

SHORT COMMUNICATION

Sequence of the WT1 Upstream Region Including the Wit-1 Gene

MANFRED GESSLER*¹ AND GAIL A. P. BRUNST†

*Institut für Humangenetik, Philipps-Universität, Bahnhofstrasse 7A, D-35037 Marburg, Germany; and †Genetics Division, Children's Hospital, 320 Longwood Avenue, Boston, Massachusetts 02115

Received February 1, 1993; revised May 10, 1993

The Wilms tumor gene WT1 encodes a Cys₂His₂-type zinc finger protein that can bind DNA and function as a transcriptional regulator. The pathological spectrum of tumorigenesis and various developmental defects produced by different WT1 alterations suggests that WT1 controls a number of subsequent effector genes. To define the role of WT1 in these developmental processes it will be important to elucidate mechanisms that govern expression of WT1 itself. To facilitate mapping of the WT1 promoter region and 5' control elements we have determined the sequence upstream of the WT1 transcription unit. This includes the Wit-1 gene that is transcribed in the opposite direction. © 1993 Academic Press, Inc.

Homozygous deletion or point mutations of the WT1 gene have been found as a key step in the development of some Wilms tumors (1, 3). In addition, genitourinary abnormalities such as cryptorchidism and hypospadias can be caused by the inactivation or deletion of one copy of the gene in germline DNA (8, 11). A much more severe phenotype of nephropathy, intersex, and a high incidence of Wilms tumors are seen in patients with Denys-Drash syndrome, which appears to be due to specific missense mutations within the zinc finger domain of the WT1 protein (9). This diversity of phenotypes indicates that the WT1 product is involved in several developmental pathways. There have been no reports, however, on the regulation of WT1 expression itself. The present study describes the sequence characteristics of the WT1 upstream region to facilitate the identification elements controlling WT1 expression.

The entire WT1 locus has been cloned in overlapping cosmid clones that have been used to determine the exon-intron structure of the gene (4). The region 5' to exon 1 of the WT1 gene was first mapped by restriction enzyme digestion of cosmid clones and suitable DNA fragments were subcloned into plasmid vectors for fine mapping and sequencing. The sequence was determined on double-stranded plasmid DNA (Sequenase, USB) using vector primers supplemented with specific oligonu-

cleotide primers in some instances. All sequences were determined at least twice and more than half of the sequence was confirmed by reading the other strand. Sequence assembly and comparison were done using the HUSAR/UWCGC package at DKFZ, Heidelberg.

Parts of the WT1 sequence upstream and downstream of exon 1 have already been published (4; X61631). An additional 4010 nucleotides have been determined in this study, giving a total contiguous sequence of 5503 bp (X69950, HSWT1WIT). The important features of the sequence are graphically depicted in Fig. 1. The entire region is characterized by a high overall G/C content of 59% with little suppression of CpG dinucleotide frequency. Especially the areas around nucleotides 400–1000 and after nucleotide 4000 likely represent true CpG islands. They fulfill the criteria of moving averages of >50% C + G and a value of >0.6 for observed vs expected frequency of CpG dinucleotides over long stretches (2, 7). In addition, they contain sites for restriction enzymes that are sensitive to CpG methylation. These sites are efficiently cut in genomic DNA from different sources (data not shown), suggesting that this area is under-methylated. The most prominent CpG island region covers the entire exon 1 of WT1 and spreads into intron 1 and the presumed promoter region.

Some of the important control elements of WT1 expression are likely located within a stretch of 419 nucleotides (3986–4404) (Fig. 2) that exhibits strong similarity (79%) to the murine DNA sequence upstream of the major transcriptional start site (10). Both sequences contain two copies of the consensus binding motif (ccgcc) for the transcription factor Sp1 at similar locations. The sequence TCTTATTT (nt 4369–4376) may represent a TATA box, although this is not typical of G/C-rich promoters. There are several additional possible binding sites for Sp1 or related factors in the conserved 5' untranslated region of the human and murine genes. Any putative factor binding site, however, will have to be confirmed by binding assays or functional studies. There are only 179 bp of additional murine sequence known beyond this region of high similarity. This sequence is only distantly related to the human sequence and it remains unclear whether there are other regions of sequence conservation further upstream.

A second divergent transcript originating from this

¹ To whom correspondence should be addressed. Telephone: 01149-6421-283584. Fax: 01149-6421-285630.

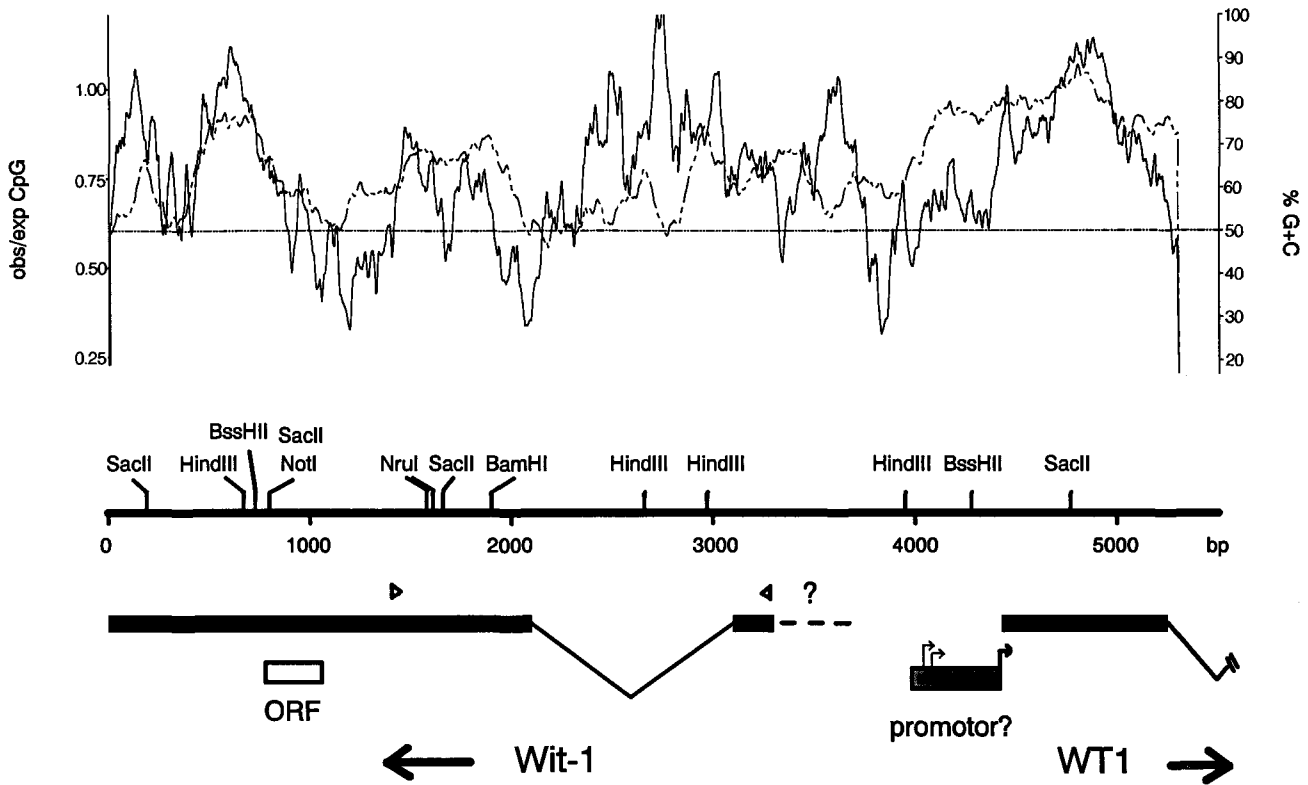
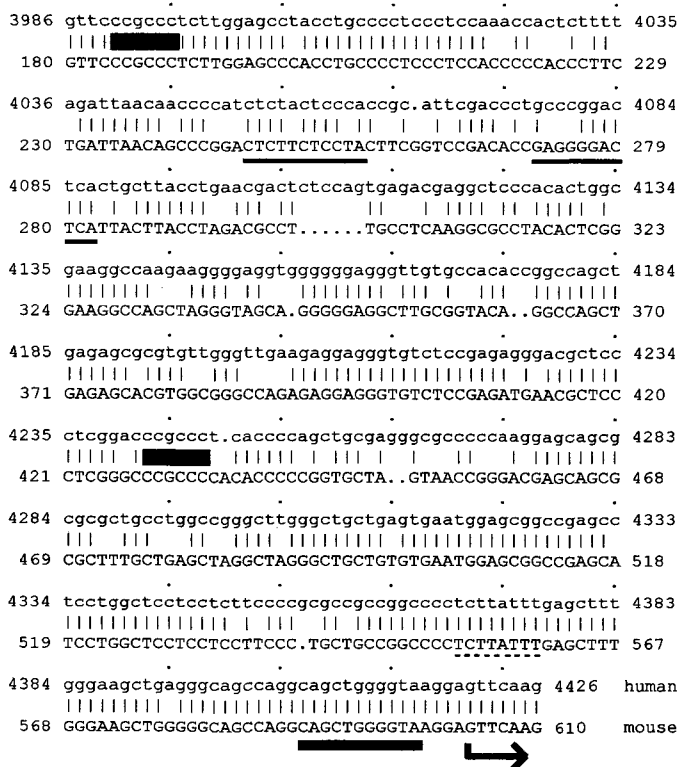


FIG. 1. Organization of the WT1 upstream region. The upper graph shows the distribution of G/C content (broken line) and observed vs expected frequency of CpG dinucleotides (solid line) along the sequence (using the basepair plot program of R. Lopez (7)). Criteria for CpG islands are $C/G > 50\%$ and $Obs/Exp\ CpG > 0.6$ over more than 200 consecutive bp. The diagram (lower graph) depicts part of the restriction map together with the location of the Wit-1 and WT1 exons. The precise start of the Wit-1 5' exon is not known. Triangles give the location of primers WIS3 and WIS5 used for the amplification of the Wit-1 cDNA fragment. The single short open reading frame for Wit-1 is highlighted by an open box. The shaded box marks the region of high sequence conservation between the human and the mouse WT1 upstream regions that likely contain some of the transcriptional control elements. The position of transcriptional start sites is deduced from the murine sequence.



genomic locus was first described by Huang *et al.* (6). The 2.1-kb Wit-1 sequence published closely resembles the one we have determined (11 deletions/insertions of single nucleotides and three substitutions in a total of 2013 nucleotides). Based on RNase protection assays, Huang *et al.* have postulated the presence of four exons for this transcript, but only the last exon including a poly(A) tract was present in their cDNA clones. To determine the precise exon-intron boundaries we have used primers complementary to sequences in the presumed first and fourth exon to isolate this transcript from kidney cDNA. A single DNA fragment could be amplified from cDNA prepared from either Wilms tumors or fetal kidney (Fig. 3). Upon direct sequencing of the PCR product and comparison to the genomic se-

FIG. 2. Alignment of the putative human and mouse WT1 promoter regions. The human sequence is shown in lowercase letters. The murine sequence (uppercase letters) and the position of the minor (thin underline) and major (thick underline) transcriptional start sites of the mouse WT1 gene are taken from Pelletier *et al.* (10). Optimal alignment was produced with the HUSAR/UWGCG program BESTFIT. The putative Sp1 binding sites are marked by boxes. A possible TATA box is highlighted by a dashed underline. The arrow denotes the start of the LK15 cDNA, which represents the longest human WT1 cDNA clone isolated (3).

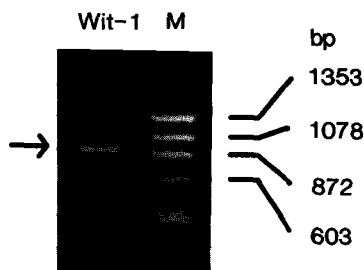


FIG. 3. Amplification of the Wit-1 transcript from human fetal kidney cDNA. Only a single DNA fragment is amplified from fetal kidney cDNA (Wit-1, marked by arrow). The size was estimated as 890 bp by comparison with a known size marker (M). There was no DNA fragment detected corresponding to that amplified from genomic DNA (≈ 1.9 kb). For the amplification reaction single-stranded cDNA was prepared from 5 μ g of total fetal kidney RNA by oligo(dT) priming using the Pharmacia first-strand kit. The PCR reaction was then carried out using $\frac{1}{20}$ of the cDNA and primers corresponding to nucleotides 1403–1422 and 3293–3274 of the genomic sequence. Amplification conditions were 30 s at 94°C, 30 s at 58°C, and 30 s at 72°C for 40 cycles. For direct sequencing the amplified DNA fragment was recovered from a 1% agarose gel by glass milk extraction (BIO101). Sequencing was performed with $\frac{1}{6}$ of the products from a 50- μ l PCR reaction using the Promega fmol kit with 32 P end-labeled primers.

quence a single splice junction was identified at nucleotides 3104/3105 and 2098/2097 (AG/gtgcgtat . . . intron . . . cattttttctcag/AA). Both splice sites also received one of the highest scores when the entire sequence was assayed for putative exons using the GeneId software (5). There was no evidence for two additional intervening exons that were postulated from the RNase protection data. The PCR assay predicts a 5' exon of at least 189 nucleotides in length, but we have not attempted to identify the 5' end of the Wit-1 transcript. The currently known combined sequence of 2.4 kb for the Wit-1 transcript only contains a short open reading frame as described previously (6), the functional significance of which is unclear at the moment.

The present study defines the sequence structure of the WT1 upstream region and will serve as a reference for future identification and mapping of regulatory elements for the expression of WT1. The promoters for both the WT1 and Wit-1 transcripts likely fall within less than 1.1 kb and there may be only a single promoter region of ≈ 420 bp with bidirectional function.

ACKNOWLEDGMENTS

This work was supported by Grants Ge539/2-2 (Deutsche Forschungsgemeinschaft) and HG00186 (National Institutes of Health).

REFERENCES

1. Call, K. M., Glaser, T., Ito, C. Y., Buckler, A. J., Pelletier, J., Haber, D. A., Rose, E. A., Kral, A., Yeger, H., Lewis, W. H., Jones, C., and Housman, D. E. (1990). Isolation and characterization of a zinc finger polypeptide gene at the human chromosome 11 Wilms' tumor locus. *Cell* **60**: 509–520.
2. Gardiner-Garden, M., and Frommer, M. (1987). CpG islands in vertebrate genomes. *J. Mol. Biol.* **196**: 261–282.
3. Gessler, M., Poustka, A., Cavenee, W., Neve, R. L., Orkin, S. H., and Bruns, G. A. P. (1990). Homozygous deletion in Wilms tumours of a zinc-finger gene identified by chromosome jumping. *Nature* **343**: 774–778.
4. Gessler, M., König, A., and Bruns, G. A. P. (1992). The genomic organization and expression of the WT1 gene. *Genomics* **12**: 807–813.
5. Guigo, R., Knudsen, S., Drake, N., and Smith, T. (1992). Prediction of gene structure. *J. Mol. Biol.* **226**: 141–157.
6. Huang, A., Campbell, C. E., Bonetta, L., McAndrews-Hill, M. S., Chilton-MacNeill, S., Coppes, M. J., Law, D. J., Feinberg, A. P., Yeger, H., and Williams, B. R. G. (1990). Tissue, developmental, and tumor-specific expression of divergent transcripts in Wilms tumor. *Science* **250**: 991–994.
7. Larsen, F., Gundersen, G., Lopez, R., and Prydz, H. (1992). CpG islands as gene markers in the human genome. *Genomics* **13**: 1095–1107.
8. Pelletier, J., Bruening, W., Li, F. P., Haber, D. A., Glaser, T., and Housman, D. E. (1991a). WT1 mutations contribute to abnormal genital system development and hereditary Wilms' tumour. *Nature* **353**: 431–434.
9. Pelletier, J., Bruening, W., Kashtan, C. E., Mauer, S. M., Manivel, J. C., Striegel, J. E., Houghton, D. C., Junien, C., Habib, R., Fouser, L., Fine, R. N., Silverman, B. L., Haber, D. A., and Housman, D. (1991b). Germline mutations in the Wilms tumor suppressor gene are associated with abnormal urogenital development in Deysn-Drash syndrome. *Cell* **67**: 437–447.
10. Pelletier, J., Schalling, M., Buckler, A. J., Rogers, A., Haber, D. A., and Housman, D. (1991c). Expression of the Wilms tumor gene WT1 in the murine urogenital system. *Genes Dev.* **5**: 1345–1356.
11. van Heyningen, V., Bickmore, W. A., Seawright, A., Fletcher, J. M., Maule, J., Fekete, G., Gessler, M., Bruns, G. A. P., Huerre-Jeanpierre, C., Junien, C., Williams, B. R. G., and Hastie, N. D. (1990). Role for the Wilms tumor gene in genital development? *Proc. Natl. Acad. Sci. USA* **87**: 5383–5386, 1990.