Structure of the Dimeric Exonuclease TREX1 in Complex with DNA Displays a Proline-rich Binding Site for WW Domains

Marina Brucet§, Jordi Querol-Audi¶, Maria Serra¶, Ximena Ramirez-Espain¶, Kamila Bertlik¶, Lidia Ruiz¶, Jorge Lloberas¶, Maria J. Macias¶, Ignacio Fita¶, and Antonio Celada §1
From the §Macrophage Biology Group, Institute for Research in Biomedicine, University of Barcelona, the ¶Institut de Biologia Molecular de Barcelona, Institute for Research in Biomedicine, Consejo Superior de Investigaciones Científicas, and the ¶Protein NMR Spectroscopy, Institute of Research in Biomedicine-ICREA, Barcelona Science Park, 08028 Barcelona, Spain

TREX1 is the most abundant mammalian 3′ → 5′ DNA exonuclease. It has been described to form part of the SET complex and is responsible for the Aicardi-Goutières syndrome in humans. Here we show that the exonuclease activity is correlated to the binding preferences toward certain DNA sequences. In particular, we have found three motifs that are selected, GAG, ACA, and CTGC. To elucidate how the discrimination occurs, we determined the crystal structures of two murine TREX1 complexes, with a nucleotide product of the exonuclease reaction, and with a single-stranded DNA substrate. Using confocal microscopy, we observed TREX1 both in nuclear and cytoplasmic subcellular compartments. Remarkably, the presence of TREX1 in the nucleus requires the loss of a C-terminal segment, which we named leucine-rich repeat 3. Furthermore, we detected the presence of a conserved proline-rich region on the surface of TREX1. This observation points to interactions with proline-binding domains. The potential interacting motif “PPPVPVRPR” does not contain aromatic residues and thus resembles other sequences that select SH3 and/or Group 2 WW domains. By means of nuclear magnetic resonance titration experiments, we show that, indeed, a polyproline peptide derived from the murine TREX1 sequence interacted with the WW2 domain of the elongation transcription factor CA150. Co-immunoprecipitation studies confirmed this interaction with the full-length TREX1 protein, thereby suggesting that TREX1 participates in more functional complexes than previously thought.

When DNA replicates, recombines, or is repaired, its ends often require remodeling to prevent the formation of aberrant structures, which might result in serious cell dysfunctions. Among these DNA remodeling enzymes the 3′ → 5′ exonuclease act by removing nucleotides at the 3′ termini and were initially found to perform proofreading functions associated with DNA polymerases. However, in the last few years, several eukaryotic 3′ → 5′ exonucleases with functions unrelated to proofreading activities have been reported. Alterations in the genes encoding these autonomous exonucleases lead to dramatic consequences, such as strong mutator phenotypes, premature aging, susceptibility to cancer, and even lack of viability (1). Despite unquestionable evidence about the in vivo importance of autonomous exonucleases, the molecular mechanisms and biological roles of these enzymes are only now beginning to be elucidated.

First detected in mammalian liver and thymus extracts (2–4), TREX1 is the most abundant mammalian 3′ → 5′ exonuclease. The inactivation of the Trex1 gene in mice revealed the relevance of this exonuclease. Trex1 knock-out mice develop inflammatory myocarditis, resulting in progressive cardiomyopathy that leads to circulatory failure and a dramatic reduction in survival (5). Recently, TREX1 has been implicated in the Aicardi-Goutières syndrome, a severe neurological brain disease that mimics a viral infection acquired in the uterus (6). TREX1, together with some members of the SET complex, has recently been implicated in DNA degradation during granzyme A-mediated cell death (7). Given that TREX1 is an autonomous non-processive robust exonuclease, several biological roles have been proposed for this enzyme, including participation in proofreading functions. In fact, in reconstituted systems with exonuclease-deficient DNA polymerases, TREX1 shows 3′-editing activity, in particular for (i) the nuclease-deficient replication DNA polymerase α (3), (ii) the repair DNA polymerase β (8), and (iii) the DNA lesion bypass polymerase η (9). Similar activity has been observed for the closely related TREX2, which interacts with the exonuclease-deficient DNA polymerase δ, thereby increasing its fidelity under adverse conditions (1).

On the basis of its sequence, TREX1 has been classified as a member of the 3′ → 5′ exonuclease family DNAQ. This family includes the proofreading exonuclease fragments Klenow of Escherichia coli and is characterized by the presence of four non-contiguous acidic residues, three aspartates, and one glutamate, which play a crucial role by binding the two catalytic metal ions (10). A fifth residue, either a tyrosine or a histidine, completes the family catalytic motif (represented as DEDDy or DEDDh, respectively) distributed among three separate sequence segments named Exo I, II, and III (11). The information available to date indicates...
that the structure is well preserved among DnaQ family members, although some have very low sequence similarities outside the Exo segments (11, 12). Members of the family are monomeric, except for TREX1 and TREX2 proteins, which are the only DnaQ deoxyribonucleases characterized as dimers (13). TREX2 is the only known homologue of TREX1 in mammals (44% sequence identity). TREX-related proteins have been identified in a number of insects and in baculovirus (14) but no homologous genes have been found in yeast. During the submission of this article, a complementary work showing the structure of TREX1 has been reported providing the molecular basis for understanding the mutations that lead to Aicardi-Goutières syndrome (15). To better understand the relationship between the structure and function of TREX1, we determined the crystal structures of the binary complexes with a nucleotide product of the exonuclease reaction and with a single-stranded DNA substrate, which we identified using binding site selection experiments.

By means of nuclear magnetic resonance experiments, we demonstrated that a group 2 WW domain can interact with a peptide selected from the polyproline-rich region in TREX1. In addition, using confocal microscopy, we also analyzed the role of the C-terminal region in regulating the subcellular localization of the TREX1 protein. To our knowledge, TREX1 interaction with its substrate (DNA) and with its product (single nucleotide) represents the only structures of these complexes described to date from a mammalian deoxyribonuclease of the DnaQ family.

**EXPERIMENTAL PROCEDURES**

**Protein Preparation**—A murine Trex1 cDNA construct (residues 9–245) was cloned into the expression vector pETM-10 (pETM-10-Trex1). The protein was overexpressed in the *E. coli* strain Rosetta (DE3) (Novagen) by adding 1 mM isopropyl 1-thio-β-d-galactopyranoside at a Δabs = 0.6 and culturing the cells overnight at 15 °C. Pelleted cells were resuspended in buffer containing 20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 5 mM MgCl2, 10 mM imidazole, 1 mM dithiothreitol, and 8% glycerol. A cell lysate was obtained after mechanical lysis using a French Press. The protein was affinity purified from the supernatant of the cell lysate by nickel-nitrilotriacetic acid-agarose resin (Qiagen), with an elution buffer containing 20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 5 mM MgCl2, 200 mM imidazole, 1 mM dithiothreitol, and 8% glycerol, and a second purification step was carried out by a Superdex-75 gel filtration column (GE Healthcare), in a Tris-based buffer (30 mM Tris-HCl, pH 8.0, 200 mM NaCl, 5 mM MgCl2, and 1 mM dithiothreitol). The purified protein was concentrated to 7 mg/ml in the same buffer using Centricon centrifugal filter units (Millipore). The cDNA for the mutant Trex1 H195A was produced from the pETM-10-TREX1 vector, using a PCR site-directed mutagenesis strategy with the QuikChange site-directed mutagenesis kit (Stratagene), and the mutant protein was purified by the same protocol.

**Crystallization, Data Collection, and Structure Refinement**—Purified TREX1 protein solution was mixed with 2 mM dTMP, left overnight at 4 °C and used for crystallization. 1 μl of this solution was mixed with 1 μl of reservoir solution (22% PEG3350, 100 mM MES, pH 6.0, 200 mM Li2SO4) and crystallization was carried out by the hanging drop vapor diffusion method at 20 °C. Single crystals were soaked for 24 h in the same crystallization solution, except that it contained 200 mM MgSO4 and 50 mM Li2SO4. Crystals were then flash-frozen with 30% ethylene glycol as cryoprotectant. For the complex with DNA, a 25-mer single-stranded oligonucleotide was used (5′-GCTAGGCAAGAACCCTCTCCCTTCC-3′, derived from a binding site selection technique). It was mixed with the purified protein at stoichiometric concentrations and incubated overnight at 4 °C. 1 μl of this preparation was mixed with 1 μl of reservoir solution (20% PEG2KMME, 100 mM imidazole, pH 6.5, 300 mM Li2SO4) and crystallized at 20 °C. These crystals were cross-linked with 25% glutaraldehyde for 30 min and then frozen with 28% glycerol as cryoprotectant. Diffraction data were collected at the European Synchrotron Radiation Facility (Grenoble, France), and processed using DENZO and SCALEPACK (16). A first model was obtained for TREX1-dTMP by molecular replacement from the structure of the TREX2 protein (17), and the model was completed and refined using the CCP4 package and manually refined with the graphic programs Turbo and Coot. Residues 166–174 were disordered and were therefore excluded from the model. The first model for TREX1-DNA was obtained by molecular replacement from the TREX1-dTMP structure, and the model was refined by the same method. The disordered loop in TREX1-DNA was partially resolved. Crystallographic data collection, processing, and refinement statistics are summarized in Table 1. Structure representations were made with Pymol, and electrostatic potential surfaces were calculated with GRASP.

**NMR Experiments**—A DNA fragment encoding the WW2 motif of mouse CA150 (corresponding to residues 442–479) was produced as described (18). This domain is the same as that of FBP28_WW2, and shares the same amino acid sequence in human and mouse. NMR samples of the WW2 domain had a concentration of ~0.2 mM for structure determination and were dissolved in 100 mM sodium phosphate buffer, 20 mM NaCl, 1 mM NaN3, and 10% D2O at pH 6.0. For binding studies (or titrations) synthetic peptides corresponding to the prolinc-rich sequence of murine TREX1 (HPHPVPRPVRV) and human TREX1 (PPPTVPPPPRV) were synthesized in-house using the Fmoc (N-(9-fluorenyl)methoxycarbonyl) strategy with a rink-amide matrix (Novabiochem). The final peptides were cleaved with 95% trifluoroacetic acid and 5% H2O and then precipitated in cold ether. The crude material was purified by preparative HPLC using a C18 column to 90% purity (as characterized by HPLC-mass spectrometry). For the binding studies, we titrated the CA150_WW2 domain with increasing amounts of TREX1-derived peptides. Peptide was added to the 15N-labeled CA150_WW2 up to a molar peptide:protein ratio of ~9:1. The NMR data corresponding to the titration were acquired on a Bruker DRX-800 NMR spectrometer at 285 K. Protein assign-
ment was performed using published homonuclear data as the starting point. The same experiment was carried out with the human and murine TREX1 proline-rich peptides.

Co-immunoprecipitation—Cell extracts were prepared as described (19), mixed with pre-immune or CA150-specific antiserum and TrueBlot anti-rabbit IgG beads (eBioscience) overnight at 4 °C. After three washes, Western blotting was performed with primary antibodies (CA150 or TREX1 specific sera) and secondary antibody (Rabbit ExactaCruz, Santa Cruz Biotechnology).

Binding Site Selection—The binding site selection was made as described (20). Trex1 was cloned into pMal-C2 vector (Promega), and overexpressed in E. coli XL1Blue cells. After lysis, proteins were purified by an amylose resin (New England Biolabs). Maltose-binding protein (MBP) was prepared with the same protocol and used for controls. Randomized oligonucleotides (5′-CGACTCTAGAGGATCC(N)24GAATTCAAGC-3′) were made double-stranded using Tag polymerase and [α-32P]dCTP. The probes (200,000 cpm) were used for electrophoretic mobility shift assay with 2-0.5 μg of MBP-TREX1. The purified DNA was PCR amplified with [α-32P]dCTP (primers: 5′-CGACTCTAGAGGATCC-3′, 5′-CGTGAAGCTTGAATTC-3′). The recovered DNA was used in 3 subsequent rounds and the affinity-selected oligonucleotides were cloned into a pCR2.1 vector (Invitrogen). Positive clones were sequenced, the sequences were aligned using DNASTar and the consensus sequences were derived after analysis.

SYBR Green-based Exonuclease Activity Assay—Exonuclease reactions were performed in a 96-well plate. First, 10× the indicated double-stranded DNA oligonucleotide (50 μM) was incubated with 10× SYBR Green (Invitrogen), heated for 10 min at 95 °C, and left for 30 min at room temperature so as to anneal the DNA probe and allow the SYBR Green to incorporate into the DNA. For each well, the reaction mixture was prepared in a final volume of 14 μl containing 20 mM Tris-HCl, pH 7.5, 5 mM MgCl2, 2 mM dithiothreitol, and 100 μg/ml bovine serum albumin and 1 ng of MBP-TREX1. 3 μl of the SYBR Green/DNA mixture was then added to the wall of each well and, after a spin of the plate, the reaction was real time followed with an ABI Prism 7700 sequence detection system (Applied Biosystems) (program: 25 °C for 5 min, 25 °C for 1 min for 90 times, and 4 °C for 2 min). TREX1 dilutions were prepared at 4 °C in 100 μg/ml bovine serum albumin. Each DNA probe was annealed with its complementary oligonucleotide by heating at 95 °C for 10 min.

Subcellular Localization Experiments—A20 cells were permeabilized with 0.2% Triton X-100. After blocking with 3% bovine serum albumin in Tris-buffered saline, samples were incubated with TREX1 antiserum and secondary antibody (Cy3-Goat anti-rabbit IgG, Amersham Biosciences). DNA was labeled with 4′,6-diamidino-2-phenylindole (Sigma). Several Trex1 cDNA constructs were cloned into a pEGFP-N3 vector (Clontech) and transfected into the murine fibroblast cell line L929: pEGFP-Trex1wt (residues 1–314), pEGFP-Trex1Δ1 (42–314), pEGFP-Trex1Δ12 (85–314), pEGFP-Trex1Δ123 (85–280), and pEGFP-Trex1Δ3 (1–280). Empty pEGFP-N3 was used as control (EGFP alone). 100,000 cells were plated into each well of a 24-well plate with each well containing a coverslip. After 24 h, cells were transfected with 1 μg of each construct using the transfection kit Lipofectamine™ 2000 (Invitrogen), following the manufacturer’s instructions. Cells were incubated for 24 h more, fixed with 3% paraformaldehyde, and visualized by confocal microscopy.

RESULTS

The Optimal DNA Sequences That Bound to TREX1 Correlate with the Exonuclease Activity—TREX1 was initially cloned by exploiting its capacity to recognize a given DNA motif (21). To further explore whether TREX1 displays DNA-binding specificity or selectivity toward certain motifs, we used the PCR binding site selection method (22). We thus generated a pool of degenerated oligonucleotides and performed binding selection cycles with recombinant TREX1 protein. Three sets of oligonucleotide sequences were selected by the protein: in 26% of the cases the consensus showed GAG (sequence A), in 25% ACA (B), and in 18% CTG (C). To examine the potential preference of TREX1 for these sequences, we performed binding assays by incubating the probes (oligonucleotides containing each of the three motifs in tandem) with increasing amounts of protein. Binding was detected by gel retardations. Under these experimental conditions, the affinity differed depending on the DNA sequence (Fig. 1A). Interestingly, when we replaced the Gly by Cys in motif A, binding was abolished. To correlate the binding with the functional activity of TREX1, we also performed a sensitive and quantitative exonuclease activity test, using SYBR Green as indicator. Detection was made by a real time PCR assay (Fig. 1B). We found a clear correlation between binding and exonuclease activity of TREX1. This observation indicates that the specificity of TREX1 for some sequences is correlated with its functional activity, a new feature in the field of deoxyribonucleases.

Regions Responsible for TREX1 Nucleus-Cytoplasm Localization—TREX1 is involved in the SET complex that degrades DNA during granzyme A-mediated cell death. This observation implies an initial cytoplasmic localization for the protein and a stimulus-mediated translocation to the nucleus (7). The observation that TREX1 binds proteins involved in nucleic acid polymerization (3, 8, 9) suggests that TREX1 is involved in a range of activities in the cell nucleus. In fact, when we localized TREX1 using polyclonal antibodies we found that present in both the nucleus and cytoplasm (Fig. 2A).

Both TREX proteins contain a pair of leucine-rich repeats at their N termini, which in TREX2 were incorporated in the protein fold. However, TREX1 contains a highly hydrophobic C-terminal region, absent in TREX2, which we named leucine-rich repeat 3 (LRR3). Indeed, bioinformatic tools predict that this region forms a putative transmembrane helix (supplemental Fig. S1). Furthermore, protein constructs lacking the LRR3 region were expressed in the soluble fractions and in higher amounts than the full-length protein. In addition, the TREX1 construct lacking the LRR3 segment maintained full catalytic capacity because it was able to bind and degrade DNA (supplemental Fig. S2). On the basis of these observations, we hypothesized that this region influences the subcellular localization of TREX1 in vivo. To test this hypothesis, cDNAs coding for distinct fragments of TREX1 were cloned in an expression vector fused to enhanced green fluorescent protein (EGFP). Four frag-
Structure of the TREX1-DNA Complex

FIGURE 1. Optimal DNA sequences that bind to TREX1. A, electrophoretic mobility shift assays with the DNA sequences that preferentially bound to TREX1, selected by the binding site selection technique. x in the sequence indicates that the selected nucleotide sequence is in tandem in the oligonucleotides used for electrophoretic mobility shift assay. The recombinant protein used was MBP-TREX1, the same used for the selection, and different amounts of protein were used to measure the binding. The quantification graph indicates the relative binding of each oligonucleotide with respect to the probe alone (arbitrary units). B, exonuclease assay using SYBR Green and a real time monitoring of the exonuclease reaction. The experiment shows how TREX1 degrades the oligonucleotides selected by binding site selection, and that there is correlation between the binding and functional exonuclease activity of TREX1. The decay in the intensity shows the degradation of the double-stranded probes, as SYBR Green binds only to double-stranded DNA. Each reaction contained 1 ng of MBP-TREX1 and 5 μg double-stranded DNA.

TREX1 Overall Structure—Despite a huge number of trials, only one construct lacking the LRR3 segment was crystallized. The protein crystals belong to space group P2₁, and the asymmetric unit contains four protein subunits organized as two molecular dimers, two magnesium ions, and a dTMP molecule per active center. This is consistent with the results obtained for the protein in solution (13) (supplemental Fig. S3). The final model was refined at 2.35-Å resolution (Table 1) and includes 190 solvent molecules.

On the basis of the data obtained from the solubility and localization experiments and also from the DNA selection technique, we designed a number of oligonucleotides and prepared several TREX1 constructs and mixtures with DNA, which were later used for crystallization trials. The mixtures contained either double-stranded or single-stranded DNAs corresponding to both the forward and reverse sequence of each oligonucleotide selected. In parallel, trials with 2 mM deoxythymidine 5’-monophosphate (dTMP) were also attempted.

TREX1 are closely related to those in TREX2 (about 40% sequence identity) (17). Despite low sequence identity with respect to monomeric exonucleases of the DnaQ family, such as the Klenow fragment of E. coli DNA polymerase I (about 20%) (23), their overall fold is quite conserved.

TREX subunits consist of a central five-stranded anti-parallel β-sheet surrounded by nine α-helices (Fig. 3A). The structures of the four subunits are very similar, with an averaged root mean square deviation between Cα atoms of 0.32 Å.

The sheet extends throughout the molecular 2-fold axis to the second subunit in the molecule, thereby giving a continuous 10-stranded anti-parallel sheet. Both polar and hydrophobic interactions act between monomers, mainly because of residues from the β3 strand and the α4 helix. TREX1 has a proline-rich segment (residues 54–63), which is absent in TREX2. This segment is located on the protein surface not far from the molecular 2-fold axis (closest distance 10.19 Å, from Cα of Pro58) and adopts a polyproline type II helix conformation (PPII). There is a disordered segment in TREX1 adjacent to the active site (residues 166–174). The corresponding loop in TREX2, also disordered, appears to play an important role in DNA binding (17). Overall, the main differences between the two TREX structures are the additional helix α1 and a long loop, which, in TREX1, corresponds to residues 47–65 including the proline-rich segment (Fig. 3B).
Active Site—The electron density map confirmed the binding to each protein subunit of one dTMP nucleotide and two magnesium ions (named MgA and MgB), defining an active site located away from the molecular 2-fold axis, in agreement with what was proposed for TREX2 (17) (Fig. 3A). The dTMP molecule would correspond to the product of the exonuclease reaction that remains bound by hydrophobic and hydrogen-bond interactions with the protein and also by interactions with magnesium ions. MgA, with an approximate trigonal bipyramidal geometry, coordinates with oxygen molecules O2P and O3P from the dTMP phosphate group and with the carboxylate oxygen from Asp18, Glu20, and Asp200. In turn, MgB, with nearly perfect octahedral geometry, coordinates with the second phosphate and with the second carboxylate oxygen from Asp18 and with four molecules of water, one of them hydrogen-bonded to the carboxylate moiety of Asp180. N6 from His195 presents interactions with dTMP oxygen molecules O1P and O2P (whereas N7 remains exposed to the solvent). Therefore, in TREX1, Asp18/Glu20/...
Asp<sup>130</sup>/Asp<sup>200</sup>/His<sup>195</sup> conforms the characteristic DEDD<sub><i>h</i></sub> exonuclease motif of the DnaQ family, presenting interactions with the product nucleotide that are similar to those determined for other members of the family. The thymine, loosely sandwiched between Leu<sup>24</sup> and Ile<sup>84</sup>, exposes its edge to the solvent without forming any other bond with the protein. This observation indicates a lack of specificity at this level as even the larger purine bases would not have steric impediments.

**FIGURE 3.** TREX1 overall structure. A, the TREX1 homodimer (monomers indicated in cyan and purple) is shown in terms of its secondary structural elements (indicated in one monomer) across the non-crystallographic 2-fold axis. The two magnesium ions (green spheres) and a dTMP nucleotide (orange atom-type sticks) are shown in each active center of the dimer. Ala<sup>165</sup> and Lys<sup>175</sup> indicate the position of the disordered loop (from 166 to 174) at the two monomers. The close-up shows the residues at the active center that participate in nucleotide binding (yellow atom-type sticks). The two magnesium ions (MgA and MgB) and coordination water molecules are also shown as green and red spheres, respectively. B, structure-based sequence alignment of TREX1 with TREX2 (Protein Data Bank code 1Y97) and the exonuclease Klenow fragment of the DNA polymerase I from E. coli (PDB code 1KSP). The following features are highlighted: red, conserved catalytic residues; blue, catalytic residues that define the subfamily DEDD<sub><i>y</i></sub>/h (see text for details); gray, identical residues; filled box in gray, proline-rich region; dots and bold letters, disordered loop; magenta, unstructured C-terminal region; underscored green, TREX1 leucine-rich repeats. Secondary TREX1 structural elements are shown above the aligned sequences.

Asp<sup>130</sup>/Asp<sup>200</sup>/His<sup>195</sup> conforms the characteristic DEDD<sub><i>h</i></sub> exonuclease motif of the DnaQ family, presenting interactions with the product nucleotide that are similar to those determined for other members of the family. The thymine, loosely sandwiched between Leu<sup>24</sup> and Ile<sup>84</sup>, exposes its edge to the solvent without forming any other bond with the protein. This observation indicates a lack of specificity at this level as even the larger purine bases would not have steric impediments.
The hydroxyl group O3 of the dTMP deoxyribose makes hydrogen bonds with Glu20 and the backbone oxygen from Ala21, and the dTMP O4 is hydrogen-bound to Tyr129 (supplemental Fig. S4A). The structure of the ISG20 exoribonuclease of the DnaQ family in complex with UMP (24) shows that the ribose hydroxyl group O2, absent in deoxyriboses, hydrogen bonds to the guanidinium group of an arginine of ISG20, equivalent to Ala81 in TREX1. Similarly to TREX1, all deoxyribonucleases from the DnaQ family lack long polar residues in this position, which might explain the preference for DNA substrates (25) (supplemental Fig. S4B).

TREX1 in Complex with DNA—Crystallization was initially attempted in the absence of magnesium as TREX1 degrades DNA in its presence. However, all trials failed, probably because the stability of the complexes in the absence of this cation was reduced. Therefore, new crystallization attempts were performed with protein constructs in which residue His195, predicted to be catalytically essential, was mutated to Ala. The catalytic role of His195 was confirmed with the mutated protein H195A, which is unable to degrade DNA (exonuclease assays in supplemental Fig. S4B).

The crystals obtained belong to space group P4_2_2_2 and contain one dimer per asymmetric unit. The structure of the complex was solved by molecular replacement using the TREX1-dTMP model. Superimposition with the dimers found in the TREX1-dTMP crystal gave an average root mean square deviation for Ca atoms of 0.75 Å.

In this complex the loop Ser166-Arg174 is also mostly disordered, although Ser166 and Arg174 were visible in subunits A and B, respectively. The conformation of the active site was nearly identical to that of the TREX1-dTMP structure but the electron density map showed extra density in the two subunits, appropriate for a four-nucleotide single-stranded DNA with the 3’-terminal nucleotide occupying the same position as the dTMP product (Fig. 4, A and B). This configuration is compatible with metal ion-catalyzed phosphoryl transfer, as proposed for DnaQ enzymes (26, 27). His195, located near the terminal phosphodiester bond, would initiate the catalytic reaction, as described for other exonucleases.

The three remaining visible nucleotides orient their bases away from the protein, thereby excluding the occurrence of specific protein-DNA interactions at this level. In turn, the phosphodiester chain runs antiparallel to the extended protein segment Lys175, Ser176, and Ser177, which interacts with phosphate groups 4, 3, and 2, respectively (Fig. 4B). The active sites of the protein are located on opposite outer edges of the whole molecule, with each active site forming a cleft of acidic, negative charged residues, as shown in the electrostatic surface representation. Adjacent to this negative cleft there are regions of positive charge that might also contribute to the interaction with DNA (Fig. 4C). The 3’ nucleotide of the DNA occupies the same position as the dTMP product, except for the rotation of the phosphate group (Fig. 4D). The disposition of the 3’ nucleotide substrate with respect to the product, together with the limited number of interactions between the remaining nucleotides and the protein, could explain the non-processive mechanism of TREX1. Effectively, in processive enzymes the removal of the product away from the active center is required before allowing entry of the next oligonucleotide substrate.

The structure of the TREX1-DNA complex is closely related to the structures of both single- and double-stranded oligonucleotides with the bacterial Klenow subunit. This is the only other member of the family with structural information available for complexes with DNA (28) (Fig. 5A). The structural similarities indicate, in particular, that TREX1 complexes with double-stranded DNA are likely to involve interactions with the second DNA strand, which is in agreement with substrate preference studies in which the optimal substrate is duplex DNA or mispaired 3’ termini structures (13). The positively charged
TREX1 Interacts with Proteins Containing WW Domains—
As indicated before, the structure of TREX1 contains a PPII helix from residue 54 to 63, which is absent in TREX2 and in monomeric exonucleases. This proline-rich sequence contains a PPXVPXPPP motif and it is conserved in mouse, rat, and human sequences but not in Anopheles or Drosophila. Only the last pair of residues (PR) is present in TREX2 (Fig. 6, A and B). Proline-rich motifs are found in many proteins and often play a critical role in protein interactions. These interactions are implicated in many cellular processes, such as signaling, cellular growth, changes in the cytoskeleton, transcription, and other biological activities (29–31).

Proline-rich sequences are targets of several protein domains, including SH3, WW, EVH1, GYF and UEV (32). These domains interact with the core motifs of their ligands through stacking of aromatic rings with the pyrrolidine ring of the prolines of the ligand. Binding specificity and affinity beyond the core motif are normally obtained through additional hydrophobic or charge interactions. The presence of several consecutive prolines, a hydrophobic residue, and an arginine in the motif indicates that WW and/or SH3 may be ideal candidates to interact with TREX1.

Regarding WW domains, a subclass known as group 2, is specialized in the recognition of this type of ligand (18). The SH3 of Sem5, a member of the growth factor receptor-bound protein 2, interacts with the PPPVPPR peptide present in the protein Son of sevenless (33). This sequence is very similar to that of mouse TREX1, thereby supporting the hypothesis that SH3 domains may interact.

PPII helices have three residues per helical turn that are arranged in a 3-fold symmetry axis. In the TREX1 structure, the PPII motif has one-third of its surface occupied with contacts to the protein, which leaves two faces of the helix still accessible. Due to the small size of WW domains they should present less steric impediments than SH3 when interacting with TREX1. In addition, given that SH3 domains are normally found in cytoplasmic proteins and that there are many nuclear proteins with WW domains, (splicing and transcription factors), we examined whether a WW domain belonging to group 2 interacts with a peptide with the proline-rich sequence of TREX1.

We used nuclear magnetic resonance to characterize the potential interaction because this technique provides a detailed description of the residues involved in the interaction at an atomic level as well as information regarding affinity. As a representative WW sequence, we selected that of the elongation transcription factor CA150_WW2 domain (formerly known as FBP28).

Addition of unlabeled peptide to the $^{15}$N-labeled CA150_WW2 domain resulted in chemical shift changes for a defined set of backbone amide chemical shifts, indicative of a specific interaction in the fast NMR time scale (Fig. 6C). With regards to the binding region, the peptide induced changes in residues localized in both XP and XP2 grooves, as previously predicted (34) and recently identified as binding regions of WW domains belonging to group 2 (35). For instance, residues 10–13 in the $\beta$1 strand, 18–21 in $\beta$2, and 28–31 in the $\beta$3 strand manifested average changes well above the threshold (numbers in the domain are maintained as in Ref. 18) (changes

region adjacent to the binding site would participate in these interactions, including the disordered loop ending in residue Arg$^{174}$, and probably also the exposed residue Trp$^{188}$, whose conformation changes between the structures with and without DNA (Fig. 5B).
FIGURE 6. TREC interacts with proteins containing WW domains. A, superimposition of the structures of exonucleases from the DEDD family. In cyan, TREC1 (PDB 2O4G); orange, TREC2 (PDB 1Y97); gray, other exonucleases: Klenow fragment (PDB 1KSP), ε subunit (1J54), Exol (PDB 1FXX), Eri1 (PDB 1W0H), WRN (PDB 2FC0), ISG20 (PDB 1WLJ), and RNase T (PDB 2F96). The superimposition reveals the presence of a loop in TREC proteins that is not contained in the monomeric proteins of the family. B, an alignment of the TREC proteins from several species shows the conservation of the polyproline region (boxed) in TREC1 proteins, but not in TREC2 or TREC from Drosophila or Anopheles. Species are: h, human; m, murine; r, rat; b, bovine; Ag, Anopheles gambiae; Dm, Drosophila melanogaster. C, 15N-HSQC of the CA150_WW2 domain with increasing amounts of HPPVPWPRPV peptide derived from TREC1. The titration shows that the addition of the peptide to the labeled domain resulted in chemical shift changes for a defined set of backbone amide chemical shifts, indicative of specific interactions with the residues localized in the XP and XP2 grooves of the domain. A picture of the domain with the interacting residues is also shown.
and with a dTMP, its enzymatic product. These complexes are reported for the first time for a mammalian member of the DnaQ 3' \rightarrow 5' exonuclease family. TREX1 complexes show close structural relationships with the complexes of the Klenow fragment of \textit{E. coli} DNA polymerase I, which confirms the architectural constancy within the DnaQ family despite large evolutionary distances and very low sequence similarities.

The structures support a metal ion-catalyzed phosphoryl transfer mechanism with a histidine, near the phosphodiester bond, that initiates the hydrolytic reaction, as described for several other exonucleases (26, 27). The structure of TREX1 in complex with DNA shows that no specific interactions take place in the active site because the bases of the nucleotides are located away from the protein. Intriguingly, we have determined, by binding assays, that some DNA sequences bind with higher affinity than others. This distinct affinity correlates with the exonuclease activity of TREX1, suggesting that TREX1 shows significant preferences for some sequences. It has been reported that TREX proteins recognize the DNA via the loop adjacent to the active site (17). In TREX1, the preference for certain DNA sequences could be explained through specific interactions between the DNA bases and the residues of this loop. However, this kind of specific interaction cannot be observed in the structure because the loop is not well ordered. Alternatively, a particular DNA secondary structure could be responsible for the binding preferences. The sequence specificity recognition is quite an unusual feature for exonucleases, and indeed there is only one other example reported, that of the exoribonuclease 3' hExo (36).

The TREX1 structures revealed the presence of a solvent-exposed proline-rich region, which is absent in all other exonucleases, including TREX2. A TREX1-derived polyproline peptide was shown to bind to a group 2 representative WW sequence, that of the elongation transcription factor CA150. This interaction was verified using NMR spectroscopy. Co-immunoprecipitation studies using both entire proteins further demonstrated that TREX1 and CA150 can interact with one another. This observation suggests that the proline-rich area exposed in the TREX1 structure and the cavity size, in principle, accommodate a group 2 WW domain.

Up to date, there is only one report of TREX1 being involved in protein-protein interactions, that of its association with the nucleosome assembly protein SET (7). However, SET protein lacks either WW or SH3 motifs, suggesting that this interaction is not mediated by the polyproline motif of TREX1. Nevertheless, we found that Fe65, a transcription-associated factor that interacts with SET, contains a group 2 WW motif (37). Other candidates to bind TREX1 could be identified by searching WW- or SH3-containing proteins in a more systematic manner.

TREX1 also contains three LRR motifs. LRRs are involved in protein-protein or protein-membrane interactions and in mediating nuclear-cytoplasm transport (38, 39). LRR3 spans about 30 residues of the C terminus of TREX1 and has been predicted to be a transmembrane segment by \textit{in silico} analysis (supplemental Fig. S1). Full-length TREX1 localizes in the cytoplasm. In contrast, removal of LRR3, which \textit{in vivo} might be performed by specific proteases, appears to trigger the translo-

**DISCUSSION**

Here we report the structure of TREX1 in complex with a single-stranded DNA, substrate of the exonuclease reaction,
cation of TREX1 toward the nucleus. This observation suggests a possible role for the LRR3 region in the interaction with hydrophobic elements of the cell, and is consistent with its observed involvement in nucleus-cytoplasmic localization.

With respect to TREX1 function, it has recently been reported that mutations in the gene encoding TREX1 are responsible for causing the Aircadi-Goutières syndrome (6). The structure of TREX1 sheds light on the effect of these clinical mutations: D201ins and V201D mutations affect the catalytic residue Asp200 involved in Mg\(^{2+}\) cation binding and thus would destabilize the active site, causing a defect in the cation binding. The cation binding is necessary for the proper interaction of the acidic residues with the negatively charged DNA substrate. A mutation (R114H) affects the α4-helix, which participates in the interaction between monomers. A mutation of this residue most likely would destabilize the dimer formation.

It is interesting to note that this residue is conserved in TREX proteins from other species but is absent in monomeric members of the DnaQ family (17). Two additional mutations were also identified that truncate the protein.

Within the DnaQ family only TREX1 and TREX2 form dimers. Residues directly involved in dimerization are highly conserved among TREX proteins and absent in other members of the DnaQ family (17). In principle, TREX1 dimers could simultaneously process two DNA termini. However, the structures of the TREX1 complexes show the infeasibility of a DNA molecule to connect the two active centers located on opposite sides of the dimer. On the other hand, the substrates of preference for TREX1 are DNA duplexes with a mispaired 3′ termini or overhanging, rather than single-stranded DNA (13). These observations raise the question “why is a dimer required if at any one time only one catalytic center seems necessary?” A number of hypotheses have been proposed (17), but a simple explanation is related to the efficacy of TREX1 to recognize 3′ termini in either of the strands when moving along a double-stranded DNA (supplemental Fig. S5). This explanation is particularly appropriate if TREX1 belongs to a supramolecular complex that would otherwise have to rotate 180° before adopting the proper orientation with respect to the free 3′ end of DNA.

Exonucleases from the DnaQ family are involved in many biological functions and for TREX1, a variety of roles have been proposed, ranging from proofreading associated with DNA polymerases (3, 8, 9) to participation in V(D) recombination, receptor editing, or class switch recombination in lymphocytes (5) and, very recently, in degradation of apoptotic DNA together with the SET complex (7). Our data provide new perspectives to interact with TREX1, thereby opening up new perspectives into the in vivo roles of this exonuclease.

Acknowledgments—We thank Carles Sanyer from CSIC in Granada, Spain, for the gift of the antibody against CA150. The technical help of Rosa Perez-Luque is highly appreciated. We also thank Tanya Yates for editing the manuscript.

REFERENCES
1. Shevelev, I. V., Ramadan, K., and Hubscher, U. (2002) Scientific World J. 2, 275
2. Lindahl, T., Gally, J. A., and Edelman, G. M. (1969) J. Biol. Chem. 244, 5014–5019
3. Perrino, F. W., Miller, H., and Ealey, K. A. (1994) J. Biol. Chem. 269, 16357–16363
4. Perrino, F. W., Mazur, D. J., Ward, H., and Harvey, S. (1999) Cell Biochem. Biophys. 30, 331–352
5. Morita, M., Stamp, G., Robins, P., Dulic, A., Rosewell, I., Hrvnak, G., Daly, G., Lindahl, T., and Barnes, D. E. (2004) Mol. Cell. Biol. 24, 6719–6727
6. Crow, Y. J., Hayward, B. E., Parmar, R., Robins, P., Leitch, A., Ali, M., Black, D. N., van Bokhoven, H., Brunner, H. G., Hamel, B. C., Corr, P. C., Cowan, F. M., Frits, S. G., Klepper, J., Livingston, J. H., Lynch, S. A., Massey, R. F., Meritet, J. F., Michaud, J. L., Ponsot, G., Voit, T., Lebon, P., Bonthron, D. T., Jackson, A. P., Barnes, D. E., and Lindahl, T. (2006) Nat. Genet. 38, 917–920
7. Chowdhury, D., Beresford, P. J., Zhu, P., Zhang, D., Sung, J. S., Demple, B., Perrino, F. W., and Lieberman, J. (2006) Mol. Cell 23, 133–142
8. Hoss, M., Robins, P., Green, N. J., Pappin, D. J., Sgouros, J., and Lindahl, T. (1999) EMBO J. 18, 3868–3875
9. Bebenek, K., Matsuda, T., Masutani, C., Hanaoka, F., and Kunkel, T. A. (2001) J. Biol. Chem. 276, 2317–2320
10. Derbyshire, V., Pinsonneault, J. K., and Joyce, C. M. (1995) Methods Enzymol. 262, 363–385
11. Moser, M. J., Holley, W. R., Chatterjee, A., and Mian, I. S. (1997) Nucleic Acids Res. 25, 5110–5118
12. Zuo, Y., and Deutscher, M. P. (2001) Nucleic Acids Res. 29, 1017–1026
13. Mazur, D. J., and Perrino, F. W. (2001) J. Biol. Chem. 276, 17022–17029
14. Yang, D. H., de Jong, J. G., Makhmoudova, A., Arif, B. M., and Krell, P. J. (2004) J. Gen. Virol. 85, 3569–3573
15. de Silva, U., Choudhury, S., Bailey, S. L., Harvey, S., Perrino, F. W., and Hollis, T. (February 9, 2007) J. Biol. Chem. 10.1074/jbc.M700039200
16. Ottowinski, Z., and Minor, W. (1997) Methods Enzymol. 267, 307–326
17. Perrino, F. W., Harvey, S., McMillin, S., and Hollis, T. (2005) J. Biol. Chem. 280, 15212–15218
18. Macias, M. J., Gervais, V., Civera, C., and Oschkinat, H. (2000) Nat. Struct. Biol. 7, 375–379
19. Xaus, J., Cardo, M., Vallerod, A. F., Soler, C., Lloberas, J., and Celada, A. (1999) Immunity 11, 103–113
20. Tang, W., and Perry, S. E. (2003) J. Biol. Chem. 278, 28154–28159
21. Klein, M. J., McKeever, S. R., Celada, A., Van Beveren, C., and Maki, R. A. (1990) Cell 61, 113–124
22. Pollock, R., and Treisman, R. (1991) Genes Dev. 5, 2327–2341
23. Shevelev, I. V., and Hubscher, U. (2002) Nat. Rev. Mol. Cell. Biol. 3, 364–376
24. Horio, T., Murai, M., Inoue, T., Hamasaki, T., Tanaka, T., and Ohgi, T. (2004) FEBS Lett. 577, 111–116
25. Mazur, D. J., and Perrino, F. W. (1999) J. Biol. Chem. 274, 19565–19566
26. Hamand, S., Carr, P. D., Brown, S. E., Ollis, D. L., and Dixon, N. E. (2002) Structure 10, 535–546
27. Beese, L. S., and Steitz, T. A. (1991) EMBO J. 10, 25
28. Beese, L. S., Derbyshire, V., and Steitz, T. A. (1993) Science 260, 352
29. Ball, L. J., Kuhne, R., Schneider-Mergener, J., and Oschkinat, H. (2005) Angew. Chem. Int. Ed. Engl. 44, 2852–2869
30. Zarrinarb, A., Bhattacharyya, R. P., and Lim, W. A. (2003) Sci STKE 2003, REB
31. Kay, J. N., Link, B. A., and Bajer, H. (2005) Development 132, 2573–2585
32. Agrawal, V., and Kishan, K. V. (2002) Protein Pept. Lett. 9, 185–193
33. Wittekind, M., Mapelli, C., Lee, V., Goldfarb, V., Friedrichs, M. S., Meyers, C. A., and Mueller, L. (1997) J. Mol. Biol. 267, 933–952
34. Kato, Y., Nagata, K., Takahashi, M., Lian, L., Herrero, J. J., Sudol, M., and Tanokura, M. (2004) J. Biol. Chem. 279, 31833–31841
35. Pires, J. R., Parthier, C., Aido-Machado, R., Wiedemann, U., Otte, L., Bohn, G., Rudolph, R., and Oschkinat, H. (2005) J. Mol. Biol. 348, 399–408
36. Yang, X. C., Purdy, M., Marzluff, W. F., and Dominski, Z. (2006) J. Biol. Chem. 281, 30447–30454
37. Telese, F., Bruni, P., Donizetti, A., Gianni, D., D’Ambrosio, C., Scaloni, A., Zambrano, N., Rosenfeld, M. G., and Russo, T. (2005) EMBO Rep. 6, 77–82
38. Katay, U., and Guttinger, S. (2005) Trends Cell Biol. 15, 121
39. Kope, B., and Kaiser, A. V. (2001) Curr. Opin. Struct. Biol. 11, 725