Alternative splicing and genomic structure of the Wilms tumor gene WTI
(alternative splicing/genomic structure)

DANIEL A. HABER*†‡, ROBERT L. SOHN*, ALAN J. BUCKLER*, JERRY PELLETIER*, KATHERINE M. CALL*, and DAVID E. HOUSMAN*

*Center for Cancer Research, Massachusetts Institute of Technology, Cambridge, MA 02139; and †Massachusetts General Hospital Cancer Center, Charlestown, MA 02114

Communicated by Kurt J. Isselbacher, July 15, 1991

ABSTRACT The chromosome 11p13 Wilms tumor susceptibility gene WTI appears to play a crucial role in regulating the proliferation and differentiation of nephroblasts and gonadal tissue. The WTI gene consists of 10 exons, encoding a complex pattern of mRNA species: four distinct transcripts are expressed, reflecting the presence or absence of two alternative splices. Splice I consists of a separate exon, encoding 17 amino acids, which is inserted between the proline-rich amino terminus and the zinc finger domains. Splice II arises from the use of an alternative 5' splice junction and results in the insertion of 3 amino acids between zinc fingers 3 and 4. RNase protection analysis demonstrates that the most prevalent splice variant in both human and mouse is that which contains both alternative splices, whereas the least common is the transcript missing both splices. The relative distribution of splice variants is highly conserved between normal fetal kidney tissue and Wilms tumors that have intact WTI transcripts. The ratio of these different WTI mRNA species is also maintained as a function of development in the mouse kidney and in various mouse tissues expressing WTI. The conservation in structure and relative levels of each of the four WTI mRNA species suggests that each encoded polypeptide makes a significant contribution to normal gene function. The control of cellular proliferation and differentiation exerted by the WTI gene products may involve interactions between four polypeptides with distinct targets and functions.

The Wilms tumor susceptibility gene WTI, mapping to the chromosome 11p13 locus, has recently been isolated (1, 2). Wilms tumor is thought to arise from the primitive cells of the developing kidney, and WTI is expressed at high levels in normal fetal kidney, both in humans and in the mouse (3–5). As for retinoblastoma, genetic evidence has pointed to the inactivation of a potential tumor-suppressor gene as a rate-limiting step in the genesis of Wilms tumor (6). Most Wilms tumor specimens contain high levels of WTI mRNA, consistent with their presumed cell of origin, and some tumors have been demonstrated to have internal deletions disrupting the WTI gene (3, 7–9). WTI, however, appears to differ from other tumor-suppressor genes like Rb1 in its relatively restricted normal tissue distribution and its developmental regulation of expression (4, 5).

The predicted WTI protein is 45 to 49 kDa in size and contains two domains that suggest it functions as a transcription factor. The carboxyl terminus of the protein consists of four Cys-His zinc finger domains, which have a high degree of homology to the early growth response 1 and 2 gene products (10, 11), and the WTI zinc finger cluster has been shown to bind the early growth response 1 DNA recognition sequence (12). The amino terminus of WTI protein contains another domain that is rich in proline and glutamine residues (13), which may be involved in mediating transactivation. However, functional analysis of the WTI gene product is complicated by the existence of four distinct transcripts, which result from alternative splicing. We have characterized the genomic structure of the WTI gene§ and of these alternative splices, and show that the relative proportion of these splice variants is conserved in all tissues that express WTI. The distinct polypeptides encoded by the WTI gene may have different targets as well as functions, contributing to its role as a regulator of cell growth.

MATERIALS AND METHODS

RNA PCR Analysis and Identification of Alternative Splices. Wilms tumor samples and normal human kidney tissues have been described (3), and mouse tissues from different stages of development have been reported (5, 14). For RNA PCR, total cellular RNA was extracted by the LiCl/urea method, reverse transcribed by using random hexanucleotide primers, and PCR amplified as described (3). Oligonucleotide primers for amplification of splice I from human RNA were no. 106 (GGAATCAGATGAACTTAGGAG) and no. 26 (CCT-GAATGCCTCTGAAACACCC). Primers for amplification of alternative splice II were no. 16 (TCCCGTGTCGAC-CACCTG) and no. 301 (GCCACCGACAGCAGAAAGGCG).

Generation of Exon/Intron Map. The WT33 cDNA clone and overlapping genomic clones have been described (1). The genomic clones were subcloned, and fragments containing WTI exons were identified by hybridization to oligonucleotides derived from the cDNA sequence. These primers were then used in dideoxynucleotide sequencing (15) to identify the exon/intron junction and the sequence of the downstream intron. Antisense primers complementary to the intron sequences were then used to determine the sequence of the upstream intron.

RNase Protection Analysis. A human cDNA fragment spanning both alternative splices (nucleotides 280–1240 of cDNA WT33 (1) was cloned into the Bluescript vector (Stratagene). With standard techniques, an RNA probe was synthesized using T7 polymerase (New England Biolabs). The probe contained both the WTI sequence and a stretch of vector sequence, allowing undigested probe to be distinguished from fully protected WTI sequence (splice variant A). Five hundred thousand counts per minute of probe were hybridized overnight to 10 μg of total cytoplasmic RNA at 55°C in 80% (vol/vol) formamide. Pipes, pH 7.4, then digested with RNase A and RNase H, precipitated, and electrophoresed on a 6% denaturing acrylamide gel. A DNA fragment comparable to the human probe was derived from nucleotides 730 to 1757 of the full-length mouse cDNA (5). RNase protection was done as described above, except for use of T3 polymerase.

To whom reprint requests should be addressed.
§The sequence reported in this paper has been deposited in the GenBank data base (accession no. M74917).

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.
ase to generate the RNA probe, and hybridization to RNA at 60°C. After electrophoresis, the gels were dried and analyzed using a Phosphorimager system (Molecular Dynamics, Sunnyvale, CA). The radioactivity emitted by each protected fragment was quantitated, corrected for gel background, divided by size of the fragment, and then expressed as a fraction of splice variant A.

RESULTS

Alternative Splices in WTI Transcript. RNA PCR analysis of the WTI transcript in normal fetal kidney and in Wilms tumors showed different PCR products resulting from the variable insertion of two sequences. PCR products spanning these insertions were cloned and sequenced. Alternative splice I consists of 51 nucleotides, encoding 17 amino acids, including 5 serines and 1 threonine, potential sites of protein phosphorylation (Fig. 1). Splice I is inserted between the proline-rich amino terminus of the predicted protein and the first zinc finger domain. We have previously shown that splice II contains 9 nucleotides inserted between zinc fingers 3 and 4 (3). This second splice encodes 3 amino acids, of which one is a serine and one a threonine.

Genomic Structure of WTI. In our earlier analysis of WTI in Wilms tumors, we defined the genomic structure of the zinc finger domains within which we had detected a small deleted segment. This analysis demonstrated that each zinc finger is separated from the next by a short intron. It also identified the nine nucleotides of splice II as derived from intron sequence adjacent to the 5' splice junction. As such, splice II reflects the use of an alternative splice donor sequence. So as to characterize the origin of the splice I sequence and to facilitate further mutational analysis of the WTI gene, we undertook to sequence all the remaining exon/intron junctions. The WTI gene spans 50 kilobases (kb) of genomic DNA, and overlapping cosmids covering this area have been described (1). The approximate position of introns was derived by comparing PCR analysis of genomic clones with that of the WT33 cDNA. However, almost all introns proved too large to reliably PCR amplify from genomic DNA. Accordingly, subclones corresponding to WTI exons were identified by hybridization to oligonucleotide primers derived from exon sequences. These were then used to sequence the exons and neighboring intron sequences in both directions. The initiation of homologous mouse sequence (5). Analysis of sequence upstream of the first ATG in mouse cDNAs shows termination codons in all three reading frames.

The exon/intron junction sequences of the human WTI gene are shown in Fig. 2, and the genomic structure is schematically represented in Fig. 3. The gene consists of 10 exons and, as noted previously, each of the four zinc finger domains is contained within a separate exon. The proline-rich amino-terminus domain, on the other hand, is encoded by the first exon alone, and the 51 nucleotides of alternative splice I compose exon 5. The molecular mechanisms resulting in alternative mRNA splicing are poorly understood but are thought to reflect both nucleotide sequence information contained in the splice junction, as well as cell type-specific regulatory factors (16). It is of interest that we have recently characterized a mutant WTI transcript from a Wilms tumor in which exon 6 is lost as a result of a genomic deletion (7). The resultant WT1 transcript nonetheless continues to demonstrate alternative splicing, with exon 5 inserted between exons 4 and 7, suggesting that the sequences mediating alternative splicing are preserved.

Relative Abundance of WTI Splice Forms. The presence of two alternative splices in the WTI transcript may reflect a degree of complexity in gene product function. Both the sequence and position of these splices are highly conserved in the mouse (5), further suggesting functional significance. As a first step toward understanding the role of these splices, we determined their relative prevalence in RNA from different tissues. WT1 mRNA is readily detected by Northern (RNA) blot in most Wilms tumors, as well as in normal fetal kidney tissue (3). In the mouse, where more detailed analysis is feasible, we have detected WT1 mRNA in developing kidney, as well as in adult testis, ovary, uterus, and spleen, and, to a lesser extent, in heart and lung (5, 14). With the more sensitive RNA PCR technique, alternatively spliced WTI transcripts can be readily demonstrated in virtually all tissues (data not shown). However, PCR analysis can only provide an approximate ratio of the various RNA species and, due to the positions and sizes of the two alternative splices, it cannot be used to distinguish various splicing combinations. The presence of two alternative splices in the WTI transcript results in the generation of four possible RNA species. We have defined these as follows: splice variant A, neither splice present; variant B, splice I present and splice II absent; variant C, splice I absent and splice II present; variant D, both splices present.

To determine the existence and relative abundance of these various forms of the WTI transcript, we developed an RNase protection assay capable of differentiating each form based on a protected fragment of distinctive length (Fig. 4a). Both human and mouse probes representing splice variant A were used in these experiments, and both generated protected fragments of expected size (Fig. 4 b and c). Nonspecific bands were also noted, as expected, given the length of the RNA probe. In addition, unexpected protected fragments were found in both human [530 base pairs (bp)] and mouse [460 bp and 560 bp] tissues. Because extensive RNA PCR analysis detected no other alternative WTI splices, these fragments may represent distinct transcripts with extensive homology to a portion of the WTI gene. To quantitate the relative amounts of WTI splice variants present in a given tissue, the radioactivity emitted by each protected band was measured on a Phosphorimager. The number of counts was then corrected for size of the body-labeled protected fragment and expressed as a ratio relative to splice variant A (see Tables 1 and 2). The RNA probe was labeled by uridine incorporation and because it did not include the highly G + C-rich sequence at the 5' end of the gene, specific activity was comparable throughout the probe.

As expected, WT1 mRNA was readily detected in human 20-week kidney and in Wilms tumors but not in adult human kidney (Fig. 4b). Analysis of these six Wilms tumors did not demonstrate any new fragments, consistent with prior analysis that did not indicate any mutation in these transcripts (3). The relative ratio of the four spliced variants in normal developing kidney is A/B/C/D, 1:2.5:3.8:8.3, which, given experimental variation, is similar to that seen in the Wilms tumors. These proportions agree with RNA PCR data suggesting a 2:1 ratio for the presence and absence of splice I and a 5:1 ratio for the presence and absence of splice II (ref. 3, and D.A.H., unpub-
Fig. 2. Nucleotide sequence of WTI exon/intron junctions. Exon/intron junctions of WTI were sequenced as described. Intron sequences are shown in lowercase letters. Exon sequences are shown in uppercase letters and are numbered, beginning with the first ATG and ending with the termination codon (underlined). The two alternative splices are shown in boxes.
lished data). The relative excess of splice form D, however, can only be appreciated by such an RNase protection analysis.

To facilitate analysis of WT1 regulation, we have isolated its mouse homolog and measured its expression as a function of development (5). In the mouse kidney, WT1 mRNA content peaks at day 17 of gestation and then rapidly declines after birth. RNase protection analysis confirms this time scale (Fig. 4c) and shows no detectable difference in the relative ratios of the WT1 splice variants as a function of kidney development. On day 3 after birth, the relative ratio of the splice variants in mouse kidney is A/B/C/D, 1:2.2:2.5:4:0. These relative ratios are generally similar to those found in human tissues. In the mouse, adult gonadal tissue and uterus have also been shown to contain high levels of WT1 mRNA (14). By RNase protection analysis, adult mouse testis, ovary, and uterus showed similar relative ratios of the four splice variants to mouse kidney. WT1 expression levels were too low in adult mouse spleen, heart, and lung for reliable quantitation of the splice forms.

**DISCUSSION**

The developmental pattern of expression of WT1 in fetal kidney and the presence of inactivating mutations in a number of Wilms tumors suggest that this gene functions as a tumor-suppressor gene. However, functional studies aimed at testing this concept depend upon the characterization of a complex pattern of transcripts resulting from alternative splicing. The presence of two alternative splices results in four distinct transcripts that can be identified by RNase protection assays. Our analysis demonstrates a remarkable consistency in the relative abundance of the WT1 splice variants in Wilms tumors versus normal fetal kidney, in mouse kidney as a function of development, and in different mouse tissues. This pattern, together with the high sequence conservation across species (5), underscores the potential functional importance of these alternative splices. One may speculate on potential regulatory roles for these two splices based on their location within the WT1 transcript and their nucleotide sequence. For instance, splice I encodes a domain rich in serines and threonines, making it a potential site for protein phosphorylation. In addition, the location of this alternative splice within a region of WT1, which may be involved in transactivation, raises the possibility that it may directly alter the function of this transcription factor. Alternatively, splice I may be involved in interactions between WT1 and other cellular proteins or among the WT1 splice variants themselves.

**Fig. 3.** Schematic representation of the WT1 genomic structure. The 10 exons of the WT1 gene are shown as boxes, and the alternative splices are darkly shaded. Exact sizes of the introns outside of the zinc finger region have not been determined. The entire gene spans 50 kb of genomic DNA, and the EcoRI restriction fragments containing each exon are outlined.

**Fig. 4.** RNase protection analysis of the WT1 splice variants. (a) Expected fragments. The four possible RNA species are shown, reflecting presence or absence of each of the two alternative splices. Probes used in the RNase protection experiments were derived from splice variant A (no splice present). Specific protected fragments corresponding to each of the four splice variants are shown in parentheses. (b) Human kidney and Wilms tumors. RNase protection analysis was done as detailed by using a RNA probe complementary to human splice variant A. Probe (P) is shown along with a tRNA control (t) and the following RNA samples: 20-week human kidney (FK), adult human kidney (AK), and five sporadic Wilms tumors (CJ, VL, BS, MH, and VG). Molecular size markers are shown at left, and the expected specific fragments are shown at right: A, 950 bp; B, 750 bp; C, 620 bp; and D, 420 bp. In addition, smaller fragments common to two alternative splice variants are seen at 330 bp (C and D) and 200 bp (B and D) (data not shown). (c) Mouse kidney development and other tissues. RNase protection was done by using an RNA probe complementary to mouse splice variant A. RNA samples are as follows: mouse kidney from day 1 to day 37 after birth, adult ovary (O), testis (T), and uterus (U). The presence of grossly undetectable implants in the mouse uterus cannot be excluded. Expected fragments specific for each splice variant are as follows: A, 1027 bp; B, 762 bp; C, 690 bp; and D, 425 bp. Fragments common to two splice variants are seen at 337 bp (C and D) and 265 bp (B and D) (data not shown).
Table 1. Relative ratio of WT1 splice variants in human kidney and Wilms tumor

<table>
<thead>
<tr>
<th>Splice variant</th>
<th>Fetal kidney FCPM</th>
<th>Wilms tumor FCPM</th>
<th>Ratio, cpm/cpm A</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>CJ</td>
<td>VL</td>
<td>BS</td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>B</td>
<td>2.5</td>
<td>1.7</td>
<td>2.7</td>
</tr>
<tr>
<td>C</td>
<td>3.8</td>
<td>2.7</td>
<td>3.9</td>
</tr>
<tr>
<td>D</td>
<td>8.3</td>
<td>5.1</td>
<td>6.4</td>
</tr>
</tbody>
</table>

The radioactivity emitted by each expected fragment shown in Fig. 4A was quantitated by using a Phosphorimager system. Number of cpm counted was then corrected for gel background, divided by size of the protected fragment, and then expressed as a ratio, relative to splice variant A.

Alternative splice II also encodes a serine and a threonine, potential sites for phosphorylation, but it is especially remarkable in that it interrupts the highly conserved spacing between zinc fingers 3 and 4. The existence of an alternative splice in the “knuckle” between two zinc fingers has not been previously reported, although alternative splicing has also been shown in the protooncogene evi-1, resulting in deletion of 2 of the 10 zinc finger domains (17, 18). As expected, disruption of the spacing between 2 WT1 zinc fingers has a dramatic effect on DNA-binding affinity. A partial WT1 polypeptide containing the zinc finger domains without splice II has been shown to bind the early growth response 1 recognition site, whereas insertion of splice II into this construct greatly reduces binding (12). This result suggests that the presence of splice II either confers an alternative DNA recognition site or prevents DNA binding altogether. The possibility that insertion of splice II produces a nonfunctional protein, incapable of DNA binding but interacting with the other splice forms of WT1, is particularly interesting in view of the recent evidence that inactive heterodimers may play a role in the physiologic regulation of some transcription factors (19).

An initial understanding of the functional domains of the WT1 protein may be derived from mutations within the WT1 transcript that are present in Wilms tumors. For instance, we have recently reported a mutated WT1 transcript with an in-frame deletion of zinc finger 3, encoding an altered but potentially functional protein (3). The presence of both mutated and wild-type transcripts in this tumor has suggested the possibility of “transdominant suppression,” leading to the functional inactivation of WT1. The characterization of additional WT1 mutations, particularly within the amino terminus of the protein, may provide further insight into other domains crucial for gene function. Our definition of the WT1 genomic structure and exon/intron junction sequences should greatly facilitate the identification of such mutations. Because gross rearrangements of WT1 detectable by Southern blotting appear rare, mutational analysis has been performed primarily by RNA PCR, an analysis restricted to tumors from which intact RNA is available. The sequencing of the introns flanking each NAG dinucleotide makes it possible to study the much larger number of tumors from which only DNA is available by using PCR-based techniques. The technique of PCR amplification and direct sequencing of each exon using oligonucleotide primers complementary to flanking intron sequences has been demonstrated in the mutational analysis of the RBL transcript (20). Indirect approaches aimed at detecting single-base pair mismatches or shifted electrophoretic mobility of altered PCR products (21) can also be used to screen Wilms tumor samples for WT1 mutations. Such PCR-based techniques can also be used to amplify DNA from formaldehyde-fixed tissue, thus extending mutational analysis to the retrospective study of specimens from tumor tissue banks. These techniques will be important to determine the fraction of Wilms tumors that have mutations in WT1 and to examine other malignancies in which WT1 may contribute to neoplasia. Elucidating the specific pathways through which WT1 exerts its effect in the cell, differentiation is likely to require the use of both biochemical and genetic analyses. The identification of naturally occurring mutations should complement functional studies of each alternative WT1 polypeptide to provide an understanding of the normal role of this tumor-suppressor gene.

This work was supported by National Institutes of Health Grant 5RO0259 (to D.E.H.). D.A.H. is a James McDonnell Scholar. A.J.B. was supported by National Institutes of Health Fellowship CA 08605. J.P. was supported by a fellowship from the Medical Research Council of Canada.