The EMBO Journal vol.4 no.8 pp.1981–1989, 1985

Structure and cell-specific expression of a cloned human retinol binding protein gene: the 5′-flanking region contains hepatoma specific transcriptional signals

Chiara D'Onofrio[1], Vittorio Colantuoni and Riccardo Cortese

European Molecular Biology Laboratory, Meyerhofstrasse 1, Heidelberg, FRG, and Istituto di Scienze Biochimiche, II Facolta' di Medicina, University of Naples, Naples, Italy

[1]Present address: Department of Experimental Medicine and Biochemical Sciences, Facolta' di Medicina II, University of Rome, Rome, Italy

Communicated by R. Cortese

Human plasma retinol binding protein (RBP) is coded by a single gene and is specifically synthesized in the liver. We have characterized a lambda clone, from a human DNA library, carrying the gene coding for plasma RBP. Southern blot analysis and DNA sequencing show that the gene is composed of six exons and five introns. Primer elongation and S1 mapping experiments allowed the definition of the initiation of transcription and the identification of the putative promoter. The 5′-flanking region of the RBP gene was fused upstream to the coding sequence of the bacterial enzyme chloramphenicol acetyl transferase (CAT): the chimeric gene was introduced, by calcium phosphate precipitation, into the human hepatoma cell line Hep G2 and into HeLa cells. Efficient expression of CAT was obtained only in Hep G2. Primer elongation analysis of the RNA extracted from transfected Hep G2 showed that initiation of transcription of the transfected chimeric gene occurs at a position identical to that of the natural gene. Transcriptional analysis of *Bal*31 deletions from the 3′ end of the RBP 5′-flanking DNA allowed the identification of the RBP gene promoter.

Key words: retinol binding protein/liver-specific gene expression/transfection/hepatoma cell line Hep G2/5′-flanking DNA

Introduction

The hepatocyte synthesizes a large number of proteins including the majority of plasma proteins (Putnam, 1975), a set of enzymes responsible for glycolysis and gluconeogenesis (Seifter and Englard, 1982), the enzymes of the urea cycle (Powers and Meister, 1982) and a large number of detoxifying enzymes (Alvares, 1982). In addition to the constitutively-expressed proteins characteristic of this cell type, numerous others are induced or repressed by hormones (Chan, 1982) dietary uptake (Greengard, 1971) or pathological conditions (Kushner, 1982). The repertoire of adult liver functions is acquired at various stages during development. Some genes are turned on at birth (Gitlin and Gitlin, 1975; Bensi *et al.*, 1985; Dente *et al.*, 1985) some only after weaning (Gitlin and Gitlin, 1975), whereas others, like α_1 fetoprotein, are expressed only in the fetal liver (Scott *et al.*, 1984). It is likely therefore, that there is a complex series of regulatory mechanisms responsible for the maintenance of the differentiated state and for the developmental and physiological variations in the pattern of gene expression.

The long-term goal of our research project is to contribute to the identification and characterization of the molecular mechanisms responsible for the regulation of gene expression in the hepatocyte. We have, therefore, cloned several liver-specific genes and studied their expression in relatively simple experimental systems. For instance, by introducing the gene into cultured cells; structural and functional comparisons will provide useful information for the understanding of this complex system.

One of the liver-specific genes that has attracted our attention is that coding for retinol binding protein (RBP) (for references on this protein see Colantuoni *et al.*, 1983). RBP is synthesized in the liver as a single polypeptide chain of 21 000 daltons, and is secreted into the blood stream where it functions as the plasma transport protein for retinol (vitamin A alcohol). Experiments in rats and differentiated hepatoma cell lines have shown that vitamin A regulates the rate of synthesis and secretion of RBP. We have cloned and sequenced a cDNA clone coding for the human RBP and have shown that this protein is coded by a single gene per haploid genome (Colantuoni *et al.*, 1983). Here we describe the cloning of the human RBP gene, its characterization and partial sequence. We show that the cloned gene, following transfection in cultured human cells, is expressed only in hepatoma cell lines and not in HeLa cells. The information for cell-specific expression is contained in the segment of DNA flanking the 5′ end of the coding region.

Results

Identification of a lambda clone carrying the human RBP gene

A human DNA genomic library, constructed as an *Mbo*I partial digest in the lambda vector EMBL 3 (Bensi *et al.*, 1985), was screened for the RBP gene, using, as a probe, the *Pst*IA restriction fragment derived from the corresponding cDNA previously isolated and sequenced (Colantuoni *et al.*, 1983). One positive clone, containing an insert of ~20 kb, lambda-RBP-1, was characterized by a combination of restriction enzyme analysis and Southern blots, using various segments derived from the cDNA clone as probes. The result of this analysis is shown in Figure 1. The gene for human RBP covers a region of ~10 kb of genomic DNA and consists of six exons and five introns. Unfortunately, the lambda-RBP-1 clone does not contain the complete gene since it is interrupted at the level of the intron following the coding sequences for amino acid 172 and lacks the last exon coding for amino acids 173–183. All our efforts to isolate more lambda clones to complete this part of the gene have been unsuccessful. However, the comparison with the rat RBP gene (P. Peterson, personal communication) shows a similar genomic organization and, in particular, the sequences coding for the last 11 amino acids are contiguous to the 3′-untranslated region present in the mRNA in rat. By analogy with the rat gene, it is likely that no additional introns are present in human and therefore, that the missing coding sequences should be included in the 6th exon.

The exons were sequenced and found to be identical to the cDNA sequence previously determined, with only two changes in the 3rd position in the codons for amino acids 46 and 116.

© IRL Press Limited, Oxford, England.

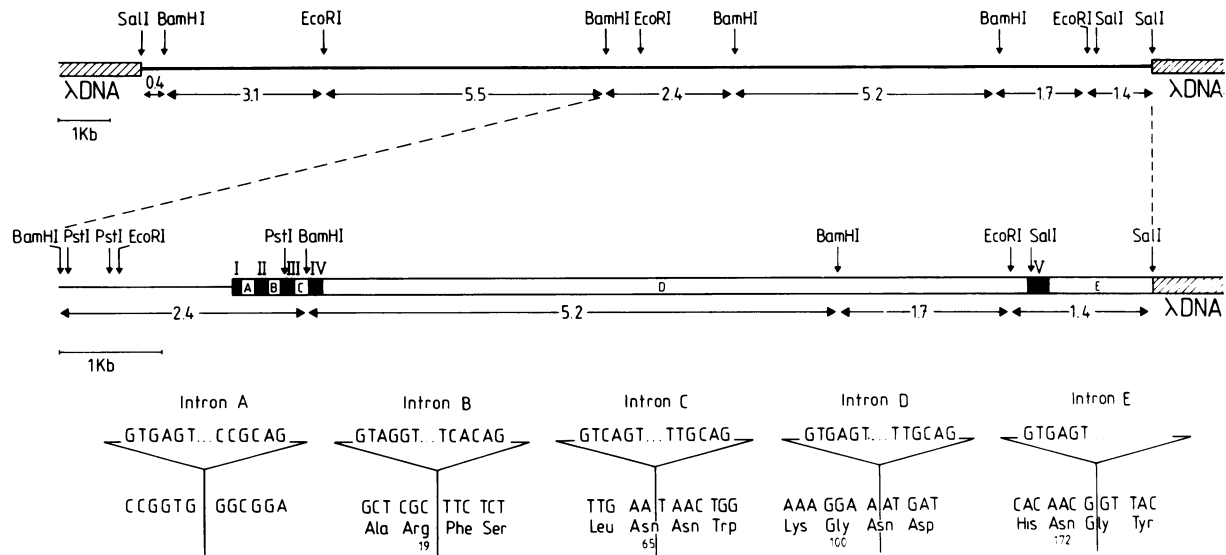


Fig. 1. Restriction endonuclease map of the lambda clone RBP-1 and organization of the human RBP gene. The upper panel shows some of the restriction enzymes used for mapping the lambda clone together with the size of the fragments in kilobases (kb). The region of the clone covering the RBP gene is enlarged, the dark blocks identify the exons (I – V), the open blocks the introns (A – E), the hatched boxes are lambda vector sequences. The lower panel shows the intron-exon junctions with their relative position to the amino acid sequence, as deduced from the alignment with the nucleotide sequence of the cDNA.

```
CCACCTTAGCTGTTTTGAAAATAAATGAAAATTCCAACATGGTTTGAAATAAAATTGCATCATAAACAATCGGTAGGTGTTTTTCAAAGTGGTTTCAGG   -716
GAAGTGCCACGGAGTAAGCAGGCGACCACCGAGGCTGCTAAAATATTTCCTGTCCTGACCAGGGTTGCGTTTCTGGAGAATATTTAACAGGGAGGGTTTT   -616
AACGCTTTTAAAGATGTTGAAACTAAAGAACAAATATTGACCAGAGGGCACCACAACGCTCCTGAAAGAGAGTAAAATACATCCTTTATAAAATGAAAAA   -516
CTACTTGGATGAATTATTCCAAAATTCTGCACAAGTGGACCTCAGAAGGCAGACGGAGGCGCCAATTTGGCATGGCCAGGGCCTGGGCACTCACGCACCA   -416
GGGAGCCTCGCGGGTCCGCTCTCGCTCTGTGAGGCCACGGTCTTCCCGCCAGGTTGACTCGAGCCTCCTGCCAGAGCCACTGGCCCCGGAGGCCACCCTA   -316
GACCGCAGCTGGCGGCCGCTGGCACGAGTGCAGGGTAACTGAGCCAGGGCCGCTGGCGCATTTTGGCCTGGCCGAGGCCACCCCGCGCGGCCGCTCCACT   -216
GTGCCCGAGGCTGTCCTGGAGGTAGGCCGGCCCACAGGGACCCTGCCCGTGCCCGGGCTCCGGTGAGTCAGGGCGCGTTATGCAAGTGCCCCCGGCGCCT   -116
                                                          -27                          +1
CCCCTTCGGTCTTTCACCCCGCGCGGTTACGAAAGCGCGACCCCCTCCCCCCGGCGC[TATAAA]GCAGCGGGGCGGCCGCGGCGCGCTCGCCTCCCTCGCT    +16
CCACGCGCGCCCGGACGCGGCGGCCAGGCTTGCGCGCGGTTCCCCTCCCGGTGGTGAGTCGCGGCCGGGGCCTCGGGGCCGGCGGCGGGATGGGCCGGGG   +116
TGGGGTGGGGTGGGGTGGGGTGAGGTGGGATACGGCGGCCCGGGGGTCGGCTGCCGGGCACTGACGGAGGCGCGTCTCCCGCAGGGCGGATTCCTGGGCA   +216
AGATGAAGTGGGTGTGGGCGCTCTTGCTGTTGGCGGCGTGGGCAGCGGCCGAGCGCGACTGCCGAGTGAGCAGCTTCCGAGTCAAGGAGAACTTCGACAA   +316
  MetLysTrpValTrpAlaLeuLeuLeuLeuAlaAlaTrpAlaAlaAlaGluArgAspCysArgValSerSerPheArgValLysGluAsnPheAspLy
GGCTCGCGTAGGTATCCGGCCCAGGGGCCAGCGCCAGGTCCCCACCCCATCCCCGCCCTCCCCGCGGCCCGCCGTGCGCCCCCGGACGCCAAACCCCTGC   +416
sAlaArg
TGTCATCCTTCTCACAGTTCTCTGGGACCTGGTACGCCATGGCCAAGAAGGACCCCGAGGGCCTCTTTCTGCAGGACAACATCGTCGCGGAGTTCTCCGT   +516
                 PheSerGlyThrTrpTyrAlaMetAlaLysLysAspProGluGlyLeuPheLeuGlnAspAsnIleValAlaGluPheSerVa
GGACGAGACCGGCCAGATGAGCGCCACAGCCAAGGGCCGAGTCCGTCTTTTGAAGTCAGTGGCCGCTGGGGCAGCTGCGCCGTTTGCGCTCCAGGGTCCC   +616
lAspGluThrGlyGlnMetSerAlaThrAlaLysGlyArgValArgLeuLeuAs
CCAAGGGCCCTGCCTGCTGATCGCCACTGGCATTGTGAAGGGAAGGGAGCACCGAATGGGTGGAGGGAGGGAGGAAGCCCTTTGCCCGGCTTGGCTGAGG   +716
ATCCCCTTGGCTTTTGCAGTAACTGGGACGTGTGCGCAGACATGGTGGGCACCTTCACAGACACCGAGGACCCTGCCAAGTTCAAGATGAAGTACTGGGG   +816
                   nAsnTrpAspValCysAlaAspMetValGlyThrPheThrAspThrGluAspProAlaLysPheLysMetLysTyrTrpGl
CGTAGCCTCCTTTCTCCAGAAAGGAAGTGAGTAGTCAGCTCTGTGGGACTGTCTTGGGAATGGGGCCCTACCGGGCCAAATCGCTGGGTTTCTTCTCCAG   +916
yValAlaSerPheLeuGlnLysGlyA
AGGTACC...TCCTGCTTCCTAAAGAAATAAAACCAGCTTCTGGAAGTCACTGTGTTCTCAGGCCTGGTCGTGAGTTTCTCCGACATCTGAGCCCATGAG
GGAAAGGCATTGATCGGCATGGTGGCTTTCTGGCTTGCAGATGATGACCACTGGATCGTCGACACAGACTACGACACGTATGCCGTGCAGTACTCCTGCC
                                        snAspAspHisTrpIleValAspThrAspTyrAspThrTyrAlaValGlnTyrSerCysA
GCCTCCTGAACCTCGATGGCACCTGTGCTGACAGCTACTCCTTCGTGTTTTCCCGGGACCCCAACGGCCTGCCCCCAGAAGCGCAGAAGATTGTAAGGCA
rgLeuLeuAsnLeuAspGlyThrCysAlaAspSerTyrSerPheValPheSerArgAspProAsnGlyLeuProProGluAlaGlnLysIleValArgGl
GCGGCAGGAGGAGCTGTGCCTGGCCAGGCAGTACAGGCTGATCGTCCACAACGGTGAGTCAGGAGCAGGGTGCCTTCTGGCTCTTCCTGCAGGGCAGTGG
nArgGlnGluGluLeuCysLeuAlaArgGlnTyrArgLeuIleValHisAsnG
GTTTGGACTAAGGGGCCCACGTGCTAGTGA
```

Fig. 2. Nucleotide sequence of the human RBP gene. The nucleotide sequence of the clone lambda RBP-1, that contains large part of the gene is illustrated. Only the sequence present in the mature RBP mRNA is underlined. The amino acid sequence is given below the corresponding codons. The proposed cap site is indicated as +1 and the sequence downstream to it is numbered positively, while the 5'-untranslated region is numbered negatively. The presumptive TATAA sequence, at position −27, is boxed. The region of intron D of ~8 kb was not sequenced, and is indicated by a dotted line. Only part of the sequence of intron E is reported.

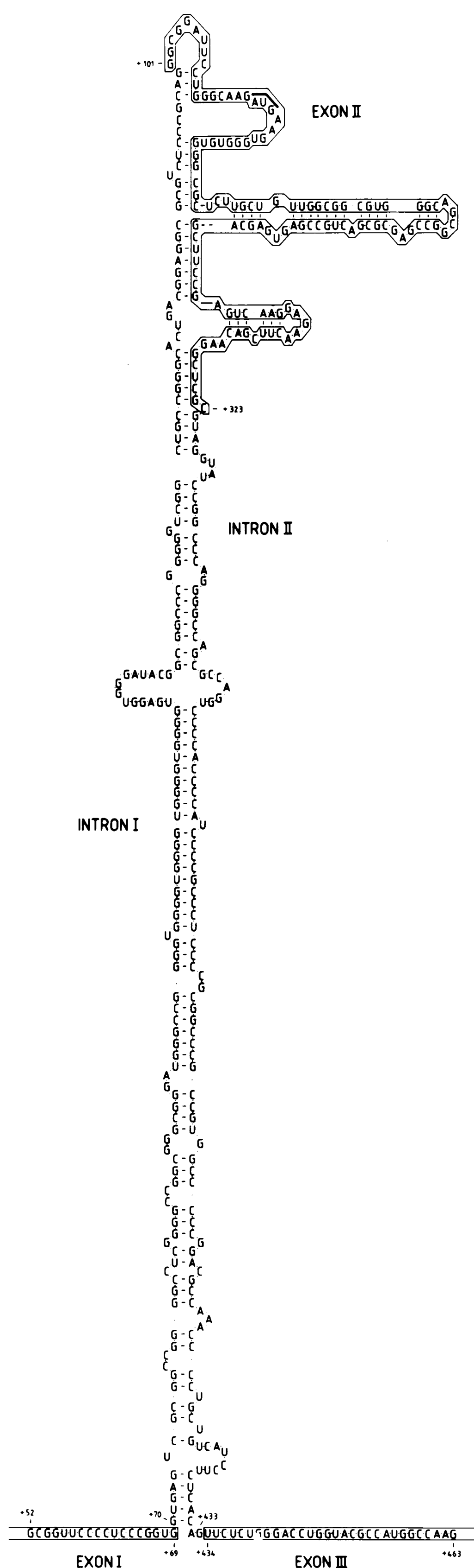


Fig. 3. Schematic representation of the predicted mRNA secondary structure. The most probable secondary structure of a region of ~400 nucleotides at the 5′ end of the RBP primary transcript is shown. The exon sequences are boxed, the AUG is marked by a bar. The numbers refer to the nucleotide sequence downstream to the cap site indicated as +1.

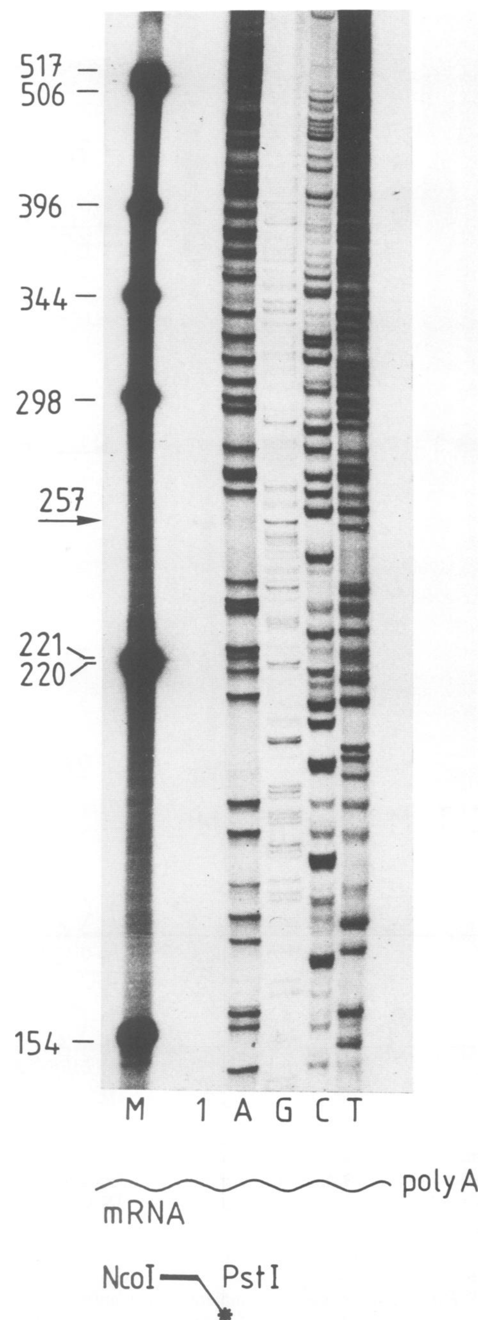


Fig. 4. Identification of the site of transcription initiation for RBP mRNA by primer extension analysis. **Lane 1** shows the product of the elongation experiment obtained after hybridization of the *Nco*I-(*Pst*I)-*Hin*dIII fragment to total liver RNA. The probe was obtained from digestion of the RBP cDNA after strand separation and 3′ end-labelling. The relative position of the primer to the mature RBP mRNA is illustrated at the bottom of the figure. pBR322 digested with *Hin*fI as mol. wt. marker is shown in **lane M**. AGCT is a DNA sequencing reaction performed by the Sanger method, and used as a ladder for a more accurate measurement of the bands. The reaction was analyzed on a 6% acrylamide/7 M urea gel.

They are relatively short, comprising 69, 123, 137, 107 and 213 nucleotides, respectively. The first exon contains only 5′-untranslated sequences, the initiating ATG is present in the second exon. The exon-intron junctions were localized by alignment with the nucleotide sequence of the corresponding cDNA coding for human RBP. They show the 5′ GT . . . . AG 3′ consensus sequence consistent with the one previously reported at the splice sites (see Figure 1, lower panel). The introns show

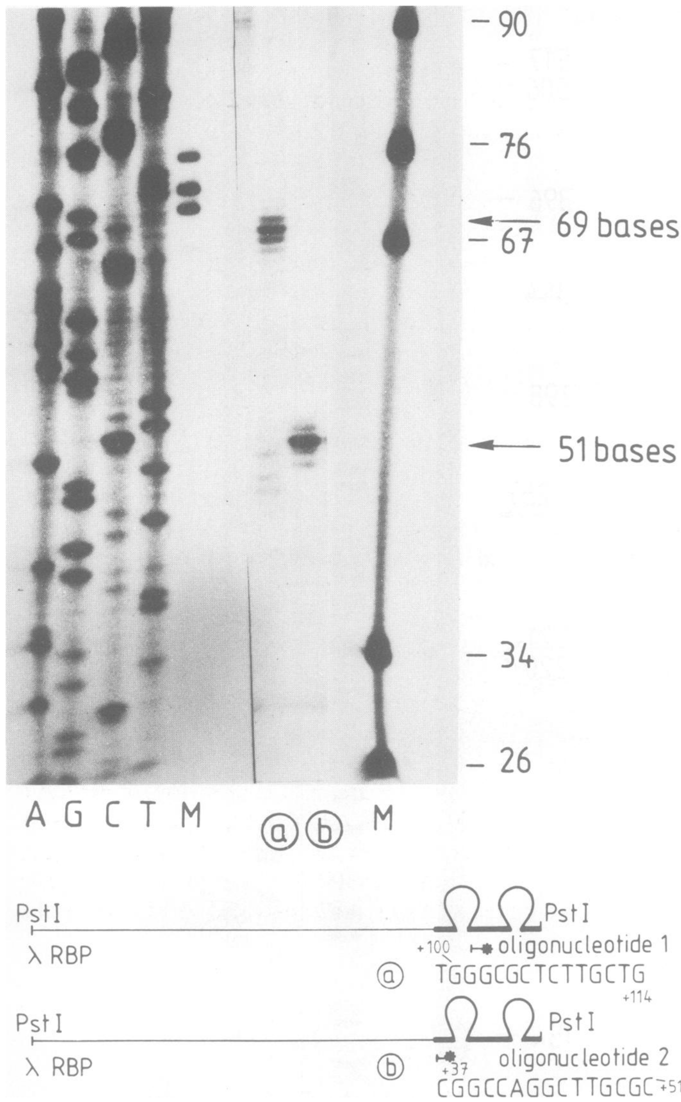


Fig. 5. Identification of the site of transcription initiation for RBP mRNA by S1 nuclease mapping. **Lanes a** and **b** show the product of S1 nuclease digestion of total liver RNA after hybridization of DNA probes obtained by annealing oligonucleotide 1 or 2 to the single strand mp9-rbp-PstI DNA and elongating in the presence of labelled deoxynucleotides. The sequences of the two oligonucleotides used are shown together with their position on the lambda RBP *Pst*I fragment used as template. The numbering refers to the proposed cap site indicated as +1. pBR322 digested with *Hin*fI (on the left) and with *Hpa*II (on the right) are used as mol. wt. markers (M). The sequence ladder on the left is used for a more accurate sizing of the bands obtained on a 6% acrylamide/7 M urea gel.

a wide range of sizes, 0.131, 0.110, 0.165 and ~8 kb respectively. The 5th intron is interrupted by the lambda arms but is at least 1.2 kb long. Figure 2 shows the partial nucleotide sequence of the gene and the deduced amino acid sequence. The RBP gene is, therefore, formed by four short exon-introns clustered at the 5′ end, with two exons (the 5th and the 6th) well separated by long intervening sequences. This arrangement is similar to that of the rat gene (P. Peterson, personal communication) which shows the same number and relative position of the exons. The three-dimensional structure of the human RBP has been determined by X-ray crystallography (Newcomer *et al.*, 1984). A comparison of the gene organization with the protein structure shows that each exon codes for a defined structural unit. The second exon, for instance, codes for the leader peptide and for the first 19 amino acids of the mature protein, which form a coil in the tertiary structure. This N terminus is very rich in charged amino acids, it is highly conserved in the protein sequence of rat, rabbit and human and has been related to the binding of RBP with the circulating prealbumin. Exons 3, 4 and 5 code for those amino acids that form the three planes of the eight-stranded anti-parallel β sheet of the barrel-shaped core.

The sequence of the first intron shows long stretches complementary to stretches of repeated sequences in the second intron. It is likely that these elements will pair as direct repeats in a secondary structure such as the one shown in Figure 3. We do not know whether this structure is formed *in vivo* or what function it might have. Possibly, the secondary structure brings the donor site of the first intron in close proximity with the acceptor site of the second intron. This could facilitate an alternative splicing reaction with consequent removal of the second exon. The resulting mRNA could be translated into a smaller protein, lacking the leader peptide. It is interesting that smaller intracellular RBPs have been described (Ong, 1984).

Definition of the transcription unit of the human RBP gene

The 5′ region of the gene included in the *Pst*I-*Pst*I fragment of 1600 bp, (Figure 1) was subcloned in mp9 (mp9-RBP-PstI) or pEMBL8 (pEMBL8-RBP-PstI). A combination of primer elongation on total liver RNA and S1 mapping allowed us to identify the possible initiation point of transcription for the RBP mRNA. In the primer elongation experiments, we used a single-stranded 45-bp DNA segment as a primer, comprising 36 bases of the RBP cDNA (from the *Nco*I site to the *Pst*I site corresponding to amino acids 27–38, see Figure 2) and nine bases from the adjacent polylinker up to the *Hind*III site. Elongation of this primer on a total liver RNA template, yields a band of 257 bp (Figure 4). The size of this band indicates that the cap site is located at ~36 bp upstream to the 5′ end of the cDNA clone (Figure 2). A similar result was obtained by S1 mapping: the DNA probe was obtained by *in vitro* DNA synthesis (Ciliberto *et al.*, 1983) using single-stranded DNA from mp9-RBP-PstI as template, and two different synthetic oligodeoxynucleotides (named 1 and 2, each 15 bases long, Figure 5) as primers. These ^{32}P-labelled DNA probes were hybridized to total liver RNA and then treated with S1 nuclease. The size of the protected band indicated again that the point of initiation of transcription is located at ~36 bp upstream to the 5′ end of the cDNA clone. Thus, the most likely initiation point is as the level of the G residue indicated in the sequence as +1 (Figure 2). A canonical TATA box is identified −27 bp upstream from the cap site. No suitable 'CAT' consensus sequence was found in the −70 bp region. The sequence of the 5′ region of the gene up to the 4th exon is given in detail in Figure 2.

Tissue and cell specific expression of the human RBP gene

Northern blot analysis of human liver RNA shows the presence of an mRNA of ~1000 bases, a similar band is observed using rat liver RNA (the human cDNA used as probe efficiently cross-hybridizes). No hybridization was observed using rat kidney or rat spleen RNA (Figure 6, panel A). This establishes that the RBP gene is not expressed in every cell type and that, therefore, there must be cell-specific control mechanisms responsible for the specificity of RBP gene expression. This specificity is also found in cultured cell lines: RBP mRNA is present in the human hepatoma cell line Hep G2, but is absent in HeLa cells (Figure 6, panel B). We decided to use these two cell lines for gene expression experiments: if the cloned RBP DNA contains sufficient information for cell-specific expression, it should direct

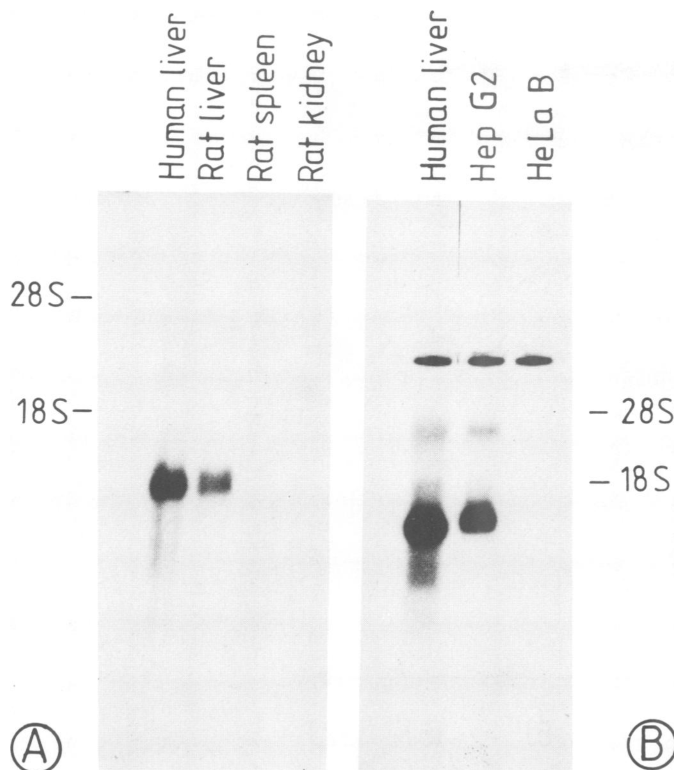


Fig. 6. Northern blot analysis. 10 μg of total RNA extracted from different tissues as indicated were loaded in each lane **(panel A)**; 15 μg were used, instead, when RNAs were extracted from Hep G2 or HeLa cells **(panel B)**. The nick-translated *Pst*IA fragment from the RBP cDNA was used as probe.

cell-specific expression of an unrelated protein in Hep G2 but not in HeLa cells. Hep G2 is a human hepatoma cell line isolated and characterized by Knowles and co-workers (Knowles *et al.*, 1980): it appears to have maintained several features of the differentiated hepatocyte, including the synthesis of several plasmaproteins.

Construction of suitable templates for in vitro expression of RBP gene

A convenient approach to study control of gene expression in eukaryotic cells, and to identify DNA sequences functioning as *cis*-acting transcriptional signals, is to fuse predetermined DNA segments upstream to the coding sequence of the bacterial enzyme chloramphenicol acetyl transferase (CAT) as shown by Gorman *et al.* (1982). These constructs are then introduced into cultured cells by transfection, and expression can be measured as CAT enzymatic activity.

To test whether sequences located in the 5′-flanking region of the human RBP gene could promote cell-specific transcription, we constructed a series of plasmids in which various DNA segments deriving from the RBP gene were inserted between the *Bgl*II and the *Hind*III site of the vector pA10-CAT2 (Figure 7), at the same time removing the SV40 promoter sequence. As a negative control we used the plasmid pA10-CAT-Sma (Ciliberto *et al.*, in preparation), which carries the coding sequences of the CAT gene without any eukaryotic promoter and is not transcribed in Hep G2 or in HeLa cells. In contrast, if the SV40 promoter-enhancer sequences are inserted upstream to the CAT gene as in the plasmid pSV2-CAT (Gorman *et al.*, 1982), the CAT gene is expressed in both cell types, giving rise to high CAT activity, which is taken as 100% activity in our assays. The RBP inserts cloned in pA10-CAT2 contained ~1200 bp in the

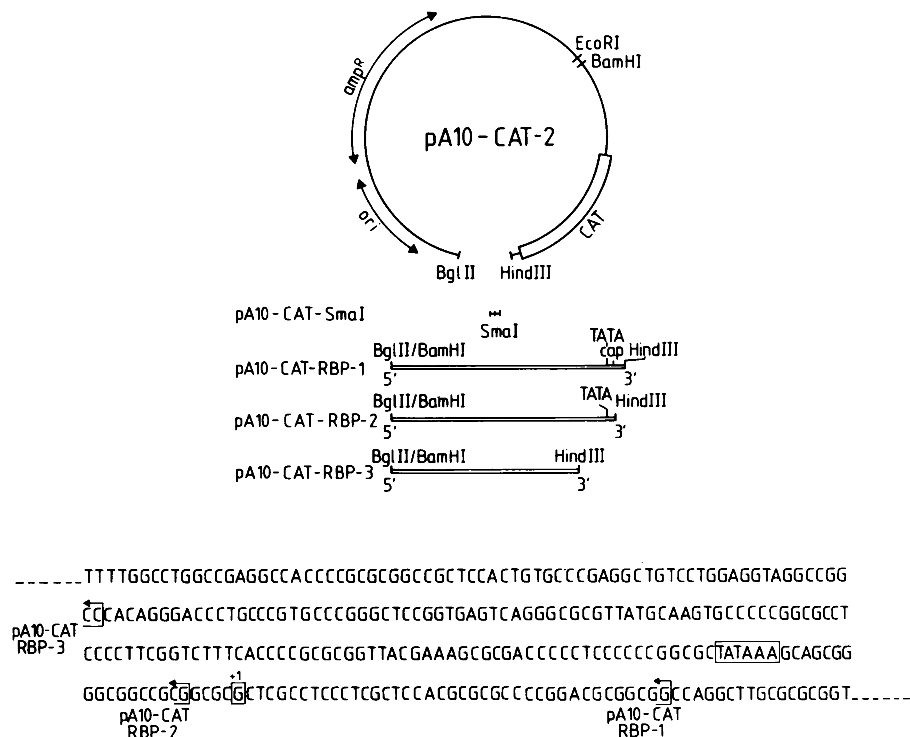


Fig. 7. Structure of the recombinant plasmids carrying the RBP-CAT fusions. The structure of the plasmids used in the transfection experiments is outlined. The *Bgl*II-*Hind*III fragment from pA10-CAT-2 plasmid, containing the SV40 early promoter was removed and replaced with DNA fragments of the 5′-flanking region of the RBP gene. The details of these constructions are given in Results. pA10-CAT-Sma is the negative control. The sequence of the 3′ most part of the RBP DNA segments used is shown.

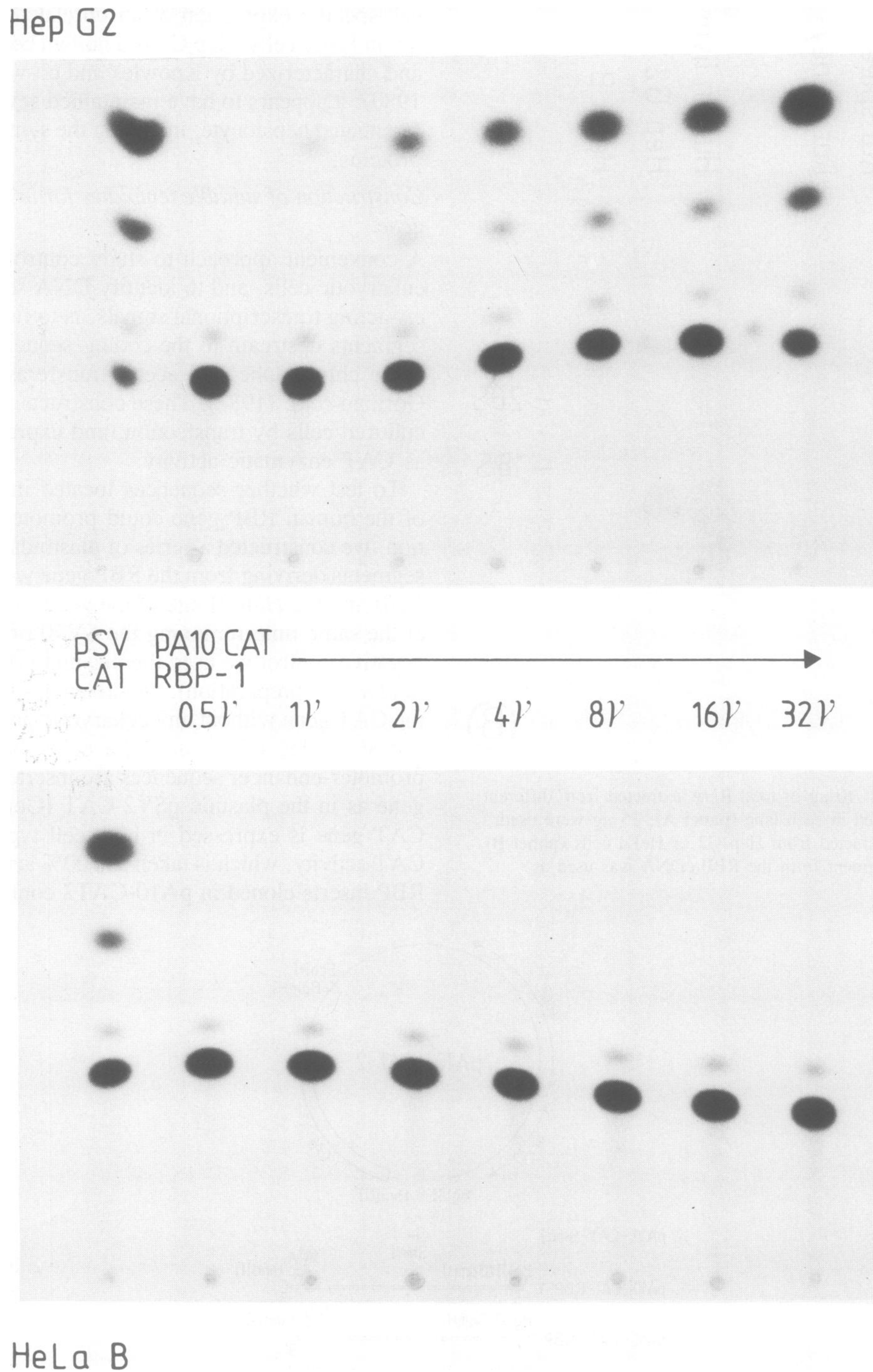


Fig. 8. Expression of the RBP-CAT fusions in Hep G2 and HeLa cells. Increasing amounts of the recombinant plasmid containing the RBP-CAT fusion (pA10-CAT-RBP1) were transfected in Hep G2 and HeLa cells, together with plasmid DNA as carrier. pSV2-CAT was used in both cell lines as positive control.

5′-flanking region upstream the cap site. *Bal*31 deletions at the 3′ end of the *Pst*I-*Pst*I 1600-bp fragment (Figure 7) progressively removed the cap site (pA10-CAT-RBP2) and the TATA box (pA10-CAT-RBP3).

Human RBP gene is efficiently transcribed in human hepatoma cells but not in HeLa B cells

pA10-CAT-RBP1, the vector carrying the 5′-flanking region of the RPB gene up to the 1st exon (from +39 bp upstream to about −1200 bp) including the cap site and the putative TATA box, was transfected by the calcium phosphate technique (Graham and van der Eb, 1973) into Hep G2 and HeLa B cells. In Hep G2 cells transcription of the CAT gene showed almost a linear correlation with the concentration of the transfected pA10-CAT-RBP1 DNA (Figure 8, upper panel) up to 32 μg DNA/5 x 10^5 cells. In contrast to Hep G2 cells, in HeLa cells no CAT activity was detected up to 32 μg of transfected DNA (Figure 8, lower panel). The expression of RBP seems, therefore, to be cell-specific and this specificity is under the control (at least in part) of the 5′-flanking region of the gene.

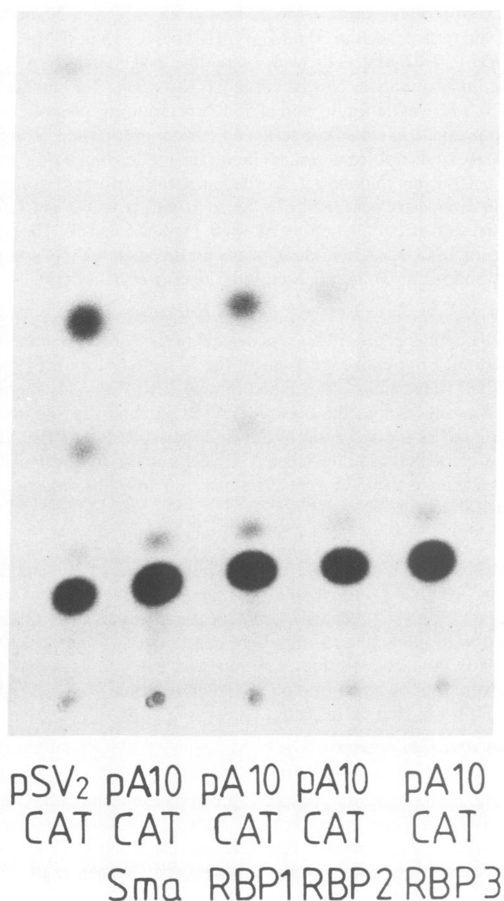
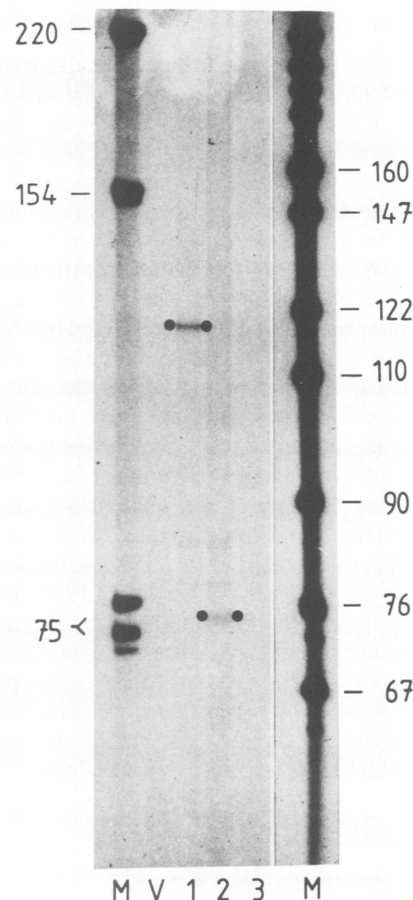


Fig. 9. Expression of different RBP-CAT fusions in Hep G2 cells. pA10-CAT-RBP 1,2 and 3 plasmids containing the RBP-CAT constructions, with deletions at the 3′ end of the RBP segment, were transfected into Hep G2 cells. CAT activity was measured as described. pSV2-CAT and pA10-CAT-Sma were used as positive and negative controls, respectively.



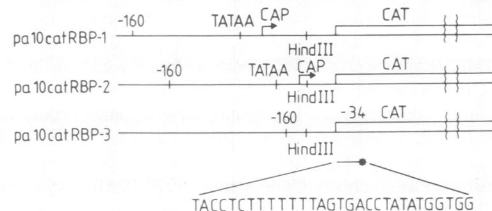


Fig. 10. Determination of the start point of transcription in pA10-CAT-RBP plasmids. The products of the extension experiment obtained after hybridization to total RNAs from transfected Hep G2 cells of a 30-mer synthetic oligodeoxynucleotide, complementary to the CAT coding region, are shown. Bands of 120 and 76 bases are visible only in **lanes 1** and **2**, corresponding to the constructions pA10-CAT-RBP1 and 2. No bands are detectable in **lanes V** and **3** which correspond to the vector and to the pA10-CAT-RBP3 plasmid. pBR322 digested with *Hinf*I (on the left) and with *Hpa*II (on the right) are used as mol. wt. markers. The sequence of the oligonucleotide used as primer is shown. The reactions were analyzed on 6% acrylamide/7 M urea gel, with 10 μg of RNA loaded in each lane.

Definition of the promoter regions of the RBP gene

The initiation point for transcription of RBP mRNA has been determined by primer elongation on total liver RNA and S1 mapping (Figures 4 and 5) and named as nucleotide +1 (Figure 2). This indicates that the TATA box of the RBP gene is likely to be the sequence TATAAA, boxed in the same figure. This was confirmed by testing the expression of the RBP gene using serial *Bal*31 deletions of this DNA region from the cap site to −152 bases upstream of it (Figure 7). These deletions, cloned in pA10-CAT2 at the *Bgl*II-*Hind*III sites in front of the CAT gene in the 5′→3′ orientation, were transfected into Hep G2 and HeLa B cells and tested for CAT activity (Figure 9). Transcription-translation of the CAT gene occurred at the highest level when the RBP insert included both the 'physiological' cap site and the TATA box of RBP (36% of pSV2-CAT activity); it was still detectable, although at a reduced level (26% of pSV2-CAT activity), when the cap site was deleted without affecting the TATA box. When the TATA box was also deleted, up to −152 bp upstream, no transcription-translation of the CAT gene was detected. Thus, the 3′ boundary of the promoter sequences of the RBP gene is contained within 152 bp upstream of the cap site.

Definition of the cap site for the transfected RBP gene

To identify the starting point of transcription of the plasmids pA10-CAT-RBP1-3 carrying the RBP gene inserts, RNA was extracted from the transfected Hep G2 cells and used as template for elongation from a 30-base synthetic oligodeoxynucleotide primer, complementary to the first 30 bases of the CAT-coding sequences (Alton and Vapnek, 1979). Primer elongation on RNA from cells transfected with plasmid pA10-CAT-RBP1 yielded a band of 120 bases (Figure 10), indicating that transcription of the cloned gene is initiated at a position identical to that of the chromosomal gene in human liver. The same initiation point is also present on the plasmid pA10-CAT-RBP2, (primer elongation yields a 76-base band, 44 nucleotides shorter than that from pA10-CAT-RBP1, as expected), whereas no elongation was observed on RNA from Hep G2 transfected with plasmid pA10-CAT-RBP3 (carrying a deletion of the putative TATA box) nor on RNA from HeLa cells transfected with pA10-CAT-RBP1 (data not shown).

Discussion

In multicellular organisms the genome is expressed in a highly specific manner during development and in the various cell types. The molecular mechanisms responsible for this specificity are poorly understood. The experiments described here show that a cloned DNA segment, coding for the human RBP synthesized in the liver, contains DNA sequences capable of controlling the adjacent gene in a cell-specific manner. In the experimental system used, we show that plasmid constructs carrying the 5'-flanking region of the RBP gene, are efficiently and accurately transcribed in the human hepatoma cell line Hep G2, but not in HeLa cells. The existence of *cis*-acting cell-specific DNA sequences located in the 5'-flanking region has also been shown for other liver-specific genes, for example the rat (Ott *et al.*, 1984) and human (Ciliberto and Cortese, unpublished) albumin genes; the human α_1-antitrypsin gene (Ciliberto *et al.*, 1985), the human haptoglobin genes (Raugei *et al.*, unpublished results). These observations, taken together with results obtained in other experimental systems (Baneriji *et al.*, 1983; Gillies *et al.*, 1983; Queen and Baltimore, 1983; Walker *et al.*, 1983; Melloul *et al.*, 1984) stress the likelihood that the molecular mechanisms responsible for the maintenance of specific gene expression throughout the lifespan of a differentiated cell might be simple and based on interaction between cell-specific *trans*-acting factors and *cis*-acting transcriptional signals, not unlike the gene control mechanisms well characterized in prokaryotes.

Materials and methods

Bacterial strains, plasmids and phage vectors

Escherichia coli K12 (strain 71/18 and K514) was used for transformation (Gronenborn and Messing, 1978). The M13 derivatives mp8, mp9 (Messing, 1981), mp18, mp19 (Messing, 1983) were used as phage vectors for subcloning and sequencing; pUC18 and pUC19 (Norrander *et al.*, 1983), pEMBL8 (Dente *et al.*, 1983) and pA10-CAT2 (Laimins *et al.*, 1982) were used as plasmid vectors for subcloning. pSV2-CAT (Gorman *et al.*, 1982) was used as positive control and pA10-CAT-Sma (Ciliberto *et al.*, in preparation) as negative control for transfections. Phage EMBL 3 (Frischauf *et al.*, 1983) was used in the construction of the human genomic library. Transformation and preparation of single- and double-stranded DNA were as described (Cortese *et al.*, 1978,1980). Large-scale preparation of EMBL 3 recombinant phages was done according to Frischauf *et al.* (1983).

Enzymes and chemicals

Restriction endonucleases, T4 polynucleotide kinase, T4 DNA ligase, DNA polymerase I holoenzyme and Klenow fragment, and *Bal*31 were purchased from B.R.L. (Neu Isenburg, FRG) and New England Biolabs (Schwalbach, FRG). ^{32}P-Labelled compounds were purchased from Amersham (Braunschweig, FRG). AMV reverse transcriptase was purchased from Boehringer (Mannheim, FRG). ^{14}C-Labelled chloramphenicol was from NEN (Dreieich, FRG) and acetyl-coenzyme A from Sigma (St. Louis, MO).

Screening of the human genomic library

The human genomic library was kindly provided by G. Bensi (Bensi *et al.*, 1985) and screened according to Frischauf *et al.* (1983). Nick-translated *Pst*I A DNA insert from the full length cDNA clone of RBP (Colantuoni *et al.*, 1983) was used as a probe in screening.

Oligodeoxynucleotide synthesis

Oligodeoxynucleotides were synthesized following the phosphite-amidite method (Winnacker and Dorper, 1982).

Blot hybridization and nucleotide sequence analysis

Restriction maps, Southern and Northern blot analysis were performed as previously described (Colantuoni *et al.*, 1983). Sequence analysis was carried out by the dideoxy-method (Sanger *et al.*, 1977).

Primer elongation and S1 mapping

The double-stranded primer (*Nco*I-*Hind*III fragment) from the RBP cDNA subcloned in mp9, was 3' end-labelled with [γ-^{32}P]ATP using the large fragment of DNA polymerase I, and strand-separated according to Maniatis *et al.* (1982). Oligodeoxynucleotide primers were labelled at the 5' end with [γ-^{32}P]ATP and T4 polynucleotide kinase. Human liver total RNA (10 μg) or cellular total RNA (10 μg) was mixed with 9 volumes of DMSO, heated at 45°C for 20 min, ethanol precipitated and resuspended in 50 mM Tris-HCl pH 8.3, 5 mM $MgCl_2$, 50 mM KCl, 1 mM DTT, 0.4 mM dNTPs, with 1 ng primer and 20 units of AMV reverse transcriptase. Incubation was carried out at 42°C for 2 h. The sample was then treated with 0.3 M NaOH for 10 min at 56°C, neutralized, phenol-chloroform-isoamylalcohol (25:24:1, v:v:v)-extracted and ethanol-precipitated. The pellet was washed with 80% ethanol, resuspended in denaturing loading buffer and run on a 7 M urea, 6% polyacrylamide gel. S1 mapping was performed using synthetic oligodeoxynucleotide primers (named oligonucleotide 1 and 2 and CAT-oligonucleotide), the sequences of which are shown in Figures 5 and 10. The experimental procedure for DNA-RNA hybridization and for digestion with S1 was according to Berk and Sharp (1977) as modified by Ciliberto *et al.* (1983).

Construction of CAT-derivative vectors with RBP inserts

pA10-CAT2 (Laimins *et al.*, 1982) was deleted of the SV40 promoter by *Bgl*II-*Hind*III digestion. Progressive deletions of the 5' region of the RBP gene were obtained by *Bal*31 digestions at the 3' end of mp9-RBP-PstI. This fragment contains the 5'-flanking region, the putative TATA box and cap site and extends up to the *Pst*I site in the 3rd exon of the RBP gene. RBP 1, RBP 2 and RBP 3 fragments were subcloned in pA10-CAT2 vector at the *Bgl*II-*Hind*III sites, in the same orientation as the CAT gene. The most 3' part of the sequences of these fragments are shown in Figure 7.

Cell cultures

Both the human hepatoma cell line, Hep G2 (Knowles *et al.*, 1980) and the human carcinoma line HeLa B were cultured as monolayers in Dulbecco's modified Eagle's medium 10% (v/v) fetal calf serum (heat-inactivated at 56°C for 30 min) supplemented with glutamine (5 mM final concentration) and penicillin/streptomycin (100 U/ml). The medium was renewed every 2 days and confluent monolayers were split by trypsinization. The doubling time was ~ 12 h for HeLa B cells and 30 h for Hep G2 cells.

Transfections and CAT assays

pSV2-CAT, pA10-CAT-Sma or pA10-CAT-RBP1, 2 and 3 were transfected into human cells by the calcium phosphate precipitation technique (Graham and van der Eb, 1973) at a concentration of 10 μg/5 x 10^5 cells. 48 h after washing out the DNA-calcium precipitate, cells were harvested and CAT enzymatic activity was assayed on cellular extracts according to Gorman *et al.* (1982).

RNA extraction from tissues and cells

RNA extraction from tissues and from cells was performed as previously described (Costanzo *et al.*, 1983; Chirgwin *et al.*, 1979).

Acknowledgements

We wish to thank Dr. Rainer Frank for providing us with the synthetic oligonucleotides, Dr. Carol Featherstone for her comments on the manuscript and Heide Seifert for secretarial help. The work done in Italy was supported by a Grant from P. F. Ingegneria Genetica e Basi Molecolari delle Malattie Ereditarie, CNR, Rome, Italy.

References

Alton,N. and Vapnek,D. (1979) *Nature*, **282**, 864-869.

Alvares,A.P. (1982) in Arias,I.M., Popper,H., Schachter,D., Shafritz,D.A. (eds.), *The Liver. Biology and Pathobiology: Oxidative Biotransformation of Drugs*, Raven Press, pp. 265-280.

Banerji,J., Olson,L. and Schaffner,W. (1983) *Cell*, **33**, 729-740.

Bensi,G., Raugei,G., Klefenz,H. and Cortese,R. (1985) *EMBO J.*, **4**, 119-126.

Berk,A.J. and Sharp,P.A. (1977) *Cell*, **17**, 721-732.

Chan,L. (1982) in Arias,I.M., Popper,H., Schachter,D., Shafrityz,D.A. (eds.), *The Liver. Biology and Pathobiology: Hormonal Control of Gene Expression*, Raven Press, pp. 169-184.

Chirgwin,J.M., Przybyla,A.E., MacDonald,R.J. and Rutter,W.J. (1979) *Biochemistry (Wash.)*, **18**, 5294-5299.

Ciliberto,G., Raugei,G., Costanzo,F., Dente,L. and Cortese,R. (1983) *Cell*, **32**, 725-733.

Ciliberto,G., Dente,L. and Cortese,R. (1985) *Cell*, **41**, 531-540.

Colantuoni,V., Romano,V., Bensi,G., Santoro,C., Costanzo,F., Raugei,G. and Cortese,R. (1983) *Nucleic Acid Res.*, **11**, 7769-7776.

Cortese,R., Melton,D.A., Tranquilla,T. and Smith,J.D. (1978) *Nucleic Acids Res.*, **5**, 4593-4611.

Cortese,R., Herland,R. and Melton,D.A. (1980) *Proc. Natl. Acad. Sci. USA*, **77**, 4147-4151.

Costanzo,F., Catagnoli,L., Dente,L., Arcari,P., Smith,M., Costanzo,P., Raugei,G., Izzo,P., Pietropaolo,T.C., Bougueleret,L., Cimino,F., Salvatore,F. and Cortese,R. (1983) *EMBO J.*, **2**, 57-61.

Dente,L., Cesareni,G. and Cortese,R. (1983) *Nucleic Acids Res.*, **11**, 1645-1655.

Dente,L., Ciliberto,G. and Cortese,R. (1985) *Nucleic Acids Res.*, **13**, 3941-3952.

Frischauf,A.M., Lehrach,H., Poustka,A.M. and Murray,N. (1983) *J. Mol. Biol.*, **170**, 827-842.
Gillies,S.D., Morrison,S.L., Oi,V.T. and Tonewaga,S. (1983) *Cell*, **33**, 717-728.
Gitlin,D. and Gitlin,J.D. (1975) in Putnam,F.W. (ed.), *The Plasma Proteins: Fetal and Neonatal Development of Human Proteins*, Vol. **II**, Academic Press, pp. 264-319.
Gorman,C.M., Moffat,L.F. and Howard,B.H. (1982) *Mol. Cell. Biol.*, **2**, 1044-1051.
Graham,F. and van der Eb,A. (1973) *Virology*, **52**, 456-457.
Greengard,O. (1971) in Campbell,P.N. and Dickens,F. (eds.), *Essays in Biochemistry: Enzyme Differentiation in Mammalian Tissues*, Vol. **7**, Academic Press, pp. 195-205.
Gronenborn,B. and Messing,J. (1978) *Nature*, **272**, 375-377.
Knowles,B.B., Howe,C.C. and Aden,D.P. (1980) *Science (Wash.)*, **209**, 497-499.
Kushner,I. (1982) *Ann. N.Y. Acad. Sci.*, **389**, 39-48.
Laimins,L., Khoury,G., Gorman,C., Howard,B.H. and Gruss,P. (1982) *Proc. Natl. Acad. Sci. USA*, **79**, 6453-6457.
Maniatis,T., Fritsch,E.F. and Sambrook,J. (1982) *Molecular Cloning, A Laboratory Manual*, published by Cold Spring Harbor Laboratory Press, NY.
Melloul,D., Aloni,B., Caluo,J., Yaffe,D. and Nudel,U. (1984) *EMBO J.*, **3**, 983-990.
Messing,J. (1981) in Walton,A. (ed.), *3rd Cleveland Symposium on Macromolecules, Recombinant DNA*, Elsevier, pp. 143-153.
Messing,J. (1983) *Methods Enzymol.*, **101**, 28-78.
Newcomer,M.E., Jones,T.A., Aquist,J., Sundelin,J., Eriksson,W., Rask,L. and Peterson,P.A. (1984) *EMBO J.*, **3**, 1451-1454.
Norrander,J., Kempe,T. and Messing,J. (1983) *Gene*, **26**, 101-106.
Ong,D.E. (1984) *J. Biol. Chem.*, **259**, 1476-1482.
Ott,M.-O., Sperling,L., Herbomel,P., Yaniv,M. and Weiss,M.C. (1984) *EMBO J.*, **3**, 2505-2510.
Powers,S.G. and Meister,A. (1982) in Arias,J.M., Popper,H., Schachter,D. and Shafritz,D.A. (eds.), *The Liver. Biology and Pathobiology: Urea Synthesis and Ammonia Metabolism*, Raven Press, pp. 251-263.
Putnam,F.W. (1975) in Putnam,F.W. (ed.), *The Plasma Protein: Perspectives – Past, Present and Future*, Academic Press, pp. 1-55.
Queen,C. and Baltimore,D. (1983) *Cell*, **33**, 741-748.
Sanger,F., Nickeln,S. and Coulson,A.R. (1977) *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.
Scott,R.W., Vogt,T.F., Croke,M.E. and Tilgham,S. (1984) *Nature*, **310**, 562-567.
Seifter,S. and Englard,S. (1982) in Arias,I.M., Popper,H., Schachter,D. and Shafritz,ad.a. (eds.), *The Liver. Biology and Pathobiology: Energy Metabolism*, Raven Press, pp. 219-249.
Walker,M.D., Edlund,T., Boulet,A.M. and Rutter,M.J. (1983 *Nature*, **300**, 557-561.
Winnacker,E.L. and Dorper,T. (1982) in Gassen,H.G. and Lang,A. (eds.), *Chemical and Enzymatic Synthesis of Gene Fragments*, Verlag Chemie, pp. 97-102.

Received on 2 May 1985; revised on 23 May 1985