Molecular Analysis of G+C-Rich Upstream Sequences Regulating Transcription of the Human Carbonic Anhydrase II Gene

LINDA HILLSTROM SHAPIRO,* PATRICK J. VENTA, AND RICHARD E. TASHIAN

Department of Human Genetics, University of Michigan Medical School, Ann Arbor, Michigan 48109-0618

Received 9 June 1987/Accepted 23 September 1987

The upstream promoter sequences of the human carbonic anhydrase II (CA II) gene have been studied by 5' deletion analysis. Promoter activity was assayed by transfection and chloramphenicol acetyltransferase assay in both human HeLa cells and murine L cells. This investigation showed that the CA II promoter is comparable in activity to that of the simian virus 40 early-region promoter and enhancer and that the CA II upstream sequences exert a different pattern of control in the two cell lines.

The analysis of gene regulation, both in general and at the tissue-specific level, requires a well-characterized gene system. The carbonic anhydrase (CA) isozyme system is well suited for such an analysis, as it has been extensively studied at both the biochemical and molecular level (6, 9). The wide distribution of the CA II isozyme lends itself to particularly interesting aspects of differential gene control, as CA II is found in almost all mammalian tissues, but it is differentially expressed in the cells of these tissues (7). The nucleotide sequences of the 5' regions of the CA II genes of humans and mice show a high degree of homology from 200 bases 5' of

G+C content and nine CCGCCC or GGGCGG boxes) and highly regulated genes (TATA box, β -globin-like tandem repeats, and limited cell-type expression) in the human CA II gene presents the possibility that this is an intermediate-type promoter (2) and may display unique regulatory mechanisms.

The nucleotide sequence of the promoter to exon 1 of the human CA gene is shown in Fig. 1. This region is extremely G+C-rich and contains a high number of transcriptional control elements which have previously been characterized in other systems. First, a TATA box (TATAAAA) is located

FIG. 1. Sequence of the 5' region of the human CA II gene. The proposed TATA (-23 through -29), CCAAT (-78 through -82), and CACC repeat elements (-99 through -125; consensus RgCCtACC-CTGg [8]) are boxed. CCGCCC or GGGCGG elements are underlined twice. Corrections of the original published sequence (8) are underlined (inserted bases). Deletion endpoints are indicated by vertical lines and are given in bold-face numbers. Numbering begins at the transcription start site (J. C. Montgomery, personal communication).

the start codon through at least 60 base pairs (bp) 5' of the second intron (8) and are also highly G+C-rich (nearly 80% for those bases 5' of the cap site). The presence of individual features characteristic of both housekeeping genes (high

101 bp upstream from the translational start site, a possible CCAAT box (CCACC) is found 48 bp upstream from the TATA box, and the 5' region contains nine CCGCCC or GGGCGG hexanucleotide sequences (GC boxes). Finally, tandemly repeated elements appear 16 and 20 bp upstream from the putative CAAT box. These elements are similar to the tandem repeats found in the β -globin gene (consensus

* Corresponding author.

FIG. 2. Scheme of the construction of parent plasmid pHCAIIcat and subsequent deletion plasmids. The *EcoRI*-*TthIII* 1 fragment from H25-3.8 (5' end of human CA II) containing 1,400 bp of the human CA II upstream region was ligated into the *HindIII* site of pSV2CAT. The resultant plasmid, pHCAIIcat, contains all SV40 promoter and enhancer sequences and 1,400 bp of the CA II gene upstream of the bacterial CAT gene. Deletion plasmids were constructed by removal of sequences between the *AccI* site and various sites within the CA II 5'-flanking region. The resultant deletion plasmids contain no SV40 promoter or enhancer sequences.

RGCCcACC-CTGg [8]), which is required for maximal levels of gene transcription in β -globin and certain other genes (5, 10, 11).

To functionally define the *cis*-acting control elements required for the expression of the human CA II gene, 1,400 bases of the 5'-flanking region of this gene (obtained from the genomic clone H25-3.8) were ligated immediately upstream from the chloramphenicol acetyltransferase (CAT) gene in the pSV2cat plasmid (3) (Fig. 1 and 2). The resultant plasmid, pHCAIIcat, contained all putative control elements of the CA II gene sequences (TATA box, CCAAT box, and GC boxes) up to the initiation codon and the simian virus 40 (SV40) promoter and enhancer elements. Subsequent deletion plasmids were constructed by partial digestion with

appropriate restriction enzymes to obtain the desired deletion. These deletion plasmids contained only CA II regulatory sequences, and all SV40 promoter and enhancer elements were deleted. Mutant test promoter plasmids are designated by the deletion endpoint (e.g., plasmid -53 contains 53 nucleotides immediately upstream from the cap site; all sequences 5' to this point have been deleted).

Because the CA II promoter region contained a high number of transcriptional regulatory elements, it was of interest to compare the efficiency of this promoter with that of the SV40 early region. Transfection of murine L-cell fibroblasts with a deletion plasmid containing 1,400 bp of CA II upstream sequence and no SV40 sequences 5' to the CAT gene showed that the CA II 5'-flanking region promotes transcription 70 to 100% as efficiently as the SV40 early region promoter-enhancer (data not shown).

A representative transfection experiment of human CA II deletion plasmids into L cells is shown in Fig. 3a. As expected, the presence of known regulatory sequences was crucial for maximal transcriptional activity of the promoter when measured by the CAT assay. More specifically, assessment of the function of the β -globin-like distal tandem repeats in the human CA II gene indicated that a deletion which removed 10 bp of the 5' repeat sequence (plus all upstream sequences) but left the entire 3' repeat sequence intact decreased promoter efficiency to 50% of the wild-type level in murine cells. Analogous experiments with the rabbit β -globin gene (1) showed that removal of the β -globin 5' repeat and all upstream sequences decreased transcription to 30% of the wild-type level, and experiments with internal deletion of either element suggested that both copies of the repeated sequence contributed to promoter function. Our results would suggest that the repeats are also required for optimal transcription in the human CA II gene. The presence of direct repeats is also somewhat characteristic of the promoters of housekeeping genes (2), although the significance of these regions is not clear. The repeat elements found in the CA II upstream region may function more as housekeeping gene promoter elements than as promoter elements operative in a gene more limited in tissue expression.

The presence of nine CCGCCC or GGGCGG hexanucleotides located throughout the human CA II promoter region (Fig. 1) suggests that the Sp1 transcription factor (4) modulates transcription of the human CA II gene. Transfection analysis of the role of each of the nine GC elements in the human CA II promoter in murine cells generally showed modest (10 to 30%) increases in transcription levels when each element was added, until maximum levels were reached, when 270 bp of 5' sequence was included in the plasmid. These results suggest that each of these potential binding sites contributes to CA II transcription, yet no single one appears notably prominent in its ability to promote transcription in the murine cell.

Quantitative results of repeated transfections of the deletion plasmids are shown in Fig. 3b. Each plasmid was tested in at least seven individual experiments of two different purified plasmid preparations and are the result of at least two independently isolated deletion plasmids.

To compare the ability of the human CA II 5' regulatory elements analyzed in the murine assay to function in the presence of autologous transcription factors, the above deletion plasmids were transfected into HeLa cells. The HeLa line is of human epithelial origin and is negative for CA II expression (unpublished results). The pattern of control in the human cells was strikingly different from that seen in the

FIG. 3. (a) A representative CAT assay demonstrating the effect of deletion of 5'-flanking sequences of the human CA II gene on promoter activity in murine Ltk⁻ fibroblasts. Cells (1.5×10^6) were transfected (12) with 3 μ g of test plasmid and assayed for CAT activity (3) at 48 h. Deletion mutants are described in the legends to Fig. 1 and 2. (b) Histogram showing the mean relative CAT activity of CA II deletion plasmids in murine Ltk⁻ cells (at least seven experiments). Solid lines in the lower half indicate sequences remaining in the plasmid 5' to the CAT gene. The positions of the various promoter elements are shown on the abscissa (numbering begins at cap site), and relative CAT activity is expressed as counts per minute of [14 C]chloramphenicol acetylated in cells transfected with the test plasmid/counts per minute of [14 C]chloramphenicol acetylated in transfectants containing the reference plasmid. The reference plasmid contains 1,400 bp of the human CA II 5'-flanking region plus an additional 50 bp of CAT vector sequence not included in the test plasmids. The standard error of the mean of each test condition was never higher than 10% of the test value. Lower levels of CAT activity in the -1400 plasmid are due to the fact that less plasmid is being transfected on a molar basis in these conditions.

FIG. 4. (a) A representative CAT assay showing the effect of 5' deletions on human CA II promoter activity in human epithelial cells (HeLa) and suggesting species specificity. Details are as described in the legend to Fig. 3. (b) Histogram showing the mean relative CAT activity of human CA II deletion plasmids in HeLa cells. Data represent the mean of six transfection experiments. The standard error of the mean of each test condition was never higher than 12% of the test value.

murine cells (Fig. 4). In the L cell transfections, addition of each regulatory element resulted in a stepwise increase in the ability of the promoter to direct transcription of the reporter gene, beginning with the TATA box through each putative element until maximal levels were reached when all known elements were included. In the human cells, however, transcription above background is not seen until 149 bases of 5' sequence are included in the test plasmid. At this point, CAT activity increased 10-fold to nearly maximal levels. This may be due to the fact that the region between -136 and -149 contains a perfect 10-base match with the sequence GCCCGCCCC, which has been determined to be the highest-affinity Sp1-binding sequence studied to date (4). Because the human CA II 5' region has similarities to other documented Sp1-responsive promoters, it is likely that human CA II transcription is being modulated by the transcription factor Sp1 in human cell lines. It is interesting that the sequence which corresponds to the -136 through -149 region in the mouse CA II gene contains a CCGCCC hexanucleotide sequence but does not fit the expanded decanucleotide sequence (8). Therefore, it is not surprising that the region between -136 and -149 is important in the transcriptional control of the human CA II gene in human cells but not in a murine system which does not possess the high-affinity binding site.

Quantitative results of six transfections of the deletion plasmids into HeLa cells are combined in Fig. 4b. As before, results reflect two independent preparations of each deletion plasmid.

We thank Myron Levine for the L and HeLa cell lines and the use of tissue culture facilities, Ya-Shiou Yu for growing some of the plasmids, Carol Laherty for help with the sequence analysis, and Anita McDermott for preparation of the manuscript.

The research was supported by Public Health Service grant GM-24681 from the National Institutes of Health. L.H.S. was supported by National Research Service Award 1F32AROF920-01.

LITERATURE CITED

1. Dierks, P., A. Van Ooyen, M. D. Cochran, C. Dobkin, J. Reiser, and C. Weissman. 1983. Three regions upstream from the cap site are required for efficient and accurate transcription of the rabbit β -globin gene in mouse 3T6 cells. *Cell* 32:695-706.
2. Dynan, W. S. 1986. Promoters for housekeeping genes. *Trends Genet.* 2:196-197.
3. Gorman, C. M., L. F. Moffat, and B. H. Howard. 1982. Recombinant genomes which express chloramphenicol acetyltransferase in mammalian cells. *Mol. Cell. Biol.* 2:1044-1051.
4. Kadonaga, J. T., K. A. Jones, and R. Tjian. 1986. Promoter-specific activation of polymerase II transcription by Sp1. *Trends Biochem. Soc.* 11:20-23.
5. Myers, R. M., K. Tilly, and T. Maniatis. 1986. Fine structure analysis of a β -globin promoter. *Science* 232:613-618.
6. Tashian, R. E., and D. Hewett-Emmett (ed.). 1984. Biology and chemistry of the carbonic anhydrases. *Ann. N.Y. Acad. Sci.* 429:1-640.
7. Tashian, R. E., D. Hewett-Emmett, S. J. Dodgson, R. E. Forster, and W. S. Sly. 1984. The value of inherited deficiencies of human carbonic anhydrase isozymes in understanding their cellular roles. *Ann. N.Y. Acad. Sci.* 429:262-275.
8. Venta, P. J., J. C. Montgomery, D. Hewett-Emmett, and R. E. Tashian. 1985. Comparison of the 5' regions of human and mouse carbonic anhydrase II genes and the identification of possible regulatory elements. *Biochim. Biophys. Acta* 86:195-201.
9. Venta, P. J., J. C. Montgomery, and R. E. Tashian. 1987. Molecular genetics of carbonic anhydrase isozymes. In M. C. Rattazzi, J. G. Scandalios, and G. S. Whitt (ed.), *Isozymes: current topics in biological and medical research*. Alan R. Liss, Inc., New York.
10. Weiher, H., and M. R. Botchan. 1984. An enhancer sequence from bovine papilloma virus DNA consists of two essential regions. *Nucleic Acids Res.* 12:2901-2916.
11. Weiher, H., M. Konig, and P. Gruss. 1983. Multiple point mutations affecting the simian virus 40 enhancer. *Science* 219:626-631.
12. Wigler, M., A. Pellicer, S. Silverstein, and R. Axel. 1978. Biochemical transfer of single-copy eukaryotic genes using total cellular DNA as donor. *Cell* 14:725-730.