDa, lower than that observed for the purified rat liver protein (87−93 kDa; Figure 1, lane 7, Figure 5A, lanes 4 and 5 and 5B) and these variations are likely to be due to the glycosylation of HNF1 (17 and our unpublished data). Examination of the complete amino acid sequence of HNF1 (Figure 4) reveals essentially two domains separated by a flexible junction rich in prolines and glycines (residues 288 to 297). The amino acid composition of the two domains is clearly different suggesting that they could have evolved separately. General features of the HNF1 protein primary sequence are schematically represented in figure 6A.

The sequence VYNWFANR (residues V264 to R271) of peptide A matches the consensus sequence I/V--WF--NRR highly conserved in the putative DNA-recognition helix of homeodomains (36). Comparison of the HNF1 sequence and homeoboxes identified to date is discussed below. To check whether this homology was functionally relevant, we analyzed the effect of progressive C-terminal deletions on HNF1 DNA-binding. Figure 5A shows that HNF1 can still bind DNA specifically when the whole C-terminal domain has been deleted (deletions T7βH-1/390, T7βH-1/281; Figure 5A, lanes 6 and 7 resp.) but not when further deletion removes a short segment containing the putative recognition helix (deletion T7βH-1/264; Figure 5A, lane 8). Thus, the C-terminal end of the HNF1 DNA-binding domain is localized within the 17 amino acids 265 to 281.

The HNF1 homeodomain is unable to bind specifically to its DNA target in the absence of a distal dimerization domain

In order to map the N-terminal end of the HNF1 DNA-binding domain, we constructed deletions using different restriction sites. The N-terminal domain encoded by the longest SmaI fragment of the HNF1 cDNA, in which the first 33 amino acids are removed and 59 extra amino acids encoded in the second phase are added in C-terminal, retained only 1−2% of the wild type activity (deletion T7βH-34/291; Figure 5A, lane 9). Deletion of 35 amino acids from E18 to T53 encoded by the single XhoI fragment of the HNF1 cDNA led to a complete loss of specific DNA-binding (deletion T7βH-Δ18/53; Figure 5A, lane 10). This deletion mutant could bind the PE56 probe with very low efficiency but only in the absence of any competitor DNA (not shown). Finally, the deletion of 175 amino acids from G34 to P208, encoded by the unique ApaI fragment of the HNF1 sequence, abolished specific as well as nonspecific DNA-binding activity (deletion T7βH-Δ34/208; Figure 5A, lane 11 and data not shown).

Thus, the HNF1 homeodomain appeared to be unable to stably bind DNA by itself and the question arose as to whether it would need to dimerize to do so, since dimers of native HNF1 had been observed as discussed above. This prompted us to analyse, by band-shift assay, the ability of the different deletion mutants to form DNA-bound heterodimers with the co-translated wild type protein. In addition, in order to analyse dimerization independantly of DNA binding, we ran the same mutants on SDS-polyacrylamide gels in nonreducing conditions, thus taking advantage of the potential spontaneous S-S bond formation between two HNF1 monomers.

The first two C-terminal deletion mutants presented, on nonreducing SDS gels, the variations in mobility that were expected for truncated HNF1 homodimers (Figure 5B, lanes 15−17), thus confirming the observations made with the native protein. Co-translation of wild-type HNF1 with those mutants revealed heterodimers by both techniques (Figure 5A, lanes 17 and 18 and 5B, lanes 23, 24). As mentioned above, deletion of the putative recognition helix abolished DNA binding, thus making it impossible to check by band-shift analysis whether or not the resulting mutant could dimerize. In fact, it could still form homodimers and more clearly heterodimers which were visualized on a SDS-polyacrylamide gel (Figure 5B: lanes 18 and 25). Thus, it appeared that amino acids C-terminal to the V264 were not required for HNF1 dimerization.

In contrast, the three deletions in the N-terminal domain strongly reduced or abolished both DNA-binding and dimerization, although in different manners. First, as mentioned above, the T7βH-Δ18/53 deletion mutant lost the ability to bind specifically to the HNF1 DNA target site; it also failed to form heterodimers with the intact protein in both the DNA binding assay and SDS PAGE in nonreducing conditions (figure 5A, lanes 10 and 21; figure 5B, lanes 20 and 27). It should be noted however, that the deletion of C50 makes the method irrelevant

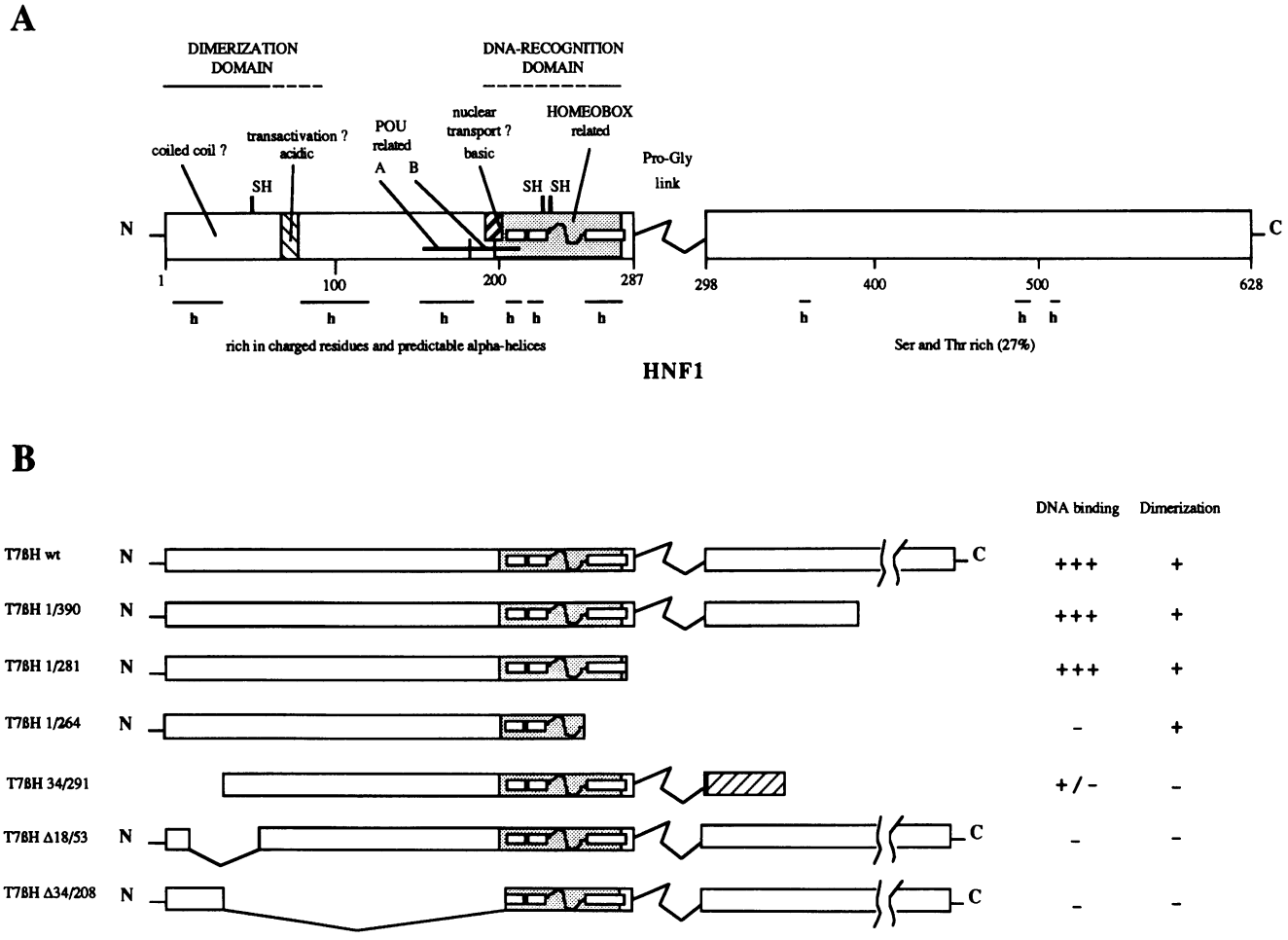

Figure 6A: Global architecture of the HNF1 protein. The domains discussed in the text are represented by boxes or bars linearly arranged along the sequence, given by the amino acid numbers used in the Figure 4. Motifs that might be involved in transactivation and nuclear transport (according to 33 and 34 resp) correspond to amino acids E71-D80 and K197-K207 respectively. **SH** indicate cysteine residues. **N** and **C** indicate the N- and C-terminal parts of the protein. Segments more than 3 amino acid long that were predicted as α-helical using the algorithm of Garnier et al. (35) are indicated by horizontal bars with an **h** letter beneath. The three helices and the 24 amino acid loop present in the HNF1 with a loop model of homeobox are also represented (see text and Figure 7). **B:** Deletion analysis of the HNF1 DNA binding domain. The 7 constructions described in text are schematically represented with same scale as in Figure 6A. Parts of the HNF1 with a loop homeodomain that are present in each of them are also represented. A hatched box indicates the 59 extra amino acids, encoded in the second phase, that are C-terminal in the T7βH-34/291 protein. **N** and **C** indicate the terminal parts of the wild type HNF1 protein when present in the deletion mutants. Results of Figure 5 are summarized in the right part of the Figure.

for this mutant without precise identification of the cysteine residues involved in S-S bond formation. In any case, these results demonstrated that the homeodomain is not sufficient for specific DNA binding and indicated that formation of the HNF1 dimers requires sequences far from the homeodomain. Second, the residual DNA binding activity observed with T7βH-34/291 suggested that the N-terminal part of the HNF1 dimerization domain (upstream of G34) is still required but perhaps less crucial than the rest of it (which includes some amino acids from G34 to T53). Indeed, the fact that DNA binding is completely lost by T7βH-Δ18/53 and that the T7βH-34/291- and T7βH-1/281-DNA complexes have similar electrophoretic mobilities makes unlikely, though does not exclude, the hypothesis of a monomeric interaction of the HNF1−34/291 molecule with DNA and rather suggests a weak residual ability to dimerize for this mutant. Nevertheless, this mutant did not form heterodimers when co-translated with wild type HNF1 and dimers including it have not yet been detected in nonreducing protein gels (Figure 5A, lane 20 and 5B lanes 19 and 26). It is possible that it dimerizes poorly in the absence of DNA and cannot compete for the dimerization process with the wild type protein, when co-translated. Finally, the third N-terminal deletion T7βH-Δ34/208, which was also unable to dimerize, confirmed that there are some amino acids crucial for dimerization downstream of P33 (Figure 5B, lanes 21, 28). In addition, it had lost any affinity for any type of DNA which, we assume, was retained by T7βH-Δ18/53 thanks to an intact homeodomain moiety. This suggested that the N-terminus of the HNF1 homeodomain could lie upstream of A209 (see also discussion below).

DISCUSSION

The HNF1 DNA-recognition domain is probably not a classical helix-turn-helix motif

To document the homology between the DNA-recognition domain defined above and the homeodomain, we aligned the predicted amino acid sequence of HNF1 in the region of peptide A with 87 of the 60 amino acid-long homeobox motifs compiled to date (37, 38). The mean overall homology was approximately

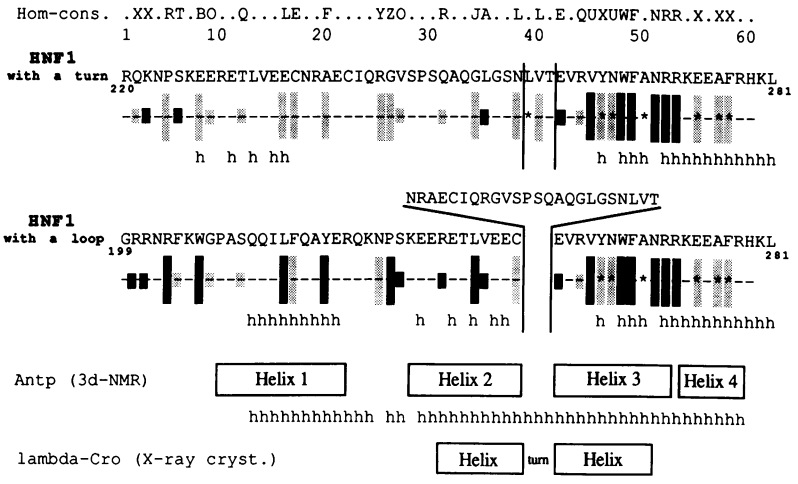

Figure 7: HNF1 homeodomain compared to the homeodomains sequenced to date. The HNF1 with a 3 amino acid turn (residues R220 to H279) and the HNF1 with a 24 amino acid loop (residues G199 to H279) sequences were aligned separately with the 31 positions where a consensus could be found among more than 60% of the homeoprotein (**hom-cons**, 37). Amino acids are single letter coded and 5 additional codes are used for the hom-cons sequence as follows: **B**= Big hydrophobic i.e. F, Y or I; **J**= I or L; **O** = T or S; **U** = V or I; **X** = K or R; **Z** = P or L. A diagram below each sequence illustrates the agreement with the consensus as follows: each of the 31 positions conserved in the homeobox consensus sequence is represented by a small or a big bar when between 60 to 80% or more than 80% of the homeoproteins agree with the consensus respectively. The bar is black when the corresponding HNF1 amino acid is in structural agreement with the consensus and pale when it differs. A star (*) indicates an atypical feature specific to HNF1 (see text). Boxes at the bottom indicate the location of helices identified by three-dimensional NMR in the *Antp* homeodomain (40) and by X-ray crystallography in the λ-Cro helix-turn-helix motif (41). **h** letters indicate the α-helical segments predicted from the HNF1 sequences and from the *Antp* homeobox sequence as a control, according to the algorithm of J. Garnier et al. (35). Numbers below the hom-cons. sequence indicate positions in the canonical homeobox sequence (18). Numbers below the HNF1 sequences indicate positions of their ends in the complete HNF1 amino acid sequence (Figure 4).

17%, clearly concentrated in the third helix. As recently proposed by Finney (39), allowing the looping out of 24 amino acids ('HNF1 with a loop') in the place of the canonical 3 amino acid turn ('HNF1 with a turn') between helices II and III in the HNF1 sequence led to significant improvement in homology (mean: 23%).

Figure 7 illustrates the degree of structural homology between the HNF1 sequence, in both configurations discussed here, with the consensus for the homeodomain (see criteria adopted in legend). The homology is clearly restricted to the third helix in the case of HNF1 with a turn and extends largely to the whole N-terminal region in the case of HNF1 with a loop. Moreover, predicted α-helices, using the algorithm that gave the best result with the Antp sequence, as compared to available structural data (40), perfectly match those of the Antp homeodomain in the case of HNF1 with a loop while they are hardly compatible in the case of HNF1 with a turn.

If further structural analysis were to confirm the validity of the HNF1 with a loop model, one might wonder about the exact nature of the selective pressure that led to an almost universal conservation of a strict 3 amino acid turn among the helix-turn-helix proteins and about the properties of the loop that allowed HNF1 to escape this selection.

Proteins of a new family, including two that are mammalian transcription factors restricted to a specific cell type, share a highly homologous homeobox and two sequences upstream of it, defined as 'POU-A' and 'POU-B' (see references 38, 42 and 43 for review). The two latter show some homology with the HNF1 sequence, however it is very weak (highest scores : 6/26 and 4/34 respectively) as compared to homologies between the already described POU proteins themselves (42, 38). Moreover, these sequences overlap the HNF1 homeodomain whereas they are separated by roughly 20−30 amino acids in POU proteins (42, 38).

Further careful examination shows that several amino acids of the segment corresponding to the putative DNA-recognition helix in the HNF1 sequence differ radically in their nature from those found in most if not all other homeobox sequences (37). These positions are outlined as stars (*) in Figure 7. Let us retain two features that are exclusive to HNF1: *An alanine at position 50* : the residue in this position is believed to determine the specificity of DNA-recognition of homeoboxes of the Antp subclass (44, 45). At this key position most known homeoboxes contain a glutamine (the other residues that have been found until now are K, C, S, H, I). An aliphatic residue at this location may seriously impair binding to DNA through hydrogen bonding or polar interactions which might explain why HNF1 should bind to DNA as a dimer. *A glutamic acid and an alanine at positions 55 and 57 respectively*: all known homeoproteins have an arginine or a lysine at these extremely conserved functional positions and almost all at the adjacent position 58 also. The replacement, exclusive to HNF1, of these three basic residues by one acidic and two hydrophobic ones is striking, since these amino acids have been recently demonstrated to be exposed on the external face of the fourth helical segment of the Antp homeodomain and proposed to be involved in general electrostatic contacts with DNA (40); they are usually taken as a distinctive character of all homeoproteins (43). More precise mutagenesis is needed to define which amino acids in the long third helix of HNF1 are involved in contacting DNA, since they are likely to differ from those of both the canonical homeodomain and prokaryotic helix-turn-helix motif, possibly in relation to the presence of the 24 amino acid loop.

The HNF1 DNA-binding domain includes a distal dimerization domain of new type

The similarities and differences between HNF1 and other homeoproteins prompted us to compare the HNF1 target site to

other homeoprotein binding sequences determined to date. Homeoprotein DNA binding sites are AT-rich and similar to the average consensus sequence for one half of the HNF1 palindromic binding site (not shown). However, two points should be mentioned: first, except in yeast, homeoprotein target sites identified so far lack dyad symmetry; second, the HNF1 monomer has never been shown to form a stable complex with DNA. Rather, taken together, the different experiments reported here, using both native and cloned HNF1, strongly suggest that HNF1 forms homodimers, in the absence of DNA as well as when bound to its DNA target and that dimerization is essential for specific and high affinity DNA-binding. A domain required for dimerization lies outside of the homeodomain in a distal N-terminal region of the HNF1 protein. This region is rich in predicted α-helical structures and shows remote homologies with sequences involved in coiled coil formation but is not included in any existing dimerization motif family (not shown). During revision of our manuscript, anoter group outlined the crucial role of the N-terminal segment of HNF1 for dimerization and DNA-binding (48) An N-terminal dimerization domain was also observed in the Mat-α2 homeoprotein (46) to which it confers a very interesting flexibility in its interactions both with DNA and with other proteins (46, 47). Mat-α2 dimers are like here stabilized *in vitro* in nonreducing conditions, the physiological relevance of which is unknown. The need to dimerize for DNA-binding has so far been observed with no other homeoprotein. By contrast, this property is shared by most of the prokaryotic factors of the λ-Cro-type (41).

HNF1 properties extend the field of the homeoprotein superfamily

The example of HNF1, in addition to that of the POU proteins, further documents the homeodomain as the DNA recognition moiety of transcription factors implicated in the control of cell specific genes (43). As we discussed before, several predicted structural features extend the properties encountered among homeoproteins; this is also true at the functional level. HNF1 was initially identified only in hepatic nuclear extracts (8–11), however, northern blot and *in situ* hybridization showed that HNF1 transcripts are present at high levels in nonhepatic tissues like intestine and kidney (32 and our unpublished observations). On the other hand, expression of exogenous HNF1 is not sufficient to drive a high rate of transcription of a co-transfected albumin promoter in several cell types (F. Tronche, unpublished results). Thus, achievement of tissue specific patterns of transcription does not appear to be based on the action of strictly tissue specific trans-acting factors, but rather on various combinations of a small number of proteins. Homeoproteins, like HNF1, might fulfill complex regulation networks by alternative homo- or hetero-dimerizations and the low specificity observed in DNA recognition by the homeoproteins might be overcome by specific protein-protein interactions.

ACKNOWLEDGEMENTS

We are grateful to S. Hirai and B. Arcangioli for advice in protein purification, to S. Pochet for providing primed affinity resin, to J. Van Damme and M. Puype for their help in protein blotting and amino acid sequence analysis, to F. Tronche for help in construction of pRSV-HNF1, to R. Treisman for the gift of T7-β-globin vector, to J. Ars for typing, to N. Dostatni, R. Sousa, J. Ham and M. Weiss for valuable comments on the manuscript. This work was supported by grants from the EEC BAP Program, from the ARC, the LNFCC, and the FMRF to M.Y. and from the Belgian National Fund for Scientific Research (N.F.W.O.) to J.V. I.B. was supported by a Boehringer Ingelheim Fonds fellowship.

REFERENCES

1. Gorski, K., Carneiro, M. and Schibler, U. (1986) *Cell*, **47**, 767–776.
2. Heard, J.M., Herbomel, P., Ott M.O., Mottura-Rollier, A., Weiss, M. and Yaniv, M. (1987) *Mol. Cell. Biol.*, 7, 2425–2434.
3. Cereghini, S., Raymondjean, M., Garcia Carranca, A., Herbomel, P. and Yaniv, M. (1987) *Cell*, **50**, 627–638.
4. Lichsteiner, S., Wuarin, J. and Schibler, U. (1987) *Cell*, **51**, 963–973.
5. Babiss, L.E., Herbst, R.S., Bennet, A.L. and Darnell, Jr., J.E. (1987) *Genes and Develpment*, **1**, 256–267.
6. Herbomel, P., Rollier, A., Tronche, F., Ott, M.O., Yaniv, M. and Weiss, M. (1989). *Mol. Cell. Biol.*, **9**, 4750–4758.
7. Tronche, F., Rollier, A., Bach, I., Weiss, M. and Yaniv, M. (1989) *Mol. Cell. Biol.*, **9**, 4759–4766.
8. Courtois, G., Morgan, J.G., Campbell, L.A., Fourel, G. and Crabtree, G.R. (1987). *Science*, **238**, 688–692.
9. Hardon, E.M., Frain, M., Paonessa, G. and Cortese, R. (1988). *EMBO J.*, **7**, 1711–1719.
10. Schorpp, M., Kugler, W., Wagner, U. and Ryffel, G.U. (1988). *J. Mol. Biol.*, **202**, 307–320.
11. Cereghini, S., Blumenfeld, M. and Yaniv, M. (1988). *Genes and Development*, **2**, 957–974.
12. Courtois, G., Baumhueter, S. and Crabtree, G.R. (1988). *Proc. Natl. Acad. Sci. USA*, **85**, 7937–7941.
13. Blumenfeld, M., Cereghini, S., Raymondjean, M., Chouard, T. and Yaniv, M. (1988). UCLA symposia on Molecular and Cellular Biology, *New Series*, Vol.**95**, Alan R. Liss Inc., New York, NY, 91–105.
14. Monaci, P., Nicosia, A. and Cortese, R. (1988) *EMBO J.*, **7**, 2075–2087.
15. Ryffel, G.U., Kugler, W., Wagner, U. and Kaling, M. (1989) *Nucl. Acids Res.*, **17**, 939–953.
16. Maire, P., Wuarin, J. and Schibler, U. (1989) *Science*, **244**, 343–346.
17 Lichtsteiner, S. and Schibler, U. (1989) *Cell*, **57**, 1179–1187.
18. Arcangioli, B., Pochet, S., Sousa, R.M. and Huynh-Dinh, T. (1989) *Europ. J. Biochem.*, 179, 359–369.
19. Laemmli, U.K. (1970) *Nature*, **227**, 680–685.
20. Bauw, G., de Loose, M., Inge, D., Van Montagu, M. and Vandekerckhove, J. (1987). *Proc. Natl. Acad. Sci. USA*, 84, 4806–4810.
21. Saïki, R.K., Gelfand, D.H., Stoffel, S., Scharf, S.J., Higuchi, R., Horn, G.T., Mullis, K.B. and Ehrlich, H.A. (1988) *Science*, **239**, 487–491.
22. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular cloning: a laboratory manual (Cold Spring Harbor, New york).
23. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463–5467.
24. Chen, C. and Seeburg, P. (1985). *DNA*, **4**, 165–170.
25. Gorman, C.M., Merlino, G.T., Willingham, M.C., Pastan, I. and Howard, B.H. (1982). *Proc. Natl. Acad. Sci. USA*, **79**, 6777–6781.
26. Chirngwin, J.M., Przybyla, A.E., Mac Donald, R.J. and Rutter, W.J. (1979) *Biochemistry*, **18**, 5294–5299.
27. Baumhueter, S., Courtois, G. and Crabtree, G.R. (1988) *EMBO J.*, **7**, 2485–2493.
28. Deschatrette, J. and Weiss, M.C. (1974) *Biochimie*, **56**, 1603–1611.
29. Ryseck, R.P., Hirai, S.I., Yaniv, M. and Bravo, R. (1988) *Nature*, **334**, 535–537.
30. Birnsteil, M., Busslinger, M. and Strub, K. (1985) *Cell*, **41**, 349–359.
31. Frain, M., Swart, G., Monaci, P., Nicosia, A., Stämpfli, S., Frank, R. and Cortese, R. (1989) *Cell*, **59**, 145–157.
32. Baumhueter, S., Mendel, D.B., Conley, P.B., Kuo, C.J., Turk, C., Graves, M.K., Edwards, C.A., Courtois, G. and Crabtree, G.R. (1990). *Genes and Development*, **4**, 372–379.
33. Ptashne, M. (1988) *Nature*, 335, 683–689.
34. Dingwall, C. and Laskey, R.A. (1986) *Ann. Rev. Cell. Biol.*, 2, 367–390.
35. Garnier, J., Osguthrope, D. and Robson, B. (1978) *J. Mol. Biol.*, **120**, 97–120.
36. Gehring, W.J. (1987) *Science*, 236, 1245–1251.
37. Scott, M.P., Tamkun, J.W. and Hartzell, GW.III (1989) *BBA Rev.*, 989, 25–48.
38. He, X., Treacy, M.N., Simmons, D.M., Ingraham, H.A., Swanson, L.W. and Rosenfeld, M.G. (1989) *Nature*, 340, 35–42.

39. Finney, M. (1990). *Cell*, **60**, 5–6.
40. Qian, Y.Q., Billeter, M., Otting, G., Müller, M., Gehring, W.J. and Wütrich, K. (1989) *Cell*, **59**, 573–580.
41. Pabo, C.O. and Sauer, R.T. (1984) *Ann. Rev. Biochem.*, **53**, 293–321.
42. Herr, W., Sturm, R., Clerc, R.G., Corcoran, L.M., Baltimore, D., Sharp, P.A., Ingraham, H.A., Rosenfeld, M.G., Finney, M., Ruvkun, G. and Horvitz, H.R. (1988) *Genes and Development*, **2**, 1513–1516.
43. Levine, M. and Hoey, T. (1988) *Cell*, **55**, 537–540.
44. Hanes, S.D. and Brent, R. (1989) *Cell*, 57, 1275, 1283.
45. Treisman, J., Gönczy, P., Vashishta, M., Harris, E. and Desplan, C. (1989) *Cell*, 59, 553–562.
46. Sauer, R.T., Smith, D.L. and Johnson, A.D. (1988) *Genes and Development*, **2**, 807–816.
47. Keleher, C.A., Goutte, C. and Johnson, A.D. (1988) *Cell*, **53**, 927–936.
48. Nicosia, A., Monaci, P., Tomei, L., De Francesco, R., Nuzzo, M., Stunnenberg, H. and Cortese, R. (1990) *Cell*, **61**, 1225–1236.