Cell-type specific activity of two glucocorticoid responsive units of rat tyrosine aminotransferase gene is associated with multiple binding sites for C/EBP and a novel liver-specific nuclear factor

Thierry Grange*, Jeanne Roux, Gildas Rigaud and Raymond Pictet
Institut Jacques Monod du CNRS, Université Paris 7, Tour 43, 2 Place Jussieu, 75251 Paris Cedex 05, France

Received August 15, 1990; Revised and Accepted December 4, 1990

ABSTRACT

The structures of two remote glucocorticoid responsive units (GRUs) that cooperatively interact to promote cell-type specific glucocorticoid induction of rat tyrosine aminotransferase gene expression have been analyzed. DNAase I footprinting and gel mobility shift analyses reveal a complex array of contiguous and overlapping sites for cell type-specific DNA binding proteins. Apart from the glucocorticoid receptor, two liver-specific nuclear factors possess multiple binding sites in each of these GRUs: C/EBP and a newly identified liver-specific factor: HNF5. C/EBP possesses four binding sites in each GRU; a DNA-binding protein with similar binding specificity has been identified in fibroblasts; this protein could be related to AP-3. HNF5 possesses two binding sites in one GRU and four in the other. There are also HNF5 binding sites in numerous regulatory regions of other liver-specific genes. The interaction of HNF5 with DNA gives a characteristic DNAase I footprint with hypersensitive sites in the middle of the recognition sequence. Some of the C/EBP and HNF5 binding sites overlap in a conserved arrangement.

INTRODUCTION

The glucocorticoid receptor is a well-characterized transcription factor which, upon binding of the hormone, regulates transcription of its target genes (for reviews see 1, 2). Sequences responsible for glucocorticoid activation of transcription can be located upstream or downstream from the promoter and thus behave as typical enhancers (1, 2). The general structure of enhancers as well as promoters is modular (3, 4, 5). They are composed of multiple overlapping binding sites for different transcriptional activator proteins (4, 5). These discrete basic units of regulatory sequences have been termed enhansons (5). Functional modules are obtained by different combinations of different, hierarchically distinguishable, types of enhansons (5). Glucocorticoid receptor binding sites (GRBS) can combine with numerous transcription factor binding sites to form a glucocorticoid responsive unit (GRU, 6). The diversity of these factors allows a very complex repertoire of GRUs with various regulatory potentials, notably cell type-specificity, to be produced.

The tyrosine aminotransferase (*TAT*, E.C.2.6.1.5.) gene is expressed specifically in liver where its transcription is increased by glucocorticoids (for a review see 7). Full glucocorticoid induction of the rat *TAT* gene is achieved through cooperative interaction of two remote GRUs (8). One, referred to as HSIII, maps around −2,500 (9); the other, referred to as HSV, maps around −5,500 (8). HSIII corresponds to a liver-specific glucocorticoid-dependent DNAase I hypersensitive site (8, 9, 10). In rat hepatoma cells it is not very active by itself and requires the presence of HSV to produce the same level of glucocorticoid stimulation as the endogenous gene (11). HSIII contains three GRBSs, termed GRE 1, 2 and 3, two of them (GRE 2 and 3) being within regions necessary for its activity (9). In contrast to HSIII, HSV corresponds to a constitutive liver-specific DNAase I hypersensitive site; furthermore it is not active in the absence of HSIII (11). It contains a GRBS essential but not sufficient for its activity since flanking sequences are also necessary (11). We show here that both GRUs have a cell-type specific activity associated with multiple binding sites for two cell-type specific transcription factors.

MATERIALS AND METHODS

Plasmid constructions, cell culture and transfections

The construction of most plasmids presented here have been described previously (8). The HSV fragment in pCR51 has been obtained using the polymerase chain reaction (12), starting from a fragment (position −5,857 to −5,180) and amplified using the oligonucleotides: GAGCTCGACGAACTGGCCTGTG and AAGCTTTGTGGCTTTTCTCTC. After self-ligation, the fragment was cleaved with Sst I and Hind III before cloning between the Sst I and Hind III sites of pKT531 (8). pGR2 was constructed

* To whom correspondence should be addressed

by insertion between the Xba I and the Sal I sites of pUTKat4 (13) of the fragment resulting from the annealing of the two following oligonucleotides: CTAGCTGTACAGGATGTTCTAGCTAC and TCGAGTAGCTAGAACATCCTGTACAG. The nucleotide sequences of all the *in vitro* synthesized fragments was verified after cloning by the dideoxy sequencing procedure (14).

Plasmid preparations, cell culture and transfections using the $CaPO_4$ precipitation procedure were exactly as described (8) except for mouse L cells which were grown in DMEM plus 8% foetal calf serum and seeded at 2.10^6 cells per 10 cm-diameter plates the day before DNA addition.

Crude nuclear extract preparation, DNAase I footprinting and gel mobility shift experiments

Crude nuclear extracts (CNE) were prepared from 10^9 cells grown in plates. The cells were recovered from the plates using trypsin and resuspended in 40 ml of buffer H (0.4 M sucrose, 10% glycerol, 10mM Hepes (pH 7.6), 60mM KCl, 15 mM NaCl, 0.15 mM spermine, 0.5 mM spermidine, 1mM EDTA, 1mM DTT, 0.1 mM PMSF). The cells were lysed with a mild detergent (NP40) as follows: The cell suspension was diluted in 40 ml of buffer H containing either 0.5% NP40 (H4 II cells) or 0.8% NP40 (L cells). After a 3 min.-incubation at 4°C the suspension was diluted with 80 ml of buffer H and the nuclei recovered by centrifugation. The pelleted nuclei were further purified on a 2M sucrose cushion (in buffer H) and the CNEs were prepared according to the procedure of Gorski *et al.* (15).

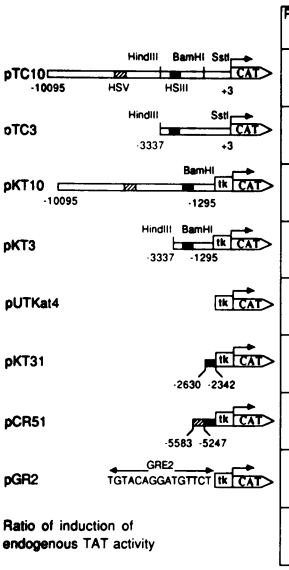
DNAase I footprinting was performed exactly as described (16). Gel shift assays were performed with blunt-end ds oligonucleotides generated by the annealing of two complementary ss synthetic oligonucleotides labeled on each strand using polynucleotide kinase and purified on acrylamide gel. The incubation of 10,000 cpm of the labeled oligonucleotide with the CNE was carried out in the same conditions as for DNAase I footprinting for 15 min. at 4°C. Gel (8% acrylamide) conditions were as described (17). The sequences of the oligonucleotides, named as in the figures, are: -s1: 5′ AAGCCC AAGGTTTACAATCTCTGC 3′; -s2: 5′ TACTTTATTTGCA ATAGAAAATC 3′; -s3: 5′ CTGAAAGTTTCCCCATGTCC AACA 3′; -s4: 5′ CTAGAACAAACAAGTCCTGCGT 3′; -s4 mut: 5′ CTAGAAAAAACCAGTCCTGCGT 3′ (mutated positions underlined); -s5: 5′CGCCTGTCGGTTTCTGGGTG TGGTGG 3′; -GRE2: 5′ CTGCTGTACAGGATGTTCTAGC TACT 3′; -HBV: 5′ CGTCAGCAAACACTTGGCACAGACC 3′; -Alb (F): 5′ GATGGCAAACATACGCAAGGGATTTA 3′; -Alb (E1): 5′ TAAGAACAAACATTCCCTGTAT 3′; -Alb (E2): 5′ CCAAGGCAAACACGTCCGGAAT 3′; -Oct Ig kappa: 5′ TGAGGGTATGCAAATTATTAAGAAGC 3′; -Core: 5′ TGCTGGGGACTTTCCACACCCTAA 3′; -NF1 site: 5′ TAT TTTGGATTGAAGCCAATATGATAATGA 3′; -NFY site: 5′ GGGGTAGGAACCAATGAAATGAAAGGTTA 3′; The last three oligonucleotides have been described by Raymondjean *et al.* (18).

RESULTS

Cell type-specific glucocorticoid response is associated with binding sites for cell type-specific nuclear proteins in TAT gene GRUs

We have used a transient expression assay to analyze the activity of the elements involved in rat *TAT* gene glucocorticoid response in two cell lines. One, a rat hepatoma cell line, expresses the *TAT* gene while the other, a mouse fibroblast line, does not. The results are shown in Figure 1. In hepatoma cells (H4II), 10 kbp of *TAT* gene 5′ flanking region confer to a reporter *CAT* gene a level of glucocorticoid induction identical to that of the endogenous *TAT* gene (pTC10). Deletion of the −10,095 to −3,337 region leads to a 70% decrease of the induction level (pTC3). Similar ratios of induction are obtained when the Herpes simplex virus tk promoter is substituted for the *TAT* gene promoter from position −1,295 to +3 (pKT10 and pKT3). In fibroblasts (L cells or an Ltk⁻ subclone), neither 10 kbp nor 3 kbp of the *TAT* gene 5′ flanking region produces a significant glucocorticoid response (pTC10 and pTC3). This cell type-specific glucocorticoid response is still observed with the tk promoter (pKT10 and pKT3). All the cell lines used contain a functional glucocorticoid receptor since *CAT* gene expression controlled by the mouse mammary tumor virus long terminal repeat is stimulated by glucocorticoids (data not shown).

The regions responsible for the effects observed were more precisely determined by the transfection of a series of constructs (8). Figure 1 also shows that in hepatoma cells, a 289 bp fragment containing HSIII (−2,630 to −2,342; pKT31) produces a glucocorticoid response similar to that of a 2 kbp fragment (−3,337 to −1,295; pKT3). Similarly, a 337 bp fragment containing HSV (−5,583 to −5,247; pCR51) enhances the HSIII-driven glucocorticoid response to the level obtained with 9 or 10 kbp fragments of the *TAT* gene 5′ flanking region (pKT10 and pTC10). These two fragments contain enough information



	Ratio of induction by dexamethasone	
	Hepatoma cells	Fibroblasts
pTC10	10.5 ± 1.0 (6)	2.4 ± 0.7 (3)
oTC3	3.1 ± 0.5 (4)	1.1 ± 0.2 (3)
pKT10	10.5 ± 0.7 (6)	2.0 ± 0.1 (4)
pKT3	4.2 ± 0.3 (10)	1.5 ± 0.1 (4)
pUTKat4	1.7 ± 0.1 (15)	1.0 ± 0.1 (9)
pKT31	3.9 ± 0.5 (6)	N. D.
pCR51	13.9 ± 1.3 (7)	2.3 ± 0.6 (4)
pGR2	2.0 ± 0.2 (6)	1.9 ± 0.3 (3)
Ratio of induction of endogenous TAT activity	10.5 ± 0.4 (48)	——

Figure 1: Cell type-specific glucocorticoid response revealed by a transient expression assay. The boundaries of the *TAT* gene 5′ flanking regions placed upstream from the *CAT* gene are numbered relative to the *TAT* gene transcription initiation site, referred to as position +1 according to the previously published nucleotide sequence of the rat *TAT* gene 5′ flanking region (51). The black box and the hatched box indicate the HSIII region and the HSV region respectively (8); tk indicates the Herpes simplex virus thymidine kinase promoter which extends here from position −207 to +56 (13); GRE2 is the sequence of the major GRBS of HSIII (9). After normalization of the CAT activity with respect to the β galactosidase activity used as an internal control, the ratio of induction of CAT activity by dexamethasone (a glucocorticoid analog) was determined for each transfection (detailed experimental procedures are described in Grange *et al.*; 8). The mean value of these independent ratios of induction is indicated ± the standard error of the mean. The number of independent experiments performed is indicated in brackets. N.D.: not done. Two clones of mouse L fibroblasts (ATCC CCL 1.2 and ATCC CCL 1.3) were used without observing significant differences in the ratio of induction.

for cell type-specific glucocorticoid response since they are not functional in fibroblasts (pCR51). The sequences flanking the GRBS contribute to the activity of HSV in hepatoma cells (8). This is also true for the HSIII region since an oligonucleotide corresponding to its major GRBS (GRE2; 9) is unable to produce a significant glucocorticoid response (pGR2 in Figure 1).

To identify the putative elements involved in the cell type-specific glucocorticoid response we looked for binding sites of cell type-specific nuclear proteins by *in vitro* footprinting analysis with DNAase I (19). Both strand of each GRU were analysed using crude nuclear extracts (CNE) prepared from both hepatoma cells and fibroblasts (Figures 2A & 2B). Protected regions cover most of the length of the fragments active in transfection experiments and surround the GRBS in HSV (8) as well as GRE2 and GRE3 in HSIII (9). Beside slight variations in signal intensity, the overall footprinting pattern of each GRUs is identical with both extracts. Most of the protected segments have the same size and boundaries with the exception of two regions located between

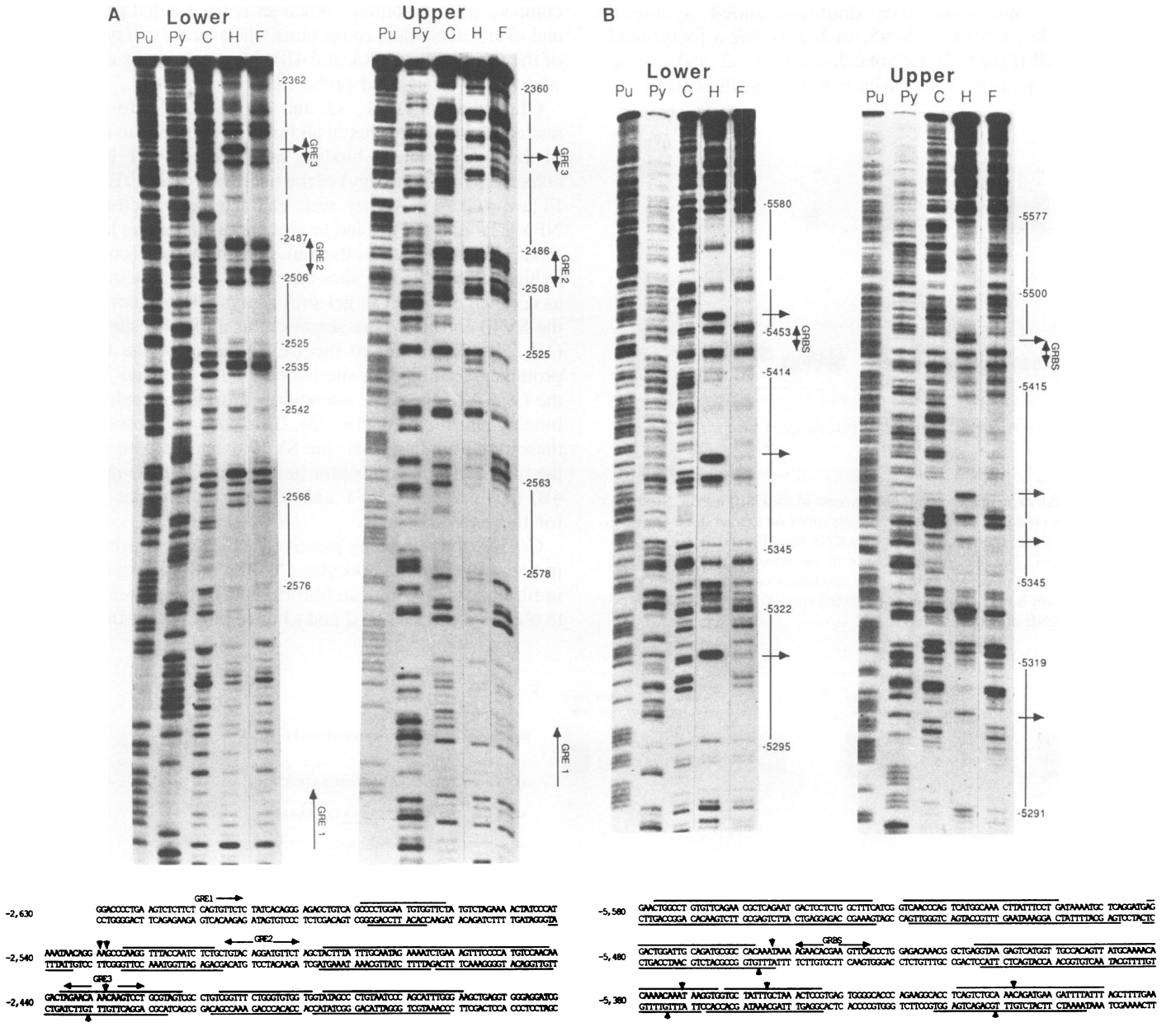


Figure 2: *In vitro* footprinting with DNAase I reveals multiple nuclear protein binding sites in the two GRUs. **A**: HSIII and **B** HSV regions. From left to right the first two lanes of each panel correspond to DNA chemically cleaved at purines (Pu) or pyrimidines (Py) (52). Lane C, control with DNAase I in the absence of CNE. Lanes H and F correspond to DNA cleaved by DNAase I after incubation of the fragments with respectively 20 μg of hepatoma cell or 30 μg of fibroblast CNE. The vertical bars on the right of each panel indicate the location of the footprints obtained with hepatoma cell CNE. We have represented the footprints obtained reproducibly at different DNAase I concentrations. The boundaries were determined on gels allowing sufficient resolution in the corresponding regions. The DNAase I hypersensitive sites specific of hepatoma cell CNE are represented by horizontal arrows. Vertical arrows indicate the location of the glucocorticoid receptor binding sites termed GRE 1 to 3 (9) and GRBS (8). End-labeling of the fragments was performed on added HindIII sites located either in HSIII close to a natural AvaII site (located upstream at −2630), or in HSV close to a natural Bgl II site (located downstream at −5184) using either the Klenow fragment of *E. coli* DNA polymerase I (lower and upper strands respectively) or T4 polynucleotide kinase (upper and lower strands respectively). When labeling was performed with the kinase the labeled fragment was blunt-ended by filling in with the Klenow fragment.

nucleotides −2,440 to −2,414 and between nucleotides −5,322 to −5,291 which are protected only by the hepatoma cell CNE. The major difference is the presence in both GRUs of DNAase I hypersensitive sites specific to hepatoma cell CNE. In HSIII these sites are located at positions −2,431 and −2,428 in the lower and upper strands respectively (Figure 2A) and in HSV at positions −5,458, −5,375 and −5,312 in the lower strand and at −5,455, −5,372, −5,355 and −5,309 in the upper strand (Figure 2B).

The nuclear proteins interacting with the *TAT* gene GRUs were analyzed by gel mobility shift assays (further designated gel shift assays; 17). We used five double-stranded synthetic oligonucleotides, termed s1 to s5, each covering a footprinted region of HSIII (Figure 3). Oligonucleotides s1, s2, and s3 give rise to similar retarded bands with both CNEs. In contrast, s4 binds (a) factor(s) only when incubated with hepatoma cell CNE. Oligonucleotide s5 binds (a) factor(s) presumably present in both extracts but in different relative amounts. The corresponding sequence includes a CACCC box homology, a motif described as the target sequence for TEF-2 (20) and maybe some other factors (21). We have restricted our investigation to the characterization of the factors binding to oligonucleotides s1 to s4.

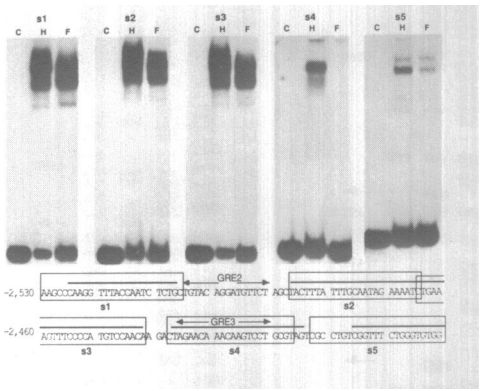


Figure 3: Gel shift experiments using subfragments of the HSIII region. Synthetic oligonucleotides (s1 to s5) were incubated with either no extract (lane C), 6 μg of hepatoma cell (lane H) or 10 μg of fibroblast CNE (lane F), and the components separated on a gel. The nucleotide sequence of the upper strand of the HSIII region analyzed is represented below. The sequences corresponding to each oligonucleotide are boxed. The regions protected from DNAase I digestion by hepatoma cell CNE are overlined.

In both GRUs there are multiple binding sites for C/EBP as well as for other factors with similar binding specificity

If s1, s2 and s3 bind to similar protein(s), each one should compete with the others. When s3 is the labeled probe, s1, s2 and s3 are all efficient competitors whatever the cell type of origin of the CNEs (Figure 4A and 4B). Identical results are obtained when s1 is the labeled probe (data not shown).

Oligonucleotides s1, s2 and s3 do not share a common nucleotide sequence, but in all three it is possible to identify one of the two consensus binding sites (the CCAAT box and the enhancer core homology) of the nuclear protein C/EBP (22; see Figure 4C). Two other well characterized nuclear proteins, NF1/CTF and NFY, bind to sequences containing the CCAAT motif (23 and references therein). Oligonucleotides corresponding to high affinity binding sites for each of these proteins were used as competitors of s3 in gel shift assays : 1) one corresponds to the SV40 enhancer core sequence, as a binding site for C/EBP (18, 24); 2) another to the CCAAT box of the rat albumin promoter, as a binding site for NFY (18, 23); and 3) a third to the CCAAT box of the adenovirus 2 origin of replication, as a binding site for NF1 (18, 25). This analysis shows that out of these three binding sites, the SV40 enhancer core sequence is the most efficient competitor in both cell CNEs (Figure 4A and 4B) thus excluding NFY and NF1 as the proteins responsible for the retarded bands.

C/EBP, the remaining putative candidate is expressed mostly in hepatocytes and adipocytes (26). Its mRNA is not detectable in fibroblasts (27); thus in fibroblast CNE the protein(s) binding to oligonucleotides s1, s2 and s3 must be different from C/EBP.

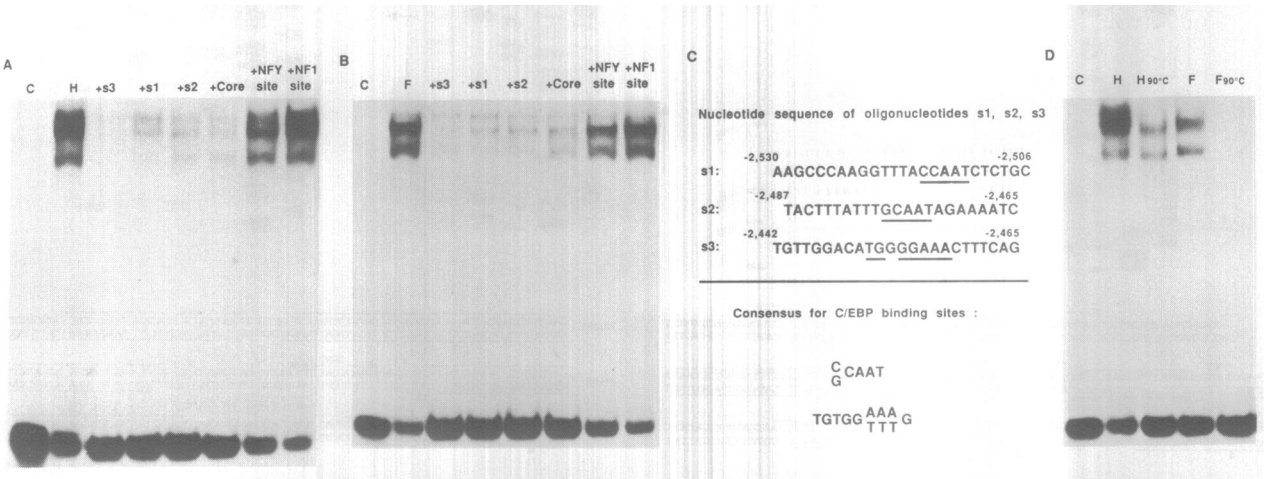


Figure 4. Similar factors bind oligonucleotides s1, s2 and s3, and one has the properties of C/EBP. **A:** Competition experiments using hepatoma cell CNE. Labeled oligonucleotide s3 was incubated either without extract (lane C) or with 6 μg of hepatoma cell CNE (all other lanes) in the absence of competitors (lane H) or in the presence of 10 ng of unlabeled oligonucleotides (all lanes marked with a +). Core, NFY site and NF1 site correspond to oligonucleotides which are preferred binding sites for C/EBP, NFY and NF1 respectively (see text and 'Materials and Methods' section). After incubation the components were separated on a gel. **B:** Competition experiments using fibroblast CNE. The data obtained with 10 μg of CNE are presented as in figure 4A. Lane F represents the incubation without competitor. **C:** Oligonucleotides s1, s2 and s3 each contain a sequence related to a described C/EBP binding site consensus (22). The sequence of the relevant strand of each oligonucleotide is represented and the location of the region presenting significant homology with one of the consensuses for C/EBP binding sites is underlined. **D:** Differential heat stability of the DNA-binding activities present in the CNEs. Before incubation with labeled oligonucleotide s3, an aliquot of each CNE was treated for 10 min. at 90°C. After centrifugation, the activity remaining in the supernatant was tested in parallel with untreated extract. In each incubation the same volume of CNE was used corresponding to 6 μg and 10 μg of protein for respectively untreated hepatoma cell (H) and fibroblast (F) CNEs.

C/EBP is a thermoresistant protein (24). Both CNEs were heated at 90°C for 10 min and the remaining DNA-binding activity was tested by gel shift assay. Figure 4D shows that half of the DNA-binding activity present in hepatoma cell CNE is thermoresistant. In contrast most of the activity present in fibroblast CNE is heat-sensitive. In conclusion one of the DNA-binding proteins responsible for the retarded bands obtained with s1, s2 and s3 is present only in hepatoma cell CNE and is likely to be C/EBP.

C/EBP indeed binds to the sequences included in the oligonucleotides s1, s2 and s3: DNAase I footprinting performed with purified C/EBP on the HSIII region shows a clear protection with boundaries identical to those obtained with both CNEs (Figures 5A and 2A). In addition it reveals a fourth C/EBP binding site, located from position −2,388 to −2,360 (Figure 5 A and data not shown). When DNAase I footprinting is performed on the HSV region, one small (position −5,481 to −5,458) and one large area (position −5,424 to −5,345) are protected (Figure 5B). Inspection of the large protected area shows two DNAase I cleavage sites at positions −5,399 and −5,364. This area can thus be divided into three subregions which are likely to represent three binding sites for C/EBP. Competition experiments performed with hepatoma cell and fibroblast CNEs and analyzed using DNAase I footprinting and gel shift assays show that the protein(s) binding to oligonucleotide s3 also bind(s) to the regions protected with purified C/EBP (data not shown). However, the boundaries of the footprints obtained with C/EBP and with the CNEs are slightly different; the footprint extends from −5,424 to −5,345 with C/EBP and from −5,415 to −5,345 with the hepatoma cell CNE (Figure 5B, see also Figure 2B). Furthermore the two internal DNAase I hypersensitive sites specific to the hepatoma cell CNE (positions −5,372 and −5,355) are not obtained with purified C/EBP (Figure 5B). These differences result from the binding of other proteins (*vide infra* and data not shown).

In conclusion there are four C/EBP binding sites in both HSIII and HSV. These binding sites are shared with other DNA-binding proteins with similar or different binding specificities.

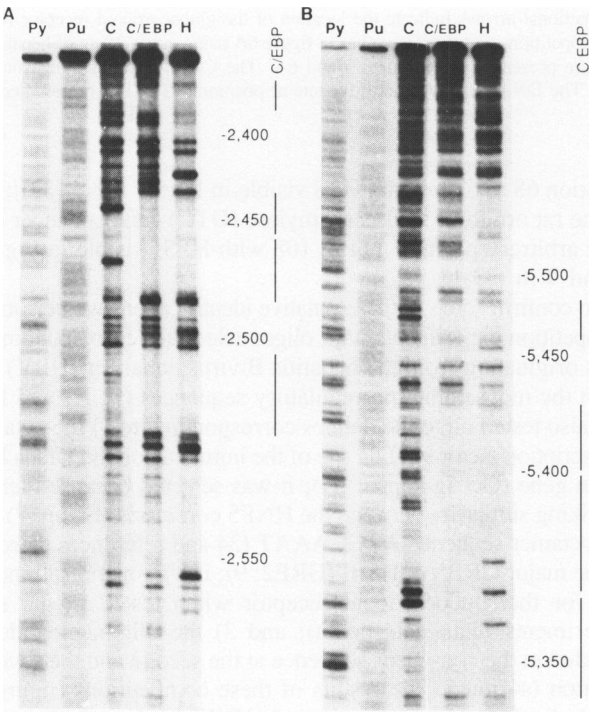


Figure 5: C/EBP possesses multiple binding sites in each GRU. The figure shows DNAase I footprinting over **A:** HSIII (lower strand) and **B**: HSV (upper strand). Before DNAase I treatment incubation was performed with either 40 ng of purified C/EBP (lane C/EBP) or 20 μg of hepatoma cell CNE (lane H). The regions protected with purified C/EBP are presented on the right. The other symbols are as in figure 2A.

The hepatoma cell-specific nuclear factor which binds oligonucleotide s4 possesses two binding sites in HSIII and four binding sites in HSV

Figure 3 shows that oligonucleotide s4 gives rise to a retarded band specific to hepatoma cell CNE. Analysis of the DNAase I footprint obtained in the corresponding region reveals the presence, only with hepatoma cell CNE, of a DNAase I hypersensitive site on each strand (at positions −2,431 and −2,428 on the lower and upper strand respectively; Figure 2A). We hypothesized that the same factor(s) was (were) responsible for the hepatoma cell-specific retarded band and the DNAase I hypersensitive sites. Since such hypersensitive sites are also obtained in the HSV region, we suspected that this factor(s) might also bind to that region. Figure 6A shows that competition with oligonucleotide s4 leads to the disappearance of the footprint extending from position −5,319 to −5,291 and of the four hepatoma cell-specific DNAase I hypersensitive sites; the resulting footprint pattern is similar to that obtained with fibroblast CNE.

Comparison of the sequences of oligonucleotide s4 and of the regions including these 4 hypersensitive sites shows that they all possess a common seven nucleotide long motif (Figure 7). This motif is likely to be partly responsible for the recognition of these five sequences by the same hepatoma cell-specific factor(s) that will be referred to as the Hepatic Nuclear Factor n°5 (HNF5). The fixed positions of the DNAase I hypersensitive sites within the motif, another feature of HNF5 binding sites, indicates that HNF5 leaves two specific phosphodiester bonds (one on each strand) accessible to DNAase I.

Examination of the sequence of HSIII reveals another HNF5 motif located around position −2,480 (Figure 7). However, there are no DNAase I hypersensitive sites at the expected corresponding positions (Figure 2A). Since this region is also a C/EBP binding site (Figure 5A), C/EBP (or factor(s) with similar binding specificity) could be responsible for their absence. To test this hypothesis we analyzed the effect of competition with oligonucleotide s3 which is a target site for C/EBP (Figure 4) but does not contain the HNF5 consensus motif (Figure 7). Figure 6B shows that this competition with hepatoma cell CNE generates a DNAase I hypersensitive site at position −2,481 (compare lanes C, H and H+s3) i.e. at a location characteristic of HNF5 binding. HNF5 is indeed responsible for this hypersensitive site since simultaneous competition with oligonucleotides s3 and s4 (an HNF5 binding site) leads to the control pattern (compare lanes C, H+s3 and H+s3+s4 in Figure 6B). As expected from the cell-type specificity of HNF5, competition with oligonucleotide s3 is sufficient to obtain the control pattern with fibroblast CNE (compare lanes F, F+s3 and C in Figure 6B). A similar HNF5 and C/EBP binding site overlap is found twice in HSV around positions −5,355 and −5,370 (Figures 5B and 6A). In these regions however the DNAase I hypersensitive sites characteristic of HNF5 binding are visible (at least on one strand) with uncompeted hepatoma cell CNE (Figure 2B).

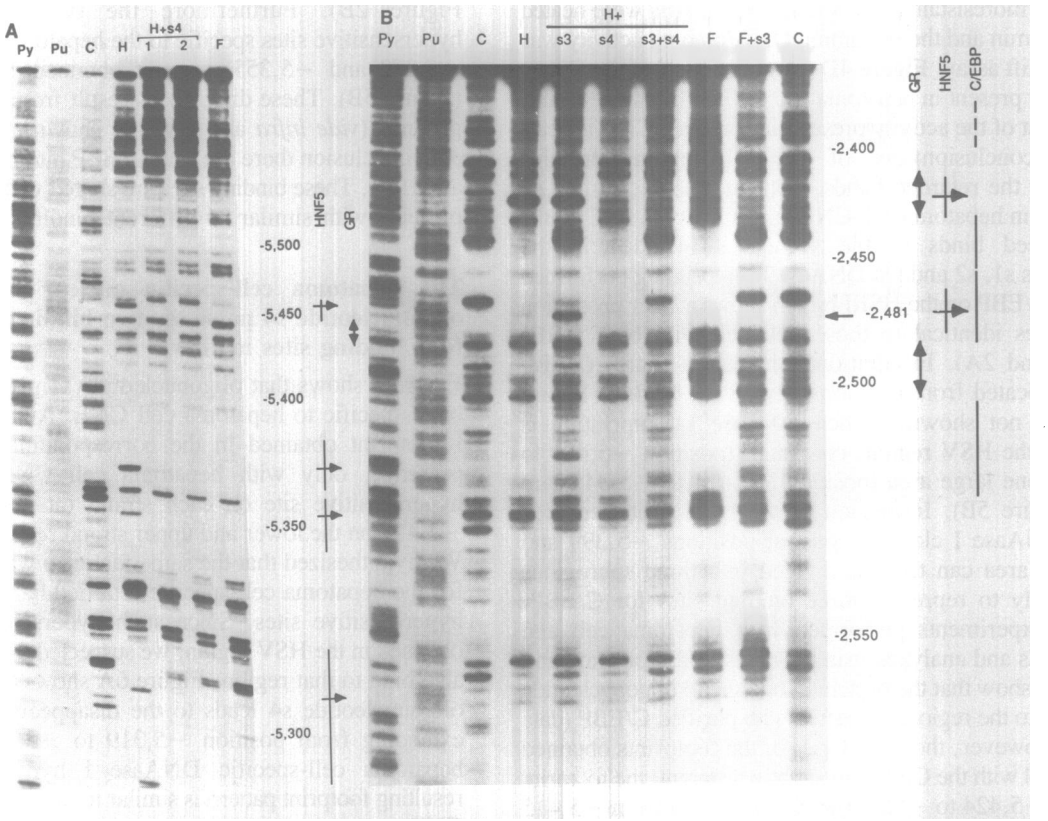


Figure 6: Competitions experiments analyzed with DNAase I footprinting reveal multiple binding sites for a novel liver-specific factor: HNF5. **A:** DNAase I footprinting experiments over the HSV region (upper strand). The labeled fragment was incubated with 20 μg of hepatoma cell CNE in the absence of oligonucleotide (lane H) or in the presence of two different amounts of oligonucleotide s4 (H+s4, 1: 10 ng, 2: 30 ng). The other symbols are as in figure 2A. The location of the HNF5 binding sites deduced from the competition-induced disappearance of the DNAase I hypersensitive sites is indicated on the right with bars. The location of these DNAase I hypersensitive sites is represented with an arrow crossing these bars. The bidirectional arrows indicate the location of the glucocorticoid receptor (GR) binding site. **B:** DNAase I footprinting experiments over the HSIII region (lower strand). Competitions were performed as in figure 6A using either 20 μg of hepatoma cell CNE or 30 μg of fibroblast CNE and 40 ng of oligonucleotides s3 and/or s4. Data are presented as in figures 5 and 6A. The C/EBP binding site which is not clearly visible upon competition with oligonucleotide s3 is indicated by a dashed line. The DNAase I hypersensitive site at position −2,481 is represented by an arrow (see text).

In conclusion there are two HNF5 binding sites in HSIII and four HNF5 binding sites in HSV. Some of these sites overlap the binding sites of other factors including C/EBP.

HNF5 binding sites are also present in multiple regulatory sequences of other genes with liver-specific expression

The presence of HNF5 binding sites within two regulatory regions conferring liver cell type-specific gene expression prompted us to search for HNF5 binding sites in regulatory regions of other liver-specific genes. Sequences matching the consensus for HNF5 binding sites can indeed be found in many of these regulatory regions, and the characteristic DNAase I hypersensitive sites observed on some of the published footprints. Such putative binding sites are present, among others, in the following regulatory regions: 1) the hepatitis B virus enhancer [position 1196 to 1190, with a hypersensitive site (HSS) visible in Figure 2B in (28)]; 2) the mouse albumin promoter [position −160 to −154, referred to as F site, with an HSS visible in Figures 1A−1B in (16)]; 3) the mouse albumin enhancer [two sites at positions −10,059 to −10,065 and −10,030 to −10,036, (29) hereafter referred to as E1 and E2 respectively]; 4) the human alpha-1-antitrypsin gene [position −100 to −106 with an HSS visible in Figure 3B in (30)]; 5) the human transferrin gene promoter [position −79 to −73, (31)] and its enhancer [arbitrary position 68 to 62 with an HSS visible in Figure 5B in (32)]; and 6) the rat ornithine transcarbamylase (OTC) gene enhancer [site II at arbitrary position 114 to 108 with HSSs visible in Figures 6 and 7 in (33)].

To confirm some of these putative identifications we performed competition experiments with oligonucleotides corresponding to sites originating from the hepatitis B virus enhancer (HBV) and from the mouse albumin regulatory sequences (F, E1 and E2). We also tested oligonucleotides corresponding to 1) the octamer transcription factor binding site of the immunoglobulin kappa light chain gene (Oct Ig kappa; 34); it was selected because there is a striking similarity between the HNF5 consensus (Figure 7) and the octamer sequence ATGCAAAT (34 and references therein); 2) the major GRBS of HSIII (GRE2; 9); it is a functional binding site for the glucocorticoid receptor when tested in gel shift experiments (data not shown); and 3) the oligonucleotide s4 mutated in the consensus sequence at the second and the seventh position (s4 mut). The results of these competitions (figure 8) lead to the following conclusions: 1) HNF5 binds to the four sites originating from liver-specific regulatory regions (the F site of the albumin promoter is a low-affinity site); 2) HNF5 is not one of the octamer transcription factors; 3) HNF5 is not the glucocorticoid receptor; and 4) the binding of HNF5 to its target sequence is abolished by the mutation of two bases located in the consensus sequence we have defined.

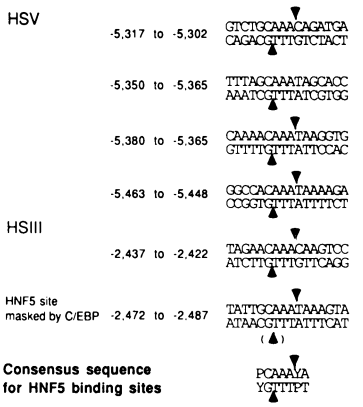


Figure 7: Nucleotide sequence of HNF5 binding sites reveals a common conserved motif. The sequences of both strands of each HNF5 binding site of the *TAT* gene GRUs are presented. The upper strand is written in the 5′ to 3′ direction. The arrows indicate the location of the phosphodiester bonds cleaved by DNAase I upon HNF5 binding. P indicates a purine residue, Y a pyrimidine residue. The consensus sequence presented does not include the purine residue conserved at the eighth position among the sites analyzed here because it is not conserved in the HNF5 binding sites found in other liver specific-genes.

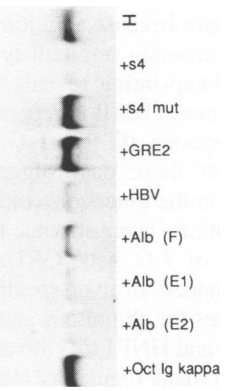


Figure 8: Analysis of putative HNF5 binding sites with the gel shift assay. Labeled oligonucleotide s4 was incubated with 6 μg of hepatoma cell CNE in the absence (lane H) or the presence of 20 ng of unlabeled oligonucleotide as competitor (all other lanes). These oligonucleotides are described in the text and their nucleotide sequence is reported in the 'Materials and Methods' section. The figure shows only the visible retarded bands.

DISCUSSION

We have analyzed the structure of two GRUs of the rat *TAT* gene. Each GRU (HSIII and HSV) is composed of multiple binding sites of a limited number of DNA-binding proteins (Figure 9). This modular organization appears to be a general feature of enhancer structure (4, 5) but the characteristic feature of these GRUs is the redundancy of the binding sites for three trans-acting factors.

One of these factors, the glucocorticoid receptor, has three binding sites in HSIII (9) and one in HSV (8). The activity of this factor, expressed in most cell-types, is dependent upon the presence of the hormone (for reviews see 1, 2). It is the key trans-acting factor for the function of these GRUs because in the absence of hormone none of these GRUs activates gene expression (8). The minimal requirement to produce a functional element conferring glucocorticoid responsivness is unclear. In some cases a GRBS alone has been reported to be sufficient (9,

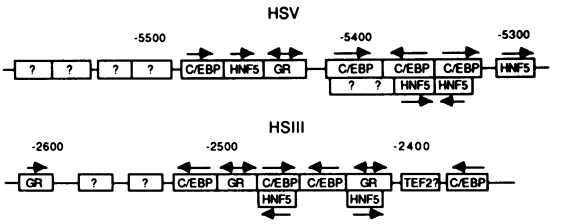


Figure 9: Recapitulation of the arrangement of the nuclear factor binding sites identified in the two *TAT* gene GRUs. The arrows above the HNF5 and C/EBP boxes indicate the orientation of the binding sites, determined arbitrarily relative to the consensuses defined in Figures 7 and 10. The bidirectional arrows indicate the palindromic GRBSs. TEF2? indicates the location of the putative TEF-2 site. ? indicates the binding sites for unidentified factors.

Gene	Position	Sequence	References
TAT GRUs (rat)	- 2510 to - 2522	AGATTGGTAAACC	This study
	- 2482 to - 2470	TATTTGCAATAGA	
	- 2449 to - 2461	CATGGGGAAACTT	
	- 2368 to - 2380	GCTGGGATTACAG	
	- 5476 to - 5464	GGATTGCAGATGC	
	- 5420 to - 5408	GCTGAGGTAAGAG	
	- 5370 to - 5382	TATTTGTTTTGTG	
	- 5360 to - 5348	TATTTGCTAAACT	
HSV tk promoter	- 88 to - 76	TCATTGGCGAATT	(53)
MSV promoter	- 75 to - 87	TGATTGGTTAGTT	
MSV enhancer	216 to 228	TCTGTGGTAAGCA	(24)
SV40 enhancer	250 to 238	GGTGTGGAAAGTC	
Polyoma enhancer	5212 to 5224	AGTGTGGTTTTGC	
HBV enhancer	1036 to 1048	AATGTGGATATCC	(22)
	1190 to 1202	TGTTTGCTGACGC	
Albumin promoter (mouse)	- 108 to - 96	ATTTTGTAATGGG	(55)
Albumin enhancer (mouse)	-10174 to -10186	TGGATGCTATAAC	(29)
	-10132 to -10144	CCTATGTTAAGTA	
TTR enhancer (mouse)	- 1918 to - 1930	TATTAGGACATGT	(55/56)
	- 1871 to - 1883	ATGTTGAGTAAGA	
A1AT enhancer (mouse)	- 462 to - 450	CAGGTGGGCACAT	(55)
	- 190 to - 202	TCAATGGAGTCTT	(54)
CONSENSUS FROM 22 C/EBP BINDING SITES		T T/G T G G T/A A/T T/A	This study
		14 11/8 18 22 12 11/8 10/6 16/6	
CONSENSUS FROM 7 AP-3 BINDING SITES		T G T G G T/A A/T T/A	(39)
		5 6 7 7 5 4/3 5/2 4/3	
CONSENSUS FOR HNF5 BINDING SITES		T G/A T T T G C/T	

OVERLAPPING OF C/EBP AND HNF5 BINDING SITES

TAT gene GRUs	-2482 to -2473	HNF5 TATTTGCAAT C/EBP
	-5360 to -5351	HNF5 TATTTGCTAA C/EBP
	-5370 to -5379	HNF5 TATTTGTTTT C/EBP
HBV enhancer	1190 to 1199	HNF5 TATTTGCTGA C/EBP
OTC enhancer	108 to 117	HNF5 TATTTGCTTA C/EBP

Figure 10: Analysis of 23 C/EBP binding sites allows the establishment of a unique consensus sequence similar to the consensus sequence for AP-3 binding sites. Analysis of five overlapping HNF5 and C/EBP binding sites reveals conservation of the relative location of the consensus sequences for both factors' binding sites.

35). However similar experiments have shown that the presence of other transcription factor binding sites may be necessary (6, 36, 37). We have not obtained significant glucocorticoid

stimulation using a GRBS originating from the HSIII region (pGR2 in Figure 1). Sequences flanking the GRBS are necessary for the activity of both HSIII (pKT31 in Figure 1) and HSV (8). These two GRUs do not function in fibroblasts (Figure 1) and their activity is extinguished in intertypic hepatoma-fibroblast hybrid cells (11). Thus the flanking sequences must bind factors responsible for this cell-type specific activity. Two factors are likely candidates because they have multiple binding sites in each GRU.

One, with four binding sites in each GRU, has the properties of C/EBP (22). C/EBP is primarily expressed in hepatocytes and adipocytes, but not in fibroblasts (26, 27). It recognizes the CCAAT box and the enhancer core sequence. Since it does so with the same DNA binding domain (38 and references therein), this suggests that it recognizes common features in seemingly unrelated sequences. Analysis of 22 C/EBP binding sites allows the identification of a common eight nucleotide long motif related to the enhancer core sequence (Figure 10). This consensus can also be found in sites containing the CCAAT box homology suggesting the existence of a single class of C/EBP binding sites. In fibroblasts, Mercurio and Karin (39) have described a DNA-binding activity (AP-3) which recognizes sequences very similar to the consensus defined here (Figure 10). Thus the activity present in fibroblast CNE which binds C/EBP sites in the *TAT* gene GRUs (Figures 3–4, data not shown) could be related to AP-3. The heat-labile fraction of the activity detected in hepatoma cell CNE (Figure 4D) could also be related to AP-3. Other regulatory sequences are target sites for multiple DNA-binding proteins, for example the CCAATT boxes (23), the TPA responsive elements (40 and references therein), and the cAMP responsive elements (41). The functional significance of this multiple recognition remains in most cases conjectural.

We have identified another cell-type specific DNA-binding activity which recognizes two sites in HSIII and four in HSV. This activity is present in hepatoma cell (lane s4, H in Figure 3) and in liver CNEs (data not shown), but neither in fibroblast (lane s4, F in Figure 3) nor in spleen, brain and heart CNEs (data not shown). Four liver-specific DNA-binding activities have been named HNF: HNF1 (42), also known as LF-B1 (43, 44); HNF2 (45), also known as LF-A1 (43); HNF3 and HNF4 (46). Since the consensus for the binding site of the factor we have identified (Figure 7) is clearly different from the others, we have named it HNF5.

The nucleotide sequence features common to all HNF5 binding sites are likely to correspond to the recognized DNA motif. Indeed mutation of two bases therein abolishes HNF5 binding (Figure 8). A characteristic feature of the interaction of HNF5 with its target sequence is revealed by DNAase I. This enzyme cleaves one phosphodiester bond on each strand, in the middle of the HNF5 recognition sequence, leaving a 3′-protruding 3 nucleotide-long stagger (Figure 7). *In vivo* footprinting analysis shows that the pattern of DNA cleavage by DNAase I on occupied HNF5 binding sites is identical to that existing *in vitro* (G.R., T.G., J.R. and R.P.; in preparation). Usually DNAase I cleaves DNA on phosphates facing one another across the minor groove (47). When DNA is set on a nucleosome, DNAase I cleaves the accessible phosphodiester bonds with a periodicity of 10 to 11 nucleotides on each strand generating a 3′ protruding 2 nucleotide-long stagger (48). Structural irregularities in the nucleosomal structure may lead to a modified cutting pattern at particular positions (49, 50) which could be due to a distortion of the DNA helix. By analogy, the DNA helix could be locally distorted upon HNF5 binding. Distortion of the target DNA may participate in the function of HNF5, for example by affecting the interaction of multiple factors binding to surrounding sequences.

In several instances, (once in HSIII, twice in HSV, once in the hepatitis B virus enhancer and once in the OTC enhancer) the HNF5 binding sites overlap C/EBP binding sites with a strict conservation of the overlap structure (Figure 10). This suggests that, on these sites, there is a specific interaction between the two factors involving either simultaneous binding or mutual exclusion. The existence of overlaps between HNF5 and other factors binding sites may be a general property of HNF5. Indeed there is also an overlap between an HNF5 and a glucocorticoid receptor binding site (GRE3) in HSIII. Another overlap, with an identical configuration, can be found in the mouse albumin enhancer at position −10,066 to −10,057, but in this case it is not known if the glucocorticoid receptor interacts with this region. Up to now we have identified only one HNF5 binding site, at position −5,317 to −5,302 in the *TAT* gene, which seems not to be shared with other nuclear factors.

The exact role of C/EBP and HNF5 in the function of the *TAT* gene GRU is unclear. C/EBP is able to activate gene expression from several promoters (38 and references therein). One of the C/EBP binding sites identified here cooperatively enhances the activity of one of the GRBSs of HSIII (37). HNF5 is presumably a transcriptional activator because a multimerized HNF5 binding site activates gene expression in a cell type-specific way (data not shown). *In vivo* footprinting reveals that the interaction of HNF5 with its target sites in HSIII is dependent upon the presence of glucocorticoid hormone (G.R., T.G., J.R. and R.P.; in preparation). Together these data suggest that these factors participate positively in the glucocorticoid-dependent activation of *TAT* gene transcription. Their absence from fibroblasts could explain the inactivity of *TAT* gene GRUs in these cells.

Promoters and enhancers of liver-specific genes are made up of multiple binding sites for ubiquitous and liver-specific nuclear factors such as C/EBP and HNF1 to 4 (46 and references therein). We have observed that many regulatory elements of liver-specific genes contain HNF5 binding sites which are located close to those of one or more of the aforementioned liver-specific factors. These observations reinforce the notion that the liver-specific function of these regulatory elements is achieved through a complex interplay of several different liver-specific factors (16, 46). Such participation of numerous trans-acting factors may allow the integration of multiple regulatory inputs by a single modular element. Indeed, the level of the active form of each trans-acting factor can be under discrete regulation, and variation in this level could affect the overall activity of the regulatory element. For example the intensity of the glucocorticoid stimulation exerted by the *TAT* gene GRUs could be modulated, not only by the level of activated glucocorticoid receptor, but also by the level of active C/EBP and HNF5 that could be under the control of various intracellular or extracellular stimuli.

Note added

A cDNA coding for an ubiquitous C/EBP like DNA-binding protein has just been isolated (57).

ACKNOWLEDGEMENTS

We thank S. McKnight for the generous gift of purified C/EBP, F. Strauss for helpful practical advise, A. Prunell for stimulating discussions, C. Dubucs and N. Desban for support in oligonucleotide synthesis and cell culture respectively, J. Oddos

and E. Droucheau for characterization of HNF5 activity in tissue nuclear extracts, M. Lakich and A.L. Haenni for critical reading of the manuscript. This work was supported in part by grants from the A. R. C. and the Ligue Nationale contre le Cancer. G.R. was supported by a fellowship from the Ligue Nationale contre le Cancer.

REFERENCES

1. Green, S. and Chambon, P. (1988) *T.I.G.* **4,** 309–314.
2. Beato, M. (1989) *Cell* **56,** 335–344.
3. Zenke, M., Grundström, T., Matthes, H., Wintzerith, M., Schatz, C., Wildeman, A. and Chambon, P. (1986) *EMBO J.* **5,** 387–397.
4. Dynan, W. S. (1989) *Cell* **58,** 1–4.
5. Fromental, C., Kanno, M., Nomiyama, H. and Chambon, P. (1988) *Cell* **54,** 943–953.
6. Schüle, R., Muller, M., Otsuka-Murakami, H. and Renkawitz, R. (1988) *Nature* **332,** 87–90.
7. Granner, D. K. and Hargrove, J. L. (1983) *Mol. Cell. Bioch.* **53/54,** 113–128.
8. Grange, T., Roux, J., Rigaud, G. and Pictet, R. (1989) *Nucl. Acids Res.* **17,** 8695–8709.
9. Jantzen, H. M., Strähle, U., Gloss, B., Stewart, F., Schmid, W., Boshart, M., Miksicek, R. and Schütz, G. (1987) *Cell* **49,** 29–38.
10. Becker, P., Renkawitz, R. and Schütz, G. (1984) *Embo J.* **3,** 2015–2020.
11. Grange, T., Roux, J., Fromont-Racine, M. and Pictet, R. (1989) *Exptl. Cell Res.* **180,** 220–233.
12. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B. and Erlich, H. A. (1988) *Science* **239,** 487–491.
13. Prost, E. and Moore, D. D. (1986) *Gene* **45,** 107–111.
14. Sanger, F., Nicklen, S. and Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **12,** 5463–5467.
15. Gorski, K., Carneiro, M. and Schibler, U. (1986) *Cell* **47,** 767–776.
16. Lichsteiner, S., Wuarin, J. and Schibler, U. (1987) *Cell* **51,** 963–973.
17. Strauss, F. and Varshavsky, A. (1984) *Cell* **37,** 889–901.
18. Raymondjean, M., Cereghini, S. and Yaniv, M. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 757–761.
19. Galas, D. J. and Schmitz, A. (1978) *Nucl. Acids Res.* **5,** 3157–3170.
20. Davidson, I., Xiao, J. H., Rosales, R., Staub, A. and Chambon, P. (1988) *Cell* **54,** 931–942.
21. Mantovani, R., Malgaretti, N., Nicolis, S., Giglioni, B., Comi, P., Cappellini, N., Bertero, M. T., Caligaris-Cappio, F. and Ottolenghi, S. (1988) *Nucl. Acids Res.* **16,** 4299–4313.
22. Landschulz, W. H., Johnson, P. F., Adashi, E. Y., Graves, B. J. and McKnight, S. L. (1988) *Genes & Dev.* **2,** 786–800.
23. Dorn, A., Bollekens, J., Staub, A., Benoist, C. and Mathis, D. (1987) *Cell* **50,** 863–872.
24. Johnson, P. F., Landschulz, W. H., Graves, B. J. and McKnight, S. L. (1987) *Genes & Dev.* **1,** 133–146.
25. Jones, K. A., Kadonaga, J. T., Rosenfeld, P. J., Kelly, T. J. and Tjian, R. (1987) *Cell* **48,** 79–89.
26. Birkenmeier, E. H., Gwynn, B., Howard, S., Jerry, J., Gordon, J. I., Landschulz, W. H. and McKnight, S. L. (1989) *Genes & Dev.* **3,** 1146–1156.
27. Friedman, A. D., Landschulz, W. H. and McKnight, S. L. (1989) *Genes & Dev.* **3,** 1314–1322.
28. Ben-Levy, R., Faktor, O., Berger, I. and Shaul, Y. (1989) *Mol. Cell. Biol.* **9,** 1804–1809.
29. Herbst, R. S., Friedman, N., Darnell, J. E. and Babiss, L. E. (1989) *Proc. Natl. Acad. Sci. USA* **86,** 1553–1557.
30. Monaci, P., Nicosia, A. and Cortese, R. (1988) *EMBO J.* **7,** 2075–2087.
31. Brunel, F., Ochoa, A., Schaeffer, E., Boissier, F., Guillou, Y., Cereghini, S., Cohen, G. N. and Zakin, M. M. (1988) *J. Biol. Chem.* **263,** 10180–10185.
32. Schaeffer, E., Boissier, F., Py, M. C., Cohen, G. N. and Zakin, M. M. (1989) *J. Biol. Chem.* **264,** 7153–7160.
33. Murakami, T., Nishiyori, A., Takiguchi, M. and Mori, M. (1990) *Mol. Cell. Biol.* **10,** 1180–1191.
34. Singh, H., Sen, R., Baltimore, D. and Sharp, P. A. (1986) *Nature* **319,** 154–158.
35. Martinez, E., Givel, F. and Wahli, W. (1987) *EMBO J.* **6,** 3719–3727.
36. Schüle, R., Muller, M., Kaltschmid, C. and Renkawitz, R. (1988) *Science* **242,** 1418–1420.
37. Strähle, U., Schmid, W. and Schütz, G. (1988) *Embo J.* **7,** 3389–3396.
38. McKnight, S. L., Lane, M. D. and Gluecksohn-Waelsh, S. (1989) *Genes & Dev.* **3,** 2021–2024.
39. Mercurio, F. and Karin, M. (1989) *EMBO J.* **8,** 1455–1460.
40. Chiu, R., Angel, P. and Karin, M. (1989) *Cell* **59,** 979–986.
41. Ziff, E. B. (1990) *TIG* **6,** 69–72.
42. Courtois, G., Morgan, J. G., Campbell, L. A., Fourel, G. and Crabtree, G. R. (1987) *Science* **238,** 688–692.
43. Hardon, E. M., Frain, M., Paonessa, G. and Cortese, R. (1988) *EMBO J.* **7,** 1711–1719.
44. Frain, M., Swart, G., Monaci, P., Nicosia, A., Stämpfli, S., Frank, R. and Cortese, R. (1989) *Cell* **59,** 145–157.
45. Li, Y., Shen, R. F., Tsai, S. Y. and Woo, S. L. (1988) *Mol. Cell. Biol.* **8,** 4362–4369.
46. Costa, R. H., Grayson, D. R. and Darnell, J. E. (1989) *Mol. Cell. Biol.* **9,** 1415–1425.
47. Drew, H. R. and Travers, A. A. (1984) *Cell* **37,** 491–502.
48. Sollner-Webb, B. and Felsenfeld, G. (1977) *Cell* **10,** 537–547.
49. Lutter, L. C. (1979) *Nucl. Acids Res.* **6,** 41–56.
50. Prunell, A. (1983) *Biochemistry* **22,** 4887–4894.
51. Oddos, J., Grange, T., Carr, K. D., Matthews, B., Roux, J., Richard-Foy, H. and Pictet, R. (1989) *Nucl. Acids Res.* **17,** 8877–8878.
52. Maxam, A. M. and Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
53. Graves, B. J., Johnson, P. J. and McKnight, S. L. (1986) *Cell* **44,** 565–576.
54. Grayson, D. R., Costa, R. H., Xanthopoulos, K. G. and Darnell, J. E. (1988) *Mol. Cell. Biol.* **8,** 1055–1066.
55. Costa, R. J., Grayson, D. R., Xanthopoulos, K. G. and Darnell, J. E. (1988) *Proc. Natl. Acad. Sci. USA* **85,** 3840–3844.
56. Costa, R. H., Lai, E., Grayson, D. R. and Darnell, J. E. (1988) *Mol. Cell. Biol.* **8,** 81–90.
57. Roman, C., Platero, J. S., Shuman, J. and Calame, K. (1990) *Genes & Dev.* 4, 1404–1415.