Molecular cloning, sequencing and functional study of the promoter region of the human α_2C4 -adrenergic receptor gene

Stéphane SCHAAK, Jean-Christophe DEVEDJIAN, Cécile CAYLA, Yolande SENDER and Hervé PARIS¹

Institut National de la Santé et de la Recherche Médicale U.317, Institut Louis Bugnard, CHU Rangueil, 31403 Toulouse Cedex 4, France

Screening of a human foetal brain genomic DNA library allowed us to isolate an *EcoRI*–*EcoRI* fragment containing 6 kb of the 5′-flanking region, the open reading frame and 4 kb of the 3′-flanking region of the α_2C4 gene. Analysis of the sequenced region (4850 bp) revealed that the first 900 bp 5′ to the start codon are very rich in GC (84%), contain several Sp1-binding sites and lack a consensus TATA box. The 5′- and 3′-ends of the α_2C4 transcript were determined by RNase-protection assays carried out with a series of antisense probes. The data obtained with cellular RNA from HepG2 cells demonstrated that transcription is initiated 891 bases upstream of the translation-start site and that the polyadenylation site is located 550 bases downstream of the stop codon. These results are consistent with the existence of a non-conventional TATA box (TTAGAAA)

and the presence of a unique polyadenylation signal (AATAAA). They also fit with the size of α_2C4 -RNA found by Northern-blot analysis (2.9 kb). The transcriptional activity of the α_2C4 promoter region was investigated by transfecting several cell types with chimaeric constructs containing various fragments of the 5′-non-coding region and luciferase as a reporter gene. The activity of the construct containing the entire 5′-non-coding region appeared to depend on the host cell. Removal of the 5′-untranslated region resulted in loss of cell specificity and a concomitant increase in luciferase activity. Transfection of HepG2 and SK-N-MC cells with constructs deleted of additional 5′-flanking fragments permitted the definition of a minimal 200 bp promoter fragment containing the pseudo-TATA box and two putative SP1-binding sites.

INTRODUCTION

The α_2 -adrenergic receptors (ARs) constitute a class of seven-transmembrane-domain receptors that mediate a number of cellular responses to catecholamines and that share the property of inhibiting adenylate cyclase using pertussis-toxin-sensitive G-proteins as signal transducers. Radioligand binding and molecular-biological approaches have now clearly established that these receptors exist as three pharmacological subtypes (α_2A , α_2B and α_2C) [1]. Like most of the other members of the AR family, α_2A -, α_2B - and α_2C -ARs are encoded by distinct intronless genes which are respectively designated α_2C10 , α_2C2 and α_2C4 in humans [2–4]. The physiological relevance of this diversity of receptors is not yet fully appreciated. The subtypes differ in their susceptibility to desensitization and down-regulation, indicating that they may be regulated differently after exposure to agonist [5,6]. This conclusion, however, is not supported by results from experiments carried out on cells that naturally express the different α_2 -AR subtypes. The subtypes of α_2 -ARs also appear to diverge in their tissue distribution. However, large discrepancies have been observed according to the species examined or the method of analysis utilized [7,8], making it as yet difficult to obtain a clear idea of the exact pattern of α_2 -AR expression and to assign a given subtype to a specific function.

The α_2C -AR can be pharmacologically distinguished from the α_2A subtype by its sensitivity to prazosin and from the α_2B subtype by its remarkably high affinity for [³H]MK912 [9]. On the basis of these properties, α_2C -ARs were found expressed in human brain [10] and kidney [11], but not in platelets, adipocytes and enterocytes which exclusively contain the α_2A subtype [12–14]. In agreement with the pharmacological recognition of α_2C -AR in human brain and kidney, these two tissues were demonstrated to contain large amounts of α_2C4 mRNAs [8]. Substantial levels of α_2C4 transcripts were also found in heart,

aorta and spleen [8] but the presence of α_2C -AR in these latter tissues remains to be established. The reason why expression of the α_2C4 gene is restricted to certain tissues is poorly understood, but it is likely that it lies in the characteristics of its promoter region. In contrast with α_2C10 , the 5′-flanking region of which has been sequenced and promoter elements studied in detail [15–18], the organization of the α_2C4 gene is completely unknown. As the single exon of the α_2C4 gene is only 1.38 kb in length [3] whereas α_2C4 mRNA is about 2.9 kb in both heart and kidney [7], one can simply assume that the α_2C4 transcript contains rather long 5′- and/or 3′-untranslated regions (UTRs).

As a first approach to understanding the molecular mechanisms governing the expression and transcriptional regulation of the α_2C -AR in human tissues, we have cloned and sequenced the 5′- and the 3′-flanking regions of the human α_2C4 gene. The transcription-initiation and polyadenylation sites were determined by RNase-protection assays (RPAs) carried out on RNA prepared from the hepatoma cell line, HepG2. The functional activity of the promoter was analysed in different cell types transfected with chimaeric constructs containing luciferase as reporter gene.

MATERIALS AND METHODS

Screening of the human genomic library

The human foetal brain genomic DNA library in the cosmid SuperCos 1 vector (Stratagene, La Jolla, CA, U.S.A.; catalogue no. 961200) was screened using the *SacI*–*SmaI* restriction fragment, corresponding to nucleotides 334/1016 of the human α_2C4 coding region [3]. Titration, hybridization and washing were carried out as recommended by the manufacturer. After two rounds of screening, a cosmid containing an approx. 40 kb insert surrounding the α_2C4 coding region was purified and charac-

Abbreviations used: ORF, open reading frame; UTR, untranslated region; FCS, foetal calf serum; AR, adrenergic receptor; RPA, RNase-protection assay.

¹ To whom correspondence should be addressed.

terized further by restriction enzyme digestion and Southern-blot analysis. This allowed us to clone in pBluescriptII KS+ (pKS⁺; Stratagene) an *EcoRI*–*EcoRI* fragment containing 6 kb of the 5'-flanking region, the open reading frame (ORF) and 4 kb of the 3'-flanking region of the α_2 C4 gene.

DNA sequencing and analysis

Restriction fragments of the *EcoRI*–*EcoRI* fragment were subcloned into pKS⁺. DNA was sequenced in both orientations using the dideoxy chain-termination method with the T7 Sequencing Kit from Pharmacia Biotech (Uppsala, Sweden) in combination with universal (T3, T7, M13-40) or specific primers. The sequences were analysed using the computer programs available at the w.w.w. server BCM Search Launcher of the Human Genome Center at Baylor College of Medicine (Houston, TX, U.S.A.).

RNA extraction, [³²P]riboprobe synthesis and RPA

Cellular RNA was isolated using the acid guanidinium thiocyanate/phenol/chloroform extraction method [19]. The integrity of the preparations was assessed by agarose-gel electrophoresis and RNA concentrations were measured by UV spectrophotometry. Antisense riboprobes for mapping of the transcription-start site were obtained by subcloning various fragments of the 5'-non-coding region of the α_2 C4 gene into pKS⁺. Five probes spanning 1367 bases 5' to the translation-start codon were used to determine the transcriptional start. Probe 1 corresponds to the *PvuII*–*NcoI* fragment (nucleotides –205/+4), probe 2 to *NotI*–*PvuII* fragment (nucleotides –620/+205), probe 3 to *NcoI*–*SmaI* fragment (nucleotides –1049/+538), probe 4 to *NcoI*–*NotI* fragment (nucleotides –1049/+793) and probe 5 to *MaeIII*–*NotI* fragment (nucleotides –1367/+793). Mapping of the 3'-end of the α_2 C4 transcript was performed using probe 6 which was prepared from the *BamHI*–*BstXI* fragment corresponding to nucleotides +1651/+2040. The plasmids were cleaved with the appropriate restriction enzyme, and antisense RNAs were synthesized in the presence of [³²P]UTP using the T3 or T7 RNA polymerase (Promega). RPA was performed as previously described but with slight modifications [20]. Lyophilized cellular RNA (100 μ g) was placed in 30 μ l of hybridization buffer [80% (v/v) formamide, 0.4 M NaCl, 1 mM EDTA, 40 mM Pipes, pH 6.7] containing the [³²P]riboprobe. The samples were heated to 95 °C for 5 min and then immediately placed at 55 °C for 14 h. Non-hybridized probe was eliminated by the addition of 0.3 ml of a mixture of RNases (40 μ g/ml RNase A and 2 μ g/ml RNase T1) in TEN buffer (10 mM Tris/HCl, 5 mM EDTA, 300 mM NaCl, pH 7.5). After 2 h at 37 °C, digestion was stopped by addition of 5 μ l of proteinase K (10 mg/ml), and the samples were further incubated for 15 min at 37 °C. Carrier tRNA (10 μ g) and 0.3 ml of solution D [4 M guanidium isothiocyanate, 0.1 M 2-mercaptoethanol, 0.5% (w/v) sarkosyl, 25 mM sodium citrate, pH 7.0] were added to each tube, and protected hybrids were precipitated with propan-2-ol. After a wash with 70% (v/v) ethanol, RNA pellets were dissolved in 10 μ l of sample buffer [97% (v/v) formamide, 0.1%

(w/v) SDS, 10 mM Tris/HCl, pH 7.0] and loaded on to a 5% (w/v) acrylamide/7 M urea gel. The gels were exposed for 48 h at –80 °C to X-ray film (Hyperfilm, Amersham, Bucks, U.K.) with an intensifying screen or analysed using a PhosphorImager (Molecular Dynamics, Sunnyvale, CA, U.S.A.).

Reporter gene constructs

The reporter gene constructs are numbered according to the translation-start site of α_2 C4 gene and were generated from the promoterless luciferase vector pGL3-Basic (Promega, Madison, WI, U.S.A.), as follows. First, the *BamHI*–*NcoI* fragment corresponding to the entire 5'-flanking region of α_2 C4 was introduced into the *BglII* and *NcoI* restriction sites of pGL3-Basic, resulting in a construct designated pGL3C4–2806/–1 where the ATG of luciferase is placed in the translational context of the α_2 C4 gene. Secondly, the *BamHI*–*PstI* (–2806/–886) restriction fragment of the 5'-flanking region of the α_2 C4 gene was subcloned into the *BamHI* and *PstI* sites of pKS⁺. This fragment was then excised with either *BamHI* and *HindIII* or *BamHI* and *EcoRV* and inserted into the *BglII*–*HindIII* sites or the *BglIII*–*SmaI* sites of pGL3-Basic. This resulted in the creation of constructs containing the –2806/–886 fragment in sense or antisense orientation which were respectively named pGL3C4–2806/–886 and pGL3C4–886/–2806.

pGL3C4–2806/–886 was then deleted using the *SmaI* site in the polylinker of pGL3 and the *SspI* (–2075), *SmaI* (–1799), *PvuII* (–1699), *StuI* (–1633) or *DraI* (–1340) blunt sites of the α_2 C4 sequence to give the pGL3C4–2075/–886, pGL3C4–1799/–886, pGL3C4–1699/–886, pGL3C4–1633/–886 and pGL3C4–1340/–886 respectively. Similarly, *SacI* deletion resulted in the generation of a shorter construct named pGL3C4–1127/–886.

Deletion of pGL3C4–2806/–886 using the *PstI* sites located at positions –1494 and –886 allowed us to create the pGL3C4–2806/–1494 construct. Insertion of the *PstI*–*FspI* fragment (–1494/–1105 of the α_2 C4 sequence) into the *PstI* and *EcoRV* sites of pGL3C4–2806/–1494 resulted in pGL3C4–2806/–1105. Finally, pGL3C4–1043/–886 and pGL3C4–1127/–961 were obtained by PCR using pGL3C4–1127/–886 as template. The –1043/–886 fragment was obtained using a sense primer 5'-CATGGTACCCCGAGCCGCCGTGCTGC-3' allowing the creation of *KpnI* site at position –1043 and an antisense primer located in the luciferase sequence. The –1127/–961 fragment was obtained using a sense primer located in the pGL3 sequence and an antisense primer 5'-CTAAGACTGCAG-CCTCCTTCGCCTC-3' allowing the creation of a *PstI* site at position –961. The products from these amplifications were digested by *KpnI* and *PstI* and cloned in pGL3 digested with the same enzymes.

Cell culture, transfections and measurements of luciferase and β -galactosidase activities

HeLa (human cervix), HepG2 and SK-Hep (human hepatocarcinomas) were routinely subcultured in Dulbecco's modified

Figure 1 Restriction map of the *EcoRI*–*EcoRI* clone and nucleotide sequence of the human α_2 C4-AR gene

An *EcoRI*–*EcoRI* fragment containing the α_2 C-AR ORF was isolated from a human foetal brain genomic DNA library, subcloned into pKS⁺ and characterized by restriction analysis as described in the Materials and methods section. The sequenced region corresponds to the *BamHI*–*BstXI* fragment and consists of 2806 nt of 5'-flanking region, 1383 nt of coding region and 658 nt of 3'-flanking region. The bases are numbered according to the translation codon start site, and the coding region is in upper case. Consensus sequences for transcription factor binding are identified as follows: the N-Myc/c-myc site is in bold and underlined by solid circles, the AP1-binding site is in italic and underlined, the pseudo-TATA box is in a rectangle and Sp1-binding sites are in bold and underlined. The two half-sites of peroxisome-proliferator-activated receptor are in bold italics. The site for initiation of transcription as determined by RPA is indicated by the arrow. The square brackets indicate the boundaries of the putative 43-codon micro-ORF. The sequence for RNA destabilization and the polyadenylation signal are underlined with broken lines and circles respectively. The location of the poly(A) tail is indicated by a solid triangle. The 21 bp direct repeat is underlined with open triangles.

Eagle's medium supplemented with 10 % foetal calf serum (FCS). SK-N-MC (human neuroblastoma) was grown in minimum essential medium containing 10 % FCS and 2 mM glutamine. All cell lines were transfected in 10 % FCS-supplemented Dulbecco's modified Eagle's medium using the calcium phosphate precipitation method [21]. Transfections were performed in 30 mm-diameter dishes using 4 μ g of the different luciferase constructs and 1 μ g of the pCMV-LacZ vector used as an internal control for transfection efficiency. Cells were collected 48 h after transfection and immediately disrupted in lysis buffer containing 100 mM potassium phosphate and 0.2 % (v/v) Triton X-100 (pH 7.5). Luciferase was measured in a 96-well plate luminometer (Luminoskan Labsystems, Labsystems Oy, Helsinki, Finland) using the luciferase reporter assay system from Promega. β -Galactosidase activity was determined using *o*-nitrophenyl β -D-galactopyranoside as substrate [22].

RESULTS

Sequence analysis of the 5'- and 3'-flanking regions of the human α_2 C4 AR gene

The use of a 0.68 kb probe (corresponding to the *SacI*–*SmaI* fragment of the α_2 C4-coding region) to screen a human genomic library allowed us to isolate an 11 kb *EcoRI*–*EcoRI* fragment containing the α_2 C4 gene. The restriction map of this clone and the nucleotide sequence of the *Bam*HI–*Bst*XI fragment covering 2806 bp of the 5'-flanking region, the ORF and 658 bp of the 3'-flanking region of the α_2 C4 gene are presented in the Figure 1.

Analysis of the nucleotide composition of the 5'-flanking sequence indicates that the first 900 nt upstream of the translation-start site is composed of 84 % G + C residues (Figure 2). Neither TATA box nor CAAT box were found in this proximal region; by contrast, numerous putative consensus binding sites for the transcription factor Sp1 (GC box) were identified. Several potential regulatory elements for controlling the transcriptional activity of many other genes were also detected. Notably a putative binding site for AP1 was detected at position –1237/–1230. The consensus sequence for transcription factors of the c-Myc family was found at position –1520/–1507. Two half-sites for peroxisome-proliferation-activated receptor were found at positions –1808/–1803 (TGACCT) and –2209/–2204 (AGGTCA). Analysis of the whole 5'-region using transcription start site Ghosh (TSSG) indicated the existence of two possible promoters with transcription-initiation sites predicted at position –490 and –910 relative to the start codon. In contrast with that at –490, the initiation-start site predicted at –910 included a pseudo-TATA box (TTAGAAA) at position –954/–949.

Figure 2 Analysis of the G + C content of the human α_2 C4 gene

Examination of the 3'-flanking region revealed the existence of a 21-base direct repeat (GGGGAGCTTTCCAGAGACCC) located 70 bases downstream of the stop codon. It also indicated the presence of a sequence for RNA destabilization (ATTTA) at position +1871/+1875 and of a single canonical polyadenylation signal (AATAAA) at position +1921/+1926. The search for sequence homology by performing a BLAST [21] against the 661-base fragment of the 3'-non-coding region (*TGA*–*Bst*XI fragment) permitted the identification of a previously partially sequenced human placental cDNA clone (Washington University–Merck EST project, clone no. T40595) which displays an almost perfect homology with the α_2 C4 (one mismatch over 344 bases). This clone ends 14 bases downstream of the polyadenylation signal, indicating that the polyadenylation site may be located at position +1940 of the α_2 C4 gene sequence.

Identification of the 5'- and 3'-ends of the α_2 C4 mRNA

To locate the promoter region, we first identified the 5'-end of α_2 C4 mRNA. A recent study from our group demonstrated that the human hepatocarcinoma cell line HepG2 expresses α_2 -ARs of the α_2 C subtype and contains large amounts of α_2 C4 transcripts [23]; RNAs prepared from this cell line were thus used to map the α_2 C4 mRNA transcriptional-start site by RPA. The size and location of the five antisense riboprobes used for this purpose are shown in Figure 3. Probes 1 and 2, which together cover 620 bases upstream of ATG, were fully protected by cellular RNA from HepG2. In contrast, probe 3 was only partially protected on about 350 bases indicating that the transcriptional-start site was located approx. 900 bases 5' to the codon start site. Probes 4 and 5 were thus used to identify the cap site more precisely. As expected from the results obtained with probe 3, probes 4 and 5 were partially protected over a region of identical size. The determination of the length of the protected region using 54-, 103- and 148-base labelled riboprobes allowed us to calculate that both probe 4 and 5 were protected on 98 ± 2 nt. Similar results were obtained using RNA prepared from human kidney (not shown). Thus, according to these data, the transcriptional-start site of the α_2 C4 gene is located 891 ± 2 bases upstream of the start codon. This conclusion is in good agreement with the computer prediction of a putative non-canonical TATA box (TTAGAAA) at position –954/–949. However, attempts to identify more precisely the cap site by means of primer extension were unsuccessful. This failure was primarily due to the impossibility of designing a suitable primer in this region and to the inherent difficulty in reverse-transcribing such GC-rich mRNA.

The functionality and unity of the polyadenylation signal located at position +1921/+1926 of the α_2 C4 sequence was also confirmed by RPA. As shown in Figure 4, probe 6, which corresponds to nucleotides +1650/+2034, was protected by HepG2 RNA on about 290 bases indicating the existence of a single polyadenylation site located 550 bases downstream of the stop codon. Taken together, the results from RPA demonstrate that α_2 C4 mRNA is about 2840 bases in length.

Analysis of promoter activity of the α_2 C4 5'-flanking region in various cell lines

We next sought to characterize the promoter activity of the α_2 C4 5'-flanking region. For this purpose we first transfected several cell lines with constructs containing either the complete 5'-flanking region (pGL3C4–2806/–1) or a fragment terminating five bases downstream of the transcription-start site detected by RPA (–2806/–886). This latter fragment was tested in sense (pGL3C4–2806/–886) and reverse orientation (pGL3C4

Figure 3 Identification of the transcription-initiation site of the human α_2 C4 gene

The 5'-end of the α_2 C4 mRNA was mapped by RPA carried out on 100 μ g of HepG2 cellular RNA. The five probes used cover 1367 bases 5' to the codon start. The sites used for their construction were as follows: *PvuII* and *NcoI* for probe 1; *NotI* and *PvuII* for probe 2; *NcoI* and *SmaI* for probe 3; *NcoI* and *NotI* for probe 4; *MaeIII* and *NotI* for probe 5. (A) RPA experiment with probe 1: lane 1, undigested full-length probe (292 nt); lanes 2–4, RNase protection with sense strand 209 nt RNA (lane 2), tRNA (lane 3) or 100 μ g of HepG2 cellular RNA (lane 4). (B) RPA experiment with probe 2: lane 1, undigested full-length probe (473 nt); lanes 2–4, RNase protection with sense strand 415 nt RNA (lane 2), tRNA (lane 3) or 100 μ g of HepG2 cellular RNA (lane 4). (C) RPA experiment with probe 3: lane 1, undigested full-length probe (548 nt); lanes 2–4, RNase protection with sense strand 353 nt RNA prepared using the *PstI*–*SmaI* fragment (–891/–539) as a matrix (lane 2), tRNA (lane 3) or 100 μ g of HepG2 cellular RNA (lane 4). (D) RPA experiment with probe 4 (257 nt): lanes 1 and 2 correspond to 54- and 103-base RNA markers; lanes 3 and 4, RNase protection with tRNA (lane 3) and 100 μ g of HepG2 cellular RNA (lane 4). (E) RPA experiment with probe 5 (575 nt): lanes 1, 2 and 3 correspond to 148-, 103- and 54-base RNA markers; lanes 4 and 5, RNase protection with tRNA (lane 4) and 100 μ g of HepG2 cellular RNA (lane 5).

Figure 4 Polyadenylation site of the human α_2 C4-AR gene

RPA was carried out using probe 6. This probe, which was prepared from a *Bam*HI–*Bst*XI fragment, covers nucleotides +1651 to +2040 of the α_2 C4 gene sequence. Lane 1, undigested full-length probe (409 nt); lanes 2 and 3, RNase protection with tRNA (lane 2) or 100 μ g of HepG2 cellular RNA (lane 3); lane 4, mixture of 103-, 148- and 323-base RNA markers.

–886/–2806). The density of α_2 C-ARs and the amounts of α_2 C4 mRNAs were also measured in these cell lines by radioligand binding and reverse transcriptase PCR experiments. Moreover

Table 1 Transcriptional activity of fragments of the 5'-flanking region of the α_2 C4 gene in various human cell lines

Fragments from the 5' region of the α_2 C4 gene were inserted into pGL3-Basic. SK-Hep, HeLa, SK-N-MC and HepG2 cells were transfected with both the indicated luciferase construct and pCMV-LacZ. Endogenous expression of the α_2 C4 gene in SK-N-MC and HepG2 but not SK-Hep and HeLa cells was previously demonstrated by reverse transcriptase PCR, RPA and binding experiments [23]. Luciferase and β -galactosidase activities were measured as described in the Materials and methods section. Data are corrected for transfection efficiency and expressed as fold increase (mean $\pm$ S.E.M.) of the activity in the corresponding cells transfected with the promoterless construct pGL3-Basic. *Significant difference at $P < 0.005$.

Endogenous α_2 C4 expression ...	SK-Hep No	HeLa No	SK-N-MC Yes	HepG2 Yes
pGL3-Basic	1.00 $\pm$ 0.07	1.00 $\pm$ 0.08	1.00 $\pm$ 0.07	1.00 $\pm$ 0.09
pGL3C4–2806/–1	0.74 $\pm$ 0.22	0.37 $\pm$ 0.31	3.33 $\pm$ 0.78*	2.52 $\pm$ 0.13*
pGL3C4–2806/–886	2.38 $\pm$ 0.22*	7.60 $\pm$ 4.15*	14.3 $\pm$ 1.62*	6.20 $\pm$ 1.03*
pGL3C4–886/–2806	0.58 $\pm$ 0.26	0.19 $\pm$ 0.05	1.44 $\pm$ 0.44	0.74 $\pm$ 0.13
pGL3-SV40	12.6 $\pm$ 1.7*	29.7 $\pm$ 5.4*	20.7 $\pm$ 3.8*	28.4 $\pm$ 3.3*

we verified that all lines tested could be efficiently transfected using pGL3-SV40 as a positive control. The results of this study are summarized in Table 1. It appears that the transfection of pGL3C4–2801/–1 in cell lines that do not express the α_2 C4 gene (SK-Hep and HeLa) led to a normalized luciferase activity similar to or less than that found with the promoterless vector, pGL3-Basic. In contrast, this construct is able to promote a

Figure 5 Transient expression of α_2C4 promoter-luciferase constructs in HepG2 and SK-N-MC cells

HepG2 cells (■) and SK-N-MC cells (□) were transfected as described in the Materials and methods section using the calcium phosphate precipitation method. Deletions were numbered according to their location relative to the start codon. Luciferase activity was measured as relative light units. The activity of the different constructs was adjusted for efficiency of transfection standardized by co-transfection of pCMV-LacZ. Results are expressed as fold increase in luciferase activity of the promoterless luciferase vector pGL3-Basic. Values are means $\pm$ S.E.M. from six independent experiments.

slight but significant luciferase activity in the two human cell lines that endogenously express α_2C4 -ARs (HepG2 and SK-N-MC) [23]. This divergence is not due to difference in transfection efficiency since pGL3-SV40 is strongly active in all cells tested. The results obtained with pGL3C4-2806/-886 demonstrate that the removal of the GC-rich 5'-UTR not only provoked a significant increase in luciferase activity in HepG2 and SK-N-MC cells but also resulted in expression in all cell types, suggesting that the -886/+1 region harbours sequences responsible for cell-specific expression of the α_2C4 gene. Finally, whatever the cell line considered, the construct containing the -2806/-886 fragment in reverse orientation (pGL3C4-886/-2806) did not promote activity, indicating that this fragment behaves as a true promoter with orientation-dependent activity.

Study of the α_2C4 promoter region in HepG2 and SK-N-MC cells

The promoter region of the α_2C4 gene was further studied by transfecting HepG2 and SK-N-MC cells with the series of deleted constructs generated as described in the Material and methods section. The results from these experiments are presented in Figure 5. Transfection of HepG2 cells indicated that the progressive removal of fragments stretching between nucleotides -2806 and -1633 resulted in a gradual increase in luciferase activity. In this cell line, the maximum activity was obtained with pGL3C4-1633/-886 and reached 17-fold greater activity than

the promoterless construct pGL3-Basic. Further deletion in the 5'-end resulted in a decrease in activity; however, the 240 bp fragment (-1127/-886) which contained the putative pseudo-TATA box and two Sp1-binding sites was still able to promote luciferase activity at a level similar to that observed with the full-length promoter construct. In agreement with a critical role for this short region, deletion of this fragment resulted in a total loss of promoter activity, HepG2 cells transfected with either pGL3C4-2806/-1494 or pGL3C4-2806/-1105 exhibiting a luciferase activity similar to or even lower than pGL3-Basic. Otherwise, the construct pGL3C4-1127/-961, from which the pseudo-TATA box was deleted, exhibited 3-fold lower activity than pGL3C4-1127/-886, suggesting that the TTAGAAA sequence functions as a TATA box. The results obtained with SK-N-MC cells differ somewhat from those with HepG2 cells. The highest level of luciferase activity in the neuroblastoma cell line was achieved by the pGL3C4-2806/-886 and pGL3C4-2075/-886 constructs, which yielded a 14-fold greater activity than the promoterless vector pGL3-Basic. The removal of the -2075/-1799 fragment resulted in a 50% attenuation of luciferase expression. Further removal of fragments between nucleotides -1799 and -1633 did not result in any significant change. As previously noted for HepG2 cells, deletion of fragment -1633/-1340 provoked a further attenuation of activity. Moreover, pGL3C4-2806/1494 and pGL3C4-2806/-1105 were also devoid of activity in SK-N-MC cells.

DISCUSSION

With the exception of the α_2 C10 gene, the promoter of which has been the subject of detailed study [15–18], the upstream regions of the genes encoding other human α_2 -ARs (namely α_2 C2 and α_2 C4) remain completely unknown. The present work was thus carried out as an initial approach to the characterization of the organization of the α_2 C4 gene and to define the elements accounting for the relatively restricted expression of the α_2 C-AR subtype in human tissues.

Screening of a human genomic library allowed us to isolate an *Eco*RI–*Eco*RI clone, the restriction map and Southern-blot analysis of which indicated that it contained the α_2 C4 ORF surrounded by 6 kb of 5'-flanking region and 4 kb of 3'-flanking region. Analysis of the sequence of the 5'-region suggested that the α_2 C4 gene exhibits many of the features of a housekeeping gene. It contains neither a conventional TATA box nor a CAAT box, but has a very high G + C content and multiple Sp1-binding sites. In contrast with what one would have expected for a TATA-less and GC-rich gene, RPA carried out on RNA prepared from HepG2 cells showed that transcription of the α_2 C4 gene is initiated from a single site 891 bases upstream of the translation-start site. Using the same method, the polyadenylation site was found to be located 550 bases downstream of the stop codon. Given these results, the size of α_2 C4 mRNA is 2840 nt. Adding on the poly(A) tail, the calculated size is in good agreement with the size of the α_2 C-AR mRNA (2.9 kb) found by Northern-blot analysis of polyadenylated RNAs from human kidney and heart [7]. The length of the 5'-leader region of α_2 C4 mRNA also matches well with the length of the 5'-end of a previously isolated rat brain cDNA (pA2d) encoding the rodent α_2 C-AR [24]. It is moreover in agreement with data obtained in gene-transfer experiments demonstrating that DNA elements that are indispensable for conveying significant levels of expression lie within the –1127/–886 fragment. This short fragment notably contains the sequence TTAGAAA, which may function as a TATA-like box. Such a role remains to be established, but the sharp diminution in promoter activity observed in the construct deleted of a 71-base segment containing this region suggests that the TTAGAAA motif may effectively serve to position the transcription-initiation complex as does a conventional TATA-box.

The transfection experiments indicate that the activity of the construct containing the entire 5'-flanking region is, at least to some extent, dependent on the host cell. They also suggest that removal of the 5'-UTR resulted in loss of cell specificity and a concomitant increase in luciferase activity. The mechanism by which the 5'-UTR of the α_2 C4 gene affects expression was not investigated in the present work. However, the 5'-leader region of α_2 C4 mRNA exhibits at least two features that may participate in the attenuation of the translation of the major ORF [25]. First, the 5'-end of the mRNA is extremely rich in G + C and may form strong secondary structures. Secondly, like other members of the AR family [26], the 5'-leader region of the α_2 C4 gene harbours a 43-codon micro-ORF (nucleotides –447 to –319) which may interfere with translation.

The transfections with the series of 5'-deleted constructs revealed some important differences in transcriptional regulation between HepG2 and SK-N-MC, cells, thus suggesting the possible presence of a repressor in one cell line but not in the other. In both models, however, deletion of the –1633/–1340 fragment provoked a large attenuation of luciferase activity. Interestingly, this region contains a putative binding site for transcription factors of the basic helix–loop–helix protein family. This group

includes c-Myc, related proteins such as Max and Mad, as well as the upstream stimulatory factor. Although ubiquitously expressed, these factors have been frequently recognized to be involved in the tissue-specific control of gene expression [27].

While the present study was in progress, Saulnier-Blache et al. [28] published a description of the promoter region of the RG10 gene which is the rodent homologue of the α_2 C4 gene. Alignment of the two sequences revealed a high degree of homology. More precisely, the first 900 bases 5' to the methionine start of RG10 displayed 80% homology with that of α_2 C4. This region was similarly rich in GC and presented several putative Sp1-binding sites. Moreover, the pseudo-TATA box (TTAGAAA) and the AP1-binding site, respectively found at positions –954/–949 and –1229/–1223 of the α_2 C4 gene, were also encountered at positions –926/–932 and –1184/–1178 of the RG10 gene sequence. Although they are highly conserved, the promoter regions of the rat and human genes appear to function very differently. Indeed, according to RPA with rat brain RNA and transfection experiments in NG-108-15 cells, RG10 gene transcription is initiated at multiple sites through the promoter activity of a 96 nt GC-rich region located 700 bases upstream of the methionine initiation site. This small fragment is moreover sufficient to specifically direct expression of rat α_2 C-AR in cells of neuronal origin. The reasons for the functional differences between human and rat genes have been ignored, but they may explain why RG10 mRNA is primarily expressed in rat central nervous system, whereas α_2 C4 mRNA distribution in human extends to peripheral tissues [7,8,29–32]. The discrepancies between the results obtained in human and rat may arise from the fact that RG10 and α_2 C4 were studied in cells of different origin (namely NG108-15 and HepG2). Thus, even if the transfection experiments in SK-N-MC do not favour such a possibility, the existence of a second promoter identical with that described in rat and preferentially active in human neuronal tissue cannot be ruled out. Finally, as suggested by the length of the 5'-end of the rat cDNA isolated by Voigt et al. [24], it is possible that the RG10 gene also contains two promoter regions. Multiple alternative promoters is a frequently used mechanism to create diversity and flexibility in the regulation of gene expression [33]. Other examples of ARs with multiple promoters have been described. The gene encoding the α_1 b subtype was shown to have three promoters, two of which are TATA-less and GC-rich [34].

In summary, the present work provides for the first time precise indications about the organization of the human α_2 C4 gene. It demonstrates that transcription is initiated 891 bases upstream of ATG through the activity of a minimal promoter region that lacks a CCAAT box but contains a pseudo-TATA box and other proximal nuclear-factor-binding sites including two GC boxes. Confirmation of the precise roles of these elements awaits definitive demonstration by footprint analysis. Our transfection experiments also indicate divergent roles for regions 5' of the minimal promoter in regulating transcription in HepG2 and SK-N-MC cells. They also suggest a role for the downstream region in cell-specific expression of the α_2 C4 gene. The issue of whether the 5'-UTR is indeed important for tissue-specific expression of α_2 C-AR *in vivo* also requires further investigation.

This work was supported by grants from the Conseil Régional Midi-Pyrénées (Toulouse, France) and the Fondation pour la Recherche Médicale (Paris, France). We thank Françoise Quinchon and Elise Fonta for technical assistance. We are also grateful to Dr. Dominique Langin and Jacques Grober for helpful assistance with genomic bank screening and DNA sequencing.

REFERENCES

- 1 Bylund, D. B. (1988) Trends Pharmacol. Sci. **9**, 356–361

- 2 Kobilka, B. K., Matsui, H., Kobilka, T. S., Yang-Feng, T. L., Franck, U., Caron, M. G., Lefkowitz, R. J. and Regan, J. W. (1987) *Science* **238**, 650–656
- 3 Regan, J. W., Kobilka, T. S., Yang-Feng, T. L., Caron, M. G. and Lefkowitz, R. J. (1988) *Proc. Natl. Acad. Sci. U.S.A.* **85**, 6301–6305
- 4 Lomasney, J. W., Lorenz, W., Allen, L. F., King, K., Regan, J. W., Yang-Feng, T. L., Caron, M. G. and Lefkowitz, R. J. (1990) *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5094–5098
- 5 Eason, M. G. and Liggett, S. B. (1992) *J. Biol. Chem.* **267**, 25473–25479
- 6 Kurose, H. and Lefkowitz, R. J. (1994) *J. Biol. Chem.* **269**, 10093–10099
- 7 Eason, M. G. and Liggett, S. B. (1993) *Mol. Pharmacol.* **44**, 70–75
- 8 Berkowitz, D. E., Price, D. T., Bello, E. A., Page, S. O. and Schwinn, D. A. (1994) *Anesthesiology* **81**, 1235–1244
- 9 Bylund, D. B., Eikenberg, D. C., Hieble, J. P., Langer, S. Z., Lefkowitz, R. J., Minneman, K. P., Molinoff, P. B., Ruffolo, Jr., R. R. and Trendelenburg, U. (1994) *Pharmacol. Rev.* **46**, 121–136
- 10 Ordway, G. A., Jaconetta, S. M. and Halaris, A. E. (1993) *J. Pharmacol. Exp. Ther.* **264**, 967–976
- 11 Trendelenburg, A. U., Limberger, N. and Rump, L. C. (1994) *Mol. Pharmacol.* **45**, 1168–1176
- 12 Bylund, D. B., Ray-Prenger, C. and Murphy, T. J. (1988) *J. Pharmacol. Exp. Ther.* **45**, 600–607
- 13 Castan, I., Devedjian, J. C., Valet, P., Paris, H. and Lafontan, M. (1995) *Fundam. Clin. Pharmacol.* **9**, 569–575
- 14 Valet, P., Senard, J. M., Devedjian, J. C., Planat, V., Salomon, R., Voisin, T., Drean, G., Couvineau, A., Daviaud, D., Cenis, C., Laburthe, M. and Paris, H. (1993) *J. Clin. Invest.* **91**, 2049–2057
- 15 Fraser, C. M., Arakawa, S., McCombie, W. R. and Venter, J. C. (1989) *J. Biol. Chem.* **264**, 11754–11761
- 16 Handy, D. E. and Gavras, H. (1992) *J. Biol. Chem.* **267**, 24017–24022
- 17 Sakaue, M. and Hoffman, B. B. (1991) *J. Biol. Chem.* **266**, 5743–5749
- 18 Shilo, L., Sakaue, M., Thomas, J. M., Philip, M. and Hoffmann, B. B. (1994) *Cell. Signal* **6**, 73–82
- 19 Chomczynski, P. and Sacchi, N. (1987) *Anal. Biochem.* **162**, 156–159
- 20 Devedjian, J. C., Fargues, M., Denis-Pouxviel, C., Daviaud, D., Prats, H. and Paris, H. (1991) *J. Biol. Chem.* **266**, 14359–14366
- 21 Ausubel, F. M., Brent, R., Kingston, R. E., Moore, D. D., Seidman, J. G., Smith, J. A. and Struhl, K. (1994) *Current Protocols in Molecular Biology*, John Wiley and Sons, New York
- 22 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- 23 Schaak, S., Cayla, C., Blaise, R., Quinchon, F. and Paris, H. (1997) *J. Pharmacol. Exp. Ther.* **281**, 983–991
- 24 Voigt, M. M., McCune, S. K., Kanterman, R. Y. and Felder, C. C. (1991) *FEBS Lett.* **278**, 45–50
- 25 Kozak, M. (1992) *Annu. Rev. Cell Biol.* **8**, 197–225
- 26 Parola, A. L. and Kobilka, B. K. (1994) *J. Biol. Chem.* **269**, 4497–4505
- 27 Parterson, J. M., Morroson, C. F., Mendelson, S. C., McAllister, J. and Quinn, J. P. (1995) *Biochem. J.* **310**, 401–406
- 28 Saulnier-Blache, J. S., Yang, Q., Sherlock, J. D. and Lanier, S. M. (1996) *Mol. Pharmacol.* **50**, 1432–1442
- 29 Blaxall, H. S., Hass, N. A. and Bylund, D. B. (1994) *Receptor* **4**, 191–199
- 30 Mandy, D. E., Flordellis, C. S., Bogdanova, N. N., Breshnahan, M. R. and Gavras, H. (1993) *Hypertension (Dallas)* **21**, 861–865
- 31 Lorenz, W., Lomasney, J. W., Collins, S., Regan, J. W., Caron, M. G. and Lefkowitz, R. J. (1990) *Mol. Pharmacol.* **38**, 599–603
- 32 Mesiter, B., Dagerlind, A., Nicholas, A. P. and Hokfelt, T. (1994) *J. Pharmacol. Exp. Ther.* **268**, 1605–1611
- 33 Ayoubi, T. A. and van de Ven, W. J. (1996) *FASEB J.* **10**, 453–460
- 34 Gao, B. and Kunos, G. (1994) *J. Biol. Chem.* **269**, 15762–15767