Structure of the human cyclo-oxygenase-2 gene

Susan B. APPLEBY, Ari RISTIMÄKI, Karen NEILSON, Kirsi NARKO and Timothy HLA*
Department of Molecular Biology, Holland Laboratory, American Red Cross, 15601 Crabbs Branch Way, Rockville, MD 20855, U.S.A.

Cyclo-oxygenase (Cox), a rate-limiting enzyme in the synthesis of prostanoids, is encoded by two genes, Cox-1 and Cox-2, which are differentially expressed and regulated. Human Cox-1 and -2 polypeptides share 61% primary sequence identity. While the expression of Cox-1 is maximal in quiescent cells, Cox-2 expression is induced by growth factors and cytokines. We have screened a human genomic library with a probe from the 5′-untranslated region (UTR) of the human Cox-2 (hCox-2) cDNA and isolated two overlapping genomic clones. We have determined the DNA sequence of 0.8 kb upstream of the transcription start site, 6 kb of protein coding region, which includes 10 exons and 9 introns, as well as 2.5 kb of the 3′-UTR. The structures of the hCox-1 and hCox-2 and the murine TIS10 (Cox-2) genes are highly conserved, with a few exceptions. The 3′-UTRs of the Cox-1 and -2 genes are distinct; for example, the largest exon in the Cox-2 gene encodes the entire 3′-UTR, containing 22 copies of the 'AUUUA' RNA instability element. Sequence analysis of the 5′-flanking region has shown several potential transcription regulatory sequences, including a TATA box, a C/EBP motif, two AP-2 sites, three SP1 sites, two NF-κB sites, a CRE motif and an Ets-1 site. These efforts serve as a basis for future studies on transcriptional and post-transcriptional mechanisms of Cox-2 gene regulation.

INTRODUCTION

The cyclo-oxygenase (Cox) enzyme (EC 1.14.99.1) catalyses the oxidation of arachidonic acid into prostaglandin H_2 [1]. The enzyme is encoded by two related genes, Cox-1 and Cox-2 [2]. The Cox-1 gene is constitutively and ubiquitously expressed, while the Cox-2 gene is only expressed at high levels upon induction by growth factors, cytokines and extracellular stimuli associated with cell activation [2]. Many cell types associated with inflammation, such as monocytes, endothelial cells and fibroblasts, express the Cox-2 gene upon induction [3]. Previous studies from our laboratory have shown that the tumour promoter phorbol 12-myristate 13-acetate (PMA) and the cytokine interleukin-1 (IL-1) induce the Cox-2 transcript in vascular endothelial cells, synovial fibroblasts and rheumatoid arthritic synovial explants [4,5]. In endothelial cells, induction by IL-1α involves both transcriptional activation and post-transcriptional regulation of Cox-2 mRNA stability [5]. In synovial fibroblasts, as well as in primary explant tissues from rheumatoid arthritic patients, the induction of Cox-2 was completely suppressed by the anti-inflammatory steroid dexamethasone [6]. However, the molecular mechanisms involved in induction by cytokines as well as in suppression by steroids are not known. To better define the structural elements involved in the regulation of expression, we now report the complete nucleotide sequence of the hCox-2 gene.

EXPERIMENTAL

Genomic Southern blot

A 300 bp fragment of the 5′-untranslated region (UTR) of Cox-2 [4] was used to probe a human genomic Southern blot. The fragment was labelled with [^{32}P]dCTP using a random-primed DNA labelling kit (Boehringer Mannheim). Human genomic DNA (Clontech) was digested with the *Sac*I restriction endonuclease. Fragments were resolved on a 1% agarose gel and transferred to a Zetaprobe membrane (Bio-Rad). The membrane was hybridized at 65 °C using the Church and Gilbert hybridization method [7] and exposed for autoradiography at −70 °C.

Library screening

The 300 bp radiolabelled cDNA probe derived from the 5′-UTR of the Cox-2 cDNA was used to screen an EMBL-3 SP6/T7 human placenta genomic library in the EMBL-3 SP6/T7 vector (Clontech). Approx. 1×10^6 plaque-forming units were plated on NM538 cells, transferred to a Colony/Plaque Screen Hybridization Transfer Membrane (New England Nuclear) and hybridized as described under Southern blot analysis. Two positive clones, 9C and 23B, were plaque-purified and further characterized by restriction analysis with *Sac*I, followed by Southern blot analysis as previously described.

Sequencing

Restriction fragments from clone 9C were subcloned into pBS SK+/− (Stratagene), and deletions were made by the exonuclease III/mung bean nuclease method (Promega Biotech). Sequence analysis was done using M13 universal primers with DyePrimer cycle sequencing chemistry [Applied Biosystems Inc. (ABI)] and sequence-specific primers with DyeTerminator cycle sequencing chemistry (ABI) followed by analysis on the ABI 373A Automated Sequencer. Oligonucleotide primer pairs were designed from the cDNA sequence of Cox-2 on both sides of each potential intron (introns 3–9). The approximate size of each intron was determined by PCR analysis using 10 ng of λ DNA of clone 23B as a template and 0.5 μg of each primer. Introns 3–9 and exon 10 were sequenced in both directions directly from the

Abbreviations used: (h)Cox, (human) cyclo-oxygenase; UTR, untranslated region; PMA, phorbol 12-myristate 13-acetate; IL-1, interleukin-1; HUVEC, human umbilical vein endothelial cells; CAT, chloramphenicol acetyltransferase.

* To whom correspondence should be addressed.

The complete nucleotide sequence of the hCox-2 gene, including the intronic sequences not shown in Figure 1, has been deposited with the GenBank, EMBL and DDBJ Nucleotide Sequence Databases under accession no. U04636.

λ DNA of clone 23B using sequence-specific oligonucleotide primers and the DyeTerminator cycle sequencing chemistry.

Primer extension analysis

An oligonucleotide was designed from the 5′-end of the hCox-2 cDNA sequence (5′ GCGGGGGTAGGCTTTGCTGTCTGA 3′) and end-labelled with T4 polynucleotide kinase (BRL). Total cellular RNA from human umbilical vein endothelial cells (HUVEC) treated with PMA and cycloheximide for 4 h was isolated using the guanidine isothiocyanate method [8]. Approx. 0.7 ng of labelled primer was hybridized with either 30 μg of HUVEC RNA or 50 μg of yeast tRNA for 16 h, then extended by Moloney murine leukaemia virus reverse transcriptase (BRL) at 42 °C for 1.5 h [9]. Products were analysed on a 6% sequencing gel.

3′-end mapping

RNA was isolated from PMA- and cycloheximide-treated HUVEC as described above and converted to cDNA using 100 ng of dT_{17}-Adaptor primer (5′ GACTCGAGTCGACATCGAT$_{17}$ 3′) [12] and 400 units of Moloney murine leukaemia virus reverse transcriptase (BRL) at 37 °C for 1 h. The cDNA was then amplified by PCR using 2.5 units of AmpliTaq polymerase (Perkin-Elmer Cetus), 500 ng of Adaptor 1 (5′ GACTCGAGTCGACATCG 3′) and 500 ng of human Cox-2 (hCox-2) 3′-UTR sequence-specific primers (primer 1, 5′ AGCTATCTGTAACCAAGATGGATG 3′; primer 2, 5′ TGTCTTATTAGGACACTATGGTT 3′). The PCR reaction consisted of 10 mM Tris/HCl, pH 8.3, 5 mM KCl, 200 μM of each of dATP, dCTP, dGTP and dTTP, and 1.5 mM $MgCl_2$, and was run for 35 cycles at 94 °C for 1 min, 58 °C for 2 min, and 72 °C for 3 min. The fragments were analysed on a 1% agarose gel. The fragment derived from primer 1 was purified, subcloned into the vector pCRII (Invitrogen) and sequenced using DyeTerminator sequencing (ABI).

RESULTS

Isolation of hCox-2 genomic clones

To isolate hCox-2 genomic clones, a 300 bp 5′ fragment of the Cox-2 cDNA [4] was utilized as a probe. Southern blot analysis of human genomic DNA digested with the *Sac*I restriction enzyme yielded a 6 kb band that hybridized to the radiolabelled Cox-2 probe. This probe was used to screen a human placenta genomic library in the phage vector EMBL3 SP6/T7. Two positive clones were plaque-purified and further characterized. Restriction analysis of the clones 9C and 23B with *Sac*I followed by Southern blot analysis with the 300 bp 5′ cDNA probe resulted in hybridization to 2.8 and 6 kb bands respectively. Southern blot analysis of these clones with labelled oligonucleotide probes derived from different regions of the Cox-2 cDNA was also performed. All of the coding region was contained within clone 23B, whereas clone 9C hybridized only to the probe derived from the 5′-end of the cDNA. These two clones were further characterized by DNA sequencing.

DNA sequence determination of Cox-2 genomic clones

The 3′-terminal 2.6 kb *Sac*I fragment of clone 9C was subcloned and the DNA sequence was determined by the exonuclease deletion method as described. This fragment encoded 0.8 kb of 5′-flanking region and exons 1–3. The λ DNA for clone 23B was used directly for DNA sequence determination. The entire sequence of both strands was determined. The composite DNA sequence of the 5′-flanking region, 10 exons encoding the entire open reading frame, the exon–intron boundaries and the 3′-UTR is shown in Figure 1.

The 5′-flanking region

The transcription start site of the Cox-2 gene was determined by the primer extension assay using total RNA from HUVEC. As shown in Figure 2, a discrete band was observed using the HUVEC RNA but not the tRNA, indicating that the initiation of transcription begins from a single site. A consensus TATA box sequence was found 25 bases upstream from the transcription start site. The nucleotide sequence of approximately 800 bp of the 5′-flanking region is indicated in Figure 1. In this 5-UTR there are several potential transcriptional regulatory sequences, including two NF-κB sites, two AP-2 sites, three SP1 sites, one CRE site, one C/EBP motif and one Ets-1 site (Figure 1).

The coding region

The open reading frame of the Cox-2 gene is encoded by 10 exons. The amino acid sequences of each exon from the hCox-1 and -2 genes are compared in Table 1 [4,11]. Cox-2 exon 1 encodes the putative signal peptide region and is only 14% identical in amino acid sequence to the corresponding exon of Cox-1. The 5′-UTR and the first four amino acids of the hCox-1 gene are encoded by an additional exon. The remaining exons are between 52 and 75% identical at the amino acid level, with the exception of exon 3, which showed only 38% identity.

The 3′-flanking region

In order to determine the 3′-end of the Cox-2 gene, we mapped the tail of the Cox-2 mRNA in HUVEC using primers 1 and 2 and the dT_{17}-Adaptor [12]. RNA from PMA- and cycloheximide-treated HUVEC was reverse-transcribed into cDNA and PCR-amplified [12]. As shown in Figure 3, a single band was amplified from each primer pair, indicating that the Cox-2 gene is terminated at a single site in PMA- and cycloheximide-induced HUVEC. The resulting fragments were cloned and the sequence was determined. This sequence was found to be identical to the corresponding region from the λ clone 23B. Thus the entire 3′-UTR is contained within exon 10. Interestingly, there are three polyadenylation motifs found in the 3′-UTR (Figure 1). However, the single band obtained from the 3′-amplification PCR (Figure 3) suggests that only the final polyadenylation signal is used by PMA- and cycloheximide-treated HUVEC. It is noteworthy that exon 10 is significantly larger than exons 1–9 and contains 410 bp of the coding region and all 2550 bp of the 3′-UTR. The 3′-UTR of the Cox-2 mRNA is approximately 1.5 kb longer than that of the Cox-1 transcript and contains 22 copies of the AUUUA motif (Figure 1) that has been associated with the RNA instability of other inducible transcripts [13].

DISCUSSION

To determine the complete nucleotide sequence of hCox-2, a portion of the cDNA sequence was used to screen a human genomic library. Two clones, 9C and 23B, were isolated and characterized. While clone 9C contains only exons 1–3, clone 23B has been found to contain at least 800 bp of 5′-flanking region, the entire open reading frame, consisting of 10 exons and 9 introns, and the entire 3′-UTR. The structural format of the

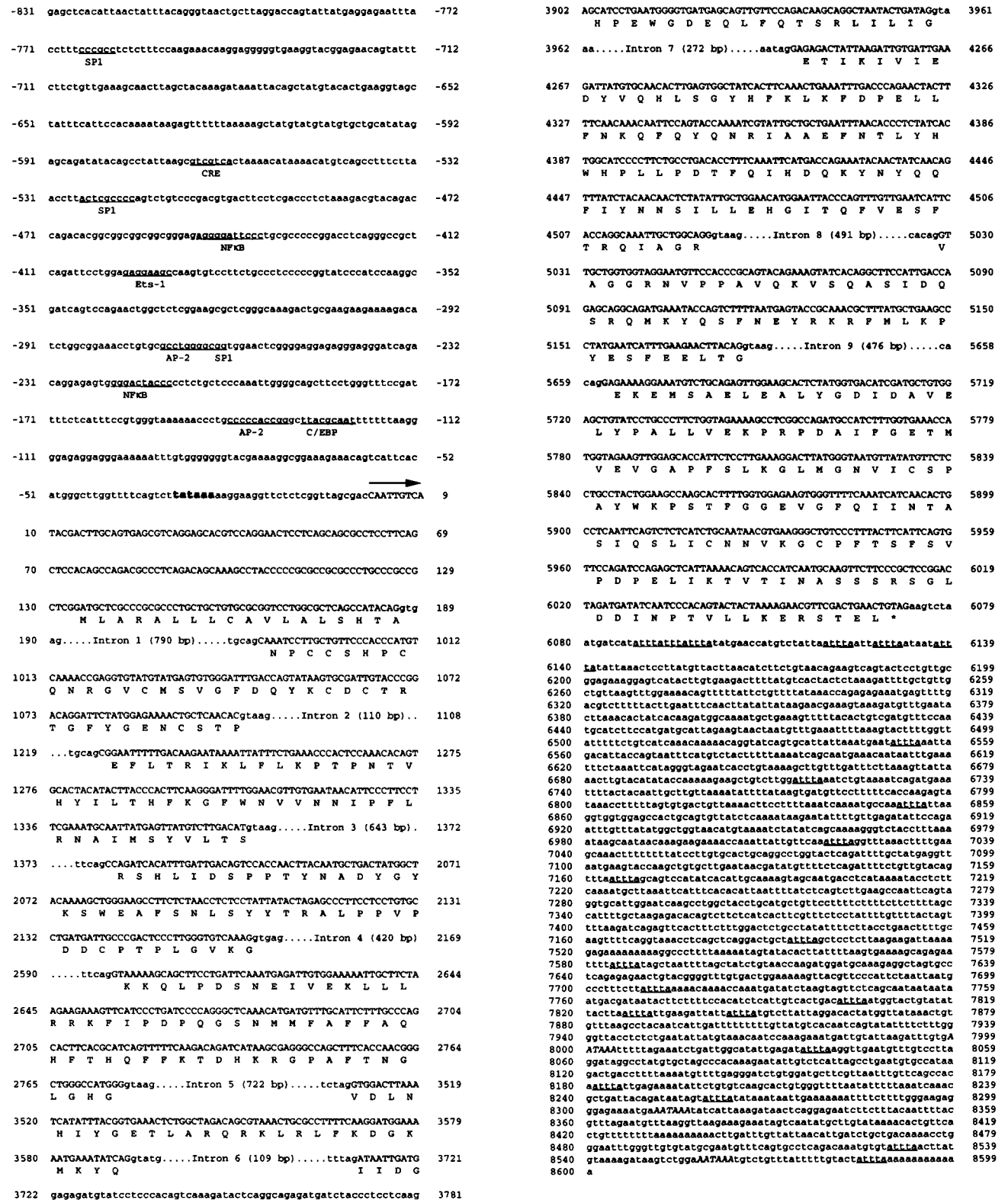

Figure 1 Nucleotide sequence of hCox-2

The nucleotide sequence of the hCox-2 gene, including the 5′-flanking region and the 3′-UTR, is shown. Upper case letters are exon sequences. Exon–intron junctions are shown with intron sequences in lower case letters and the number of nucleotides omitted is indicated in parentheses. Potential transcriptional regulatory sequences are underlined, the TATA box is outlined, and the transcriptional start site is indicated with a closed arrow. The amino acid sequence is designated directly below the first nucleotide of the corresponding codon. The termination codon is indicated with an asterisk. In the 3′-UTR, the 22 copies of the sequence ATTTA, which has been associated with rapidly degraded RNA messages, have been underlined, and the three AATAAA polyadenylation sequences are in capitalized italics.

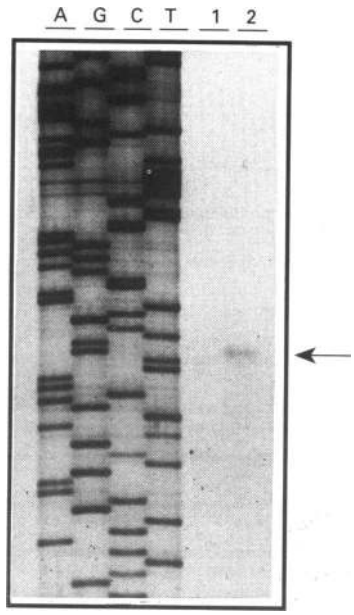

Figure 2 Primer extension analysis

Control yeast tRNA or total cellular RNA isolated from PMA- and cycloheximide-treated HUVEC was hybridized to a radiolabelled oligonucleotide corresponding to the hCox-2 sequence approximately 120 bp from the TATA consensus sequence, reverse-transcribed, and analysed on a 6% sequencing gel. Lanes A, C, G and T contain a sequencing ladder, lane 1 contains the yeast tRNA control, and lane 2 contains the HUVEC RNA. The single band seen corresponds to the first base of transcription, 25 bp downstream of the TATA box.

Table 1 Amino acid identity of hCox-1 and hCox-2

Exon number refers to the hCox-2 sequence; the numbers below each gene refer to the amino acid position of each corresponding exon. The percentage identity was determined from the amino acid sequence.

Exon	hCox-1	hCox-2	Identity (%)
1	4–32	1–18	14
2	33–71	19–57	59
3	72–118	58–105	38
4	119–166	106–153	71
5	167–226	154–213	73
6	227–254	214–241	75
7	255–337	242–324	70
8	338–432	325–419	65
9	433–482	420–469	52
10	483–599	470–604	57

coding region of the hCox-2 gene is highly similar to that of the mouse counterpart [14]. However, the genomic region of the murine 3′-UTR has not been reported.

The transcription start site, determined by primer extension analysis, was found 25 bp downstream of a conventional TATA box. The 800 bp of the 5′-flanking region has several potential transcriptional regulatory sequences, including two NF-κB sites, two AP-2 sites, three SP1 sites, one C/EBP motif, one Ets-1 site and one CRE site. Transfection of the 5′-flanking region of the Cox-2 gene fused to the chloramphenicol acetyltransferase (CAT) reporter gene into Cox cells results in significant expression of CAT activity (results not shown). This suggests the functionality

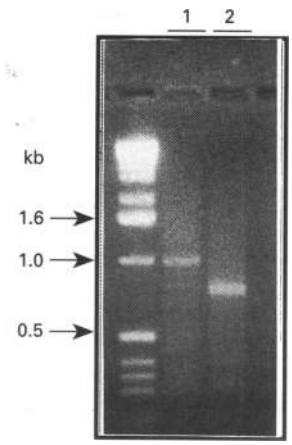

Figure 3 3′-end mapping of hCox-2

Total cellular RNA was isolated from PMA- and cycloheximide-treated HUVEC and reverse-transcribed using dT_{17}-Adaptor 1 primer. The cDNA was amplified using Adaptor 1 and 3′-UTR specific primers 1 and 2. Lane 1 contains the PCR product from 3′-UTR primer 1, and lane 2 the PCR product from 3-UTR primer 2. The 1 kb band from primer 1 and the 0.7 kb band from primer 2 correspond to the expected fragment length from the final polyadenylation site.

of the 5′-flanking region for efficient transcription initiation. The Cox-2 gene is known to be transcriptionally induced by serum, phorbol esters, forskolin and IL-1 in a variety of cell types [2,3]. It will be of interest to determine the contribution of the potential *cis*-acting sequences in the transcriptional induction of the gene in human cells.

The Cox-2 gene consists of 10 exons and 9 introns; it lacks, similar to the murine TIS10 [15], the exon in Cox-1 which encodes the putative signal peptide. The remaining exons are identical in length to those in both Cox-1 and TIS10, with the exception of the varying lengths of the 3′-UTR [11,14,16]. Amino acid residues that were shown to be critical for enzymic catalysis by the sheep Cox-1 protein are also conserved in hCox-2: the axial haem binding site (His^{295}) in exon 7, the distal haem binding site (His^{374}) in exon 8, the active-site tyrosine (Tyr^{371}) in exon 8 and the aspirin acetylation site (Ser^{516}) in exon 10 [17,18]. The N-terminal active-site region of exon 10-encoded polypeptide is highly similar between the two isoenzymes; however, hCox-2 contains a unique 18-amino-acid insertion in the C-terminal region which contains a potential N-linked glycosylation site. hCox-2 exons 2 and 4–10 show between 52% and 75% identity to Cox-1 exons. Exon 3 in the Cox-2 gene possesses the lowest sequence identity with the corresponding region of the hCox-1 gene (38% identity at the amino acid level). A recent crystallographic model of the corresponding region of the ovine Cox-1 protein suggested that it encodes four short amphipathic helical segments (A–D) which comprise the membrane attachment site [19]. In addition, the model predicts that the helices A–D form the entrance for substrate molecules into the 'cyclo-oxygenase channel' from the phospholipid bilayer [19]. It will be of interest to determine the contribution of the highly divergent exon 3-encoded membrane-attachment site in the differential polypeptide stabilities and enzyme activities of the two Cox isoenzymes [20].

One notable feature of the Cox-2 gene structure is that exon 10 is very large. It includes not only the final 410 bp of the coding region, but also 2550 bp of the 3′-UTR. There are no introns present in the 3′-UTR. The 3′-UTR has three potential polyadenylation sequences, approximately 280 bp apart from each other. PCR analysis of cDNA from HUVEC treated with PMA and cycloheximide using a primer upstream of the first site

suggested only the distal polyadenylation signal is used by HUVEC. The large number of AU-rich elements [13], which has been associated with rapidly degraded messages in the 3′-UTR of Cox-2, suggests a high turnover of Cox-2 mRNA. Indeed, the Cox-2 transcript in Cos-7 cells has been shown to be less stable than the Cox-1 mRNA [4]. In endothelial cells, Cox-2 mRNA is induced by both PMA and IL-1 [4,5] Recent data have suggested that the IL-1α induction mechanisms include both transcriptional activation and post-transcriptional mRNA stabilization [5]. Future studies will focus on the structural elements of RNA that are involved in the rapid and regulated degradation of the Cox-2 transcript.

In conclusion, we have determined the complete gene structure of the hCox-2 gene. This work will serve as a basis for studying the transcriptional and post-transcriptional controls of Cox-2 expression.

We thank Mark Evans for computer assistance, Dave DeWitt for transcription factor database analyses, and Javier G. Barriocanal and Jeff Winkles for critical reading of the manuscript. This work is supported by NIH grant HL49094 to T.H.

REFERENCES

1 Needleman, P., Turk, J., Jakshick, B. A., Morrison, A. R. and Lefkowith, J. B. (1986) Annu. Rev. Biochem. **55**, 69–102
2 Smith, W. L. (1992) Am. J. Physiol. **263**, F181–F191
3 Hla, T., Ristimäki, A., Appleby, S. and Barriocanal, J. G. (1993) Ann. N.Y. Acad. Sci. **696**, 197–204
4 Hla, T. and Neilson, K. (1992) Proc. Natl. Acad. Sci. U.S.A. **89**, 7384–7388
5 Ristimäki, A., Garfinkel, S., Wessendorf, J., Maciag, T. and Hla, T. (1994) J. Biol. Chem., **269**, 11769–11775
6 Crofford, L. J., Wilder, R. L., Ristimäki, A., Remmers, E., Epps, H. R. and Hla, T. (1994) J. Clin. Invest., **93**, 1095–1101
7 Church, G. M. and Gilbert, W. (1984) Proc. Natl. Acad. Sci. U.S.A. **81**, 1991–1995
8 Hla, T. and Maciag, T. (1990) J. Biol. Chem. **265**, 9308–9319
9 Sambrook, J., Fritsch, E. F. and Maniatis, T. (1989) Molecular Cloning: A Laboratory Manual, 2nd edn., p.11.39, Cold Spring Harbor Press, Cold Spring Harbor, NY
10 Gorman, C. M., Moffat, L. F. and Howard, B. (1982) Mol. Cell. Biol. **2**, 1044–1051
11 Yokoyama, C. and Tanabe, T. (1989) Biochem. Biophys. Res. Commun. **165**, 888–894
12 Hla, T. and Maciag, T. (1990) Biochem. Biophys. Res. Commun. **167**, 637–643
13 Shaw, G. and Kamen, R. (1986) Cell **46**, 659–667
14 Fletcher, B. S., Kujubu D. A., Perrin D. M. and Herschman, H. R. (1992) J. Biol. Chem. **267**, 4338–4344
15 Kujubu, D. A., Fletcher, B. S., Varnum, B. C., Lim, R. W. and Herschman, H. R. (1991) J. Biol. Chem. **266**, 12866–12872
16 Kraemer, S. A., Meade, E. A. and DeWitt, D. L. (1992) Arch. Biochem. Biophys. **293**, 391–400
17 Shimokawa, T. and Smith, W. L. (1991) J. Biol. Chem. **266**, 6168–6173
18 Shimokawa, T. and Smith, W. L. (1992) J. Biol. Chem. **267**, 12387–12392
19 Picot, D., Loll, P. J. and Garavito, R. M. (1994) Nature (London) **367**, 243–249
20 Holtzman, M. J., Turk, J. and Shornick, L. P. (1992) J. Biol. Chem. **267**, 21438–21445

Received 21 December 1993/4 March 1994; accepted 11 March 1994