The rat hepatic leukemia factor (HLF) gene encodes two transcriptional activators with distinct circadian rhythms, tissue distributions and target preferences

E.Falvey, F.Fleury-Olela and U.Schibler[1]

Department of Molecular Biology, University of Geneva,
30, Quai Ernest Ansermet, CH-1211 Geneva, Switzerland

[1]Corresponding author

Hepatic leukemia factor (HLF) is a member of the PAR family of transcription regulatory proteins. We have characterized the rat HLF gene and studied its expression and activity. The rat HLF gene is transcribed from two alternative promoters, α and β, with different circadian amplitudes and tissue specificities. The α RNA isoforms produce a 43 kDa protein, HLF43, abundant in brain, liver and kidney, like the previously described human HLF RNA. The β RNA HLF isoforms use a CUG codon to initiate translation of a novel 36 kDa protein, HLF36, which is shorter at its N-terminus relative to the 43 kDa form. HLF36 is expressed uniquely in the liver, where it is the most abundant HLF protein. Surprisingly, the two proteins accumulate in the liver with different circadian amplitudes and have distinct liver-specific promoter preferences in transfection experiments. Thus, HLF43 stimulates transcription from the cholesterol 7α-hydroxylase promoter much more efficiently than from the albumin promoter, while the converse is true for HLF36.
Keywords: bZIP proteins/circadian rhythm/leukemia/transcription factors

Introduction

The hepatic leukemia factor (HLF) was originally isolated due to its presence in chimeric transcripts in certain human acute lymphoblastic leukemias (Hunger *et al.*, 1992; Inaba *et al.*, 1992). These chimeric transcripts resulted from a chromosomal translocation in which sequences of the HLF gene encoding the DNA binding and dimerization domains were fused with sequences of the E2A gene encoding the N-terminal transactivation domain of the basic helix–loop–helix proteins E12/E47. Although normal HLF mRNA was detected in the liver, kidney and lung of normal adult tissues (Hunger *et al.*, 1992; Inaba *et al.*, 1992), nothing is yet known about its protein expression *in vivo*, or its physiological function and regulation in either normal or transformed cells.

HLF is a member of the PAR family of transcription factors, so named because of a common proline- and acid-rich domain (for review, see Lavery and Schibler, 1994). In addition to HLF, the PAR family also includes the albumin D-site binding protein (DBP; Mueller *et al.*, 1990), the thyrotroph embryonal factor (TEF; Drolet *et al.*, 1991) and the vitellogenin gene binding protein (VBP),

the putative chicken TEF ortholog (Iyer *et al.*, 1991). All three factors, DBP, HLF and TEF, accumulate in liver cell nuclei (Mueller *et al.*, 1990; P.Fonjallaz and U.Schibler, in preparation) and may therefore participate in hepatocyte-specific gene expression. A hallmark of DBP is its robust circadian expression pattern in the liver, where DBP RNA and protein accumulate to remarkably higher levels in the evening than in the morning (Wuarin and Schibler, 1990), and thus may influence circadian variations in liver function (Wuarin *et al.*, 1992).

PAR proteins comprise a subfamily of the basic leucine zipper proteins (bZIP), a major family of transcription factors characterized by a positively charged α-helical region involved in DNA binding and an adjacent amphipathic α-helical region that allows homodimer formation or dimerization with other related bZIP proteins (Kouzarides and Ziff, 1988; Landschulz *et al.*, 1988). This choice of dimerization partners by bZIP proteins can influence the selection of DNA binding sites (Busch and Sassone-Corsi, 1990; Jones, 1990). All PAR family members can form stable homo- and heterodimers with each other, and no dimerization partners other than PAR family members have been identified so far. A likely reason for this dimerization fidelity is that the leucine zippers of all three PAR proteins contain charged amino acids at consecutive e and g α-helical positions, using the nomenclature whereby the seven amino acids of each α-helical repeat are designated by the letters a–g (O'Shea *et al.*, 1992; Vinson *et al.*, 1993). These charged amino acids at e and g positions potentially can establish eight electrostatic interactions between the dimerized coiled-coils. In addition, DNA binding site preference of bZIP proteins is established by amino acid side chains in the α-helical basic region that interact with nucleotides in a DNA recognition site (O'Shea *et al.*, 1991). Due to their highly conserved basic regions, all PAR proteins recognize identical or nearly identical binding sites *in vitro* (Drolet *et al.*, 1991; Hunger *et al.*, 1994b; E.Falvey, L.Marcacci and U.Schibler, in preparation).

This similarity in binding and dimerization properties among PAR family members raises the question of how these factors discriminate between their respective targets. Alternatively, they may represent a transcription factor subfamily with overlapping physiological roles. However, even weakly cooperative interactions with other sequence-specific or general transcription factors may significantly influence the selection of target sites. In addition, other less characterized elements within or near the trans-activation domain may contribute to target specificity. Such protein–protein interactions may allow PAR family members with identical *in vitro* binding specificities to regulate different target genes *in vivo*.

In order to gain insight into the expression and function of HLF, we have examined its role in tissue-specific and
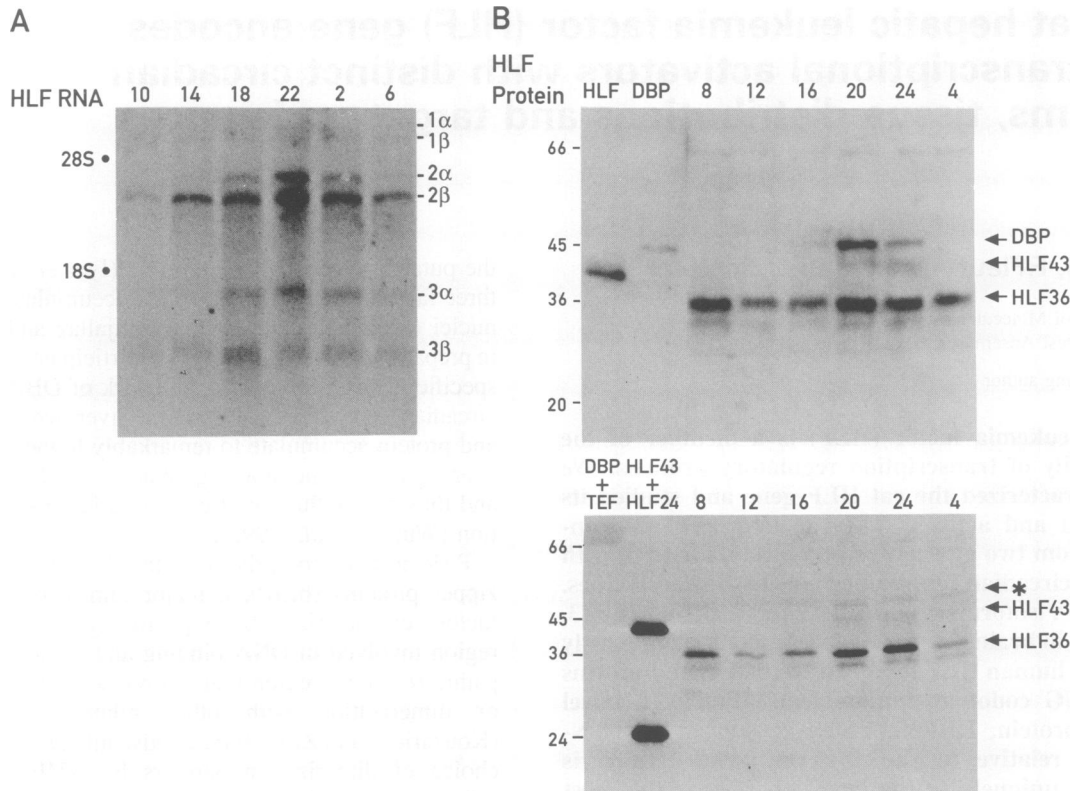
Fig. 1. Circadian accumulation of HLF RNA and proteins in the rat liver. (**A**) Northern blot analysis of poly(A)$^+$ RNA (10 µg) isolated from rat liver nuclei: autoradiogram of membrane hybridized with an hHLF cDNA probe. Rats were sacrificed at 4 h intervals for 24 h, as indicated above each lane. The position of rRNAs are shown on the left and the positions of HLF RNA isoforms are shown on the right. (**B**) Western blots. Equivalent amounts of liver nuclear extracts from rats sacrificed at 4 h intervals. Upper panel: Western blot incubated with α-hHLF serum. Lanes labeled HLF and DBP contain 10 ng of recombinant HLF (human) or recombinant DBP, respectively. Lower panel: Western blot of identical samples incubated with α-hHLF serum that had been depleted of DBP and TEF epitopes. DBP+TEF, a mixture (10 ng each) of recombinant DBP plus recombinant TEF. HLF43+HLF24, a mixture (10 ng each) of recombinant HLF43 (human) plus recombinant HLF24 (rat). The position of molecular weight size standards is shown on the left in kilodaltons. The protein migrating at 46 kDa (∗) on this blot is most likely non-specific cross-reactivity and not DBP, since this serum does not react to recombinant DBP and this band does not show the characteristic DBP cycle (see Materials and methods).

circadian gene expression in the rat. We describe several unique features of HLF gene expression and activity that distinguish HLF amongst the PAR family members. The HLF gene specifies six mRNAs that can be translated into at least four different proteins; the translation of two of these is initiated using a CUG codon. The two major proteins detected in the liver, HLF43 and HLF36, differ in their N-termini by the presence or absence of 49 amino acids. Interestingly, these proteins accumulate with different circadian amplitudes and have distinct tissue distributions. Moreover, they appear to have different promoter preferences since they differentially activate two liver-specific circadian promoters, albumin and cholesterol 7α-hydroxylase, linked to reporter genes in transfection experiments.

Results

HLF mRNAs and proteins display a distinctive circadian rhythm

DBP, the founding member of the PAR family, exhibits a striking circadian expression pattern with at least 100-fold higher levels of its mRNA at its 6 p.m. peak than at its 6 a.m. minimum (Wuarin and Schibler, 1990). DBP protein expression follows accordingly, with a peak at 8 p.m. and

a trough at 8 a.m. In order to examine the possibility that other PAR family members might be regulated in a similar manner, we analyzed HLF mRNA isolated from adult rat liver at 4 h intervals throughout the day by Northern blot hybridization with a human HLF (hHLF) cDNA probe (Figure 1A). Indeed, the mRNA displays a marked circadian rhythm profile, with a maximum at 10 p.m. and a minimum at 10 a.m. Intriguingly, the HLF pattern is offset from the cycle of DBP RNA by a delay of ~4 h. To verify that this is correct, the same membrane used in the experiment in the upper panel was stripped and hybridized with a DBP probe (data not shown).

Figure 1B shows Western blots of liver proteins extracted from nuclei at different times of the day. The predominant protein detected with anti-HLF serum is a 36 kDa species (HLF36) whose accumulation cycles with a weak amplitude, reaching peak expression levels between 8 and 12 p.m. (upper panel). A minor form migrates as 43 kDa (HLF43) and displays a more pronounced fluctuation throughout the day. Since the anti-HLF serum reacted strongly to DBP in addition to HLF, the serum was purified to remove Ig molecules reacting with epitopes common to PAR proteins other than HLF. As shown in the lower panel of Figure 1B, this DBP- and TEF-depleted serum, which does not react with recombinant DBP or TEF

protein, also detects the major 36 kDa and the minor 43 kDa proteins. (The latter is barely visible since the purified antiserum was weakened during purification by the loss of shared epitopes among the different PAR family members.)

Multiple HLF mRNAs accumulate in rat liver

In contrast to the single mRNA detected for the DBP gene, we observed a complex pattern of HLF mRNA transcripts. The hHLF gene has been reported to specify 3.5 and 4.0 kb mRNA species encoding a 43 kDa protein in liver cells (Hunger et al., 1992; Inaba et al., 1992). However, in the rat, we detect six discrete RNA species, three pairs of doublets, that migrate with apparent sizes of 6.1, 5.5, 4.2, 3.6, 1.8 and 1.2 kb (see Figure 1A). While all isoforms display the same phase of the rhythm, the upper form (α form) of each RNA doublet has a more pronounced amplitude than the lower form (β form).

Cloning and characterization of the rat HLF gene and cDNAs

In order to understand the complex expression pattern of HLF and to analyze the transcripts in greater detail, we characterized the structure of the rat HLF gene. Clones spanning the HLF gene were isolated from a rat liver genomic library using the full-length human cDNA as a probe. A map of the gene was constructed by Southern analysis with cDNA clones (first with the human cDNA, generously provided by Dr T.Look, then with the rat cDNAs described below). Restriction fragments encompassing exons and intron/exon boundaries were subcloned and sequenced. The previously described hHLF cDNA mapped to four exons, and we infer that polyadenylation occurs in the final exon, either 2.7 or 4.6 kb downstream of the termination codon. No genomic clone was obtained containing the entire gene, presumably because of the large size of the second intron (>15 kb).

The structure of the mRNAs was inferred from Northern hybridization with genomic HLF probes. A restriction fragment containing the first exon hybridized only to the upper bands of each doublet, the α RNAs (Figure 2A). To characterize the 5' ends of the three β RNAs, a 5' RACE (rapid amplification of cDNA ends) PCR was performed. Using oligonucleotide primers from within the first and second exons, two distinct PCR products were isolated. One fragment corresponded to the expected length of the transcripts initiated at a start site ~600 bp upstream of the previously defined initiator AUG (Hunger et al., 1992; Inaba et al., 1992). Cloning and sequencing of the other fragment defined a new 5' end lying within the first intron, 111 nucleotides upstream from the intron I/exon II boundary. S1 nuclease mapping analysis using liver poly(A)+ RNA and genomic exon I and exon II probes positioned the precise transcriptional start sites on the two promoters (see Figure 3). The difference between the length of the two start sites corresponds to the difference in length between the α form and the β form of each RNA doublet (~600 bp), as seen on Northern blots.

A novel 1.2 kb cDNA was isolated from a rat liver cDNA library that contains sequences corresponding to the first two exons spliced to an alternative third exon and terminating at a unique 3' end. This alternative exon was located in the genomic clone between exons II and IV
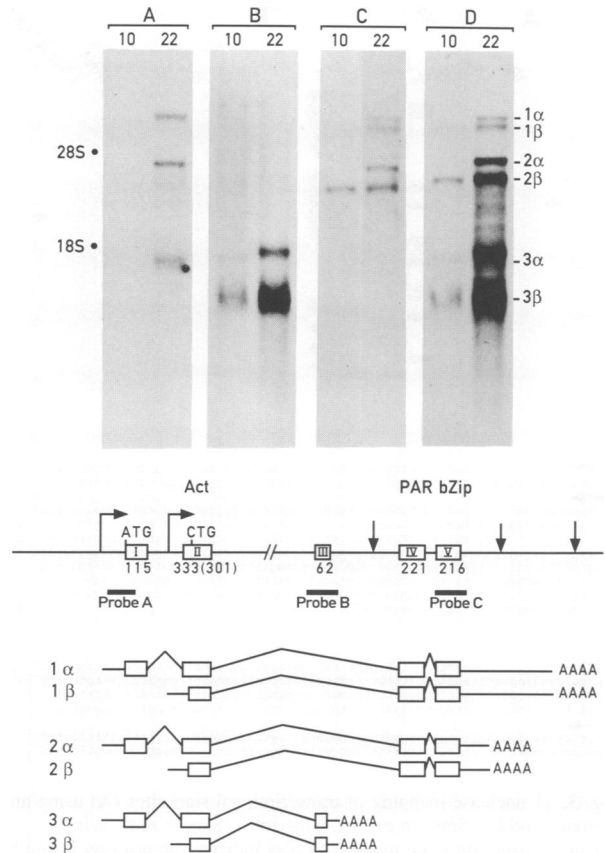

Fig. 2. Architecture of the HLF gene and its RNA transcripts. Top: Northern blot analysis of poly(A)+ RNA (10 µg) isolated from rat liver nuclei at 10 h and 22 h, probed with either (A) the first HLF exon, (B) the alternative third exon, (C) the fifth exon (the bZIP domain) or (D) the full-length hHLF cDNA. Bottom: a schematic representation of the HLF gene and its RNA transcripts. The intron/exon structure of the HLF gene is shown together with the six RNA isoforms. The location of the probes used for Northern analysis are indicated. Arrows mark the relative positions of transcriptional start and polyadenylation sites. The number of base pairs in the ORF of each exon and the positions of the initiation codons are shown. The following protein domains are identified above the corresponding exons: Act, the activation domain; PAR, the proline acid-rich domain; and bZip, the DNA binding and dimerization domain. Map distances are not to scale.

(see Figure 2, lower panel). By Northern analysis, this exon hybridizes to the 1.8 and 1.2 kb species, 3α and 3β (Figure 2B). A restriction fragment containing the fifth exon, which encompasses the bZIP domain, hybridizes to the upper four RNA species, 1α, 1β, 2α and 2β, but not 3α and 3β (Figure 2C). Thus, 3α and 3β RNAs contain sequences encoding the N-terminus of the hHLF, but lack sequences corresponding to the PAR region and the bZIP domain.

After determining the sequences of the RNAs by Northern hybridization to other genomic fragments (data not shown), we isolated the rat 2β cDNA (corresponding to 2β RNA) by PCR amplification of poly(A)+ RNA using primers at the 3' end of the final exon and at the 5' end of the RNA, as identified by RACE. 2α cDNA (the equivalent of hHLF cDNA) was assembled from 3α cDNA and 2β cDNA by subcloning (see Materials and methods).

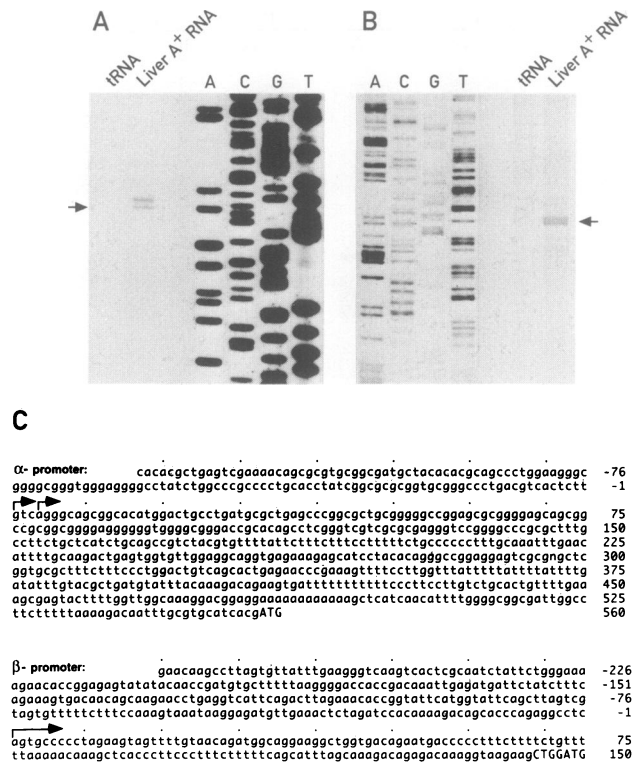A schematic map of the HLF gene and its RNA

Fig. 3. S1 nuclease mapping of transcriptional start sites (A) using an oligonucleotide primer in exon I, adjacent to Sanger sequencing reactions using the same primer. Arrows indicate the position of the 5' end. (B) The same as in (A), using an oligonucleotide primer in exon II. (C) Sequence of the α and β promoter regions.

transcripts is shown in the lower panel of Figure 2. Two promoters direct overlapping transcription units. Three distinct transcripts are produced from each promoter by alternative splicing and/or polyadenylation, yielding the six RNA isoforms. The gene has a modular structure corresponding to the modular domains of the protein. The activation region, by analogy with other PAR proteins (Drolet *et al.*, 1991; Ossipow and Schibler, in preparation) is encoded by exon II, the PAR region by exon IV and the bZIP domain by exon V.

Analysis of HLF proteins encoded by the HLF gene

The six transcripts shown in Figure 1A potentially give rise to four overlapping but distinct protein products, since 1α RNA and 2α RNA, as well as 1β RNA and 2β RNA, have identical open reading frames (ORFs). From the expression pattern and rhythmicity of RNAs and proteins, we anticipated that the markedly cyclic α RNAs would produce HLF43, while the less cyclic β RNAs would produce HLF36. 2α RNA, the rat ortholog of hHLF, has an ORF of 295 amino acids and, like the HLF cDNA isolated from HepG2 cells, yields a translation product of 43 kDa in a rabbit reticulocyte lysate (data not shown; Inaba *et al.*, 1992), considerably larger than the predicted mol. wt of 33.1 kDa calculated from its amino acid sequence. The translation product of 3α cDNA, with an ORF of 170 amino acids and a predicted size of 18.7 kDa, migrates at 24 kDa (data not shown), hence HLF24. This discrepancy between the values for observed and predicted molecular weights is similar to other PAR members and

is probably due to their high proline content (Mueller *et al.*, 1990).

The translation of the major liver HLF protein, HLF36, initiates at a CUG codon in exon 2

The predominant *in vitro* translation product of 2β RNA migrates at 36 kDa, and several minor products of 30, 23 and 18 kDa are detected (Figure 4A, lane 5). Upon inspection of the coding capacity of this transcript, the first AUG is not in-phase with the HLF ORF and potentially initiates an out-of-frame protein of 154 amino acid residues and a predicted mol. wt of 17 kDa. The first in-frame AUG would initiate an ORF of 210 amino acids, corresponding to a protein with a predicted mol. wt of 23.6 kDa. Even after compensating the predicted molecular weight for high proline content (see above), it appeared unlikely that the 36 kDa protein could be initiated at this AUG. We therefore searched the cDNA sequence for non-AUG initiation codons. Such alternative initiation codons are rare, but utilized occasionally, in eukaryotic mRNA translation. Most notably, CUG is used in the cellular proto-oncogenes *myc* (Hann *et al.*, 1988), *int-2* (Acland *et al.*, 1990) and *pim-1* (Saris *et al.*, 1991), as well as in murine retroviruses (Prats *et al.*, 1989). Indeed, a putative CUG initiation codon lies immediately upstream of the first AUG of the HLF β RNAs. Moreover, this CUG is in a favorable context for non-AUG initiation codons (Boeck and Kolakofsky, 1994; Grünert and Jackson, 1994) and could potentially initiate an in-frame ORF for a protein with a predicted mol. wt of 27.5 kDa. The translated protein would contain the bZIP domain, the PAR region and the activation domain, but would lack 49 amino acids at its N-terminus. We therefore decided to test the possibility that 2β RNA employs this CUG as an initiator codon to produce the 36 kDa protein observed in liver and in *in vitro* translations.

Translation products of a series of mutations at and around the CUG were analyzed in a rabbit reticulocyte lysate translation system (Figure 4A) and by transfection into HepG2 cells and Western blotting (Figure 4B). When the first 236 nucleotides of HLF 2β cDNA were deleted, thereby removing sequences including the CTG but not the first in-frame ATG (Mut 1), production of the 36 kDa protein was suppressed (Figure 4A, lane 1). Proteins presumably initiated at downstream AUGs were present at higher levels than from the wild-type 2β RNA. When the CTG was changed to a CTT (Mut 2), production of the 36 kDa protein was again suppressed (Figure 4A, lane 2), although downstream proteins appeared in similar quantities to the wild-type. A mutation from CTG to ATG (Mut 3) resulted in increased production of HLF36, with an accompanying decrease in the quantities of downstream products both *in vitro* (Figure 4A, lane 3) and in HepG2 cells (Figure 4B, lanes 1 and 2). This confirms that translation is indeed initiated at the CUG. The CUG is inefficient, however, thereby allowing some ribosomal 'leakage'.

Mut 2, the CTG to CTT mutation that eliminated HLF36 (Figure 4A, lane 2), produced only short products, including an 18 kDa species. Interestingly, a pair of AUGs in position +5 and +8 relative to the CUG opens an alternative reading frame that could produce a protein with a predicted mol. wt of 17 kDa. The use of this
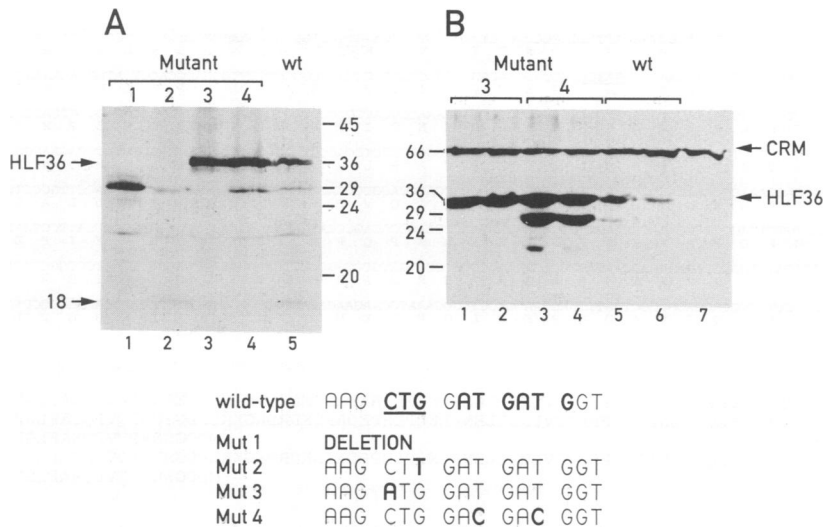
alternative frame in *in vitro* translations was demonstrated by one additional mutation, Mut 4. When the two ATGs were eliminated by replacing them with ACGs only this 18 kDa product disappeared in translation reactions (Figure 4A, lane 4). When this mutant was expressed in HepG2 cells, more HLF36 protein accumulated compared with wild-type (Figure 4B, lanes 3 and 4). In addition, shorter proteins, presumably initiated at downstream in-phase AUGs, were also observed at higher levels when compared with proteins produced from wild-type 2β mRNA (see Discussion).

The ORFs of the HLF cDNAs and their protein-coding capacities are summarized in Figure 5. The rat and human HLF43 proteins differ in only nine amino acid positions, as shown in the lower panel. The basic region and the leucine zipper are absolutely conserved and only one amino acid change is located in the PAR region—the other eight amino acid differences are located in the N-terminal region of the protein. The amino acid sequences of HLF36 and the alternative protein forms, HLF24 and HLF17, translated from 3α and 3β RNAs, respectively, are also shown.

HLF isoforms have characteristic tissue distributions

The expression profile of HLF isoforms is tissue specific, as seen in the Western blots in Figure 6A. HLF36 is detected mainly in the liver. HLF43 is most abundant in kidney, liver and brain, weakly expressed in lung, and virtually absent in spleen and testes. We did not detect HLF24 or HLF17 in any tissues examined.

The regulation of this tissue-specific expression profile is controlled at the level of RNA transcription or RNA stability, as demonstrated by Northern analysis of whole cell RNA (Figure 6B) and by S1 analysis and RNase

mapping with polyadenylated RNAs from different tissues (data not shown). HLF RNAs are highly expressed in liver, kidney and brain, and weakly in spleen, testes and pancreas. In the liver, the 1β, 2β and 3β RNAs are the predominant forms, in agreement with the high liver expression of HLF36. 1α, 2α and 3α RNAs are the more abundant forms in the kidney where HLF43 is highly expressed. The kidney HLF RNA levels are 3.5-fold higher at 10 p.m. (22 h) than 10 a.m. (10 h). 1α and 2α (but not 3α) RNA are also expressed highly in the brain, where HLF43 is also the major protein form, but there is very little difference, if any, in expression between 10 a.m. (10 h) and 10 p.m. (22 h). Although 3α and 3β RNAs are present in the liver, and 3α RNA exists in the kidney, we have never detected their respective protein products, HLF17 or HLF24. For example, the purified α-HLF serum detects the recombinant form of HLF24 in the Western blot experiment in Figure 1B, whereas no cellular counterpart is detected in the liver. This suggests that these proteins may be particularly unstable or sequestered in a cellular compartment other than the nucleus. However, we cannot rigorously exclude that, for technical reasons, these proteins escape detection by immunoblotting.

The average cell size varies greatly among the tissues we examined. Cell sizes are accompanied by corresponding changes in RNA:DNA ratios, as recently shown by Schmidt and Schibler (1995). Therefore these tissue-specific RNA:DNA ratios should be taken into account when assessing tissue differences of RNAs, particularly those encoding nuclear proteins. We have quantitated the HLF mRNA distribution by phosphorimager analysis and normalized our results to cellular equivalents with the known RNA:DNA ratios (Schmidt and Schibler, 1995). The chart in Figure 6C shows that the relative levels of HLF mRNA isoforms, after adjusting for cell size,
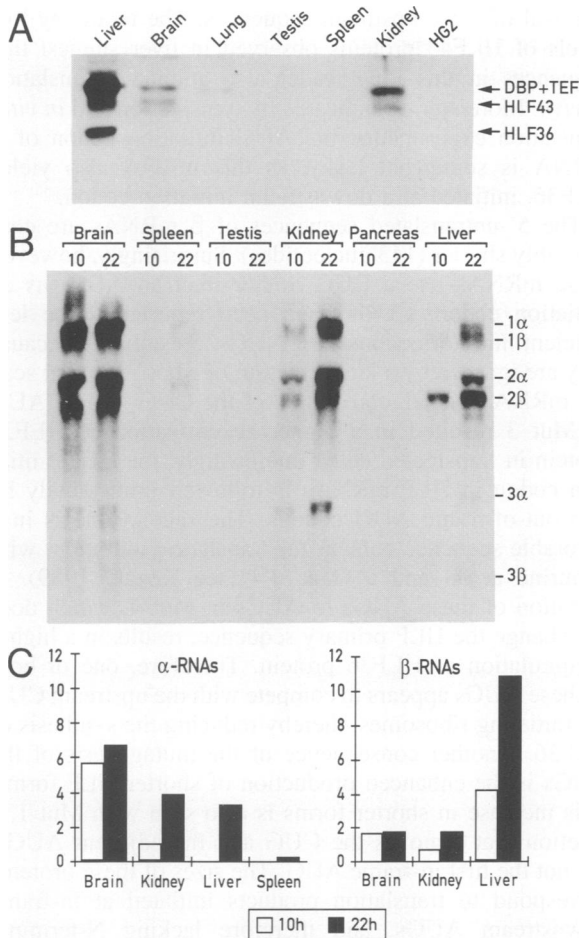
HLF43 HLF36

```
ATGGAGAAAATGTCCCGACAGCTCCCCCTGAACCCTACCTTTATCCCGCCTCCCTACGGCGTGCTCAGGTCCCTGCTGGAGAACCCGCTGAAGCTCCCCCTTCATCCTGAA    668
 M  E  K  M  S  R  Q  L  P  L  N  P  T  F  I  P  P  P  Y  G  V  L  R  S  L  L  E  N  P  L  K  L  P  L  H  P  E     37
GACGCATTTAGCAAAGACAGAGACAAAGGTAAGAAGCTGGATGATGGTAGCAACAGCCGACCGTCCCCCAGTCGGCCTTCTTGGGACCCACCTTATGGGACAAAACCCTT    779  218
 D  A  F  S  K  D  R  D  K  G  K  K  L  D  D  G  S  N  S  P  T  V  P  Q  S  A  F  L  G  P  T  L  W  D  K  T  L      74   25
CCCTATGACGGAGATACTTTCCAGTTGGAATACATGGACCTGGAGGAGTTTCTGTCAGAAAATGGCATCCCCCCTAGTCCATCGCAGCACGACCACAGCCCTCACCCTCCT    890  329
 P  Y  D  G  D  T  F  Q  L  E  Y  M  D  L  E  E  F  L  S  E  N  G  I  P  P  S  P  S  Q  H  D  H  S  P  H  P  P     111   62
GGTTTGCAACCTGCTTCCTCCACTGCTCCTTCAGTCATGGATCTCAGCAGCCGGGCCACGGCACCCCTCCACCCTGGCATCCCGTCTCCGAACTGTATGCAGAACCCCATC   1001  440
 G  L  Q  P  A  S  S  T  A  P  S  V  M  D  L  S  S  R  A  T  A  P  L  H  P  G  I  P  S  P  N  C  M  Q  N  P  I    148   99
AGACCAGGCCAGCTGTTGCCAGCAAACCGAAACACACCGAGTCCCATTGACCCCGACACCATCCAGGTCCCAGTCGGGTATGAACCAGACCCGGCAGACCTTGCCCTCTCC   1112  551
 R  P  G  Q  L  L  P  A  N  R  N  T  P  S  P  I  D  P  D  T  I  Q  V  P  V  G  Y  E  P  D  P  A  D  L  A  L  S    185  136
AGCATCCCGGGGCCAGAAATGTTTGACCCTCGCAAACGGAAGTTCTCCGAGGAAGAACTGAAGCCACAGCCCATGATTAAGAAAGCTCGCAAAGTCTTCATTCCTGATGAT   1223  662
 S  I  P  G  P  E  M  F  D  P  R  K  R  K  F  S  E  E  E  L  K  P  Q  P  M  I  K  K  A  R  K  V  F  I  P  D  D    222  173
TTGAAGGATGACAAGTACTGGGCGAGGCGCAGAAAGAACAATATGGCGGCCAAGCGCTCCCGTGATGCCCGGCGGCTGAAGGAGAACCAGATCGCAATCCGGGCCTCATTC   1334  773
 L  K  D  D  K  Y  W  A  R  R  R  K  N  N  M  A  A  K  R  S  R  D  A  R  R  L  K  E  N  Q  I  A  I  R  A  S  F    259  210
CTGGAGAAGGAGAACTCGGCCCTCCGCCAGGAGGTGGCTGATTTAAGGAAGGAGCTGGGCAAATGCAAGAACATACTTGCCAAGTACGAGGCCAGGCACGGGCCCCTGTA   1445  884
 L  E  K  E  N  S  A  L  R  Q  E  V  A  D  L  R  K  E  L  G  K  C  K  N  I  L  A  K  Y  E  A  R  H  G  P  L  .    295  246
```

```
human HLF43: MEKMSRPLPLNPTFIPPPYGVLRSLLENPLKLPLHHEDAFSKDKDKEKKLDDESNSPTVPQSAFLGPTLWDKTL    74
  rat HLF43: MEKMSRQLPLNPTFIPPPYGVLRSLLENPLKLPLHPEDAFSKDRDKGKKLDDGSNSPTVPQSAFLGPTLWDKTL    74
  rat HLF36:                                                 MDDGSNSPTVPQSAFLGPTLWDKTL    25
  rat HLF24: MEKMSRQLPLNPTFIPPPYGVLRSLLENPLKLPLHPEDAFSKDRDKGKKLDDGSNSPTVPQSAFLGPTLWDKTL    74
  rat HLF17:                                                 MDDGSNSPTVPQSAFLGPTLWDKTL    25


             PYDGDTFQLEYMDLEEFLSENGIPPSPSQHDHSPHPPGLQPASSAAPSVMDLSSRASAPLHPGIPSPNCMQSPI   148
             PYDGDTFQLEYMDLEEFLSENGIPPSPSQHDHSPHPPGLQPASSTAPSVMDLSSRATAPLHPGIPSPNCMQNPI   148
             PYDGDTFQLEYMDLEEFLSENGIPPSPSQHDHSPHPPGLQPASSTAPSVMDLSSRATAPLHPGIPSPNCMQNPI    99
             PYDGDTFQLEYMDLEEFLSENGIPPSPSQHDHSPHPPGLQPASSTAPSVMDLSSRATAPLHPGIPSPNCMQNPI   148
             PYDGDTFQLEYMDLEEFLSENGIPPSPSQHDHSPHPPGLQPASSTAPSVMDLSSRATAPLHPGIPSPNCMQNPI    99

                                         PAR Region
             RPGQLLPANRNTPSPIDPDTIQVPVGYEPDPADLALSSIPGQEMFDPRKRKFSEEELKPQPMIKKARKVFIPDD   222
             RPGQLLPANRNTPSPIDPDTIQVPVGYEPDPADLALSSIPGPEMFDPRKRKFSEEELKPQPMIKKARKVFIPDD   222
             RPGQLLPANRNTPSPIDPDTIQVPVGYEPDPADLALSSIPGPEMFDPRKRKFSEEELKPQPMIKKARKVFIPDD   173
             RPESSLALLQSVTMRPIEEYDA                                                      170
             RPESSLALLQSVTMRPIEEYDA                                                      121

                Basic Region                         Leucine Zipper
             LKDDKYWARRRKNNMAAKRSRDARRLKENQIAIRASFLEKENSALRQEVADLRKELGKCKNILAKYEARHGPL   295
             LKDDKYWARRRKNNMAAKRSRDARRLKENQIAIRASFLEKENSALRQEVADLRKELGKCKNILAKYEARHGPL   295
             LKDDKYWARRRKNNMAAKRSRDARRLKENQIAIRASFLEKENSALRQEVADLRKELGKCKNILAKYEARHGPL   246
```

Fig. 5. Sequence of ORF and the proteins encoded by the rat HLF gene. Upper panel: HLF43 cDNA ORF. The sequences involved in translation initiation of HLF36 are underlined. The positions of splice junctions are marked with double arrowheads. Lower panel: comparison of rat and human HLF proteins. Positions where the rat HLF amino acid sequence differs from human HLF are in bold type.

correspond reasonably well with the levels of HLF36 and HLF43 proteins observed on Western blots in brain, liver, kidney and spleen. This suggests that the expression of HLF protein, like that of DBP, may be regulated predominantly by RNA accumulation rates in these tissues.

HLF36 and HLF43 have different activation targets in HepG2 cells

To examine the activation potentials of HLF36 and HLF43, we studied their activities using the albumin and cholesterol 7α-hydroxylase (C7αH) promoters linked to reporter genes in co-transfection experiments. These promoters were chosen because both contain recognition sequences for PAR proteins, and transcription from both promoters was stimulated by DBP in co-transfection experiments (Mueller et al., 1990; Lavery and Schibler, 1993). The transcription of both the albumin (Wuarin and Schibler, 1992) and C7αH genes is circadian (Noshiro et al., 1990; Lavery and Schibler, 1993), and highly restricted to the liver (Tilghman and Belayew, 1982; Jelinek et al., 1990) which accumulates both HLF36 and HLF43.

Reporter plasmids, harboring either the albumin or C7αH promoter, were transfected into human hepatocarcinoma (HepG2) cells with or without the expression vectors pCMV-HLF43 or pCMV-HLF36, using pSV2CAT (Gorman et al., 1982) as a positive control. A luciferase expression vector was used as an internal control to normalize samples for transfection efficiency. As shown in Figure 7, co-transfection with CMV-HLF43 results in

very strong activation of the C7αH promoter, even at the lowest levels of this expression vector. In contrast, cotransfection of CMV-HLF36 with this reporter construct yields only a relatively weak activation. Conversely, cotransfection with CMV-HLF36 strongly stimulates CAT expression from the albumin promoter, whereas co-transfection with CMV-HLF43 results in very weak activation of the alb–CAT fusion. In this case, 1 μg of CMV-HLF36 resulted in higher CAT activity than did 5 μg of this plasmid (compare lanes 10 and 11), presumably due to a 'squelching' mechanism in the presence of excessive amounts of this strong activator.

Nuclear extracts of transfected cell nuclei were analyzed by Western blotting to monitor HLF protein expression. Surprisingly, cells transfected with CMV-HLF43 plasmids produced some HLF36 in addition to HLF43 in HepG2 cells, suggesting some ribosome 'leakage' beyond the AUG initiation codon (data not shown). As expected (see above), cells transfected with CMV-HLF36 produced predominantly HLF36. Although this makes it more difficult to distinguish between the effect of HLF43 alone and the combined effect of HLF43 and HLF36, we can make the following conclusions. Production of HLF43 is necessary (and perhaps sufficient) for activation of the C7αH promoter. Although HLF36 is impotent on this promoter, it does not impede the stimulation of transcription by HLF43. Conversely, HLF36 is a much more potent activator of the albumin promoter than HLF43 and HLF36 together.

Fig. 6. Tissue distribution of HLF isoforms in the rat. (**A**) Western blot analysis of lysates from an equal amount of nuclei of various tissues, or HepG2 cells, HG2. (**B**) Northern blot analysis of total RNAs from various tissues. (**C**) Relative distribution of RNA isoforms in different tissues, corrected for tissue-specific RNA:DNA ratios. RNA:DNA ratios used were: liver, 4.71; kidney, 1.24; brain, 1.44; spleen, 0.42.

Fig. 7. Activation of C7αH and albumin promoter–CAT fusion constructs by HLF43 or HLF36 in co-transfected HepG2 cells. (**A**) A representative CAT analysis of cells co-transfected with the following: pSV2CAT, which expresses the CAT gene under the control of the SV40 promoter/enhancer (lane 1); CH–CAT (lane 2); CH–CAT with 1 μg (lane 3) or 5 μg (lane 4) of pCMV-HLF43; CH–CAT with 1 μg (lane 5) or 5 μg (lane 6) of pCMV-HLF36; alb–CAT (lane 7); alb–CAT with 1 μg (lane 8) or 5 μg (lane 9) of pCMV-HLF43; alb–CAT with 1 μg (lane 10) or 5 μg (lane 11) of pCMV-HLF36. (**B**) Graphic representation of relative activations. Lanes as defined in (**A**). Values were derived as described in Materials and methods. (**C**) Co-transfection constructs. pCMV-HLF43 and pCMV-HLF36 (see Materials and methods); –340 CH–CAT (Lavery and Schibler, 1993), Alb–CAT (Mueller *et al.*, 1990).

Discussion

Three members of the PAR subfamily of bZIP proteins have thus far been characterized: DBP (Mueller *et al.*, 1990), TEF/VBP (Drolet *et al.*, 1991; Iyer *et al.*, 1991) and HLF (Hunger *et al.*, 1992; Inaba *et al.*, 1992). All of these factors accumulate to high levels in liver cell nuclei and may therefore play key roles in liver-specific gene expression. In this study, we describe several unique features of HLF expression and activity that distinguish HLF amongst the PAR family members. The HLF gene is transcribed from two alternative promoters, α and β, with different cell type specificities. Most interestingly, the HLF mRNAs initiated at these two promoters accumulate according to circadian rhythms with different amplitudes and encode transcriptional activators with different target gene specificities in co-transfection experiments.

We have several reasons to believe that the observations described here in the rat may be paralleled in other species, such as man. Inaba and co-workers (1992) have identified two hHLF mRNAs of 3.5 and 4.0 kb in normal human liver cells, although only a single 4.0 kb mRNA was detected in kidney cells, similar to the expression of the rat α and β mRNAs. The rat gene is very similar in

structure to the human gene, and the sequences at the intron/exon junctions are highly conserved, even when they differ considerably from consensus splice site sequences (E.Falvey and U.Schibler, unpublished data; Hunger *et al.*, 1994b). At the nucleotide level, there is 92% homology between the rat 2α and the human cDNAs, with most differences occurring at third codon positions. Indeed, the CUG and the neighboring sequences important for translation initiation of the β RNAs (Figure 4; Boeck and Kolakofsky, 1994; Grünert and Jackson, 1994) are also strictly conserved between the two species.

Circadian expression of HLF isoforms

DBP, the founding member of the PAR family, was shown to accumulate according to a robust circadian rhythm (Wuarin and Schibler, 1990). As shown here, the HLF gene is also expressed in a circadian fashion. Interestingly, the transcripts initiated at the upstream α promoter exhibit a much higher circadian amplitude than transcripts initiated at the downstream β promoter (Figure 1A). These different

amplitudes are clearly reflected in the proteins issued from the α and β mRNAs. Thus, HLF43 levels oscillate much more strongly during the day than do HLF36 levels. The amplitude of cyclic RNA and protein accumulation depends on both the rate of synthesis and the stability of these macromolecules (Wuarin et al., 1992). While we do not know the transcription initiation rates at the two promoters and the half-lives of the resulting transcripts, the synchronous phases of HLF α and β mRNA accumulation can be explained more readily by postulating a concerted transcriptional mechanism. Conceivably, the synthesis of α and β mRNAs is governed by a single circadian mechanism, such as an enhancer regulating both promoters. If so, these transcripts would be unlikely to have dramatically different half-lives, since this would lead to asynchronous accumulation of α and β transcripts (for mathematical model, see Wuarin et al., 1992). More probably, the α promoter is more strongly dependent upon the putative circadian enhancer than is the β promoter.

Interestingly, the peak of circadian HLF mRNA accumulation is delayed by ~2 h with respect to maximal DBP protein expression in the rat. It is conceivable, therefore, that DBP is involved in controlling the daily cycle of HLF expression. This would also be consistent with the concomitant tissue distribution of HLF and DBP. Both transcription factors are observed in liver, kidney and brain, but are nearly absent in lung, spleen and testis. Moreover, in brain, where DBP is expressed throughout the day and cycles with a low amplitude, HLF mRNA levels are constitutively high and do not fluctuate significantly during the day. Given the different phases of DBP and HLF accumulation in liver and kidney, it is tempting to speculate that these transcriptional regulatory proteins fulfil different physiological roles. They might either sustain transcription of the same target genes during different time periods or regulate the activity of different target genes whose products are required at different times.

The cyclic accumulation of HLF, like that of DBP (Wuarin and Schibler, 1990), is free running; that is it does not depend on external time cues once the circadian cycle has been entrained (data not shown). Thus, the rhythmic expression of PAR proteins appears to be an output of the circadian clock or a component of the clock itself. In mammals, the circadian pacemaker resides in a cluster of neurons within the suprachiasmatic nucleus (SCN) of the hypothalamus (for review, see Moore, 1992). Circadian rhythms are entrained by light signals that are transmitted directly via the optic nerve from photoreceptor cells in the retina to neurons in the SCN. The SCN, in turn, influences expression of a variety of endocrine systems, including the pineal gland and the major output route, the hypothalamus adrenal pituitary axis (as reviewed in Krieger, 1979). It is therefore likely that the rhythmic expression of HLF is governed directly or indirectly by cyclic hormones, such as glucocorticoids, melatonin and thyroid hormones.

Translation of HLF isoforms

In vitro, translation of HLF43 and HLF36 from full-length α and β mRNAs, respectively, is inefficient (data not shown). α mRNAs harbor a GC-rich 560 nucleotide 5'-untranslated region. In reticulocyte translation extracts, detectable amounts of proteins are only obtained after removal of these upstream sequences. The relatively low levels of HLF43 proteins observed in liver suggest that sequences in this long leader also attenuate translation *in vivo*. Moreover, as indicated by transfection and *in vitro* translation experiments, the AUG initiation codon of α mRNA is somewhat leaky, as this mRNA also yields HLF36, initiated at a downstream initiation codon.

The 5'-untranslated sequences of β mRNAs are considerably shorter (143 nucleotides). Surprisingly, however, these mRNAs use a CUG, rather than an AUG, as an initiation codon. CUGs have been reported to be less efficient initiator codons than AUGs, presumably because they are ignored by a large fraction of ribosomes that scan the mRNA. Indeed, conversion of the CUG into an AUG in Mut 3 resulted in a higher accumulation of HLF36 protein in transfected cells. Interestingly, the CUG initiation codon in HLF mRNAs is followed immediately by two out-of-frame AUG codons. The second one is in a favorable sequence context for translation initiation, with a purine at −3 and a G at +4 (see Kozak, 1989). A mutation of these AUGs to AUCs in Mut 4, which does not change the HLF primary sequence, results in a higher accumulation of HLF36 protein. Therefore, one or both of these AUGs appears to compete with the upstream CUG for initiating ribosomes, thereby reducing the synthesis of HLF36. Another consequence of the mutagenesis of the AUGs is the enhanced production of shorter HLF forms. This increase in shorter forms is also seen with Mut 1, a deletion that removes the CUG and the adjacent AUGs, but not the first in-frame AUG. The sizes of these proteins correspond to translation products initiated at in-frame downstream AUGs, and therefore lacking N-terminal domains. Since wild-type HLF mRNA does not produce such truncated proteins at significant levels, the out-of-frame AUGs may recruit most of the scanning ribosomes that bypass the CUG initiation codon. We do not yet know whether translation of the 462 nucleotide alternative reading frame from these out-of-frame AUGs produces a functional protein, or whether the sole function of this translation is to prevent synthesis of short HLF forms. Since such products would contain DNA binding and dimerization domains but lack activation domains, they could be antagonistic to activators (for review, see Foulkes and Sassone-Corsi, 1991). It is noteworthy to mention in this context that a leaky ribosome scanning mechanism is used to produce both activators and repressors from both C/EBP α and β mRNAs (Descombes and Schibler, 1991; Ossipow et al., 1993) while, in the case of HLF, a mechanism may exist that inhibits the production of N-terminally truncated proteins.

Translation of the 3α and 3β mRNAs would produce C-terminally truncated proteins bearing the activation domain but lacking the PAR and bZIP domains. In co-transfection experiments, such proteins attenuate activation by HLF, probably by a squelching mechanism (E.F. and U.S., unpublished data). However, we have not been able to detect the expression of such C-terminally truncated proteins in liver cells by immunoblotting. These proteins are therefore unlikely to accumulate to levels similar to those observed for HLF43 and HLF36. Nevertheless, we cannot exclude that small amounts of these C-terminally truncated 'HLF squelcher proteins' are produced *in vivo*.

In the course of this work, we encountered several

different features of the HLF gene that might down-regulate the expression and activities of HLF proteins. First, splicing and polyadenylation create an alternative third exon that reduces the proportion of transcripts that could otherwise produce functional activator proteins. Second, the putative proteins specified by these prematurely polyadenylated and alternatively spliced 3α and β mRNAs could potentially reduce the activation potential of HLF (and related factors). Third, translation initiation may be dampened by either the long GC-rich untranslated sequences (in the case of α mRNAs) or the use of an inefficient CUG initiation codon (in the case of β mRNAs). We do not yet know whether these attenuating mechanisms are physiologically meaningful. However, it is interesting to note, in this context, that CUG initiation codons have been found in cellular proto-oncogenes, such as c-myc (Hann et al., 1988), whose aberrant expression has been associated with the transformed phenotype. The involvement of HLF in human leukemias may suggest that expression of HLF, like c-myc, must be tightly regulated to maintain proper growth and development. An additional possibility is that multiple levels of regulation are required to effect the complex circadian expression pattern of HLF. Moreover, it appears reasonable to postulate that the concentrations of circadian transcription factors would have to be rate limiting in order to ensure the cyclic expression of downstream genes.

HLF36 and HLF43 target different promoters

HLF36 and HLF43 share the DNA binding and dimerization domains and bind identical DNA sequences *in vitro*, yet these two proteins have different activation potentials for the promoters of the C7αH gene and the albumin gene. Apparently, complex interactions in the context of a natural promoter element, such as the ones chosen for the present study, are required to distinguish between the activities of HLF36 and HLF43. Protein–protein interactions of N-terminal regions with the transcription machinery or other sequence-specific transcription factors may therefore be necessary to guide the HLF isoforms to their appropriate destination. In contrast to these studies using natural promoters, recent transfection studies with artificial promoters containing simple or multimerized PAR binding sites have not detected strong differences in activation potentials of PAR proteins, despite the use of radically different activation domains, such as E2A–HLF and HLF (Hunger et al., 1994a; Inaba et al., 1994) or two VBP fusion proteins with different N-terminal domains (Burch and Davis, 1994). Binding studies with hHLF and the chimeric fusion protein E2A–HLF suggest that these two activators, which contain identical DNA binding and dimerization domains, have identical binding activities and can perhaps only be distinguished by their preferences for subtle deviations from the consensus sequence (Hunger et al., 1994b).

At present we do not know how the 49 N-terminal residues determine the different target promoter specificities of HLF36 and HLF43. One possible mechanism is that this stretch of amino acids contains an additional activation domain making HLF43 a stronger activator than HLF36, but HLF43 is excluded from the albumin binding site by steric hindrance. Alternatively, the extra N-terminal amino acids could contact a co-activator intrinsic to the

cholesterol hydroxylase promoter. A third possibility is that the N-terminal residues induce a conformational change in HLF43 that facilitates its binding to the cholesterol hydroxylase promoter. These considerations stress that promoter geometry and contacts with ancillary factors may be as important as DNA recognition domains in directing transcription factors to their target sites.

While numerous cases have been described in which activator and repressor proteins are encoded by the same gene (e.g. C/EBPα and β, CREM; for review, see Foulkes and Sassone-Corsi, 1991), we are not aware of other activators specified by a single gene that have different target promoter preferences. Future experiments aimed at identifying the precise residues of HLF43 involved in making contacts with the transcriptional machinery at each promoter should help resolve the models proposed above and lead to a clearer understanding of the mechanism of tissue-specific activation by the PAR family of transcription factors.

Materials and methods

Cloning of the HLF gene and rat HLF cDNA isoforms

An EMBL-3 genomic library derived from rat liver (generous gift of A.Kahn) was screened with the entire hHLF cDNA (generous gift of A.T.Look). Three phages were purified to homogeneity and inserts were characterized by restriction mapping and Southern hybridization to cDNA clones. The inserts of all three clones hybridized to 3' portions of the human cDNA. Two additional phages were isolated by screening an independently prepared EMBL-3 library from rat liver genomic DNA with a probe containing only the 5' end of the rat cDNA (see below) and characterized. The largest clones from each screening, λIII₁ and λ5'₁, 16 and 18 kb respectively, were characterized in greater detail by subcloning and sequencing the exons and intron/exon boundaries. Together, these two clones contained all the sequences present in the rat HLF cDNAs. No clone was obtained that contained the entire gene. Standard techniques were used for molecular cloning and hybridization (Sambrook et al., 1989), and sequencing, using a set of primers homologous to various portions of the rat HLF cDNA and a commercially prepared modified T7 polymerase sequencing kit (Biofinex).

The rat HLF24 cDNA was isolated by screening a λgt11 cDNA library constructed from RNA isolated from 11 p.m. rat liver (generous gift of D.Lavery) with a hHLF cDNA. Three phages were purified to homogeneity, inserts were subcloned into pKS+ (Strategene), and characterized by sequencing and Southern hybridization to genomic clones.

RACE-PCR and RT-PCR

The 5' ends of β RNAs were identified using a commercially prepared 5' RACE kit (BRL-Life Technologies). An HLF-specific oligonucleotide primer (H5, 5'-GATGGGGTTCTGCATACAGTT-3'), containing sequences homologous to the exon II, which we had found to be common to all HLF mRNA species, was used as a primer for cDNA synthesis. Separate reactions contained either 10 h or 22 h liver poly(A)+ RNA. H5 and the 'anchor' primer (supplied with the kit) were used for 25 rounds of PCR amplification of the cDNA, and a second primer (H3, 5'-GTCCAGGTACATAAGGTTGACCT-3'), also lying within the second exon but internal to H5, was used with the nested 'UAP' (supplied with the kit) for a subsequent 25 rounds of PCR amplification, using one-tenth of the product of the first round as a template. The final products of these reactions ran as a collection of bands on an agarose gel of the approximately correct molecular weight of 200 bp for 10 h RNA and 200 and 650 bp for 22 h RNA. RACE products were cloned into the EcoRV site of a pKS+ T vector (Marchuk et al., 1991) and sequenced.

HLF36 cDNA was cloned by reverse transcription (RT) PCR using 1 μg of 10 h poly(A)+ liver RNA and an oligonucleotide primer (H7, 5'-ACAGATGGCAGGAAGGCT-3'), containing sequences homologous to the extreme 5' end of the 1/2α mRNA ORF for cDNA synthesis, followed by 25 rounds of PCR amplification with primer H7 and primer H6 (5'-AAAATGCCATCCTACAGG-3') that was homologous to the

extreme 3' end of the cDNA, as determined by S1 analysis. This was followed by an additional 15 rounds of amplification using one-tenth of the first-round product and an internally nested set of oligonucleotide primers for 5' and 3' ends (H8, 5'-GAAGGCTGGTGACAGAATG-3' and H71, 5'-AGCTGGATCCTACAGGGGCCCGTGCCT-3', respectively). The final products of this reaction, which ran as a single band of the expected mol. wt of 851 bp, were cloned into a pKS+ T vector (Marchuk et al., 1991) and sequenced. The correct clones, as determined by comparison with the genomic sequence, were used for subsequent cloning into eukaryotic and bacterial expression vectors (see below).

PCR mutagenesis and plasmid constructions

The rat HLF36 mutants 1–4 (see Figure 4C) were constructed by PCR mutagenesis using pKS+HLF36 as template. Deletion mutant 1 was generated by a single 25-cycle PCR amplification using primer H6 and a synthetic oligonucleotide starting at position +237 of the HLF36 cDNA (5'-TCCAGTTGGAATCCATGGACCTG-3') with 100 ng of pKS+HLF36 template. Mutants 2–4 were constructed by double PCR amplifications. For the first round, KS primer (Strategene) and a synthetic oligonucleotide containing corresponding base change(s) were used as primers in a PCR with 100 ng of pKS+HLF36 template. The 148 bp product of this reaction was isolated and purified from an agarose gel (Quiagen) and, with oligonucleotide H5, used as primers for a second PCR amplification on the same template. The four mutants were identified by cloning the PCR products into the EcoRV site of a pKS+ T vector and sequencing.

The mammalian expression vectors pSCT-hHLF43, pSCT-rHLF43, pSCT-HLF36, and the mutants pSCT-rHLF36.1–4 were constructed by inserting the following fragments into the vector pSCT-GalX556 (Rusconi et al., 1990) by standard cloning methods: for pSCT-hHLF43, ScaI–BgIII and BgIII–HinfI fragments (containing, respectively, bp 1–638 and 638–1105 of the hHLF cDNA sequence; Hunger et al., 1992) into pSCT (BamHI–PvuII); for pSCT-rHLF43, the 230 bp ScaI–Van91I fragment of pKS+HLF24 and the 734 bp Van91I–EcoRI fragment of pKS+HLF36 into pSCT (BamHI–PvuII); for pSCT-HLF36, the 866 bp HindIII–EcoRI fragment of pKS+HLF36, containing the entire ORF plus 107 bp of untranslated leader, into pSCT (HindIII–EcoRI); for pSCT-rHLF36.1–4, the 222 bp HindIII–PvuII fragment of pKS+HLF36.1 or the 420 bp HindIII–PvuII fragments of mutants pKS+HLF36.2–4 into pSCT-HLF36 (HindIII–PvuII).

The bacterial expression vector pET3c-hHLF43 was constructed by cloning the 980 bp blunted HinfI fragment of the human HLF43 cDNA into the blunted BamHI site of pET3c (Studier et al., 1990).

RNA isolation and analyses

RNA was isolated from 10 g of rat liver at 4 h intervals during a 24 h period, and from 2 g of kidney, brain, spleen, testes and pancreas at 10 and 22 h, and purified as described previously (Chomczynski and Sacchi, 1987). Poly(A)$^+$ RNA was isolated from liver RNA by two rounds of chromatography on oligo(dT)-cellulose. RNA was separated on 1.5% agarose gels containing 0.6 M formaldehyde, transferred to Nytran filters (Schleicher and Schuell) and hybridized to random-primed cDNA probes using standard methods (Sambrook et al., 1987). For consecutive hybridization to DBP RNA, the filter was stripped by incubating it for 30 min at 85°C in 10 mM Tris-HCl pH 7.4; 1 mM EDTA; 0.1% SDS, and autoradiographed to assay the signal before reprobing with an in vitro-transcribed RNA probe from full-length DBP cDNA.

For the Northern analysis of RNA from different tissues, whole cell RNAs were used. Nytran filters were incubated with a random-primed full-length HLF36 cDNA probe for 18 h at 65°C in 3× SSC; 10× Denhardt's solution; 0.1% SDS; 10% dextran sulfate; 50 μg/ml denatured salmon sperm DNA, using standard hybridization methods (Sambrook et al., 1987). The filter was washed at 65°C for 2.5 h, with a final wash solution of 0.5× SSC; 0.1% SDS. The hybridization signal was quantitated using a Bio-Rad phosphorimager.

Oligonucleotide primers within the first (H2, 5'-CGGCTGCAGATG-AGCA-3') and second exon (H3, see above) were used to prepare quasi-end-labeled, single-stranded hybridization probes for S1 nuclease mapping of liver poly(A)$^+$ RNA (Sambrook et al., 1989).

Antibody preparation and Western blot analyses

Rabbit HLF antibodies were raised according to standard procedures (Harlow and Lane, 1988) against a purified hHLF fusion protein, expressed with bacterial expression vector pET3c-hHLF43, and purified over a heparin–agarose column. HLF-specific antisera were purified by passing rabbit sera sequentially over an Affigel-10 column (Bio-Rad) coupled to recombinant DBP, followed by passing the eluate over a

recombinant TEF column (gift of P.Fonjallaz), thereby depleting the rabbit HLF antibodies of other PAR protein determinants. The flow-through fraction was tested for activity and used directly in Western blot experiments. The α-HLF serum purification method resulted in a reduced affinity antisera, requiring more serum to detect HLF, and thus giving a higher non-specific reactivity (with abundant proteins not related to PAR proteins).

Nuclei were isolated from rat liver, kidney, lung, testes and brain tissues and HepG2 cells and extracted with NUN (0.3 M NaCl; 1% NP-40, 1 M urea), as described previously (Tian and Schibler, 1991; Lavery and Schibler, 1993). Western transfers were incubated with a 1:1000 dilution of rabbit polyclonal anti-HLF (human) serum or a 1:100 dilution of purified serum (see below) and detected with either [^{125}I]protein A (0.2 μCi/ml, Amersham) and autoradiography or with peroxidase-conjugated goat anti-rabbit IgG and the ECL detection kit (Amersham).

Transient transfection assays

HepG2 cells were transiently transfected by calcium phosphate co-precipitation as previously described (Mueller et al., 1990) with 2–5 μg of expression vector, 8 μg of reporter CAT vector, 0.4 μg of RST-luciferase vector (de Wet et al., 1987) and pBS KS+ as carrier DNA to a total of 15 μg of plasmid DNA/10 cm plate. Isolation of nuclei and cell extracts were carried out as described previously (Descombes and Schibler, 1991). CAT activity was assayed by standard methods (Gorman et al., 1982) and measured by thin layer chromatography using a Berthold TLC linear analyzer. Values were normalized for transfection efficiency by assaying luciferase activity with a Bio-Orbit luminometer. CAT activity derived from independent transfection experiments (2) and independent CAT assays (4) were quantitated and corrected for transfection efficiency. Average values for fold activation were determined by dividing the value of CAT activity obtained with the indicated expression construct and the co-transfected promoter–CAT fusion by the activity derived from the promoter construct alone. The value of 'fold activation' for pSV2-CAT was calculated relative to the background level of CH–CAT activity.

In vitro translations

pSCT-HLF36 (1 μg) and mutant plasmids were linearized with EcoRI and transcribed using T7 RNA polymerase. Equivalent quantities of RNAs (~1 μg) were added to a 50 μl reaction containing 35 μl of nuclease-treated rabbit reticulocyte lysate (Promega) and labeled with [^{35}S]methionine. Five mircrolitres of translation reaction were added to 1 vol of 2× SDS loading buffer, heated for 5 min at 90°C and loaded on a 12% SDS–polyacrylamide gel (Laemmli, 1970). Gels were fixed with 50% methanol; 7% acetic acid; 5% glycerol and dried for autoradiography.

References

Acland,P., Dixon,M., Peters,G. and Dickson,C. (1990) Subcellular fate of the Int-2 oncoprotein is determined by choice of initiation codon. Nature, 343, 662–665.

Boeck,R. and Kolakofsky,D. (1994) Positions +5 and +6 can be major determinants of the efficiency of non-AUG initiation codons for protein synthesis. EMBO J., 13, 3608–3617.

Burch,J.B.E. and Davis,D.L. (1994) Alternative promoter usage and splicing options result in the differential expression of mRNAs encoding four isoforms of chicken VBP, a member of the PAR subfamily of bZIP transcription factors. Nucleic Acids Res., 22, 4733–4741.

Busch,S.J. and Sassone-Corsi,P. (1990) Dimers, leucine zippers and DNA-binding domains. Trends Genet., 6, 36–40.

Chiang,J.Y.L, Miller,W.F. and Lin,G.-M. (1990) Regulation of cholesterol 7α-hydroxylase and the immunochemical evidence for the induction

of cholesterol 7α-hydroxylase by cholestyramine and circadian rhythm. *J. Biol. Chem.*, **265**, 3889–3897.

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

Descombes,P. and Schibler,U. (1991) A liver-enriched transcriptional activator protein, LAP, and a transcriptional inhibitory protein, LIP, are translated from the same mRNA. *Cell*, **67**, 569–579.

deWet,J.R., Wood,K.V., DeLuca,M., Helinski,D.R. and Subramani,S. (1987) Firefly luciferase gene: structure and expression in mammalian cells. *Mol. Cell. Biol.*, **7**, 725–737.

Drolet,D.W., Scully,K.M., Simmons,D.M., Wegner,M., Chu,K., Swanson,L.W. and Rosenfeld,M.G. (1991) TEF, a transcription factor expressed specifically in the anterior pituitary during embryogenesis, defines a new class of leucine zipper proteins. *Genes Dev.*, **5**, 1739–1753.

Foulkes,N.S. and Sassone-Corsi,P. (1992) More is better: activators and repressors from the same gene. *Cell*, **68**, 411–414.

Gorman,C.M., Moffat,L.F. and Howard,B.H. (1982) Recombinant genomes which express chloramphenicol acetyl transferase in mammalian cells. *Mol. Cell. Biol.*, **2**, 1044–1051.

Grünert,S. and Jackson,R.J. (1994) The immediate downstream codon strongly influences the efficiency of utilization of eukaryotic translation initiation codons. *EMBO J.*, **13**, 3618–3630.

Hann,S.R., King,M.W., Bentley,D.L., Anderson,C.W. and Eisenman,R.N. (1988) A non-AUG translational initiation in c-*myc* exon 1 generates an N-terminally distinct protein whose synthesis is disrupted in Burkitt's lymphomas. *Cell*, **52**, 185–195.

Harlow,E. and Lane,D. (1988) *Antibodies: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Hunger,S.P., Ohyashiki,K., Toyama,K. and Cleary, M.L. (1992) HLF, a novel hepatic bZIP protein, shows altered DNA-binding properties following fusion to E2A in t(17;19) acute lymphoblastic leukemia. *Genes Dev.*, **6**, 1608–1620.

Hunger,S.P., Brown,R. and Cleary,M.L. (1994a) DNA-binding and transcriptional regulatory properties of hepatic leukemia factor (HLF) and the t(17;19) acute lymphoblastic leukemia chimera E2A–HLF. *Mol. Cell. Biol.*, **14**, 5986–5996.

Hunger,S.P., Devaraj,P.E., Foroni,L., Secker-Walker,L.M. and Cleary,M.L. (1994b) Two types of genomic rearrangements create alternative E2A–HLF fusion proteins in t(17;19)-ALL. *Blood*, **83**, 2970–2977.

Inaba,T., Roberts,W.M., Shapiro,L.H., Jolly,K.W., Raimondi,S.C., Smith,S.D. and Look,A.T. (1992) Fusion of the leucine zipper gene HLF to the E2A gene in human acute B-lineage leukemia. *Science*, **257**, 531–534.

Inaba,T., Shapiro,L.H., Funabiki,T., Sinclair,A.E., Jones,B.G., Ashmun,R.A. and Look,A.T. (1994) DNA-binding specificity and trans-activating potential of the leukemia-associated E2A–hepatic leukemia factor fusion protein. *Mol. Cell. Biol.*, **14**, 3403–3413.

Iyer,S.V., Davis,D.L., Seal,S.N. and Burch,J.B. (1991) Chicken vitellogenin gene-binding protein, a leucine zipper transcription factor that binds to an important control element in the chicken vitellogenin II promoter, is related to rat DBP. *Mol. Cell. Biol.*, **11**, 4863–4875.

Jelinek,D.F., Andersson,S., Slaughter,C.A. and Russell,D.W. (1990) Cloning and regulation of cholesterol 7α-hydroxylase, the rate limiting enzyme in bile acid biosynthesis. *J. Biol. Chem.*, **265**, 8190–8197.

Jones,N. (1990) Transcriptional regulation by dimerization: two sides of an incestuous relationship. *Cell*, **61**, 9–11.

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

Kozak,M. (1989) The scanning model for translation: an update. *J. Cell Biol.*, **108**, 229–241.

Krieger,D.T. (1979) *Endocrine Rhythms.* Raven Press, New York.

Laemmli,U.K. (1970) Cleavage of structural proteins during the assembly of the head of bacteriophage T4. *Nature*, **227**, 680–685.

Landschulz,W.H., Johnson,P.F. and McKnight,S.L. (1988) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science*, **240**, 1759–1764.

Lavery,D.J. and Schibler,U. (1993) Circadian transcription of the cholesterol 7α hydroxylase gene may involve the liver-enriched bZIP protein DBP. *Genes Dev.*, **7**, 1871–1884.

Lavery,D.J. and Schibler,U. (1994) DBP and related transcription factors of the PAR family. In Tronche,F. and Yaniv,M. (eds), *Liver Gene Expression.* R.G.Landes Co., Austin, TX, pp. 259–275.

Marchuk,D., Drumm,M., Saulino,A. and Collins,F.S. (1991) Construction

of T-vectors, a rapid and general system for direct cloning of unmodified PCR products. *Nucleic Acids Res.*, **19**, 1154.

Moore,R.Y. (1992) The suprachiasmatic nucleus and the circadian timing system. *Disc. Neurosci.*, **8**, 26–32.

Mueller,C.R., Maire,P. and Schibler,U. (1990) DBP, a liver-enriched transcriptional activator is expressed late in ontogeny and its tissue specificity is determined posttranscriptionally. *Cell*, **61**, 279–291.

Noshiro,M., Nishimoto,M. and Okuda,K. (1990) Rat liver cholesterol 7α-hydroxylase: pretranslational regulation for circadian rhythm. *J. Biol. Chem.*, **265**, 10036–10041.

O'Shea,E.K., Klemm,J.D., Kim,P.S. and Alber,T. (1991) X-ray structure of the GCN4 leucine zipper, a two-stranded parallel coiled coil. *Science*, **254**, 539–544.

O'Shea,E.K., Rutkowski,R. and Kim,P.S. (1992) Mechanism of specificity in the Fos–Jun oncoprotein heterodimer. *Cell*, **68**, 699–708.

Ossipow,V., Descombes,P. and Schibler,U. (1993) CCAAT/enhancer-binding protein mRNA is translated into multiple proteins with different transcription activation potentials. *Proc. Natl Acad. Sci. USA*, **90**, 8219–8223.

Prats,A.C., de Billy,G., Wang,P. and Darlix,J.L. (1989) CUG initiation codon used for the synthesis of a cell surface antigen coded by the murine leukemia virus. *J. Mol. Biol.*, **205**, 363–372.

Rusconi,S., Severne,Y., Georgiev,O., Galli,I. and Wieland,S. (1990) A novel expression assay to study transcriptional activators. *Gene*, **89**, 211–221.

Sambrook,J., Maniatis,T. and Fritsch,E.F. (1989) *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Saris,C.J.M., Domen,J. and Berns,A. (1991) The *pim*-1 oncogene encodes two related protein-serine/threonine kinases by alternative initiation at AUG and CUG. *EMBO J.*, **10**, 655–664.

Schmidt,E.E. and Schibler,U. (1995) Cell size regulation, a mechanism that controls cellular RNA accumulation: consequences on regulation of the ubiquitous transcription factors Oct1 and NF-Y, and the liver-enriched transcription factor DBP. *J. Cell Biol.*, **128**, 467–483.

Studier,F.W., Rosenberg,A.H., Dunn,J.J. and Dubendorff,J.W. (1990) Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enymol.*, **185**, 60–89.

Tian,J.-M. and Schibler,U. (1991) Tissue-specific expression of the gene encoding hepatocyte nuclear factor 1 may involve hepatocyte nuclear factor 4. *Genes Dev.*, **5**, 2225–2234.

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

Vinson,C.R., Hai,T. and Boyd,S.M. (1993) Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: prediction and rational design. *Genes Dev.*, **7**, 1047–1058.

Wuarin,J. and Schibler,U. (1990) Expression of the liver-enriched transcriptional activator protein DBP follows a stringent circadian rhythm. *Cell*, **63**, 1257–1266.

Wuarin,J., Falvey,E., Lavery,D., Talbot,D, Schmidt,E., Ossipow,V., Fonjallaz,P. and Schibler,U. (1992) The role of the transcriptional activator protein DBP in circadian liver gene expression. *J. Cell Sci.* (Suppl.), **16**, 123–127.