Human dipeptidyl peptidase IV gene promoter: tissue-specific regulation from a TATA-less GC-rich sequence characteristic of a housekeeping gene promoter

Stephan K. BÖHM, James R. GUM, JR., Roger H. ERICKSON, James W. HICKS and Young S. KIM*

Gastrointestinal Research Laboratory (151M2), Department of Veterans Affairs Medical Center and Departments of Medicine, Pathology and Anatomy, University of California, San Francisco, CA 94121, U.S.A.

The dipeptidyl peptidase IV gene encodes a plasma-membrane exopeptidase that is highly expressed in small intestine, lung and kidney. In order to better understand the mechanisms responsible for this tissue-specific expression we cloned, sequenced and functionally characterized the 5'-flanking region of the human dipeptidyl peptidase IV gene. The first 500 bases of the 5'-flanking sequence constituted an unmethylated CpG island, contained several Sp1-binding sites and lacked a consensus TATA box, all characteristics of gene promoters lacking tissue-specific expression. RNase-protection analysis using both small intestinal and Caco2 cell RNA indicated that the dipeptidyl peptidase IV transcript was initiated from no fewer than six major and 12 minor start sites. The 5'-flanking sequence also exhibited functional promoter activity in transient transfection

experiments. Here, various lengths of the sequence were cloned upstream of a luciferase gene and introduced into cultured cells using lipofectin. A region located between bases –150 and –109 relative to the start of translation was found to be important for high-level promoter activity in both Caco2 and HepG2 cells. Moreover, Caco2 cells and HepG2 cells, which express high levels of dipeptidyl peptidase IV activity, exhibited much higher normalized luciferase activity after transfection than did 3T3, Jurkat or COS-7 cells, which have low enzyme levels. Sodium butyrate was found to increase both enzyme activity and normalized luciferase in HepG2 cells. Thus the dipeptidyl peptidase IV promoter possesses the ability to initiate transcription in a tissue-specific fashion in spite of having the sequence characteristics of a housekeeping gene promoter.

INTRODUCTION

Our laboratory has long been interested in the structure and function of intestinal brush-border-membrane peptidases and other hydrolases [1–4]. Analysis of these enzymes has contributed greatly to our understanding of the overall processes involved in digestion and absorption in the small bowel. The study of the genes that encode these enzymes likewise has the potential to enhance our understanding of the dual processes of gene expression and differentiation in the gut. Dipeptidyl peptidase IV (DPP IV; EC 3.4.14.5) is an especially attractive candidate for studies of gene regulation. This enzyme is a dipeptidase that preferentially cleaves peptides from the N-terminus with the sequence Gly-Pro-Xaa or Ala-Pro-Xaa in the penultimate position. DPP IV is expressed relatively abundantly in several tissues including small intestine, lung and kidney [5]. Moderate levels are found in the liver, associated with bile canaliculi, and in the spleen, where it is probably associated with thymocytes [5–7]. Other tissues have been found to contain either low or undetectable levels of DPP IV [5]. Thus DPP IV is expressed in a select range of tissues. Several functional roles have been proposed for DPP IV, including digestion [1–4], participation in the thymocyte-activation cascade [8, 9], participation in HIV-induced apoptosis [10], interactions with the substratum [11], modulation of hormonal activity [12], and hydrolysis of renal peptides [13].

In the intestine, DPP IV is differentially expressed along the crypt/villus axis, with mature villus cells expressing about seven times that found in the crypts [14]. Moreover, DPP IV levels can be induced by feeding a gelatin (high-proline-containing) diet

[15]. Thus intestinal DPP IV exhibits both inducibility and spatial regulation. As part of our overall goal of characterizing regulatory processes in the gut, we initiated our characterization of the regulatory sequences of the DPP IV gene.

In this paper, we describe the sequence of the 5'-flanking region of the human DPP IV gene. This stretch was shown to have cell-specific promoter activity in transfection experiments using cultured cells. We also identify the promoter regions necessary for high level transcription and show that the DPP IV promoter resides in an unmethylated CpG island. Finally, we demonstrate the presence of multiple start sites for the DPP IV transcript. Thus we have characterized the 5'-region of the human DPP IV gene and its transcripts, demonstrated that this region of the gene has tissue-specific promoter activity and identified a region of the 5'-flanking sequence necessary for this activity.

EXPERIMENTAL

Genomic library screening, subcloning and DNA sequencing

A human placental genomic DNA library in the λFixII vector was obtained from Stratagene. A total of 600 000 plaques were screened in 24 150 mm Petri dishes using a full-length rat DPP IV cDNA as a probe [6]. Standard procedures for hybridization and washing were used [16]. Two positives were obtained and one was determined to contain 5'-terminal sequence by screening with an 822 bp *EcoRI*–*PstI* fragment derived from the 5'-terminal portion of the cDNA. λFixII phage DNA was isolated and the 5'-terminal region was mapped to a single *XbaI*–*XbaI* fragment,

Abbreviations used: DPP IV, dipeptidyl peptidase IV; DMEM, Dulbecco's modified Eagle's medium; FCS, fetal calf serum; PMA, phorbol 12-myristate 13-acetate; PHA, phytohaemagglutinin; SV40, simian virus 40; BRE, butyrate-responsive element; HNF-1, hepatocyte nuclear factor-1.

* To whom correspondence should be addressed.

Table 1 Primers employed for amplifying fragments of the DPP IV sequence by PCR

Name and sequence of the primers are given. Restriction sites introduced by changing DPP IV sequence are underlined.

Name	Sequence	Reaction site
M13A	5'-GCGCGTTGGCCGATTCATT-3'	—
DPP-3	5'-GAATGCCTCCATTGCTTCAC-3'	—
DPP-4	5'-GAGAAGGAGCGCAGGCAGAAGTCACC-3'	—
DPP-7	5'-GAAGGATCCAGCGCAGAAG-3'	<i>Bam</i> HI
DPP-8	5'-ACCCTGCAGCCTCCTAGC-3'	<i>Pst</i> I
DPP-9	5'-TGAGGTTGGAGGATCCGAA-3'	<i>Bam</i> HI
DPP-10	5'-GAAGAGGTGACACAGCTC-3'	<i>Sal</i> I
DPP-11	5'-GGAAGTGAGGATTCAGAGAAG-3'	<i>Bam</i> HI
DPP-12	5'-GCGCTAGCTAATGTTAACTC-3'	<i>Nhe</i> I
DPP-13	5'-CGAGTTAAAGCTTAGTGAGCG-3'	<i>Hind</i> III
DPP-14	5'-GAGGCCGCTAGCCCTTTCC-3'	<i>Nhe</i> I
DPP-15	5'-CTCGGATCCGACAGCCCTG-3'	<i>Bam</i> HI
DPP-16	5'-CTGTGTGAGCTAGCCAAAGTC-3'	<i>Nhe</i> I
DPP-17	5'-GGTACCCCGGCCGGGAT-3'	<i>Kpn</i> I
DPP-18	5'-CGAAACTTGCTAGCGCGCA-3'	<i>Nhe</i> I

which was excised, cloned into M13mp19 and sequenced in both directions using the exonuclease III method [17]. DNA sequences were compiled and analysed using MacVector (International Biotechnologies) and Intelligence Suite software.

Anchored PCR

This was performed as described using the DPP-3 primer (Table 1) and small intestinal total RNA [16]. The resulting product (approx. 500 bp) was digested with *Eco*RI, which cleaves the anchor primer, and cloned into *Eco*RI–*Sma*I-digested M13mp19. A total of 12 plaques was sequenced from their 5'-start site using the M13 universal primer and 11 were found to contain sequences derived from the DPP IV message.

RNase-protection assay

Constructs for riboprobe preparation spanning the putative 5'-terminus of the DPP IV gene were prepared in pBluescript SK(–). Figure 1 contains a map of the constructs. The PCR was used to produce fragments for subcloning using one of the DPP IV (*Xba*I–*Xba*I) M13mp19 clones as template. Standard conditions were used for the PCR [16]. Construct 1 was prepared using the M13A primer and primer DPP-8. Subcloning into pBluescript was performed using the *Pst*I site introduced by primer DPP-8 and an upstream *Sal*I site remaining from the original λ FixII clone. Construct 1 resulted in probes spanning bases –952 to –399 of Figure 2. The second construct covering bases –434 to +45 was amplified using primers DPP-9 and DPP-10. Construct 2 ends 39 bases downstream of the 5'-splice site of the first intron.

³²P-labelled riboprobe preparation and RNase-protection assays were performed as before [18]. RNA from human small intestine and 21-day post-confluent Caco2 cells was isolated by extraction with guanidinium thiocyanate and centrifugation through CsCl [19]. Yeast tRNA, used as a control, was purchased from Sigma. The protected fragments were analysed on a standard sequencing gel using an M13 sequencing reaction as a standard, applying a correction factor of 1.0/0.98 to correct for differences in the mobility of RNA compared with DNA as described [18].

Figure 1 Map of human DPP IV genomic clone λ hD1

M13hD1/2.5 is the designation given to a 2.5 kb *Xba*I–*Xba*I fragment of the genomic clone that was inserted into M13mp19 and sequenced. Sequence 5' of the *Sal*I site in M13hD1/2.5 is part of the λ FixII phage. The relative locations of exons 1 and 2 are shown as filled boxes. Exon 1 has multiple start sites; hence the left edge of the box is jagged. Below the map of M13hD1/2.5 the relative location of the constructs used for RNase-protection assays are given.

DNA blot analysis

DNA was isolated from human peripheral blood leucocytes, human seminal fluid, Jurkat cells and Caco2 cells by digestion with RNase and proteinase K followed by phenol/chloroform extraction [20]. DNA was double-digested with *Hinc*II and *Eag*I, *Hinc*II and *Nae*I or *Hinc*II and *Nar*I, and electrophoresis was conducted on 1% agarose gels. Capillary transfer to Biotrans nylon membranes was performed as before [21]. The blots were hybridized with a probe spanning bases –952 to +1432 in Figure 2 (5'-terminus to an internal *Hinc*II site). Hybridization and washing were performed as previously [21]; the final wash was with 0.1 $\times$ SSC (1 $\times$ SSC is 0.15 M NaCl/0.015 M sodium citrate)/0.1% SDS for 30 min at 55 °C.

Construction of DPP IV promoter/luciferase reporter gene plasmids

Vectors pGL2-Basic (without promoter or enhancer) and pGL2-Enhancer [containing the simian virus 40 (SV40) enhancer] were purchased from Promega. These vectors contain a luciferase reporter gene located immediately downstream of a polylinker for analysis of promoter activity of cloned fragments. In pGL2-Enhancer, the SV40 enhancer is located downstream of the luciferase gene. pGL2-Control (containing the SV40 promoter and enhancer) was also obtained from Promega and was used as a positive control. The pRSV-Luc vector (rous sarcoma virus promoter and enhancer, luciferase reporter) was a gift from Dr. Sandra Tong, VAMC, San Francisco. Genomic DNA fragments containing restriction sites were prepared using PCR. Primers used are listed in Table 1, and the plasmids produced are listed in Table 2. The fidelity of all amplifications was confirmed by sequencing.

Cell culture and transfection

Caco2 cells, 3T3 cells, COS-7 and HepG2 cells were obtained from the American Type Culture Collection. Jurkat cells were a gift from the Division of Immunology, VAMC, San Francisco. Caco2 cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 20% fetal calf serum (FCS), 100 units/ml penicillin and 100 units/ml streptomycin in

- 952 gatcttttattcttccaaaacccttgaaagaactatgagaggcgcgcagatagatgacccattaaagacct
 - 882 gaactcttaatttaggatctacaacttctagctgaatggcttgccaaagtattttattctctgtgtaca
 - 812 gtcttcttctctgtaaaacacagactaaatgcttagagcagcaccctgggaaaaagtcaagttccacat
 - 742 aagtgacttactgtcattatcttctgcatctgtgatgagtggtgaagcctgcaagacaatcctggaaattccct
 - 672 agggagggtgaatccagatactcagctgtccaaaggccattcaagagggaacaatcaccttttctcttgc
 - 602 ctttcatactgcttttagaccctcactgcttctcaataactttcactggcgaagagcggagtcctggg
 - 532 tttcagttcagttcctgctggtgggtgtgtgagtttgccaaagtcctgcccctctctgggtctcgggt
 - 462 tccctcgctgtcagctgaggttgaggagctgaacgcgcagctcatttttagctaaaggaggcaggg
 - 392 tcccgagctcgccggccaggtctgtgcctcagcagcgcgccttccctcccccaggggtctctc
 - 322 cggcccccagctctgtcgcgcgcgcctcagcagcgcctcagcagcgcctcaggggcccctcgcggggt
 - 252 ccccccggcgaggatgccagtgcccgccgcaacgcgcgcctcctcgcgcgcctcctgcccctcagcctgc
 - 182 cggcgccgcctttatatacagcgggctcggcgctcactaatgttaactggggccgaaacttgcacgag
 - 112 gctgctgactccaccccgagcagcgggtgcagcagcgcgcctcctcccgcccgccggaacttgcct
 - 42 CGCTCTCTCTGAACGCTCACTTCGAGGAGACGCCGACGATGAAGtgagtgccgcgccacgtacgt
 + 29 ccccgggcgagctgtggggacctcttgcgcaaacctcccaaatggcccaaaagctggcgaaacgggacc
 + 99 agggactccccctaggcggaagcccaaggccaccctctgacaggtgaaggagcgtccagagctctcgc
 + 169 ttagctttcggaccctggtctctgctaggaccctggcgtgagctttgaatgctccggagagccactcac
 + 239 agaacgcacgcccgaacccggcgaggggtagagaaacacagcctaattgtctcgaaggcacaagttaa
 + 309 gggcaccgccttagagaactccaactggcgactccagacgcccacccccaccccgagccggtctcc
 + 379 ctcttctaagcactccactgcgcaaatcctgctgctcggagcagccccgaccctcctacggaactg
 + 449 gggccacagccttccctctaattggaactgctcgtcgttcgagcACACGCTGGAAGGTTCTCTGGG
 + 519 ACTGCTGGGTGCTGCTGCTGTGTACCATCATCACCGTGGCGTGTCTGCTGAACAAAGGCAGtaag
 + 589 tttgaaacatttgaaagaataggacagctggagctcaggtcaggtccgcttaccacgctaaaggaaa
 + 659 cgagggttttgcgtgggttggaacagtcgcccgagtgtagctgagaaggcttccagttgtgtgtgc
 + 729 tgctttagagcagcagctggatcactcgcggggtcttcagtggttaggctgaaactcttgcggcttt
 + 799 cagctcttgtatgataataatgcatgcttctgaaattaggagagtcgggatttttaaaatttcaa
 + 869 ttttaatttaatttgaatttttaattttacatgtaatttttgagaaaagtttttgggtgtgacgaa
 + 939 aaaggaaaaagcagctgtttaaagagcccaagacagcagtaaggatattctgaaatggaaggaaaaat
 +1009 gagaaagcaagcccaaaatcagtttgagaatgatgttccacagggaacccagactgtttccctca
 +1079 aggggtgtgagaattcagttgggggtggggaggcagggaagtaattgaccaatttagcagatttgcaga
 +1149 gctaaataataaagctggtatacagcataatatttaagaagggtctggaatcagtgatgttttgggtgat
 +1219 tctttttatttaaggacttcatgtcttttttagcttgccataatgtataaacactcacataatggaata
 +1289 ctcaactgtaggatgcaagcaagtgaagactactagtttcagttcccccctggccttacttctcttaaac
 +1359 acttgcttttcttcagcattacacactataaattagtcagccagttaaagcattctctctggaagattt
 +1429 agtcaactgaggaatgtatagtcagtgcttgcaactcaaggagcttattctaga

Figure 2 Sequence of the 5'-flanking region of the human DPP IV gene

Sequence identified as message by anchor PCR is indicated by uppercase lettering. The arrow indicates the initial base of the longest anchor PCR clone obtained. Open and closed circles under the sequence indicate minor and major start sites respectively identified by RNase-protection analysis. As multiple start sites were observed, the first nucleotide of the initiation codon was designated position 1. The sequence of the DPP IV protein encoded is given; the initiation codon is underlined. BRE, butyrate-responsive element. Potential nuclear-protein-binding sites were identified using the algorithms and databases contained in MacVector and Intelligenetics Suite software and are enclosed in boxes.

5% CO₂. Caco2 cells were passaged every 6–8 days. HepG2 cells and 3T3 cells were grown in DMEM supplemented with 10% FCS and antibiotics in 5% CO₂. COS-7 cells were maintained in DMEM supplemented with 20% FCS and antibiotics in 10% CO₂. Jurkat cells were grown in RPMI 1640 supplemented with 10% FCS, 2 mM glutamine and antibiotics in 5% CO₂.

For transfection, adherent cells were seeded in 35 mm dishes at 5×10^6 cells per dish. All adherent cells were transfected at 80% confluence, a density typically obtained 3–5 days after seeding. Plasmids for transfection were prepared by alkaline lysis followed

Table 2 Plasmids used for transfection experiments

The name of the plasmid and the 5'- and 3'-primers employed to generate them are given.

Plasmid	5'-Primer	3'-Primer
pGL-952/-41	M13A	DPP-7
pGL-952/-41E	M13A	DPP-7
pGL-497/-41	DPP-16	DPP-7
pGL-434/-41	DPP-9	DPP-7
pGL-352/-41	DPP-14	DPP-7
pGL-250/-41	DPP-17	DPP-7
pGL-150/-41	DPP-12	DPP-7
pGL-109/-41	DPP-18	DPP-7
pGL-497/-361	DPP-16	DPP-15
pGL-352/-140	DPP-14	DPP-13
pGL-434/-41	DPP-9	DPP-7

by CsCl banding and purification over Sepharose CL-4B columns. Liposome-mediated transfections were performed using lipofectin reagent (Gibco-BRL) as described by Felgner et al. [22]. A given amount of lipofectin and plasmid was each diluted in 100 μ l of serum- and antibiotic-free medium in polypropylene tubes. The solutions were combined and incubated for 15 min at room temperature. Then 800 μ l of serum- and antibiotic-free medium was added and the solution carefully dropped on the cells, which had been washed three times with the same medium. After incubation for a fixed period of time the transfection solution was removed and replaced with serum-containing growth medium. Cells were harvested 48 h after transfection.

The amounts of lipofectin (20–50 μ g) and plasmid (1.5–20 μ g) used in each transfection as well as incubation time (5–20 h) with transfection solution were optimized for each cell line in preliminary experiments. For transfection of suspended cultures of Jurkat cells, the cells were washed once with serum- and antibiotic-free medium and placed in six-well plates at a density of 3×10^6 per well in 800 μ l of serum- and antibiotic-free medium. Luciferase reporter plasmid (5 μ g) and lipofectin (20 μ g) were diluted as described above, combined and incubated for 15 min at room temperature. The transfection mix was then added to the cells. After an incubation period of 5 h, 1 ml of complete growth medium containing twice the normal concentration of serum was added to each well. To stimulate Jurkat cells, 50 mg/ml phorbol 12-myristate 13-acetate (PMA) and 1 μ g/ml phytohaemagglutinin (PHA) were added per well at this point. Cells were harvested after 48 h. Transfected HepG2 cells were stimulated with 5 mM sodium butyrate for 24 h before harvest.

Assays of protein, DPP IV and luciferase

For all assays, cells were harvested by washing three times in Ca²⁺ and Mg²⁺-free PBS and lysed on the plate by the addition of 1 ml of solution containing 1% Triton, 25 mM glycylglycine, pH 7.8, 15 mM MgSO₄, 4 mM EGTA and 1 mM dithiothreitol. The plates were scraped, and the lysates pelleted by centrifugation at 12000 *g* for 5 min at 4 °C. The supernatants were decanted and briefly vortexed before assay.

Protein was measured by the method of Lowry et al. [23]. DPP IV activity was measured with glycyl-proline- β -naphthylamide as substrate [24]. A unit of enzyme activity is equivalent to 1 μ mol of substrate hydrolysed/min. Analysis of luciferase activity was performed as described [25] using a Monolith 2001 TM Luminometer from Analytical Luminescence Laboratory (San Diego, CA, U.S.A.).

Figure 3 RNase-protection analysis

Probe preparation, hybridization, digestion and electrophoresis was performed as described in the Experimental section. Yeast tRNA (negative control), small intestinal RNA and Caco2 cell RNA were examined. A standard M13mp18 sequencing reaction was included for size standards. Probes used are indicated as 1-, 1+, 2- or 2+ representing antisense (-) or sense (+) probes made from constructs 1 or 2. The scale indicates initiation sites for transcription calculated from the size of the protected fragments.

RESULTS

Cloning and sequence analysis of the 5'-region of the human DPP IV gene

The screening described in the experimental section resulted in the isolation of a single clone that hybridized to the rat 822 bp *EcoRI*-*PstI* fragment. This clone spanned 18 kb but a single *XbaI*-*XbaI* fragment of 2.5 kb was found to contain all sequences that hybridized to the 5'-probe (Figure 1). This fragment was cloned into M13mp19 in both directions for sequencing (Figure 2). Comparison of this sequence with the human DPP IV coding sequence [26] indicated that we had obtained a portion of the DPP IV gene apparently spanning the first two exons and

Figure 4 CpG island characterization

(a) Graph of percentage C + G content (—) and CpG frequency (· · ·) per 200 bp segment (step of 10 bp) through the DPP IV genomic clone M13hD2.5. Zero on the abscissa marks the start of translation. (b) Observed over expected (O/E) frequency of CpG per 200 bp. The ratio was calculated as follows: (number of CpG $\times$ N/number of C $\times$ number of G), where N is the total number of nucleotides in the sequence being analysed. Moving average values for % (G + C) and for O/E CpG were calculated using a 100 bp window ($N = 100$) moving along the sequence at 10 bp intervals. (c) Methylation of CpG enzyme sites in DPP IV genomic DNA. The line at the top of the map represents the DPP IV genomic clone M13hD2.5, with the locations of exons 1 and 2 indicated. The 2.4 kb *HincII*-*HincII* probe described in the Experimental section is shown as the thick line. The expected restriction fragments using three different methylation-sensitive enzymes and *HincII* as a 3'-limiting enzyme are indicated for each enzyme.

flanking sequences. We next performed anchor PCR using primer DPP-3 (Table 1) (which spans bases 380–361 of the human cDNA coding sequence [26]) in order to obtain the extreme 5'-sequence of the human message. The longest cDNA thus obtained extended 110 bases upstream of the initiation codon (Figure 2). Other clones were initiated at bases -21, -53, -63, -68, -71 and -106 (three individual clones each initiated at both base -53 and -71). Thus the first exon contains an extended 5'-untranslated region and encodes only two amino acids. The first intron spans 489 bases and is flanked by consensus splice sites. The second exon encodes 29 amino acids, consisting primarily of the putative transmembrane segment identified previously [6]. The second intron is apparently larger than 16 kb, as the third exon was not present in the λ phage clone. The structure presented here for the first two exons and introns is consistent with the structure of the DPP IV gene deduced by Abbott et al. [27].

The 5'-flanking region of the DPP IV gene contained potential binding sites for many transcription factors, as indicated in Figure 2. These include several GC Boxes (Sp1-binding sites) and AP2-binding sites. No consensus TATA box was present, although two AT-rich TATA-like sequences were found. Other potentially important transcription-factor-binding sites were found including those that many play a role in cell proliferation and differentiation (*c-myc*, *c-myb*) [28,29], gene induction

Figure 5 Southern-blot analysis of genomic DNA digests with methylation-sensitive restriction enzymes

(a) Southern blot of genomic DNA isolated from human blood leucocytes (lane 1), human spermatozoa (lane 2), Jurkat cells (lane 3) and Caco2 cells (lane 4) probed after double digestion with *EagI* + *HincII* (left half) and *NaeI* + *HincII* (right half). The sizes of DNA standards are indicated on the left side, the sizes of the detected fragments on the right side. (b) Southern blot of genomic DNA [same sources as in (a)] probed after double digestion with *NarI* + *HincII*. The probe was described in Figure 4(c).

[butyrate-responsive element (BRE and NK κ B)] [30,31] or tissue-specific gene expression [hepatic nuclear factor-1 (HNF-1)] [32].

Identification of transcriptional start sites

Analysis by anchor PCR suggests that the DPP IV gene expresses a transcript with an extended 5'-untranslated region. In order to examine this in more detail, we performed RNase-protection analysis using riboprobes spanning the entire cloned 5'-untranslated region of the DPP IV gene (Figure 3). Construct 1 spanned bases -952 to -399 and construct 2 spanned bases -434 to +45 (Figure 1). Total RNA from both human small intestine and Caco2 cells was used as were sense and antisense riboprobes. Yeast tRNA was used as a negative control. As can be seen in Figure 3, no protected fragments were observed with either sense or antisense riboprobes from construct 1 or with sense riboprobes from construct 2. In contrast, at least 18 protected fragments of various intensities and sizes were obtained when antisense probe from construct 2 was used. No bands were observed with the yeast tRNA control. The pattern obtained was similar with both small intestine and Caco2 cell RNA. The most intense band appeared at nucleotide -92 of Figure 2. Other bands of major intensity appeared at positions -109, -106, -105, -93 and -75, as indicated in Figure 2. Other extremely faint bands can be discerned as well, although these are barely above background and their significance is unclear. These data

demonstrate that DPP IV gene transcription is initiated from multiple sites in a region spanning approx. 65 nucleotides.

CpG island characterization

The finding of numerous GC-rich transcription-factor-binding sites within the 5'-flanking region of the DPP IV gene prompted us to evaluate the DPP IV genomic sequence for the presence of a CpG island. CpG islands are regions containing an unusually high content of unmethylated CpG dinucleotides. These regions frequently coincide with gene promoters, and especially with promoters that are active in a variety of cell types, i.e. those present in housekeeping genes [33,34]. Empirically, CpG islands are defined as unmethylated regions larger than 200 bp with an average GC content above 50% and a ratio of observed CpG doublets to expected CpG doublets greater than 0.6 [33-35] (CpG doublets are rare in vertebrate DNA [33]). Figure 4(a) shows the moving average percentage GC content and CpG dinucleotide content for the 2.5 kb DPP IV gene sequence. It is apparent that the region immediately upstream of the DPP IV coding sequence contains both a high GC content and significant numbers of CpG dinucleotides. Figure 4(b) shows the observed-to-expected ratio for the rare CpG dinucleotide. Again, a broad peak in the observed-to-expected CpG content occurs in this region. Thus the 5'-flanking sequence of the DPP IV gene

Table 3 Transfection of different cell lines

The human DPP IV 5'-flanking region of -952/-41 relative to the translational start site was either linked to the luciferase reporter gene in pGL2-Basic (pGLB) or in pGL2-Enhancer (pGLE), which contains the SV40 enhancer, as described in the Experimental section. These constructs (pDPPIVGLB and pDPPIVGLE) were used for transfection. As a positive control (pGLC), either the pGL2-Control vector (3T3, COS-7, Caco2, HepG2) or the pRSV-Luc vector (Jurkat) was used. Cells were also transfected with the promoter-less constructs pGL2-Basic and pGL2-Enhancer and a mock transfection was performed for every set of transfections. The values for luciferase activity are expressed as light units per mg of protein; approx. 0.1 mg of protein was used per assay. The ratios of activities of the promoter-containing vectors (pDPPIVGLB, pDPPIVGLE, or pGLC) to the appropriate controls (pDPPIVGLB, pDPPIVGLE, or pGLC) are also given. + after Jurkat and HepG2 cells indicates that these cells were stimulated with PMA/PHA (5 h after transfection for about 43 h) and *n*-butyrate (24 h after transfection for 24 h) respectively. DPP IV enzyme activity and protein concentration were determined as described in the Experimental section. All values are expressed as means $\pm$ S.E.M. for the number of individual experiments given in parenthesis.

Cell line	DPP IV activity nmol/min per mg	pDPPIVGLB (units/mg)	pGLB (units/mg)	pDPPIVGLB/ pGLB	pDPPIVGLE (units/mg)	pGLE (units/mg)	pDPPIVGLE/ pGLE	pGLC (units/mg)	pGLC/ pGLE
3T3	4.5 $\pm$ 0.1 (15)	1180 $\pm$ 80 (3)	630 $\pm$ 20 (3)	1.9	2310 $\pm$ 30 (3)	870 $\pm$ 20 (3)	1.9	13280 $\pm$ 1220 (3)	15.3
COS-7	3.3 $\pm$ 0.1 (15)	19260 $\pm$ 4180 (3)	3400 $\pm$ 590 (3)	5.7	142990 $\pm$ 11530 (3)	9920 $\pm$ 1900 (3)	14.4	1858060 $\pm$ 41810 (3)	187.3
Jurkat	2.5 $\pm$ 0.4 (10)	1180 $\pm$ 370 (2)	910 $\pm$ 210 (2)	1.3	1040 $\pm$ 300 (2)	990 $\pm$ 410 (2)	1.1	6110 $\pm$ 850 (2)	6.2
Jurkat +	1.4 $\pm$ 0.2 (10)	1770 $\pm$ 530 (2)	1200 $\pm$ 370 (2)	1.1	1490 $\pm$ 270 (2)	1300 $\pm$ 450 (2)	1.1	36360 $\pm$ 1870 (2)	24.4
Caco2	103.7 $\pm$ 2.2 (28)	199750 $\pm$ 39080 (5)	6420 $\pm$ 1100 (5)	31.1	551790 $\pm$ 112710 (5)	19350 $\pm$ 1850 (5)	28.5	1329290 $\pm$ 126230 (6)	68.7
HepG2	87.2 $\pm$ 5.7 (29)	6030 $\pm$ 960 (9)	410 $\pm$ 30 (9)	14.9	23550 $\pm$ 5480 (6)	610 $\pm$ 50 (6)	38.6	41300 $\pm$ 6840 (6)	67.7
HepG2 +	95.7 $\pm$ 3.6 (28)	18260 $\pm$ 6040 (6)	400 $\pm$ 10 (6)	45.7	20000 $\pm$ 2770 (6)	550 $\pm$ 60 (6)	36.4	63340 $\pm$ 15030 (4)	115.2

contains (i) a high GC content and (ii) significant numbers of CpG dinucleotides, well above the defining observed-to-expected value of 0.6 for a span of approx. 500 nucleotides.

Having established that the sequence fits the criteria for a CpG island, we next sought to determine whether this region contains unmethylated CpG doublets. Here, genomic DNA from human blood leucocytes, spermatozoa, Jurkat cells and Caco2 cells was first digested to completion with *HincII*. This material was then digested with several enzymes that fail to cut when their recognition site includes a methylated CpG doublet [36]. These enzymes were chosen on the basis of having recognition sites at strategic locations within the DPP IV gene (Figure 4c). The digested DNA was then subjected to blot analysis (Figure 5) using the probe described in the Experimental section (see also Figure 4c). With every enzyme and with every DNA sample, a band of the predicted size was observed (1678 bp for *EagI*, 1811 bp for *NaeI* and 1608 bp for *NarI*; Figure 5). This indicates that CpG doublets contained within the recognition sites are unmethylated in all cases. Thus the DPP IV 5'-flanking sequence clearly resides within a substantially unmethylated CpG island.

Functional analysis of the 5'-flanking region of the human DPP IV gene

We next sought to characterize possible functional promoter activity of the DPP IV 5'-flanking sequence. For this purpose, we prepared a variety of 5'-flanking sequence/luciferase reporter gene constructs using pGenelight (Promega) vectors (Table 2). Transient transfections were then performed as described in the Experimental section. We first transfected several cell lines with constructs containing bases -952 to -41 of Figure 2. These constructs, prepared using vectors both with and without the SV 40 enhancer, stretched from the 5'-terminus of the cloned segment to nucleotide -41, thus terminating downstream of all transcription start sites detected by RNase-protection analysis (Figure 3). We also performed DPP IV activity assays in these cell lines; moreover we confirmed that all tested lines could be transfected using positive control vectors.

Cell lines 3T3, COS-7 and Jurkat were found to express low levels of DPP IV activity. These cell lines also expressed only low levels of luciferase activity after transfection with pGL2-Basic containing the DPP IV 5'-flanking region, as compared with results obtained with the promoter-less vector (Table 3). This can be observed most conveniently in the column giving the ratio pDPPIV/pGLB. Results with the positive control vectors are also given for comparison. Slightly higher levels of activity were obtained with the constructs containing the DPP IV 5'-flanking region cloned into the pGL2-Enhancer vector, although the levels obtained were still low relative to those obtained with the positive control vectors. No increase in DPP IV promoter activity was observed in Jurkat cells stimulated with PMA/PHA, although significant activation was observed with the control vector in these cells. Caco2 and HepG2 cells, on the other hand, expressed high levels of DPP IV activity (Table 3). These cell lines also expressed high levels of luciferase activity with the DPP IV/pGL2-Basic construct, thus mirroring the enzyme activity measurements. In particular, DPP IV activity was induced approximately five times in HepG2 cells by 4 days treatment with 5 mM sodium butyrate (Table 4). Similarly, an approximate 3-fold increase in transcriptional activity was observed with the DPP IV/pGL2-basic construct when 5 mM sodium butyrate was included in the medium after transfection (Table 3). Thus cell cultures that express high levels of DPP IV also express high levels of transcriptional activity, suggesting that the DPP IV 5'-flanking region harbours a tissue-specific promoter.

Table 4 Induction of DPP IV enzyme activity in HepG2 cells by n-butyrate

HepG2 cells were seeded at a density of 500 000 cells per 35 mm dish and grown for 24 h in regular medium. After that period, cells were either maintained as controls or 5 mM sodium butyrate was added. Medium was changed every 24 h thereafter. Cells were harvested after 1, 2, 3, 4 and 5 days of treatment and cell lysates were assayed for protein concentration and DPP IV enzyme activity. Each value represents three dishes and values are given as means $\pm$ S.E.M. Asterisks mark values that differ significantly from controls in Student's *t*-test (**P* < 0.01, ***P* < 0.001).

Days of treatment	DPP IV activity (nmol/min per mg of protein)	
	– Butyrate	+ Butyrate
1	23.9 $\pm$ 5.9	25.4 $\pm$ 1.0
2	30.4 $\pm$ 5.7	44.2 $\pm$ 1.4
3	41.7 $\pm$ 2.3	79.7 $\pm$ 7.6*
4	49.8 $\pm$ 0.5	143.2 $\pm$ 9.4**
5	53.9 $\pm$ 2.3	143.2 $\pm$ 10.3**

Next we generated a set of deletion constructs as described in the Experimental section and transfected them into Caco2 and HepG2 cells. In Caco2 cells the maximum luciferase activity was produced by construct pGL–497/–41 (Figure 6). It reached

43-fold greater activity than the promoter-less construct, pGL2-Basic. In HepG2 cells the highest luciferase activity was achieved by construct pGL–352/–41, yielding 80-fold greater activity than pGL2-Basic. The data shown in Figure 6 indicate that luciferase activity fell off precipitously going from construct pGL–150/–41 to construct pGL–109/–41 in both cell lines, perhaps more so with HepG2 cells than with Caco2 cells. The region between base –150 and base –109 includes a potential binding site for HNF-1 and the proximal TATA-like sequence. With longer constructs, differences in transcriptional regulation between Caco2 and HepG2 cells became evident. Addition of bases –434 to –353 resulted in a marked decrease in transcription in HepG2 but not in Caco2 cells (Figure 6), indicating the possible presence of a repressor in one cell line but not in the other. Constructs not ending in the 5'-untranslated region produced only very low activity levels with both cell lines [constructs pGL–497/–361, pGL–352/–140 and pGL–41/–434 (reverse orientation)]; Figure 6].

We also examined the effects of sodium butyrate on the promoter activity of the various constructs in HepG2 cells (Table 5). Under these conditions we found that two of the shortest constructs, pGL–250/–41 and pGL–150/–41, had the highest activity. Addition of bases –352 to –251 actually had a negative effect on sodium butyrate-induced activity. Activity in the presence of sodium butyrate with the longer constructs paralleled activity observed in the absence of this compound.

Figure 6 Transfection of deletion constructs into Caco2 and HepG2 cells

The different deletion constructs of the DPP IV 5'-flanking region were generated and linked to the luciferase reporter gene as described in the Experimental section and are shown in the left panel with their respective 5' and 3' end relative to the translational start site. Above each construct, potential binding sites for transcriptional factors are indicated as boxes. Transfections were conducted as described in the Experimental section. Values are expressed as means $\pm$ S.E.M. □, Caco2 cells; ■, HepG2 cells.

Table 5 Transfection of deletion constructs into HepG2 cells stimulated with sodium butyrate

The experimentation was performed as described in the legend of Figure 6 except that 24 h after transfection half the dishes were treated with medium containing 5 mM sodium butyrate, the other half with control medium. Incubation was continued for another 24 h before harvesting and assay. Results are expressed as activity with a given construct divided by the activity of pGL-Basic ('fold pGL-Basic').

Construct	Luciferase activity (fold pGL-Basic)		Fold increase
	– Butyrate	+ Butyrate	
–999/–41	37.3 ± 6.6	222.5 ± 51.7	6.0
–497/–41	52.2 ± 6.6	235.8 ± 69.2	4.5
–434/–41	25.8 ± 7.3	120.4 ± 65.6	4.7
–352/–41	79.6 ± 18.8	210.3 ± 57.3	2.6
–250/–41	47.5 ± 13.0	362.7 ± 115.7	7.6
–150/–41	76.5 ± 22.9	341.3 ± 100.6	4.5
–109/–41	4.1 ± 1.4	7.1 ± 1.6	1.7

DISCUSSION

DPP IV is expressed at various levels in a number of seemingly disparate tissues. As such, the factors and mechanisms that control DPP IV gene expression are likely to be complex. In the present study, we isolated a genomic DNA clone containing the 5'-flanking region of the DPP IV gene. This enabled us to establish that this segment of the gene has promoter activity in cells that express DPP IV enzyme and to otherwise characterize the transcriptional regulation of DPP IV.

It is clear from Figure 4 that the 5'-flanking region of the DPP IV gene is GC-rich. The first 500 bases upstream of the initiation codon contains 71% G+C residues. Moreover, it contains a high content of the rare CpG dinucleotide (Figure 4). Southern-blot analysis using methylation-sensitive restriction enzymes indicates that at least some of these dinucleotides are unmethylated in all four cell types tested. Thus the DPP IV promoter resides within an unmethylated CpG island. This is characteristic of the promoters found in housekeeping genes, i.e. genes that are expressed in almost all cell types [33]. The DPP IV 5'-flanking sequence has other features of a housekeeping gene promoter [37,38]. It contains several Sp1-binding sites and lacks a consensus TATA box. These characteristics may play a role in the expression of the DPP IV transcript in multiple cell types.

Although the DPP IV promoter lacks a consensus TATA box, it contains two AT-rich sequences that are deemed to be TATA-like. The proximal TATA-like sequence (5'-TTTAACTC) may well function to bind transcription factor IID (TATA-binding protein [39–42]. Deletion of a 41 bp segment containing this region leads to a sharp diminution of promoter activity in transient transfection experiments (Figure 6). Also, the presence of major transcription start sites at positions 24, 27 and 28 bases downstream of this motif (Figures 2 and 3) suggests that it may serve to position the transcription initiation complex as does a conventional TATA box [39]. Still, the absence of a strong TATA box may lead to a large number of transcription initiation sites seen in Figure 3. Other promoters without consensus TATA boxes also exhibit multiple start sites [43–45].

While this study was in progress, Abbott et al. [27] published a description of the structure of the human DPP IV structural gene. It was shown to contain 26 exons. Notably, these authors found the second intron to be large, i.e. not entirely encompassed by two λ phage clones, one containing exon 2 and the other

containing exon 3. This structure is consistent with our work which indicates that the second intron is at least 16 kb (Figure 1). Abbott et al. also published a sequence for 299 bases of 5'-flanking sequence. Our sequence differs significantly from this sequence with 30 base substitutions and eight gaps of one or two nucleotides. It is not known if this represents a polymorphism. Our sequence reading was determined in both directions using single-stranded M13 templates and was clear.

Figure 2 indicates that the 5'-flanking sequence of the DPP IV gene contains potential binding sites for many transcription factors. Several of these sites may be functionally important. NF κ B was originally described as a B-lymphocyte transcription regulator. Later work demonstrated its presence in several cell types including T-lymphocytes and HepG2 cells [46,47]. Thus NF κ B may be important in maintaining the high constitutive levels of DPP IV found in these cell types. Moreover, NF κ B levels appear to be modulated in several cell types and to play a role as a second messenger in certain processes including T-cell activation [46]. Interestingly, immunohistochemical analysis indicates that DPP IV levels increase during T-cell activation [9]. Thus alterations in NF κ B levels may play a role in regulating DPP IV levels *in vivo* or during induction in cultured cell systems.

There are four potential AP2-binding sites in the 5'-flanking region of the DPP IV gene (Figure 2). AP2 was originally identified as a transcription factor that bound the enhancer regions of SV40 and the human metallothionein IIA gene [48]. Subsequent work has revealed its interaction with many promoters and its presence in many cell types. Surprisingly, however, AP2 has been reported to be absent from HepG2 cells [49]. At least one AP2 site is present in every forward construct used in Figure 6. This may at least partially explain the different activity profiles obtained with the various deletion constructs after transfection into HepG2 compared with Caco2 cells (Figure 6).

Two BREs as described by Fregeau et al. [30] were observed in Figure 2. Interestingly, however, even the –150/–41 construct (which lacks a BRE) possesses butyrate-inducible promoter activity in HepG2 cells (Table 5). This suggests that the BREs defined above play a rather insignificant role in mediating DPP IV induction by butyrate, at least in HepG2 cells. Other promoters have been described that contain butyrate-responsive DNA segments [50–52] different from the one described by Fregeau et al. [30]. Thus it is clear that several, perhaps many, BREs exist. Others have suggested that butyrate exerts its pleiotropic effects through general alterations in chromatin structure [53], histone acetylation [53–55], histone phosphorylation [56] and/or phosphorylation of high-mobility-group proteins [57]. Our data do not allow us to distinguish between possible mechanisms of action of butyrate. It is also of interest that butyrate had no discernible effect on the DPP IV/pGLE construct in HepG2 cells (Table 3). Previous work has ascribed the SV40 enhancer as being itself sensitive to butyrate [58]. Perhaps this effect can be mediated only in conjunction with certain promoters.

Finally, a potential binding site for HNF-1 is apparent in Figure 2 located between bases –147 and –135. This sequence matches 10 of 13 nucleotides of the HNF-1 consensus sequence [32]. Moreover, it is contained within the minimal region of the promoter needed for high-level expression in both Caco2 and HepG2 cells (Figure 6). This element, once thought to be liver-specific, is now known to play a role in gene regulation in many tissues including intestine, liver, kidney and lung [59,60]. Recently, HNF-1 sites were shown to be important in regulating the intestine-specific expression of the sucrase-isomaltase gene [61]. The proximal positioning of this sequence in the DPP IV

promoter overlapping a TATA-like site is interesting. Clearly, the possibility that HNF-1 may modulate DPP IV gene expression merits further investigation.

In summary, the DPP IV 5'-flanking sequence has many characteristics of a housekeeping gene promoter including its residence in an unmethylated CpG island, the presence of Sp1-binding sites and the absence of a consensus TATA box. We demonstrated that this promoter initiates transcription from at least 18 start sites over a 65 bp region. Moreover, the 5'-flanking region contains consensus binding sites for several transcription factors that may be functionally important including NF- κ B, AP2 and HNF-1. Thus the DPP IV 5'-flanking region exhibits widely varying promoter activity which mirrors DPP IV enzymic activity in several cell types. On the other hand, this region has the physical characteristics of a housekeeping gene promoter. Future studies will be directed towards determining the mechanisms responsible for achieving specificity in such an interesting setting.

This work was supported by the National Institutes of Health Grant DK17938 and by the Department of Veterans Affairs Medical Research Service.

REFERENCES

- Morita, A., Chung, Y. C., Freeman, H. J., Erickson, R. H., Slesinger, M. H. and Kim, Y. S. (1983) *J. Clin. Invest.* **72**, 610–616
- Erickson, R. H., Bella, A. M., Brophy, E. J., Kobata, A. and Kim, Y. S. (1983) *Biochim. Biophys. Acta* **756**, 258–265
- Erickson, R. H. and Kim, Y. S. (1990) *Annu. Rev. Med.* **41**, 133–139
- Erickson, R. H., Suzuki, Y., Sedlmayer, A. and Kim, Y. S. (1992) *J. Biol. Chem.* **267**, 21623–21629
- Hong, W., Petell, J. K., Swank, D., Sanford, J., Hixson, D. C. and Doyle, D. (1989) *Exp. Cell Res.* **182**, 256–266
- Ogata, S., Misumi, Y. and Ikehara, Y. (1989) *J. Biol. Chem.* **264**, 3596–3601
- Marguet, D., Bernard, A.-M., Vivier, I., Darmoul, D., Naquet, P. and Pierres, M. (1992) *J. Biol. Chem.* **267**, 2200–2208
- Schon, E., Jahn, S., Kiesel, S. T. et al. (1987) *Eur. J. Immunol.* **17**, 1821–1826
- Gorrell, M. D., Wickson, J. and McCaughan, G. W. (1991) *Cell. Immunol.* **134**, 205–215
- Morimoto, C., Lord, C. I., Zhang, C., Duke-Cohan, J. S., Letvin, N. L. and Schlossman, S. F. (1994) *Proc. Natl. Acad. Sci. U.S.A.* **91**, 9960–9964
- Piazza, G. A., Callanan, H. M., Mowery, J. and Hixson, D. C. (1989) *Biochem. J.* **262**, 327–334
- Frohman, L. A., Downs, T. R., Heimer, E. P. and Felix, A. M. (1989) *J. Clin. Invest.* **83**, 1533–1540
- Tiruppathi, C., Miyamoto, Y., Gunapathy, V., Roesel, R. A., Whitford, G. M. and Leibach, F. H. (1990) *J. Biol. Chem.* **265**, 1476–1483
- Darmoul, D., Rouyer-Fessard, C., Blais, A. et al. (1991) *Am. J. Physiol.* **261**, G763–G769
- Suzuki, Y., Erickson, R. H., Sedlmayer, A., Chang, S.-K., Ikehara, Y. and Kim, Y. S. (1993) *Am. J. Physiol.* **264**, G1153–G1159
- Gum, J. R., Hicks, J. W., Toribara, N. W., Rothe, E. M., Lagace, R. E. and Kim, Y. S. (1992) *J. Biol. Chem.* **267**, 21375–21383
- Henikoff, S. (1984) *Gene* **28**, 351–359
- Gum, J. R., Hicks, J. W., Toribara, N. W., Siddiki, B. and Kim, Y. S. (1994) *J. Biol. Chem.* **269**, 2440–2446
- Chirgwin, J. M., Przybyla, A. F., MacDonald, R. J. and Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299
- Borello, M. G., Pierotti, M. A., Tamborini, E. et al. (1992) *Oncogene* **7**, 269–275
- Gum, J. R., Byrd, J. C., Hicks, J. W., Toribara, N. W., Lampion, D. T. A. and Kim, Y. S. (1989) *J. Biol. Chem.* **264**, 6480–6487
- Felgner, P. L., Gadek, T. R., Holm, M. et al. (1987) *Proc. Natl. Acad. Sci. U.S.A.* **84**, 7413–7417
- Lowry, O. H., Rosebrough, N. J., Farr, A. L. and Randall, R. J. (1951) *J. Biol. Chem.* **193**, 265–275
- Bella, A. M., Jr., Erickson, R. H. and Kim, Y. S. (1982) *Arch. Biochem. Biophys.* **218**, 156–162
- Brasier, A. R., Tate, J. E. and Habener, J. F. (1989) *BioTechniques* **7**, 1116–1122
- Darmoul, D., Lacasa, M., Baricault, L. et al. (1992) *J. Biol. Chem.* **267**, 4824–4833
- Abbott, C. A., Baker, E., Sutherland, G. R. and McCaughan, G. W. (1994) *Immunogenetics* **40**, 331–338
- Lüscher, B. and Eisenman, R. N. (1990) *Genes Dev.* **4**, 2025–2035
- Roy, A. L., Carruthers, C. and Roeder, R. G. (1993) *Nature (London)* **365**, 359–361
- Fregeau, C. J., Helgason, C. D. and Bleackley, R. C. (1992) *Nucleic Acids Res.* **20**, 3113–3119
- Baeuerle, P. A. (1991) *Biochim. Biophys. Acta* **1072**, 63–80
- Mendel, D. B. and Crabtree, G. R. (1991) *J. Biol. Chem.* **266**, 677–680
- Bird, A. P. (1986) *Nature (London)* **321**, 209–213
- Bird, A. P. (1992) *Cell* **70**, 5–8
- Gardiner-Garden, M. and Frömmer, M. (1987) *J. Mol. Biol.* **196**, 261–282
- Lindsay, S. and Bird, A. P. (1987) *Nature (London)* **327**, 336–338
- Seghal, A., Patil, N. and Chao, M. (1988) *Mol. Cell. Biol.* **8**, 3160–3167
- Smale, S. T. and Baltimore, D. (1989) *Cell* **57**, 103–113
- Drapkin, R., Merino, A. and Reinberg, D. (1993) *Curr. Opin. Cell. Biol.* **5**, 469–476
- Brandl, C. J., Martens, J. A., Liaw, P. C.-Y., Furlanetto, A. M. and Wobbe, C. R. (1992) *J. Biol. Chem.* **267**, 20943–20952
- Wiley, S. R., Kraus, R. J. and Mertz, J. E. (1992) *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5814–5818
- Yoganathan, T., Horikoshi, M., Roeder, R. G. and Sells, B. H. (1993) *FEBS Lett.* **326**, 163–166
- Corden, J., Wasyluk, B., Buchwalder, A., Sassone-Corsi, P., Kedinger, C. and Chambon, P. (1980) *Science* **209**, 1406–1414
- Grosschedl, R. and Birnstiel, M. L. (1980) *Proc. Natl. Acad. Sci. U.S.A.* **77**, 1432–1436
- McKnight, S. L. and Kingsbury, R. (1982) *Science* **217**, 316–324
- Leonardo, M. J. and Baltimore, D. (1989) *Cell* **58**, 227–229
- Malter, J. S. and Gerber, M. S. (1989) *Hepatology* **10**, 1017–1018
- Mitchell, P. J., Wang, C. and Tijan, R. (1987) *Cell* **50**, 847–861
- Imagawa, M., Chiu, R. and Karin, M. (1987) *Cell* **51**, 251–260
- Bohan, C. A., Robinson, R. A., Luciw, P. A. and Srinivasan, A. (1989) *Virology* **172**, 573–583
- Tang, D.-C. and Taylor, M. W. (1990) *J. Virol.* **64**, 2907–2911
- Gill, R. K. and Christakos, S. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 2984–2988
- Vidali, G., Boffa, L. C., Bradbury, E. M. and Allfrey, V. G. (1978) *Proc. Natl. Acad. Sci. U.S.A.* **75**, 2239–2243
- Riggs, M. G., Whittaker, R. G., Neumann, J. R. and Ingram, V. M. (1977) *Nature (London)* **268**, 462–464
- Sealy, L. and Chalkley, R. (1978) *Cell* **14**, 115–121
- Boffa, L. C., Gruss, R. J. and Allfrey, V. G. (1981) *J. Biol. Chem.* **256**, 9612–9621
- Levy-Wilson, B. (1981) *Proc. Natl. Acad. Sci. U.S.A.* **78**, 2189–2193
- Gorman, C. M., Howard, B. H. and Reeves, R. (1983) *Nucleic Acids Res.* **11**, 7631–7649
- Kuo, C. J., Conley, P. B., Hsieh, C. L., Franke, U. and Crabtree, G. R. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 9838–9842
- Baumhueter, S., Mendel, D. B., Conley, P. B. et al. (1990) *Genes Dev.* **4**, 372–379
- Wu, G. D., Chen, L., Forslund, K. and Traber, P. G. (1994) *J. Biol. Chem.* **269**, 17080–17085