Thirteen-exon-motif signature for vertebrate nuclear and mitochondrial type IB topoisomerases

Hongliang Zhang, Ling-Hua Meng, Drazen B. Zimonjic¹, Nicholas C. Popescu¹ and Yves Pommier*

Laboratory of Molecular Pharmacology and ¹Laboratory of Experimental Carcinogenesis, Center for Cancer Research, National Cancer Institute, NIH, DHHS, Bethesda, MD, USA

Received February 6, 2004; Revised and Accepted March 11, 2004

ABSTRACT

DNA topoisomerases contribute to various cellular activities that involve DNA. We previously identified a human nuclear gene that encodes a mitochondrial DNA topoisomerase. Here we show that genes for mitochondrial DNA topoisomerases (type IB) exist only in vertebrates. A 13-exon topoisomerase motif was identified as a characteristic of genes for both nuclear and mitochondrial type IB topoisomerases. The presence of this signature motif is thus an indicator of the coexistence of nuclear and mitochondrial type IB DNA topoisomerases. We hypothesize that the prototype topoisomerase IB with the 13-exon structure formed first, and then duplicated. One topoisomerase specialized for nuclear DNA and the other for mitochondrial DNA.

INTRODUCTION

DNA topoisomerases play important roles in cellular activities that involve DNA. We previously identified a human nuclear gene that encodes a mitochondrial DNA topoisomerase. Here we show that genes for mitochondrial DNA topoisomerases (type IB) exist only in vertebrates. A 13-exon topoisomerase motif was identified as a characteristic of genes for both nuclear and mitochondrial type IB topoisomerases. The presence of this signature motif is thus an indicator of the coexistence of nuclear and mitochondrial type IB DNA topoisomerases. We hypothesize that the prototype topoisomerase IB with the 13-exon structure formed first, and then duplicated. One topoisomerase specialized for nuclear DNA and the other for mitochondrial DNA.

MATERIALS AND METHODS

Cloning of Top1mt

DNA manipulation, PCR and DNA sequencing were performed according to standard protocols. We obtained clone BF139529 from Incyte Genonics (St Louis, MO), IMAGE clone 2601221 from ATCC (Manassas, VA) and clone pgf2n.pk002.c13 from Delaware Biotechnology Institute (Newark, DE), and sequenced them on a 377 DNA sequencer using ABI Prism Big Dye Terminator (PE Applied Biosystems). The missing 5¢ end portions of Top1mt genes were amplified using a GeneRacer kit (Invitrogen, Carlsbad, CA). The 5¢ end was joined to the corresponding clones to generate full-length Top1mt. All oligonucleotide sequences used for cDNA identification are available upon request.

Fluorescence microscopy, FISH localization, DNA relaxation assays and DNA cleavage assays

These procedures were carried out as described previously (6).

Database searches and alignment

We identified putative homologous genes using the contiguous Mega BLAST (http://www.ncbi.nlm.nih.gov) to search all available NCBI databases. We aligned DNA sequences and corresponding amino acid sequences with available Top1 and Top1mt genes using the ClustalW in MacVector (Accelrys, San Diego, CA).

To whom correspondence should be addressed. Tel: +1 301 496 5944; Fax: +1 301 402 0752; Email: pommier@nih.gov

*AF362952, AF503620, AY429654, AY429655 and TPA BK001786

Nucleic Acids Research, Vol. 32 No. 7 © Oxford University Press 2004; all rights reserved

Published online April 19, 2004

DOI: 10.1093/nar/gkh525
RESULTS

Identification of a mouse mitochondrial topoisomerase I gene (Mm-TOP1mt)

Screening of the NCBI database with the BLAST search engine and human mitochondrial topoisomerase I (Hs-top1mt) as the bait yielded a mouse cDNA sequence (DDBJ/EMBL/GenBank accession no. BF139529). Sequencing of BF139529 revealed an open reading frame encoding a polypeptide with high homology to Hs-top1mt. The GeneRacer protocol (Invitrogen, Carlsbad, CA) was used to determine the sequence of the 5′ end of the gene. The combination of both approaches yielded a 2011 bp cDNA sequence that encodes a 593 amino acid protein. This protein, which we have designated Mm-top1mt, shares 73% sequence identity and 84% similarity with Hs-top1mt (6) (Fig. 1).

We next designed two sets of PCR primers based on the 5′ and 3′ ends of the Mm-TOP1mt cDNA for the purpose of screening a mouse genomic library. A bacterial artificial chromosome clone containing the full-length Mm-TOP1mt gene was obtained. We then used this clone as a probe to determine the chromosomal location of Mm-TOP1mt by fluorescence in situ hybridization. In two independent experiments with biotin- or digoxigenin-labeled probes, most metaphase spreads with informative signals and minimal nonspecific background fluorescence yielded symmetrical fluorescent spots on a small chromosome. Furthermore, 27 out of a total of 30 labeled spreads recorded in the two experiments exhibited a specific signal at the same site, bands E2–E3, on both chromosomes 15, to which we therefore assign Mm-TOP1mt (Fig. 2). This region of mouse chromosome 15 is homologous to human chromosome 8q24.3 (8), the site of Hs-top1mt (6).

Both Hs-TOP1mt and Mm-TOP1mt are positioned between locus H of the lymphocyte antigen 6 complex and the rhophilin (Rho GTPase binding protein 1) gene. The region of the mouse genome containing Mm-TOP1mt is thus syntenic to that of the human genome containing Hs-TOP1mt, suggesting that these regions share a common ancestor.

To determine the structure of Mm-TOP1mt, we sequenced the 5′ end of the gene and combined the resulting sequence with that available in the NCBI database. Like Hs-TOP1mt, Mm-TOP1mt contains 14 exons. This 14-exon structure is also shared by other TOP1mt genes (Table 1; see below). All TOP1mt genes also exhibit the same intron phases (Table 1). Furthermore, the corresponding introns of the human and mouse TOP1mt genes are similar in size, with the exception of intron 7 which is larger in human (Homo sapiens) than in mouse (Mus musculus) (Table 1).

**TOP1mt genes are present only in vertebrates**

We examined the available eukaryotic DNA sequences to determine which species possess genes for both mitochondrial and nuclear topoisomerases I. With Hs-top1mt and Mm-top1mt as baits, we detected TOP1mt genes in all the vertebrate genomes: zebra fish (Danio rerio) (Dr), chicken (Gallus gallus) (Gg) and rat (Rattus norvegicus) (Rn).

For chicken top1mt, we derived most of the sequence from a cDNA clone (clone ID, pgf2n.pk002.e13) and used GeneRacer to obtain the remaining 5′ sequence. For zebra fish, the cDNA sequence was directly derived from a single clone (IMAGE clone ID, 2601221). Expression experiments revealed that both the recombinant chicken (Gg-top1mt) and zebra fish (Dr-top1mt) proteins possess topoisomerase I activity. Cleavage assays (6) also confirmed that Gg-top1mt is a type IB topoisomerase, given that it forms a covalent bond with the 3′ end of the cleaved DNA (data not shown).

The sequences and structures of the rat, chicken and zebra fish TOP1mt genes were derived from the recently released databases (NCBI). The rat (Rn-TOP1mt) and chicken (Gg-TOP1mt) genes, like the human and mouse genes, comprise 14 exons (Table 1). For the zebra fish gene (Dr-TOP1mt), we were able to compile only 11 exons from the incomplete genomic sequence (Table 1). The exon sizes for these five vertebrate TOP1mt genes vary for the first exon but are identical for the remaining 13 exons, with the minor exception that exons 2 and 13 of the rodent genes are 3 bp shorter (corresponding to deletion of one amino acid and likely a characteristic of the common rodent ancestor).
The NH2-terminal portion of Hs-top1mt encoded by exon 1 contains the mitochondrial localization signal (6). Alignment of the corresponding NH2-terminal regions of the vertebrate top1mt polypeptides revealed that they share little sequence homology (Fig. 1A). To verify that the newly identified top1mt proteins are indeed mitochondrial enzymes, we transfected M059J human neuroblastoma cells with green fluorescent protein (GFP) and then examined the transfected cells by fluorescence microscopy, as previously described for Hs-top1mt (6). Both of the GFP fusion proteins localized to mitochondria (data not shown), demonstrating the presence of a functional mitochondrial targeting sequence in both mouse and chicken top1mt.

When the sequences encoded by the first exons were removed, the probability of mitochondrial targeting was high for all five top1mt genes: 92, 98, 99, 99 and 98% for zebra fish, chicken, mouse, rat and human top1mt, respectively. The catalytic residues, including the critical basic amino acids (RKR, marked with an asterisk), and tyrosine residue (Y, marked with a plus sign) and proline residue of the corresponding NH2-terminal regions of the vertebrate TOP1mt genes exhibited a higher level of identity (83.43 ± 6.87%) than did the TOP1mt genes (73.98 ± 7.91%); the level of identity between the TOP1 and TOP1mt genes was lower (67.72 ± 2.08%).

Both the 13-exon motif and the presence of two type IB topoisomerase (mitochondrial and nuclear) genes are restricted to vertebrates

We next investigated the existence of genes for type IB topoisomerases in nonvertebrate eukaryotes. The 13-exon topoisomerase motif was not detected in budding yeast (Saccharomyces cerevisiae), fission yeast (Schizosaccharomyces pombe), fruit fly (Drosophila melanogaster), nematode (Caenorhabditis elegans), rice (Oryza sativa) or thale cress (Arabidopsis thaliana). In budding yeast, the TOP1 gene contains no introns. The fission yeast TOP1 gene contains two introns at its 5' end. The fruit fly TOP1 gene consists of eight exons, and the nematode TOP1 gene comprises five exons. Both the rice and the two thale cress TOP1 genes share a common 13-exon structure unrelated to the vertebrate 13-exon topoisomerase motif (not shown). The existence of a distinct but shared (in length and phase) 15-exon structure in these plant species indicates that they are derived from a common ancestor.

The sea squirt (Ciona intestinalis) (Ci) has a single TOP1 gene that is markedly similar to those of vertebrates (Table 2).
We determined the structure of exons 3–21 of Ci-\textit{TOP1}, assuming that the gene consists of 21 exons (Table 2). For exons 3–8, the homology with vertebrate \textit{TOP1} genes is low. In contrast, the homology (in terms of size and phase) is high for exons 10–18 and for exon 21 of Ci-\textit{TOP1} and vertebrate \textit{TOP1} genes. Moreover, the cumulative length of exons 19 and

For exons 1 and 21, the first numbers correspond to the coding region and those in parentheses include the noncoding region. Bold indicates differences in exon size. The dashed line separates the conserved last 13 exons. The intron sizes of human and mouse were derived from the NCBI and our own sequence data. Vert., vertebrates; Hs, \textit{H.sapiens}; Rn, \textit{R.norvegicus}; Mm, \textit{M.musculus}; Gg, \textit{G.gallus}; Dr, \textit{D.rerio}; Ci, \textit{C.intestinalis}. Exon and intron sizes are in base pair units.

\textit{a}Exon size deduced from neighbors.

\textit{b}Exon sizes and intron phases for Ci could not be determined from the available sequence data.

Table 3. Comparison of the nucleotide and predicted amino acid sequences of the last 13 exons of the genes for vertebrate mitochondrial (mt) and nuclear (n) type IB topoisomerases

\begin{tabular}{|c|cccc|cccc|}
\hline
\textbf{Exon no.} & \textbf{Exon size (vertebrates)} & \textbf{Hs} & \textbf{Rn} & \textbf{Mm} & \textbf{Gg} & \textbf{Dr} & \textbf{Ci} & \textbf{Intron phase} & \textbf{Intron size} & \textbf{Hs} & \textbf{Mm} \\
\hline
1 & 33 (280) & 33 & 33 (246) & 33 & 33 (143) & \_\_ & \_\_ & 0 & 330 & 345 \\
3 & 97 & 103 & 103 & 97 & 104 & 0 & 2 & 14 680 & 12 448 \\
4 & 124 & 124 & 124 & 127 & \_\_ & \_\_ & 131 & 0 & 2 & 1 287 & 1 673 \\
5 & 56 & 56 & 56 & 59 & 47 & \_\_ & \_\_ & 40 & 0 & 2 & 2 447 & 3 144 \\
6 & 96 & 96 & 96 & 96 & 90 & \_\_ & 44 & 2 & 2 & 9 84 & 2 533 \\
7 & 76 & 76 & 76 & 76 & \_\_ & \_\_ & 82 & 0 & 0 & 3 221 & 3 343 \\
8 & 107 & 107 & 107 & \_\_ & \_\_ & \_\_ & 98 & \_\_ & 45 & 0 & 2 & 7 903 & 4 321 \\
10 & 122 & 122 & 122 & 122 & 122 & \_\_ & \_\_ & 122 & 0 & 0 & 8 73 & 74 & 8 \\
11 & 123 & 123 & 123 & 123 & 123 & \_\_ & \_\_ & 123 & 0 & 0 & 1 718 & 1 088 \\
12 & 188 & 188 & 188 & 188 & 188 & \_\_ & \_\_ & 188 & 2 & 2 & 9 65 & 2 06 \\
13 & 145 & 145 & 145 & 145 & 145 & \_\_ & \_\_ & 145 & 0 & 0 & 1 1 428 & 7 323 \\
14 & 144 & 144 & 144 & 144 & 144 & \_\_ & \_\_ & 144 & 0 & 0 & 1 044 & 1 292 \\
15 & 186 & 186 & 186 & 186 & 186 & \_\_ & \_\_ & 186 & 0 & 0 & 1 215 & 7 79 \\
16 & 69 & 69 & 69 & 69 & 69 & \_\_ & \_\_ & 69 & 0 & 0 & 8 38 & 5 98 \\
17 & 115 & 115 & 115 & 115 & 115 & \_\_ & \_\_ & 115 & 1 & 1 & 1 776 & 1 559 \\
18 & 128 & 128 & 128 & 128 & 128 & \_\_ & \_\_ & 128 & 0 & 0 & 3 399 & 3 064 \\
19 & 95 & 95 & 95 & 95 & \_\_ & \_\_ & \_\_ & \_\_ & \_\_ & \_\_ & 174 & 0 & 0 & 2 15 & 2 53 \\
20 & 150 & 150 & 150 & 150 & 150 & \_\_ & \_\_ & 71 & 2 & 2 & 1 039 & 3 75 \\
21 & 103 (1298) & 103 & 103 (1310) & 103 (1411) & 103 (130) & 103 & \_\_ & \_\_ & \_\_ & \_\_ & 103 & \\
\hline
\end{tabular}
DISCUSSION

From zebra fish to human, all the vertebrate type IB topoisomerases examined possess a common terminal 13-exon motif, suggesting that this motif is characteristic of vertebrates. This 13-exon motif, encoding the topoisomerase activity, corresponds to the portion of human topoisomerase I resolved by crystallography (9). Interestingly, the exon–intron boundaries do not occur at the boundaries of the domains identified in the crystal structures of human topoisomerase I.

Given that the organisms with this 13-exon motif possess nuclear genes for both a nuclear and a mitochondrial topoisomerase IB, it is likely that both genes evolved from a common ancestor gene. During evolution, gene duplication might thus have resulted in the emergence of one gene for an enzyme targeted to nuclear DNA and of another gene for an enzyme targeted to mitochondrial DNA. The common topoisomerase I catalytic domain is encoded by the last 13 exons of each gene, and the targeting sequences are encoded by the first exon of the genes for the mitochondrial enzymes and by the first eight exons of the genes for the nuclear enzymes. The structure of the sea squirt TOP1 gene shares similarities with the vertebrate genes in its last 12 exons, suggesting that this gene might share a common ancestor with an early precursor of the vertebrate TOP1 and TOP1mt genes, but failed short of vertebrates.

The absence of a specific TOP1mt gene in the other euksyartic species raises the question of how these organisms perform mitochondrial DNA metabolism functions. It is possible that other topoisomerases (types II or IA) perform such functions in these species. However, the only specific mitochondrial topoisomerase enzymes identified to date are type IB enzymes. We cannot exclude the possibility that other types of topoisomerases contain mitochondrial targeting sequences that are short and not readily recognizable. Alternatively, a single gene might encode two polypeptides that are targeted either to the nucleus or to mitochondria. The human TOP3α gene, for example, contains two start codons that yield two distinct enzymes, one for the nucleus and the other for mitochondria (7). The sequence data from this study have been submitted to GenBank under the following accession numbers: Mm-TOP1mt cDNA, AF362952; the 5' end of Mm-TOP1mt gene, AF503620; Gg-TOP1mt cDNA, AY429654; Dr-TOP1mt cDNA, AY429655; Rn-TOP1mt cDNA, TPA BK001786.

ACKNOWLEDGEMENT

We wish to thank Dr Kurt W. Kohn for insightful discussions.

REFERENCES


