Characterization of the Promoter Region of the Human Transforming Growth Factor-β1 Gene*

(Received for publication, June 7, 1988)

Seong-Jin Kim‡, Adam Glick§, Michael B. Sporn, and Anita B. Roberts
From the Laboratory of Chemoprevention, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892

The 5'-end of the human transforming growth factor-β1 gene (TGF-β1) was isolated from a human leukocyte genomic DNA library. Analysis of the transcriptional start sites of human TGF-β1 mRNAs by S1 mapping and primer extension revealed two major start sites 271 nucleotides from one another; several minor sites were also identified. DNA sequence analysis showed that the promoter region contains neither a "TATA" box nor a "CAAT" box, is very G+C rich, and contains 11 CCGCCC repeats. Seven putative binding sites for the transcription factor Sp1 were also identified. To determine the location of sites that may be important for the function of the TGF-β1 promoter, we joined the 5'-end of the TGF-β1 gene to the coding region for chloramphenicol acetyltransferase. The chimeric gene produced high levels of chloramphenicol acetyltransferase activity in transfected HT-1080, A-2B-2, and A-549 cells. Sequences responsible for both promotion and inhibition of transcription were located in the region extending from 1400 to 300 base pairs upstream of the first major TGF-β1 transcriptional start site. The 130-base pair fragment located between 453 and 323 base pairs upstream of this start site contains positive regulatory activity in all cells tested. A second promoter activity was identified in the region between the two major transcriptional start sites. These findings revealed a complex pattern of regulation of human TGF-β1 gene expression.

Materials and Methods

Cloning and Nucleotide Sequence Determination of TGF-β1 5'-Flanking Region—A human leukocyte genomic library in phage EMBL3 (obtained from Clontech) was screened by plaque hybridization (20) using a 0.6-kb BamHI fragment of the 5'-untranslated region of the porcine TGF-β1 cDNA (pTGF-BS3) as probe. The 5.2-kb BamHI-BamHI fragment of clone hTBG103 which hybridized with the porcine TGF-β1 probe by Southern blot analysis (22) was used to clone the human TGF-β1 promoter region. Two overlapping clones were isolated which extended from a HindIII site located 1300 bases upstream of the transcription initiation site to an EcoRI site located 260 bases downstream.

The nucleotide sequence(s) reported in this paper has been submitted to the GenBank™/EMBL Data Bank with accession numbers J04451.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

* The abbreviations used are: TGF-β1, transforming growth factor-β1; kb, kilobase(s); CAT gene, chloramphenicol acetyltransferase gene; HEPES, 4-(2-hydroxyethyl)-1-piperazinethanesulfonic acid; bp, base pair(s); NF-1, nuclear factor-1.
isolated and subcloned into the pUC 18 vector (pTBGl01). This fragment was digested with Nael, and each of four fragments generated was then subcloned into the pUC 18 vector. The nucleotide sequences of the inserts were determined by the dideoxy chain termination method.

Construction of Plasmids—The BamHI-Nael (−1362 to +101) fragment subcloned into pUC18 was digested with BamHI and SstII. After filling in the ends, the resulting fragment (−1362 to +11) was inserted into the Smal site of the multiple cloning site of pGEM4-SVOCAT (the CAT gene was inserted at the SacI site in the pGEM4-SVOCAT vector). All other deletion mutants were constructed by inserting blunt-ended fragments into the Smal site of pGEM4-SVOCAT. The nucleotide sequences of the five 5′-deletion mutants of the BamHI-SstII (−1362 to +11) fragment by use of various restriction endonucleases (see Fig. 3) were determined. A sequence of the 1.463-kb BamHI-Nael, 0.8- and 1.7-kb Nael-Nael fragments, and 1.23-kb Nael-BamHI fragments were cloned into pUC18 and designated phTBGl04, phTBGl05, phTBGl06, and phTBGl07, respectively. A sequence of the 1.463-kb fragment was obtained following subcloning into the mp18 derivative of M13 (Fig. 1). The sequence includes 1362 nucleotides upstream and 101 nucleotides downstream of the 5′-most residue of the human TGF-β1 precursor cdna (12). The nucleotide sequence obtained is in complete agreement with the reported partial genomic sequence (28) and extends that published sequence 225 nucleotides in the 5′-direction.

Transcription Initiation Sites and Analysis of the 5′-Flanking Region of the Human TGF-β1 Gene—Both the S1 nuclease protection assay and primer extension reactions were used to identify the probable transcription initiation sites of human TGF-β1 gene. The transcriptional start sites were located by S1 analysis using the 1180-bp HincII-TaqI fragment extending from the nucleotide −453 to the nucleotide +727 (see Fig. 1). This fragment was end-labeled at the 5′ terminus of the TaqI site prior to digestion with HindII and subsequently isolated from the agarose gels. Using this probe, we were able to observe five or more S1 nuclease-resistant fragments (Fig. 2A) after hybridization with HT1080 mRNA. The two most prominent species of mRNA protected with this probe represented transcriptional start sites at the 5′-most end of human TGF-β1 cdna, as described elsewhere (12), and 271 bp downstream from that site; these sites were also confirmed by primer extension analysis using two different primer fragments (Fig. 2B). The other two sites protected from S1 digestion with this probe gave rise to less intense bands; these were located 470 and 525 bp downstream from the 5′-most end of human TGF-β1 cdna. Other minor start sites are also seen (Fig. 2A).

Examination of the human TGF-β1 gene sequences 5′ to the first major transcription initiation site designated as +1 reveals several notable features. Within this region, the sequence TTCCAAA beginning at −71 and ATTTA at −356 are the closest approximations to the consensus TATA box (29). The sequence between nucleotides −262 and −1 which contains the putative promoter has a G + C content of 80%; moreover, there are nine hexanucleotide repeats (CCGGCC or GGGCGG) at positions, −32, −43, −78, −108, −120, −176, −218, −265, and −313, similar to those found in promoters of viral and cellular housekeeping genes (30–34). These sequences are also found at positions +88 and +175 in the region between the two major transcriptional start sites. The sequence GGGGGCGG and its reverse complement, corresponding to sequences identified as binding sites for the transcription factor SP1 (35), are found in the human TGF-β1 gene at positions +175, −43, −78, −108, −120, −218, and −265. The role of these multiple sites in the regulation of TGF-β1 gene transcription remains to be determined. We also identified the sequences homologous to FSE2 (fat-specific elements) which acts as a negative element in preadipocytes (36). These sequences are found at positions, −1241 and −556 and its reverse complement is present at −1147. The positions of these negative elements are well matched with the CAT activities of human TGF-β1-CAT chimeric genes; however,
the actual function of the FSE2 sequences in the TGF-β1 promoter is still untested. The reverse complement of the nuclear factor 1 (NF-1) consensus sequence (37) is also present at -267. Rossi et al. (38) reported that a NF-1 binding site mediates the transcriptional activation of a mouse α(2)I collagen promoter by TGF-β1. The identification of a NF-1 site in the TGF-β1 promoter supports the suggestion that a similar mechanism might be operative in the TGF-β1 dependent increase in the steady-state levels of its own message in many different normal and transformed cells in cultures (17). The consensus sequences of the transcription factor AP-1 are also identified at positions -418 and -371. The significance of these sequences remains to be determined.

**Cellular Expression Directed by the 5′-Flanking Region of the Human TGF-β1 Gene**—In an attempt to delineate the sequences essential for transcription of the human TGF-β1 gene, we proceeded with a deletion analysis in which various portions of the 5′-flanking region were fused with the coding region of the bacterial chloramphenicol acetyltransferase (CAT) gene in the plasmid pHSG-SVOCAT. Fig. 3 depicts the structure of the 5′ flanking region of the human TGF-β1 gene and the various restriction sites used to generate the fragments cloned into pGEM-SVOCAT. A similar plasmid containing the SV40 early promoter (pSV2CAT) was used as a control (25). These constructs were transfected into HT-1080, AKR-ZB, NIH-3T3, CCL-64, and A-549 cells using the calcium phosphate coprecipitation method. After 48 h, the cells were harvested and the CAT activities in the cell extracts were determined. As shown in Table I, the pHSG5, TGF-β1 promoter-CAT chimeric construct expressed chloramphenicol acetyltransferase activity in both normal and transformed cells, as expected (25). These results clearly indicate that the putative promoter region of the human TGF-β1 gene contains sequences that control the expression of this gene.

**Enhancer-like Element and Putative Negative Regulatory Region in the 5′-Flanking Region of the Human TGF-β1 Gene**—To elucidate the regulatory elements of the TGF-β1 gene, a series of deleted fragments of the 5′-promoter region of the human TGF-β1 gene fused to pGEM-SVOCAT were
transfected into A-549, NIH-3T3, and HT-1080 cells, and the expression of the CAT activity was tested for each plasmid (Table II and Fig. 3). The pattern of expression of the deletion mutants was similar in all of the cell lines examined. Thus, the CAT activity of phTG2 was greater than that of phTG1, and further deletion to phTG3 again reduced the activity. Deletion mutant phTG2 reproducibly resulted in a 2- to 10-fold higher expression of CAT activity than did phTG3. The CAT activity was again increased 5- to 10-fold over that of phTG3 when the deletion reached -453 (phTG5). Further studies with A-549 cells showed additionally that the activity was reduced when the deletion reached -175 (Fig. 3 and Fig. 4A). Taken together, these data suggest that the 5' -flanking sequences of the human TGF-β1 gene contain two negative regulatory regions (-1362 to -1132 and -731 to -453) which repress the strong transcription unit very efficiently. The negative regulatory regions (-1362 to -1132 and -731 to -453) which affect the activity of the human TGF-β1 gene were used to construct plasmids, phTG1 to phTG7, represented on the upper panel by the solid bar and bounded by the indicated restriction endonuclease sites (B, BamHI; P, PstI; X, XbaI; G, HgiAI; H, HincII; Bs, BstEI; Ba, BanII; S, SstI). Arrow (p1) denotes the position of one of the major transcription initiation sites (+1) for human TGF-β1 that was identified by S1 mapping. The column on the right gives the average CAT enzymatic activity obtained in several experiments with each plasmid in A-549 cells. The data are expressed as a percentage of the mean activity obtained with phTG5 (-453 to +11), the plasmid with the strongest promoter activity.

**TABLE I**

Promoter activities of the human TGF-β1 gene

<table>
<thead>
<tr>
<th>Host cell lines</th>
<th>TGF-β1 expression</th>
<th>Relative CAT activity directed by promoter of TGF-β1</th>
</tr>
</thead>
<tbody>
<tr>
<td>A-549</td>
<td>Human lung carcinoma +++</td>
<td>9.7</td>
</tr>
<tr>
<td>NIH-3T3</td>
<td>Murine fibroblast ++</td>
<td>8.0</td>
</tr>
<tr>
<td>CCL-64</td>
<td>Mink lung epithelial cell +</td>
<td>5.0</td>
</tr>
<tr>
<td>AKR-2B</td>
<td>Hamster fibroblast +++</td>
<td>33.0</td>
</tr>
<tr>
<td>HT-1080</td>
<td>Human fibrosarcoma ++++</td>
<td>54.0</td>
</tr>
</tbody>
</table>

Each cell line was transfected with pSV2CAT and phTG5. Relative CAT activities in each cell line are shown as a percentage of pSV2CAT. Values represent the mean of three independent experiments. Relative expression of TGF-β1 mRNA in these five cell lines was determined by Northern RNA blotting (data not shown).

level of CAT mRNA was examined by quantitative S1 nuclease analysis (Fig. 4B). The size of the undigested probe was 458 nucleotides, whereas the size of the CAT-specific fragment protected from S1 nuclease digestion would be 256 nucleotides. Consistent with the results of CAT activities presented above, the CAT mRNA level was higher in cells transfected with phTG5 than that of phTG2 (Fig. 4B, lanes a and b).

**Evidence for Two Active Promoter Regions at the 5'-End of the Human TGF-β1 Gene**—Studies using S1 nuclease analysis indicated that the human TGF-β1 gene directed transcripts from two major start sites, located about 271 bp apart in a segment of DNA that includes a long untranslated leader sequence (see Figs. 1 and 5). To determine whether the region
Position of enhancer-like element (E) and negative elements (N1 and N2) as well as positive element (P) are shown as a percentage of that of pSV2CAT. The numbers in parentheses indicate the activity relative to that of the strongest fragment, phTG5, which is taken as 1.

<table>
<thead>
<tr>
<th>Constructs</th>
<th>Relative CAT activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>phTG1 (-1362 to +11)</td>
<td>0.4 (0.02) 1.2 (0.02) 0.1 (0.006)</td>
</tr>
<tr>
<td>phTG2 (-1131 to +11)</td>
<td>8.77 (0.44) 25.5 (0.47) 5.4 (0.34)</td>
</tr>
<tr>
<td>phTG3 (-731 to +11)</td>
<td>1.4 (0.07) 3.6 (0.06) 1.0 (0.06)</td>
</tr>
<tr>
<td>phTG5 (-453 to +11)</td>
<td>20.0 (1.00) 54.8 (1.00) 15.8 (1.00)</td>
</tr>
</tbody>
</table>

**DISCUSSION**

In this report, we have identified the TGF-β1 promoter and several regulatory elements. We have shown that the 5'-flanking sequences of the TGF-β1 gene contain five distinct regulatory regions including a region having enhancer-like activity, two with negative regulatory activity, and two different promoter regions. The negative regulatory regions (−1362 to −1132 and −731 to −453) strongly repress the activity of the transcriptional unit (Table II). However, enhancer-like sequences at nucleotides −1132 to −731 efficiently overcome the activity of the more downstream (−731 to −453) negative regulatory region. We have also identified sequences located between nucleotides −453 to −323 which have positive regulatory activity; removal of this region completely abolishes the transcriptional capacity of the upstream TGF-β1 promoter. This same 130-nucleotide fragment seems to be responsible for positive regulation of the TGF-β1 gene upon stimulation by the serum or by TGF-β1 itself. Thus, for maximum expression of the hybrid gene, at least 453 bp upstream from the +1 transcription start site are required. Interestingly, our data show that a second promoter (+1 to +271) is also very active. This suggests that sequences downstream from the +1 start site are also required for expression of the human TGF-β1 gene and that one of the major human TGF-β1 mRNAs is likely to be independently regulated and to be transcribed from this second promoter region. c-myc mRNA is also transcribed from two active promoter sites, both of which are followed by an unusually long, untranslated first exon, and in lymphoblastoid cells c-myc transcripts from

3 S. J. Kim, manuscript in preparation.
both promoters appear in differing amounts (39). These results suggest that multiple transcriptional start sites may be common to genes which have an unusually long, untranslated first exon and that multiple promoter sites can be regulated differently.

The profiles of transcriptional activities of the deletion mutant constructs in HT-1080, NIH-3T3, and A-549 cells are similar. Thus, even though the levels of expression of the constructs vary over 10-fold when compared to pSV2CAT (see, for example, expression of pHTG5 in CCL-64 compared to HT-1080 cells, Table I), the relative increases and decreases of activity resulting from the deletions are nearly identical in each of the three cell lines examined in Table II. This leads us to believe that the regulatory mechanisms of control of TGF-β1 gene expression are similar in these cells.

The TGF-β1 gene does not possess typical TATA or CAAT boxes, and the initiation of RNA transcription occurs at multiple sites; the major sites are at +1 and +271, although two less frequently used sites have been identified at +470 and +525. Analysis of the 5'-flanking region of the human TGF-β1 gene has revealed the presence of several sequences which are homologous to the sequences of elements thought to initiate transcription in vitro. The FSE2 element at position -556 can be correlated with the consensus sequence of NF-1, a factor discovered for its ability to initiate replication of adenovirus DNA (46); this sequence is highly homologous to a common sequence found in the promoter region of several genes responsive to 12-O-tetradecanoylphorbol 13-acetate (21).

The extent to which the mechanism of control by TGF-β1 of expression of the mouse α(2)I collagen gene (38) can be generalized to control by TGF-β1 of not only its own expression but also other genes involved in control of extracellular matrix, like those for fibronecin and plasminogen activator inhibitor, remains to be determined. Although it seems reasonable to suspect that the NF-1 site in the 5'-flanking region of the TGF-β1 gene might be involved, we cannot exclude the possibility that alternative mechanisms might be operative. It is clear from our data that the effects of not only positive regulatory regions but also negative regulatory regions must be considered; future experiments will be directed at determination of the precise sequences that mediate the activity of inducers and whether trans-acting factors are involved.

Acknowledgments—We thank Drs. P. Rossi and B. de Crombrugge for pSVOCAT and pSV2CAT, K. T. Jeong for M13 CAT, and P. Kondaisa for porcine TGF-β1 cDNA. We are grateful to Drs. P. Kondaisa, S. Watanabe, and C. Baker for helpful discussions.

REFERENCES

Cell. Biol. 2, 1044–1051
26. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J., and Rutter,
W. J. (1979) Biochemistry 18, 5294–5299
6, 2020–2026
Nucleic Acids Res. 15, 3188–3189
50, 349–383
30. McGrogan, M., Simonsen, C. C., Smouse, D. T., Farnham, P. J.,
31. Melton, D. W., Konecki, D. S., Brennand, J., and Caskey, T.
32. Reynolds, G. A., Basu, S. K., Osborne, T. F., Chiu, D. J., Gill, G.,
38, 275–285
33. Yamaguchi, M., Hirose, F., Hayashi, Y., Nishimoto, Y., and
34. Valerio, D., Duyvesteyn, M. G. C., Dekker, B. M., Weeda, G.,
Berkvens, T. M., van der Voorn, L., van Ormondt, H., and van
36. Distel, R. J., Ro, H. S., Rosen, B. S., Groves, D. L., and Spiegel-
37. Nowock, J., Borgmeyer, U., Puschel, A. W., Rupp, R. A. W., and
38. Rossi, P., Roberts, A. B., Roche, N. S., Karsenty, G., Sporn, M.
39. Battey, J., Moulding, C., Taub, R., Murphy, W., Stewart, T.,
Viruses (Tooze, J., ed) part II, pp. 799–841, Cold Spring
Harbor Laboratory, Cold Spring Harbor, NY
992
601–610
365