The Human T-Cell Transcription Factor-4 Gene: Structure, Extensive Characterization of Alternative Splicings, and Mutational Analysis in Colorectal Cancer Cell Lines

Alex Duval, Sandra Rolland, Emmanuel Tubacher, Hung Bui, Gilles Thomas, and Richard Hamelin

Institut National de la Sante ´ et de la Recherche M´edicale U454, Centre d’Etude du Polymorphisme Humain, 75010 Paris, France

ABSTRACT

The human T cell transcription factor-4 (hTF-4) interacts functionally with β-catenin in the Wnt signaling pathway, which regulates many developmental processes. Moreover, inappropriate reactivation of this pathway attributable to APC or β-catenin mutations has been described in colorectal cancers. Because only the human TCF-4 cDNA sequence was known, we determined its genomic structure. A total of 17 exons, of which 5 were alternative, were identified. Moreover, four alternative splice sites were observed either experimentally or in silico by a BLAST approach in expressed sequence tag databases. The alternative use of three consecutive exons localized in the 3′ part of the hTF-4 gene changes the reading frames used in the last exon, leading to the synthesis of a number of expressed sequence tag databases. The alternative use of three consecutive exon-intron boundaries in this gene. Because alternative splicing properties, we characterized the different hTF-4 mRNA species and lead to transactivation of several target genes. This occurs during embryonic development by binding of the Wnt protein to its receptor on the cell surface, as well as during the progression of some cancers. On the other hand, recent evidence suggests that in the absence of β-catenin, TCF/LEF factors could also act as transcriptional repressors of Wnt-responsive genes (10). Other proteins were shown to act as corepressors in this process (11, 12), particularly CtBP that binds to the COOH-terminal end of the xTCF-3 protein by two binding domains well conserved throughout evolution and present in hTF-4 (13).

Involvement of the Wnt signaling pathway has been demonstrated notably in colorectal cancers. TCF-4 (also known as TCF7L2) is the most highly expressed member of the TCF/LEF family in colonic epithelium (9). Inactivating or activating mutations of the APC or β-catenin (CTNNB1) genes, respectively, have been shown to result in aberrant accumulation of transactivating β-catenin/TCF-4 complexes (14), causing up-regulation of the expression of a number of genes including c-MYC (15). Moreover, we have recently reported frequent truncating mutations within a coding repeat sequence of hTCF-4 in a subset of colorectal cancers characterized by microsatellite instability (MSI-H tumors; Ref. 16).

TFC-4 belongs to a family of transcription factors with homology to a region of the HMG I proteins, termed the HMG box (17, 18). To modulate transcription of target genes, TCF/LEF proteins are thought to be factors that facilitate assembly of multiprotein enhancer complexes. The HMG box of these different transcription factors is the DNA binding domain, which binds to the A/T A/T CAAAG consensus (19). The β-catenin binding domain is localized upstream, at the beginning of the coding sequence (8, 20). TCF/LEF mRNAs are subject to alternative splicings at their 3′ ends, which has been proposed to be important in regulating transactivational properties of the corresponding protein isoforms (21, 22). Except for the closely related TCF-1 gene, only the mRNA sequences for the other TCF/LEF factors are available.

To better understand its multiple roles in development and in human carcinogenesis, as well as to allow more accurate screening for mutations, we analyzed the genomic structure of hTCF-4. Sequencing of two BACs and RT-PCR experiments has allowed us to detect 17 exon-intron boundaries in this gene. Because alternative splicing mechanisms could be essential in regulating its transactivating properties, we characterized the different hTCF-4 mRNA species and deduced alternative exons by RT-PCR experiments. Moreover, the complete hTCF-4 coding sequence was screened for mutations in a series of 24 colorectal cancer cell lines.

INTRODUCTION

Recent molecular and cell biology research in the field of development and cancer has shown that the Wnt signaling pathway, also called APC/β-catenin/TCF pathway, is one of the key developmental and growth regulatory mechanisms of the cell (1–3). TCF-4 knock-out mice die shortly after birth and show absence of a proliferative compartment in the prospective crypt regions between the villi (4). The level of cytosolic β-catenin is normally down-regulated by the APC tumor suppressor gene product, which promotes its degradation (5–7). In contrast, activation of the Wnt pathway causes accumulation of free β-catenin in the cytoplasm, which can then bind TCF/LEF factors (8, 9), and lead to transactivation of several target genes. This occurs during embryonic development by binding of the Wnt protein...
DNA sequences were obtained using the Phred/Phrap/Consed system (23–25), and PRISM 377 DNA sequencer (Applied Biosystems). Contigs of shotgun sequences were interrupted, probably because of the presence of repetitive elements in genomic DNA. One internal region of the hTCF-4 cDNA was missing from these contigs and was shown to contain a single exon by “primer walking” on both strands using direct screening of the BAC clone I. hTCF-4 cDNA sequences localized to the 5' part of the gene were also missing in BAC clone I. Another BAC clone containing these sequences (hTCF-4 BAC clone II) was isolated from the CEPH human BAC library using additional probes, and intron-exon junctions of four additional hTCF-4 exons were determined by primer walking.

A total of 14 exons were thus identified, each with canonical acceptor and donor sites, as shown in Fig. 1. These exon sequences matched perfectly to the previously published hTCF-4E mRNA sequence (9). Additional intron sequences are deposited in the EMBL database (accession nos. AJ270770 to AJ270778).

Characterization of Alternative Exons and Splice Sites. Using different primer sets chosen in hTCF-4 exons as shown in Fig. 2, we amplified a number of mRNAs by RT-PCR in a series of colorectal
cancer cell lines and analyzed the different products on agarose gels. In most cases, a unique PCR product with the predicted size was observed. In cases indicated by framing on Fig. 2, longer fragments were also reproducibly obtained in addition to the predicted DNA fragments. These longer DNA fragments were excised from agarose gels and sequenced. One of these sequences was localized in the 5' part of the gene and was extended by primer walking on the BAC clone II. Another was found to be present in one of the genomic DNA contigs. In both cases, these sequences were flanked by consensus splicing sequences, and hence we deduced that they corresponded to alternative exons 4 and 15 expressed in all of the cell lines tested. A third additional cDNA sequence was present in one of the genomic contigs just upstream of exon 9, determined by the use of an alternative splice acceptor site in this exon.

Finally, an additional cDNA fragment was shown to be present in the COLO 320 cell line and absent in all other cell lines, except when more sensitive techniques were used (data not shown). The corresponding sequence was present in one of the contigs, flanked by canonical splice sites, and named exon 16.

Fig. 3 is a summary of the hTCF-4 genomic structure, which is thus composed of 17 exons (with exon 16 mainly expressed in the COLO 320 cell line). In previously published work (16, 30), when the genomic structure of the hTCF-4 gene was only partially known, we numbered some selected putative exons as deduced from the nomenclature of the closely related hTCF-1 gene. We now indicate in Fig. 3 the relation between this previous numbering and the proposed new one.

Characterization of hTCF-4 3' Alternative Splicing Products. Because the profiles of RT-PCR products between exons 12 and 17 showed a complicated pattern on agarose gel, we decided to analyze
A total of 12 DNA variants was observed. As shown previously (16), four cell lines with microsatellite instability (LS 174T, LoVo, TC71, and TC7) had a 1-bp deletion in an (A)9 repeat localized at the beginning of exon 17 (exon 10 according to the previous TCF-I-like numbering). In the cell line LS 1034, we observed a 2-bp deletion in the GAGA nucleotides at coding position 112–115 within exon 1. The FET cell line contained a C to A transversion at nucleotide 62 of exon 15, leading to a stop codon in one of the used reading frames. We were unable to find a mutation in the flanking intronic sequences of exon 16 of COLO 320, which could explain the abnormally high level of this alternative splicing in this cell line. A C to T transition at the nucleotide 35 of exon 4 was present in the SW 48 cell line, leading to an alanine to valine change. Finally, four noncoding alterations were observed: a T to G transversion and a deletion of one T in a 7T-run in the polypyrimidine tracts of the splice acceptor sites of exon 13 in HCT-15 and exon 14 in Co-115 cell lines, respectively, and a 1-bp deletion in the (A)9 repeat preceding (and including) the initiation ATG codon in two MSI-H cell lines (TC71 and HCT-116). In all cases but two (LoVo and TC71 in the exon 17 coding repeat), only one allele was altered.

**DISCUSSION**

A link was recently established between TCFs and Wnt signaling, a pathway that plays a crucial role in many developmental processes as well as in human carcinogenesis (1, 2). In the present work, we describe the genomic structure of an important member of the TCF/LEF gene family, the human TCF-4 gene that we have recently mapped on chromosome 10q25.3 by FISH analysis (31). This study also highlights the complexity of hTCF-4 mRNA transcripts. Indeed, we report here a large number of alternative splicings in the 3’ part of this gene, which generate several TCF-4 isoforms differing at their COOH-terminal ends. This mechanism is expected to be of functional significance because it has been shown that TCF factors were able to interact functionally by COOH-terminal binding domains with CtBP, a corepressor protein required to mediate transcriptional repression of the TCF family activity. Finally, because the APC/β-catenin/TCF
pathway is known to play an important role in colorectal carcinogenesis, we screened the entire hTCF-4 gene for mutations in a series of 24 colorectal cancer cell lines and found DNA coding alterations in 33% of the cases ($n = 8$).

We have obtained experimental evidence for a total of 17 exons within hTCF-4 of human cell lines. Two different human hTCF-4 mRNA transcripts, named hTCF-4b and hTCF-4e, have already been described (9). According to our proposed exon nomenclature, they contain exons 1, 2, 3, 5–12, 17 and 1, 2, 3, 5–14, 17, respectively. We therefore describe three additional exons (4, 15, and 16), although a polypeptide potentially encoded by exon 15 has been reported previously with no further details (32). It has to be noted that exons 14 and 15 have the same size (73 bp) and show 66% homology at both the nucleotide and amino acid levels, suggesting duplication during evolution.

Exons 4, 13, 14, 15, and 16 are alternative exons, whereas three alternative splice acceptor sites are used in exons 9, 16, and 17. This creates 256 theoretically different splice products, making hTCF-4 expression more complicated than hTCF-1 with its 96 potential different mRNA species (20). This number may even be increased to >500 different forms if the in silico-detected alternative splice site in exon 7 doubles the number of different mRNA transcripts. The use of exon 4 and the alternative exon 9 acceptor site appear as frequent as their skipping; however, we did not examine whether they were exclusive of each other and whether they were conserved with alternative splicings downstream of exon 12. Moreover, their presence or absence does not change the reading frame.

Alternative mRNAs between exons 12 and 17 were more extensively analyzed by RT-PCR experiments. Eight mRNAs were easily detected and sequenced. They contain: exons 12, 17 (hTCF-4b); 12, 13, 17; 12, 14, 17; 12, 15, 17; 12, 13, 14, 17 (hTCF-4e), 12, 13, 15, 17; 12, 14, 15, 17 and 12, 13, 14, 15, 17 (Fig. 4 and 6). As a consequence of this baroque splicing mechanism, exon 17 is either noncoding ($n = 2$; because of the presence of an upstream in-frame stop codon when exons 14 and 15 are spliced together) or encodes for medium (25 amino acids; $n = 2$) or long (138 amino acids; $n = 4$) hTCF-4 COOH-terminal isoforms, depending of which reading frame is used.

TCF-4 and other members of the TCF/LEF family have conserved sequences corresponding to functional domains (2). The $\beta$-catenin binding domain (in exon 1, according to the present exon numbering) and DNA-binding HMG box (exons 10 and 11) of hTCF-4 are highly conserved with other TCF/LEF family members in human as well as other species including Drosophila (Pangolin), Xenopus (XTCf-3), chicken (chTCF), and mouse (mTCFs). Outside of these domains, a high conservation is observed for the final exon 17. For this exon, there is: (a) a complicated splicing pattern that has been conserved in murine TCF-4 as well as in the related human TCF-1 gene; (b) two reading frames, long or medium in size, that are also well conserved in Pangolin (Drosophila) and POP-1 (C. elegans) or in chicken TCF, respectively (2); (c) two CtBP binding motifs only...
present in the long-size COOH-terminal hTCF-4 isoforms that were conserved throughout evolution in the TCF family and may play a role in down-regulating TCF transactivating properties. Together, these observations suggest that alternative use of different reading frames in the 3' part of hTCF-4 should generate different protein isoforms with agonist or antagonist transactivating activities.

We detected 12 hTCF-4 DNA variants in a series of 24 colorectal cancer cell lines. Four are noncoding with unknown consequences, if any, on the TCF-4 protein function. The other eight variants were in coding sequences. A missense alanine to valine alteration in the alternative exon 4 was observed in the SW48 cell line. It is of unclear significance and could be a coding polymorphism. Another variant is a frameshifting, 2-bp deletion within the β-catenin binding domain in exon 1 in the LS 1034 cell line. It is known that the closely related hTCF-1 gene has two different initiation sites, leading to the synthesis of long and short NH2-terminal TCF-1 isoforms, and it has been reported that the short NH2-terminal isoforms act as negative regulators of transcription (20). In the hTCF-4 gene, no initiation at internal ATG has been demonstrated, although a DN-TCF4 truncated construct has been shown to have a dominant-negative effect over the full-length hTCF-4 product bound to β-catenin (9). We noticed that another in-frame ATG fulfilling the minimum Kosak criteria of an A at −3 is present in exon 3 of the hTCF-4 gene. Preliminary observations indicate that the hTCF-4 gene could also have an alternative transcriptional initiation site. Indeed, the sequence corresponding to intron 2 of the already known hTCF-4 mRNAs is 208 bp long and has a GC content of 82% with a CpG/GpC = 1 and a putative TATA box (see AJ270771). Moreover, we were able to amplify and sequence a hTCF-4 cDNA product containing exons 3 to 6 with a sense primer localized between the TATA box and the beginning of exon 3 in the 5' untranslated region of the putative alternative transcript (data not shown). In that context, the LS1034 frameshift mutation should be interpreted as a selected event that favors the expression of hTCF-4 isoforms lacking a β-catenin binding domain with an unexpected dominant-negative effect.

The other hTCF-4 mutations are a 1-bp deletion in an (A)9 repeat within exon 17 in four cell lines, which have been already reported in a large series of primary MSI-H colorectal cancers (16, 30), a nonsense mutation in exon 15 in FET, and a high usage of alternative exon 16 in COLO 320. They are all localized in the 3' part of the hTCF-4 gene and are truncating mutations. The common result of all these mutations is to change the reading frames used in exon 17 in the

---

Fig. 5. ESTern e-blot analysis of the hTCF-4 gene. The result of a BLAST search on EST databases is represented, and accession numbers of the ESTs giving a score >200 are shown. The query was the longer hTCF-4 mRNA sequence, including all alternative exons and splice sites as deduced from our experiments. An interruption of the line representing an EST indicates a sequence present in the query sequence and absent in the EST. Sequence alignments were also checked for sequences present internally in ESTs and absent in the query sequence; however, none were observed.

Fig. 6. Reading frames used in alternative exons between exons 12 and 17. The reading frames used in each exon are indicated in gray. Exons 13, 14, 15, and 16 have 51, 73, 73, and 25 bp, respectively. The alternative splicing events of exons 14 and 15 change the reading frames used in exon 17 as indicated. For the different COOH-terminal ends, encoded by the corresponding alternative mRNA species: S, short; M, medium; L, long.
corresponding cell lines. As a consequence, hTCF-4 isoforms with either a noncoding, a short, or a medium COOH terminus are enriched, whereas the long hTCF-4 isoforms either disappear or are reduced in their expression level (Fig. 7). If the relative proportion of these different COOH-terminal hTCF-4 isoforms plays a role in regulating hTCF-4 activity, such changes may alter the resulting hTCF-4 activity in cell lines harboring these mutations. Notably, all of these alterations favor the synthesis of hTCF-4 isoforms lacking their capacity to interact with CtBP, a protein implicated in the transcriptional repression mediated by the TCF COOH-terminal domain. It should be noted that all cell lines containing hTCF-4 mutations in coding sequences were also mutated in either the APC or the b-catenin genes. Thus, hTCF-4 alterations could constitute a second mutational event frequently selected in these cell lines to reactivate Wnt signaling. Further analysis will be necessary to confirm this hypothesis and to functionally characterize the consequences of each of these alterations on TCF/b-catenin target gene expression.

ACKNOWLEDGMENTS

We thank Drs. Barry Iacopetta and Elizabeth Newcomb for critical reading of the manuscript.

REFERENCES


Note Added in Proof

During the review process of this manuscript, an additional chromosome 10q genomic sequence was reported by the Sanger Center, Hinxton, under GenBank Accession number AL135792. This sequence contains exons 6–17 of hTCF-4 and shows that intron 5 is at least 60 kb long.

Fig. 7. Consequences of hTCF-4 mutations on the eight alternative TCF-4 isoforms encoded by the different mRNA species. The main functional domains of hTCF-4 are indicated. COOH-terminal ends of the eight isoforms encoded by the eight differently spliced mRNA species are framed. The S (short), M (medium), and L (long) COOH-terminal ends of each isoform are indicated. The two CtBP binding sites present in long isoforms are represented by gray circles. The total content of short, medium, and long isoforms in nonmutated and mutated cell lines are indicated in boxes with the same lettering system. The mRNA encoding the L* form in the COLO 320 cell line has not been experimentally detected.


The Human T-Cell Transcription Factor-4 Gene: Structure, Extensive Characterization of Alternative Splicings, and Mutational Analysis in Colorectal Cancer Cell Lines

Alex Duval, Sandra Rolland, Emmanuel Tubacher, et al.

Cancer Res 2000;60:3872-3879.

Updated version
Access the most recent version of this article at:
http://cancerres.aacrjournals.org/content/60/14/3872

Cited articles
This article cites 31 articles, 15 of which you can access for free at:
http://cancerres.aacrjournals.org/content/60/14/3872.full.html#ref-list-1

Citing articles
This article has been cited by 34 HighWire-hosted articles. Access the articles at:
/content/60/14/3872.full.html#related-urls

E-mail alerts
Sign up to receive free email-alerts related to this article or journal.

Reprints and Subscriptions
To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org.

Permissions
To request permission to re-use all or part of this article, contact the AACR Publications Department at permissions@aacr.org.