Biochemical properties of mammalian TREX1 and its association with DNA replication and inherited inflammatory disease

Tomas Lindahl*, Deborah E. Barnes*, Yun-Gui Yang† and Peter Robins*

*Clare Hall Laboratories, Cancer Research UK London Research Institute, South Mimms, Herts. EN6 3LD, U.K., †Beijing Institute of Genomics, Chinese Academy of Sciences, 7 Beilucheng West Road, Chaoyang District, Beijing 100029, China

Abstract
The major DNA-specific 3′–5′ exonuclease of mammalian cells is TREX1 (3′ repair exonuclease 1; previously called DNase III). The human enzyme is encoded by a single exon and, like many 3′ exo- and endonucleases, exists as a homodimer. TREX1 degrades ssDNA (single-stranded DNA) more efficiently than dsDNA (double-stranded DNA), and its catalytic properties are similar to those of Escherichia coli exonuclease X. However, TREX1 is only found in mammals and has an extended C-terminal domain containing a leucine-rich sequence required for its association with the endoplasmic reticulum. In normal S-phase and also in response to genotoxic stress, TREX1 at least partly redistributes to the cell nucleus. In a collaborative project, we have demonstrated TREX1 enzyme deficiency in Aicardi–Goutières syndrome. Subsequently, we have shown that AGS1 cells exhibit chronic ATM (ataxia telangiectasia mutated)-dependent checkpoint activation, and these TREX1-deficient cells accumulate ssDNA fragments of a distinct size generated during DNA replication. Other groups have shown that the syndromes of familial chilblain lupus as well as systemic lupus erythematosus, and the distinct neurovascular disorder retinal vasculopathy with cerebral leukodystrophy, can be caused by dominant mutations at different sites within the TREX1 gene.

Introduction
In an early biochemical survey of mammalian enzymes that act on strand breaks in DNA, we discovered and characterized two major exo- and endonucleases, TREX1 (3′ repair exonuclease 1) and FEN1 (flap endonuclease-1) (initially called DNase III and DNase IV respectively). Until recently, TREX1 was by far the least studied of these enzymes. However, this situation changed as a consequence of the cloning and characterization of the human TREX1 gene, which is located to chromosome 3p21 [1,2], the establishment of an unexpected autoimmune phenotype in Trex1-knockout mice [3], and the finding of inactivating mutations of TREX1 in a recessively inherited human disease related to systemic lupus erythematosus known as AGS (Aicardi–Goutières syndrome) [4].

Enzymology functions of TREX1
TREX1 functions as a 3′→5′ exonuclease specific for DNA; it acts 3–4-fold more efficiently on ssDNA (single-stranded DNA) than on dsDNA (double-stranded DNA), but is totally devoid of endonuclease activity [5]. TREX1 has sequence similarities to several other 3′ exo- and endonucleases in prokaryotes and eukaryotes and is related to the Escherichia coli DnaQ/MutD protein, which contains the editing function of the multisubunit replicative DNA polymerase III. However, TREX1 does not remove mismatched nucleotides during DNA replication, but plays a different role [3]. In biochemical assays, TREX1 is more similar to the little-studied E. coli exoX (exonuclease X) [6] than DnaQ, exoI and other E. coli exo- and endonucleases; both TREX1 and exoX display catalytic preference for ssDNA and bind unusually tightly to ssDNA. E. coli exoX has an accessory and apparently redundant role in DNA repair and homologous recombination and may serve to confer genetic stability to tandem repeat sequences [6,7]. Surprisingly, a similar 3′ exonuclease has not been detected in budding yeast (Saccharomyces cerevisiae) or fission yeast (Schizosaccharomyces pombe); TREX1 is only found in mammalian cells. These cells also contain a closely related enzyme, TREX2, which is smaller than TREX1 [2]. Other higher eukaryotes such as Xenopus laevis have only one enzyme of this type, which appears to be more similar to TREX2 than TREX1, at least with regard to the C-terminal region [8,9].

The three-dimensional structure of the catalytic domain of TREX1 was solved recently by two independent groups [10,11]. The active enzyme occurs as a homodimer. Similarly to several other 3′ exo- and endonucleases, TREX1 contains three distinct conserved sequence motifs for binding of the obligatory Mg2+ (or Mn2+) cofactor through aspartate residues. Furthermore, a polyproline II helix occurs, which, among related exonucleases, is unique to TREX1. This flexible region might be employed for interactions with other proteins. However, no distinct protein partners of TREX1 have been demonstrated.
identified with certainty to date. In this regard, it was reported that one function of TREX1 may be as a component of the large SET (suppressor of variation 3–9, enhancer of zeste, trithorax protein) complex active in caspase-independent apoptosis [12]. The structure of the deoxynucleotide-binding pocket in TREX1 explains the previously observed inability of the enzyme to excise an altered 3′-terminal DNA residue at a single-strand interruption, such as a phosphoglycolate caused by ionizing radiation. Recently, a detailed assessment of the hydrolytic mechanism of the 3′ exonuclease activity of TREX1 and the *E. coli* DnaQ protein was made by quantum mechanics/molecular mechanics calculations, considering conserved important amino acid residues, with the reasonable assumption that a water molecule bound to the catalytic bivalent metal acts as the nucleophile for hydrolysis of the phosphodiester bond [13].

The unique C-terminal hydrophobic region of TREX1

TREX1 differs from TREX2, exoX and other members of the DnaQ/MutD family of 3′ exonucleases in having an extended C-terminal region of approx. 70 amino acid residues [1,2]. The origin of this C-terminal region of the mammalian TREX1 enzyme is unclear, because no similar sequence has been detected in available protein databases. Surprisingly for an enzyme acting on DNA, TREX1 is predominantly cytosolic and localizes to the ER (endoplasmic reticulum). Only a very small proportion of the protein is present in the cell nucleus, although this proportion increases in S-phase and after apoptotic or genotoxic stress. Other groups have also shown the ER localization and its dependence on the C-terminal region of TREX1 [12,14]. In the present article, we provide a further detailed analysis of studies that systematically mutated the TREX1 C-terminal region [1,15].

A series of N-terminally-tagged GFP (green fluorescent protein) constructs encoding C-terminal deletions of TREX1 were generated to identify the residues that are required for its ER association (Figure 1). The subcellular localization of these GFP–TREX1 mutant proteins was investigated following transfection of the deletion constructs into MEFs (mouse embryonic fibroblasts) (Figure 1, upper panel). The expression of similar levels of protein of expected molecular mass was verified by Western blotting (results not shown). As observed for the full-length GFP–TREX1 (amino acids 1–314), GFP–TREX1-(1–307) lacking the C-terminal seven amino acids still localizes in a perinuclear pattern. However, deletion of a further 14 amino acids abolished this characteristic TREX1 localization and the GFP–TREX1-(1–293) protein was diffusely distributed; cellular fractionation experiments indicated that GFP–TREX1-(1–293) was still entirely in the cytoplasm. The truncated TREX1-(1–293) enzyme remained in the cytoplasm even in cells exposed to ionizing radiation, whereas the full-length enzyme translocated from the ER to the nucleus after radiation (results not shown). Thus the region between amino acids 293 and 307 plays a critical function in the ER localization of TREX1. There is very limited homology of the C-terminal region in all the available mammalian TREX1 cDNA sequences, except for the most C-terminal 30 amino acids, which are almost entirely.

![Figure 1](image-url)

Figure 1 | Site-specific mutagenesis analysis of the contribution of the TREX1 C-terminal domain to the subcellular distribution of the protein

Five key hydrophobic leucine residues occur within residues 293–307 of TREX1. Upper panel: subcellular localization of GFP–TREX1 truncations and point mutants. Lower panel: schematic diagram of TREX1 protein. Conserved exonuclease motifs are indicated (E1–E3). The putative TMD is shown as a boxed sequence, and the core hydrophobic leucine residues are highlighted in bold italics; this sequence is highly conserved in humans (h) and mice (m). The five leucine residues (Leu293, Leu294, Leu295, Leu297 and Leu299) were sequentially mutated to alanine in GFP–TREX1 [15]. All mutants were transfected into MEFs; 24 h later, TREX1 localization was observed by immunofluorescence microscopy. Note the loss of ER localization in the TREX1 TMD Leu(x5)-Ala mutant in which all five leucine residues were converted into alanine.
TREX1 deficiency and human disease

Insights into the functional role of TREX1 were initially obtained by the construction and investigation of Trex1 knockout mice [3]. These mice showed a greatly reduced lifespan and inflammatory myocarditis. By re-derivation of pathogen-free animals, it was demonstrated that the cause of the disease was not a virus or some other infectious agent, but rather an autoimmune condition. Human genetics and biochemical studies on relevant families identified TREX1 deficiency as a major cause of AGS [4, 16]. The Trex1-knockout mice provide an excellent animal model system for human AGS. Inherited systemic lupus erythematosus, variant forms of AGS and related clinical diseases with TREX1 dominant mutations have also been described [17, 18]. These studies have been reviewed in [19–21]. In a subsequent development, cloning of the gene deficient in patients with the neurovascular disorder RVCL (retinal vasculopathy with cerebral leukodystrophy) showed that this was also due to dominant mutations in the TREX1 gene [14, 21]. Interestingly, mutations in AGS1 are usually located to the catalytic domain of TREX1 and lead to loss of enzymatic activity, whereas TREX1 mutations in RVCL are found in the C-terminal domain and lead to loss of the perinuclear localization of catalytically active TREX1, as shown in the model system in Figure 1. The relative activity of mutated forms of TREX1 on ssDNA compared with dsDNA might also affect the disease pattern [22].

Checkpoint activation in TREX1-knockout cells

Murine and human TREX1-knockout cell lines and primary fibroblasts exhibit chronic checkpoint activation dependent on the ATM (ataxia telangiectasia mutated) protein kinase, similar to that seen transiently in normal cells exposed to ionizing radiation. The properties and features of this response have been reported and reviewed recently [15, 19, 20]. Our data indicate that the checkpoint activation in TREX1-knockout cells is due to intracellular accumulation of ssDNA fragments. These would be efficiently degraded in normal TREX1-positive cells, which may explain why they have not been detected previously. The accumulated ssDNA was found to be of a discrete length of 60–65 nt. Our data [15] indicate that these ssDNA fragments seen in AGS1 cells are generated during lagging-strand DNA synthesis. An additional source of ssDNA was proposed recently by Stetson et al. [23], who reported that the ssDNA in TREX1-negative mouse heart tissue may largely represent activated retroelements. It is noteworthy, however, that the cytoplasmic accumulation of ssDNA of 60–65 nt that we detect in TREX1-negative cells could be sufficient to trigger the innate immune system in mice and humans; this observation provides a possible link between the pathological inflammatory response in AGS and TREX1 deficiency.

References


Received 13 February 2009
doi:10.1042/BST0370535