

# Identification of two conserved aspartic acid residues required for DNA digestion by a novel thermophilic Exonuclease VII in *Thermotoga maritima*

Andres A. Larrea, Ilene M. Pedroso, Arun Malhotra and Richard S. Myers\*

Department of Biochemistry and Molecular Biology, University of Miami Miller School of Medicine, Miami, FL 33136, USA

Received April 23, 2008; Revised August 31, 2008; Accepted September 2, 2008

## ABSTRACT

Exonuclease VII was first identified in 1974 as a DNA exonuclease that did not require any divalent cations for activity. Indeed, *Escherichia coli* ExoVII was identified in partially purified extracts in the presence of EDTA. ExoVII is comprised of two subunits (XseA and XseB) that are highly conserved and present in most sequenced prokaryotic genomes, but are not seen in eukaryotes. To better understand this exonuclease family, we have characterized an ExoVII homolog from *Thermotoga maritima*. *Thermotoga maritima* XseA/B homologs TM1768 and TM1769 were co-expressed and purified, and show robust nuclease activity at 80°C. This activity is magnesium dependent and is inhibited by phosphate ions, which distinguish it from *E. coli* ExoVII. Nevertheless, both *E. coli* and *T. maritima* ExoVII share a similar putative active site motif with two conserved aspartate residues in the large (XseA/TM1768) subunit. We show that these residues, Asp235 and Asp240, are essential for the nuclease activity of *T. maritima* ExoVII. We hypothesize that the ExoVII family of nucleases can be sub-divided into two sub-families based on EDTA resistance and that *T. maritima* ExoVII is the first member of the branch that is characterized by EDTA sensitivity and inhibition by phosphate.

## INTRODUCTION

Misincorporation of DNA bases during the replication cycle creates base mismatches that must be quickly repaired lest they be passed on to the next generation as

mutations. The DNA mismatch repair [methyl-directed mismatch repair (MMR)] pathway is commonly used by bacteria to repair base mutations thereby keeping genome mutation frequencies low. A similar MMR pathway is found in eukaryotes as well. Although the core features and proteins of MMR are conserved through all branches of life, there are differences in how newly synthesized DNA is discriminated from 'old' DNA and how the nicked, mismatched strand is excised (1,2). Mutations in MMR proteins have been linked to increased rates of frameshift mutation, decreased fidelity of homologous recombination and increased rates of certain types of cancer, including colon, endometrial and rectal cancer (3–5). In *Escherichia coli*, single nucleotide DNA insertion/deletion mutations are repaired by proteins that first recognize the mismatch (MutS) and distinguish the parental DNA from the daughter DNA strand. Next, a nick is created 5' or 3' to the mismatch (MutL and MutH), and the nicked strand is unwound (DNA helicase II UvrD/MutU) and digested (Exonuclease I, RecJ, Exonuclease VII or Exonuclease X) back towards the mismatch to remove the mispaired bases (6–8). It has been proposed that these four nucleases (ExoI, RecJ, ExoVII and ExoX) serve redundant functions in MMR. However, these enzymes are dissimilar and possess distinct mechanisms; their individual deletions affect other features of DNA repair in different ways (8–16) suggesting that they are not maintained in genomes as 'backup' activities for MMR.

Exonuclease VII, one of the nucleases implicated in resection of the nicked mismatched strand during bacterial MMR, was first identified and characterized in 1974 (17,18). Early work showed that *E. coli* ExoVII: (i) is a hetero-pentameric protein composed of one large subunit (XseA) and four identical small subunits (XseB), (ii) is specific for linear single-stranded DNA (ssDNA),

\*To whom correspondence should be addressed. Tel: +1 305 243 2056; Fax: +1 305 243 3955; Email: r.myers@miami.edu

Present address:

Andres A. Larrea, Laboratory of Structural Biology and Laboratory of Molecular Genetics, National Institute of Environmental Health Sciences, NIH, Research Triangle Park, NC 27709, USA

(iii) can initiate digestion from ssDNA ends of either polarity, introducing endonucleolytic nicks to release oligonucleotide products of various sizes, (iv) has no apparent requirement for a divalent cation [completely active in the presence of 10 mM EDTA and only slightly (25%) increasing activity in the presence of MgCl<sub>2</sub>] and (v) is strongly stimulated by the presence of phosphate (a 5- to 10-fold increase in activity up to 67 mM phosphate) (17–19). While the release of oligonucleotide digestion products implies that ExoVII cleaves DNA ‘endonucleolytically’, the requirement for free DNA ends led to its classification (EC 3.1.11.6) as an ‘exonuclease’ (17). Surprisingly, little is known about the cellular role of ExoVII, even though it is highly conserved throughout the bacterial and archaeal kingdoms. In fact, ExoVII is the most highly conserved of the four nucleases that have been implicated in MMR (our unpublished data). ExoVII is present in almost 90% of sequenced bacterial genomes, but has only been studied in *E. coli*.

Here, we present a characterization of the *Thermotoga maritima* homolog of Exonuclease VII. We chose to study the homolog from *T. maritima* in part because this organism shows extensive evidence of lateral gene transfer from both bacteria as well as from archaea (20,21). The *T. maritima* homolog was overexpressed and purified from *E. coli* and characterized. Briefly, our results show that *T. maritima* ExoVII is composed of two subunits (TM1768 and TM1769), is a linear ssDNA-specific nuclease releasing oligonucleotide products, and is processive just like its *E. coli* counterpart. Surprisingly, the *T. maritima* ExoVII requires Mg<sup>2+</sup> and is strongly inhibited in the presence of either phosphate or sulfate. These last two properties set it apart from *E. coli* ExoVII, but are consistent with standard hydrolytic DNases.

Furthermore, we have leveraged sequence homology between the large number of bacterial ExoVII orthologs to identify a conserved core motif within the C-terminus of the XseA (TM1768) subunit of ExoVII that resembles metal-binding sites of other DNA hydrolases. This putative active site was experimentally verified by site-directed mutagenesis that showed that the residues D235 and D240 are essential for ExoVII catalytic activity. We propose that this highly conserved set of amino acids comprise a previously undiscovered active site motif that possesses properties consistent with other hydrolytic nucleases that coordinate catalytic divalent metal ions. Therefore, we propose that ExoVII is a metal-dependent hydrolase and that what distinguishes *E. coli* ExoVII and *T. maritima* ExoVII is their susceptibility to inactivation by EDTA and modulation by DNA-mimetic anions.

## MATERIALS AND METHODS

### Bacterial strains, plasmids and growth conditions

A plasmid for the co-expression of the *T. maritima* genes TM1768 and TM1769 was provided by the Joint Center for Structural Genomics (Michael DiDonato, JCSG, personal communication). This plasmid is based on the pET-DUET-1 expression system (Novagen, San Diego, CA), and puts these genes under the control of an IPTG-inducible

promoter with an N-terminal His-tag on TM1769. This plasmid (CoEx\_TM1768/TM1769) was transformed into *E. coli* BL21(DE3) cells for overexpression. Cells were grown in Terrific Broth (Difco) with glycerol (0.4% final concentration) and ampicillin (100 µg/ml).

### Bioinformatics

Pairwise sequence alignments were done using Sequence Analysis for Mac OS X (<http://informagen.com/SA/>) along with minor manual adjustments. Multiple sequence alignments were done by first retrieving all sequences identified as homologous to *E. coli* XseA and *E. coli* XseB using BLASTp against all completed bacterial genomes on the NCBI site ([http://www.ncbi.nlm.nih.gov/sutils/genom\\_table.cgi](http://www.ncbi.nlm.nih.gov/sutils/genom_table.cgi)). These sequences were then aligned using either MUSCLE (22,23) or ClustalW ver. 1.8 (24–26). Since this procedure identified a large number of XseA homologs (345), we used an in-house developed script to search for conserved motifs in the multiple sequence alignment. This script, PAL (described in detail at <http://hdl.nature.com/10101/npre.2007.917.1> and available by request), allows rapid identification of highly conserved cores in large multiple sequence alignment files. These conserved cores often identify residues that are important to the activity of the enzyme. Although many programs have been written to analyze sequence alignments (27–31), PAL provides improvements in speed and ease of use.

### Overexpression and purification of wild-type and mutant *T. maritima* ExoVII

BL21(DE3) [CoEx\_TM1769/TM1768] cells were grown to mid exponential phase and induced with 1 mM IPTG for 4 h at 37°C. Cells were then harvested by centrifugation, resuspended in 5 ml of Buffer I (300 mM NaCl, 0.25 mM TCEP, 20 mM Tris pH 8, 20 mM imidazole and 10% glycerol) per gram of wet cell pellet, lysed by two passes through a high-pressure homogenizer (Avestin Inc. Ottawa, Canada), and centrifuged at 35 000g for 60 min to pellet insoluble cell debris. The soluble fraction was heated to 65°C for 20 min and centrifuged at 35 000g for 60 min to remove aggregated *E. coli* proteins. The soluble fraction, heavily enriched in *T. maritima* ExoVII, was applied on a 5 ml HisTrap nickel-NTA column (GE Healthcare Inc. Piscataway, NJ), washed with 10 column volumes of Buffer I and eluted with a linear imidazole gradient in Buffer II (Buffer I with 500 mM imidazole). Fractions containing both subunits eluted at ~200 mM imidazole, were pooled, concentrated and further fractionated on an S-200 gel filtration column (GE Healthcare Inc.) pre-equilibrated with Buffer III (300 mM NaCl, 1 mM DTT, 10% glycerol, 20 mM Tris acetate pH 7.5). Since an N-terminal His-tag had been engineered on TM1769, we obtained fractions enriched in TM1769 as well as fractions containing TM1768 in a complex with TM1769. Fractions containing both subunits of *T. maritima* ExoVII (at an approximate ratio of 1 TM1768: 4 TM1769 as judged by Coomassie staining of proteins separated by SDS polyacrylamide gel electrophoresis) were pooled and concentrated to 20 mg/ml in Buffer III

and stored at  $-80^{\circ}\text{C}$ . The preparation of *T. maritima* ExoVII was >99% pure as determined by Coomassie stained SDS polyacrylamide gels. For activity assays, protein stocks were diluted in  $1\times$  reaction buffer to 2.5–5 mg/ml (30–60  $\mu\text{M}$ ) with reaction concentrations typically between 1.5  $\mu\text{M}$  and 3  $\mu\text{M}$ .

### ***Escherichia coli* ExoVII**

Experiments using *E. coli* ExoVII were carried out using commercially available preparations of this enzyme (USB). The enzyme was supplied by the manufacturer in a 50 mM Tris HCl (pH 8.0), 200 mM NaCl, 0.5 mM EDTA, 10 mM DTT, 50% glycerol buffer at  $\sim 1$  mg/ml concentration.

### **Dynamic light scattering**

Dynamic light scattering (DLS) analysis employed a Protein Solutions DynaPro system using a Microsampler cell and running DYNAMICS software ver 6.2. Proteins were analyzed at concentrations between 1 mg/ml and 5 mg/ml after the samples were centrifuged to remove insoluble particles. All samples were analyzed a minimum of 10 times and the resulting data were analyzed to estimate apparent molecular weight assuming a globular protein in an aqueous solution.

### **Preparation of $^3\text{H}$ -labeled substrate**

Bacteriophage  $\lambda$  genomic DNA (Invitrogen, Carlsbad, CA) was tritiated by incubating DNA with SssI CpG methylase (New England Biolabs) and  $^3\text{H}$ -SAM (GE Healthcare) according to the manufacturer's (New England Biolabs, Ipswich, MA) reaction conditions. After a 3-h incubation, DNA was purified using QIAEX II according to the manufacturer's instructions (Qiagen, Germantown, MD). Radioactive DNA was diluted with unlabeled  $\lambda$  DNA to a final concentration of  $\sim 100$   $\mu\text{g}/\text{ml}$  and specific activity of  $10^5$  c.p.m./ $\mu\text{g}$ . In all cases where ssDNA was used as the substrate, the DNA was first boiled for 10 min followed by rapid chilling in an ice-water bath.

### **Preparation of $^{32}\text{P}$ -labeled substrate**

A 76-base DNA oligonucleotide (5'-CGAGATCGGTTG CGTCTCAGCGTCGTCATCGTCTCATGCGCTACA GCACAGATTCACAATTAAGCTCTGCCATCAG) was 5'-end labeled by treatment for 16 h with T4 polynucleotide kinase (New England Biolabs) using  $\gamma$ - $^{32}\text{P}$ -ATP (MP Biomedicals, Solon, OH) according to the manufacturer's (New England Biolabs) reaction conditions.

### **ExoVII activity assayed by nucleotide solubilization ( $^3\text{H}$ -labeled substrates)**

Reactions contained 100 mM potassium acetate and 20 mM Tris acetate pH 7.9. The concentrations of protein, substrate DNA and magnesium were altered depending on the experiment and as detailed in the figure legends. Reactions were incubated for 20 min at  $80^{\circ}\text{C}$ . After digestions were complete, reactions were quenched by diluting into 10 volumes of ice-cold 10% TCA and incubated on

ice for 15 min followed by centrifugation at  $4^{\circ}\text{C}$  for 20 min to pellet residual substrate DNA and large DNA hydrolysis products. The soluble fractions containing ssDNA fragments less than 15 nt were analyzed by mixing 100  $\mu\text{l}$  of the soluble fraction with 10 ml scintillation cocktail and 300  $\mu\text{l}$  Tris base and then determining the amount of radioactivity released by ExoVII using a Beckman LS 6000 scintillation counter. Under standard reaction conditions with 200 ng of  $^3\text{H}$ -labeled  $\lambda$ -DNA (24 fmol DNA ends) and 2.5 mM magnesium acetate in a total volume of 20  $\mu\text{l}$  at  $80^{\circ}\text{C}$ , the specific activity of *T. maritima* ExoVII was  $\sim 221$  U/mg protein where 1 U corresponds to 1 nmol of nucleotides solubilized in 30 min. This is similar to the previously published activity of *E. coli* ExoVII (267 U/mg) assayed on a partially ssDNA substrate created by exonuclease III treatment of double-stranded DNA (dsDNA) (17).

### **Metal dependence of *T. maritima* ExoVII activity**

The standard reaction condition was employed, except 200 ng of  $^3\text{H}$ -labeled  $\lambda$ -DNA (24 fmol DNA ends) was incubated with 1.5  $\mu\text{M}$  *T. maritima* ExoVII in a total volume of 20  $\mu\text{l}$  at  $80^{\circ}\text{C}$ . Each reaction contained 100 mM potassium acetate, 20 mM Tris acetate pH 7.9 and one of a series of divalent cations ( $\text{Mg}^{2+}$ ,  $\text{Ni}^{2+}$ ,  $\text{Mn}^{2+}$ ,  $\text{Zn}^{2+}$  or  $\text{Co}^{2+}$ ) at 5 mM.

### **Determination of anion sensitivity of *T. maritima* ExoVII activity**

The standard reaction condition was employed, except 200 ng of tritiated  $\lambda$  DNA (24 fmol DNA ends) was incubated with 1.5  $\mu\text{M}$  *T. maritima* ExoVII in a total volume of 20  $\mu\text{l}$  at  $80^{\circ}\text{C}$ . Each reaction contained 100 mM potassium acetate, 20 mM Tris acetate pH 7.9, 5 mM magnesium acetate and between 12.5 and 125 mM of either potassium phosphate or potassium sulfate.

### **Processivity of *T. maritima* ExoVII**

The standard reaction was assembled on ice with 100 mM potassium acetate, 20 mM Tris acetate pH 7.9, 2.5 mM magnesium acetate, 8 fmol DNA ends and 1.5  $\mu\text{M}$  *T. maritima* ExoVII (final concentration). Digestion was initiated by transferring the reaction to  $80^{\circ}\text{C}$ . After allowing reactions to proceed for 60 s, >150-fold unlabeled single-strand  $\lambda$  DNA was added as a trap and the reactions were allowed to proceed. As a mock trap control, 20  $\mu\text{l}$  of EB buffer (Qiagen) was added after 60 s in place of unlabeled DNA.

### **ExoVII activity assayed by PAGE ( $^{32}\text{P}$ -labeled substrates)**

Reactions were assembled on ice with 100 mM potassium acetate, 20 mM Tris acetate pH 7.9 and 2.5 mM magnesium chloride (for *T. maritima* ExoVII) or with the same plus additional 50 mM potassium phosphate and 5 mM EDTA (for *E. coli* ExoVII) and 80 pmol oligonucleotide DNA ends. To these mixtures, either 1.1  $\mu\text{M}$  *E. coli* ExoVII or 10.4  $\mu\text{M}$  *T. maritima* ExoVII (final concentration) were added in a total reaction volume of 10  $\mu\text{l}$  and incubated for 1 min, 5 min or 10 min at the appropriate



temperature to capture intermediates and products (*E. coli* ExoVII, 37°C; *T. maritima* ExoVII, 65°C). Reactions were quenched by addition of 10 µl 2 × formamide loading dye (deionized formamide in Tris Base Boric acid EDTA (TBE) with 0.01% bromophenyl blue and 0.01% xylene cyanol) followed by incubation at 95°C for 10 min. To separate the substrate and product species, 7 µl of each sample were loaded onto denaturing polyacrylamide gels (18% solution at 19:1 polyacrylamide:bis-acrylamide, 7-M urea) containing Tris Taurine EDTA (TTE) buffer. Electrophoresis was performed in 0.5 × TTE buffer for 6 h at 20 V/cm. Gels were fixed for 30 min in a mixture of methanol and acetic acid and then dried for 1.5 h at 55°C before exposing them to a phosphorimager screen (Bio-Rad, Hercules, CA).

### Mutagenesis

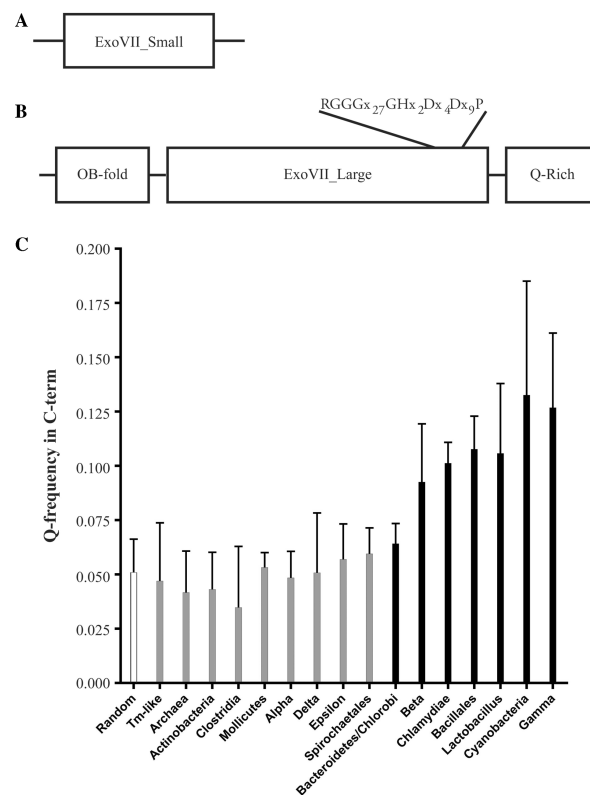
Standard PCR reactions containing 2 U KOD (Novagen), 10 pmol of each primer (5'-GGG ATA GGA CAC GAG ATA GCC CGT GTG ATA GCG GCT TTC GTG GCA GAT GTA TCG-3' and its complement 5'-CGA TAC ATC TGC CAC GAA AGC CGC TAT CAC ACG GGC TAT CTC GTG TCC TAT CCC-3'), 10% DMSO, 1 mM Mg<sup>2+</sup> and 200 µM dNTPs were cycled 18 times using 95°C as the denaturation temperature, 55°C as the annealing temperature and 68°C as the extension temperature. Template DNA was then digested using DpnI (New England Biolabs) to destroy non-mutated template DNA and reactions were transformed into *E. coli* XL10-Gold cells (Stratagene, Cedar Creek, TX). Transformations were verified by DNA sequencing to confirm the presence of mutations. Plasmids that were confirmed to contain the D235A/D240A mutations were then transformed into BL21(DE3) cells (Novagen) for protein expression which was carried out using the same protocol as for wild-type *T. maritima* ExoVII.

## RESULTS

### Genomic distribution of ExoVII

Bioinformatic analysis shows that the *xseA* and *xseB* genes responsible for ExoVII are highly conserved throughout the microbial world. These genes are readily identified in most bacteria and in many archaea but, to date, not in eukaryotes. XseA from *E. coli* and *T. maritima* (TM1768) are remarkably similar with 32% identity (59% similarity) even though the two microbes are not closely related (20). The ExoVII small subunit from *E. coli* (XseB) is also very similar to TM1769 from *T. maritima*, with 47% identity (59% similarity). The domain structure of these two ExoVII subunits is shown as a schematic representation in Figure 1A and B.

Although the small subunit plays an important function in regulating ExoVII activity (32–34), it lacks any recognizable sequence motifs or domains. The structure of the XseB ortholog from *Bordetella pertussis* has been solved, revealing it to be an elongated ‘paper clip’ structure (PDB ID: 1VP7; doi: 10.2210/pdb1vp7/pdb). Several conserved residues are interspersed throughout the length of XseB



**Figure 1.** Bioinformatic analysis of the ExoVII subunits. (A) Domain structure of the small subunit of ExoVII. Analysis shows that XseB homologs look like other XseB homologs with few defining motifs. (B) Domain structure of the large subunit of ExoVII. XseA has two well-defined domains (N-terminal OB-fold and central ExoVII\_Large domain which contains a highly conserved core) and a putative C-terminal region which is highly enriched for glutamine residues. (C) A comparison of the frequency of glutamine occurrence in the CTD domain of XseA from a representative set of bacteria. The Q-rich CTD domain is only present (striped bars) in a subset of bacteria including Bacteroidetes, Chlamydiae, Cyanobacteria, Bacillales, Lactobacillus and  $\beta$ - and  $\gamma$ -Proteobacteria but is absent (gray bars) in Actinobacteria, Clostridia, Mollicutes,  $\alpha$ -,  $\delta$ - and  $\epsilon$ -Proteobacteria, Spirochaetales and Archaea. Interestingly, the Q-rich CTD domain is present in *E. coli* but absent in *T. maritima*.

homologs and these appear to cluster together on one end of this elongated protein (Supplementary Figure S1).

The large subunit, XseA has two distinct domains. The N-terminus of XseA has an oligosaccharide/oligonucleotide binding (OB) domain, which covers approximately the first 100 amino acids. This domain is often found in nucleotide-binding proteins (35), and its presence strongly suggests that DNA recognition and binding is accomplished by the large subunit of ExoVII. The central region of XseA homologs comprises another domain, termed the ExoVII\_Large domain. This domain contains a highly conserved core and is the signature motif of XseA. This core has features evocative of a metal-coordinating hydrolytic active site, but it is exclusively found in XseA homologs and does not share substantial homology with other nucleases (data not shown). The region at the C-terminus of the ExoVII\_Large domain is more variable, both in sequence and length, but is highly enriched in glutamine residues in a number of bacteria. We refer to

this domain as the 'Q-rich C-terminal domain (CTD)' (Figure 1B, Supplementary Figure S1). Analysis of all the XseA homologs from completely sequenced bacterial genomes shows that this C-terminal domain possesses a statistically significant enrichment of glutamine residues and is present in the genomes of Bacteroidetes, Chlamydiae, Cyanobacteria, Bacillales, Lactobacillus and  $\beta$ - and  $\gamma$ -Proteobacteria but absent from the genomes of Actinobacteria, Clostridia, Mollicutes,  $\alpha$ -,  $\delta$ - and  $\epsilon$ -Proteobacteria, Spirochaetales, *T. maritima*-like bacteria and Archaea (data not shown, Figure 1C and Supplementary Figure 1). The unusual charge distribution in the Q-rich CTD suggests that this region of the XseA may be poorly ordered. This hypothesis is supported by analysis of selected CTD sequences using predictors of protein disorder (36).

### Purification and solution properties

Purification of the *T. maritima* ExoVII homolog was facilitated by the 6X-His tag on the N-terminal end of the small subunit, TM1769, and by the thermostability of the enzyme. Heating of the crude lysate to precipitate most *E. coli* proteins, followed by metal affinity chromatography routinely yielded between 5 mg and 10 mg of purified *T. maritima* ExoVII per liter of cells grown in enriched medium. During the final purification step of size exclusion chromatography on a Superdex S-200 column, there were always two major peaks. The first contained functional ExoVII with both subunits, while the second peak contained oligomers of the small subunit with an apparent molecular weight corresponding to a tetrameric form. Based on Coomassie staining, the fraction containing *T. maritima* ExoVII activity has TM1768 and TM1769 subunits in approximately a one to four molar ratio, similar to the ratio of the orthologous subunits reported for *E. coli* ExoVII (32) and in commercial preparations of *E. coli* ExoVII obtained from USB (Figure 2A).

Previous studies have shown that the small subunit serves an important role in ExoVII activity, but the mechanism for this is not well understood (34). Overexpression of the large subunit alone leads to an overall decrease in ExoVII activity in *E. coli* (32). Recent studies have shown that contact of *Neisseria meningitidis* with host cells induces *xseB* expression, leading to a subsequent increase in ExoVII activity. The authors propose that this mechanism is used to regulate the nuclease activity while maintaining constant levels of the XseA subunit (34). Genetic studies have also suggested that the small subunit may serve a regulatory role *in vivo* (33,34). Taken together, these studies suggest that stoichiometry of small to large subunits regulates ExoVII activity, and the ratio of these proteins in cells may be variable and respond to environmental signals. Our *in vitro* activity assays similarly show that addition of exogenous TM1769 to the standard reaction can slightly increase the apparent activity of purified *T. maritima* ExoVII (data not shown), suggesting that the quaternary structure of ExoVII is dynamic and can be driven to the active ratio of 4:1 when the small subunit is in excess. All of the data shown in this article are from

experiments performed in the absence of additional TM1769.

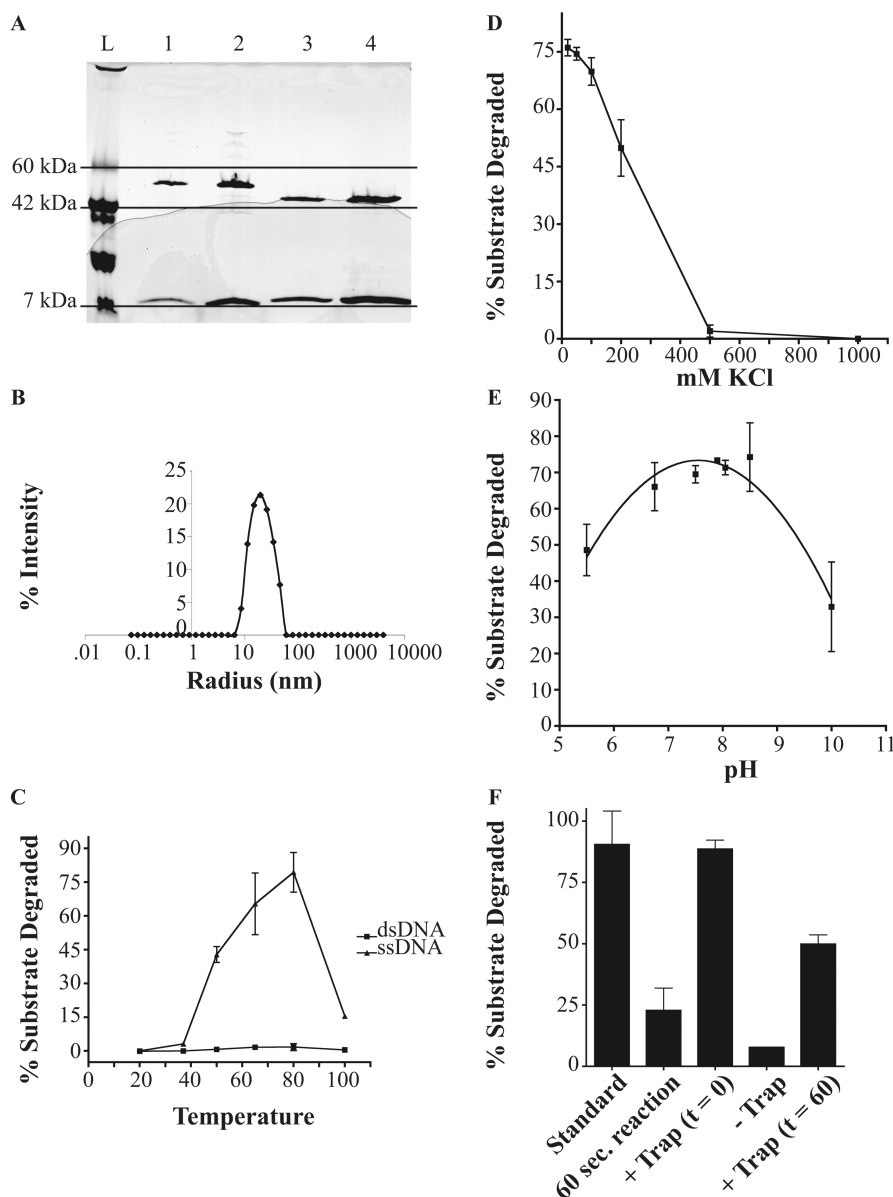
Purified *T. maritima* ExoVII (Figure 2A) was analyzed by DLS (Figure 2B). Time-dependent fluctuations in light scattering were analyzed using regularization analysis functions in the DYNAMICS software package (Wyatt Technology Corp. Santa Barbara, CA), which indicated that the protein has a very large apparent hydrodynamic radius (20 nm), consistent with our gel-filtration data and previous observations that *E. coli* ExoVII is highly elongated (17). This large hydrodynamic radius remained constant over wide salt (0–1 M) and temperature ranges (25–95°C, data not shown) suggesting that the small and large subunits of *T. maritima* ExoVII are in a stable complex. The structure of the XseB homolog from *B. pertussis* also showed an elongated structure (doi: 10.2210/pdb1vp7/pdb). Although no structure is yet available for the ExoVII complex, we anticipate that the ExoVII holoenzyme will also have a very elongated structure.

### *Thermotoga maritima* ExoVII is a ssDNA-specific nuclease

The nuclease activity of *T. maritima* ExoVII was tested, as described in the 'Materials and methods section', by incubating the enzyme with labeled DNA under conditions in which DNA digestion was linear with respect to enzyme concentration. *Thermotoga maritima* ExoVII showed robust degradation of ssDNA with maximal activity at 80°C (Figure 2C), as expected for an ExoVII ortholog from a thermophilic organism. No significant activity was seen on dsDNA (Figure 2C). *Thermotoga maritima* ExoVII activity was maximal at low salt (<250 mM KCl, Figure 2D) and pH 7.9 as reported for *E. coli* ExoVII (17); however, the *T. maritima* ExoVII pH profile was much broader (Figure 2E) than that observed for *E. coli* ExoVII, which has been reported to be very pH sensitive and to lose 50% of its activity when assayed 0.5 pH units above or below its optimal pH (17).

### *Thermotoga maritima* ExoVII is processive

To determine if *T. maritima* ExoVII, like *Ec* ExoVII, was processive, we used unlabeled  $\lambda$  DNA as a trap for any ExoVII that dissociated from the substrate during the reaction. When this trap was added at time  $t = 0$ , no activity was observed indicating that the amount of trap was sufficient to capture all of the enzyme in the reaction (Figure 2F). If, however, the reaction was allowed to proceed for 60 s to allow ExoVII to bind DNA ends before adding the trap, we estimated that approximately 2500 additional nt were released per ExoVII–DNA complex before the enzyme dissociated. This value was obtained by multiplying the acid soluble counts after a time period sufficient for the reaction to reach completion (estimated at  $t = 1200$  s, see Figure 6A) by the specific activity of the substrate (nucleotides per counts per minute), and dividing this by the number of ssDNA ends initially present at  $t = 0$ . The counts at  $t = 1200$  were corrected for the counts released during the initial 60 s. The counts released after 60 s were significantly higher ( $P = 0.03$ ) than the amount released during the first 60 s alone, but significantly lower than what was seen in the absence of



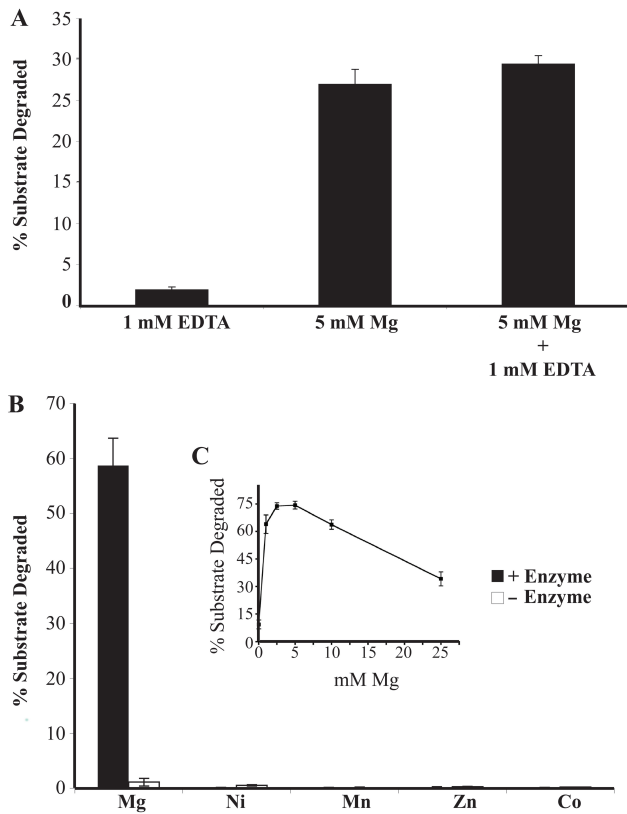
**Figure 2.** Purification and enzymatic properties of *T. maritima* ExoVII. (A) SDS-polyacrylamide gel of 8–25% (PhastGel, GE Biosciences) showing 5 U and 10 U of commercially prepared *E. coli* ExoVII (USB) in lanes 1 and 2, respectively, and 1 and 2  $\mu$ g of purified *T. maritima* ExoVII in lanes 3 and 4, respectively. (B) Hydrodynamic radius distribution of purified *T. maritima* ExoVII derived using dynamic light scattering. Although the apparent radius is very large, especially for the predicted molecular weight, the presence of a single defined peak strongly suggests that ExoVII is highly elongated or may oligomerize, rather than aggregating nonspecifically. The calculated long axis is  $\sim$ 200 Å. (C) Plot showing activity of *T. maritima* ExoVII as a function of temperature on ssDNA (filled triangle) and dsDNA (filled square) substrates. Although some weak activity can be seen on the dsDNA at the higher temperatures, this is probably related to small amounts of thermal breathing. (D) Plot showing activity of *T. maritima* ExoVII as a function of increasing KCl concentration. Activity is completely lost  $>$ 500 mM KCl. (E) Plot showing activity of *T. maritima* ExoVII as a function of pH. Although the same pH optimum is seen with *T. maritima* ExoVII as was reported for *E. coli* ExoVII, the shape of the curve is much broader for the *T. maritima* enzyme. (F) Processivity experiment showing the effects of addition of unlabeled trap DNA at  $t = 60$  s on the reaction (see text for details).

trap ( $P = 0.003$ ) (Figure 2F). These results demonstrate that *T. maritima* ExoVII activity is highly processive.

***Thermotoga maritima* ExoVII requires  $Mg^{2+}$  for activity, and is strongly inhibited by phosphate and sulfate**

Two interesting features of *E. coli* ExoVII are: (i) its activity in the presence of large amounts of EDTA and (ii) the strong stimulation of its activity by phosphate ions.

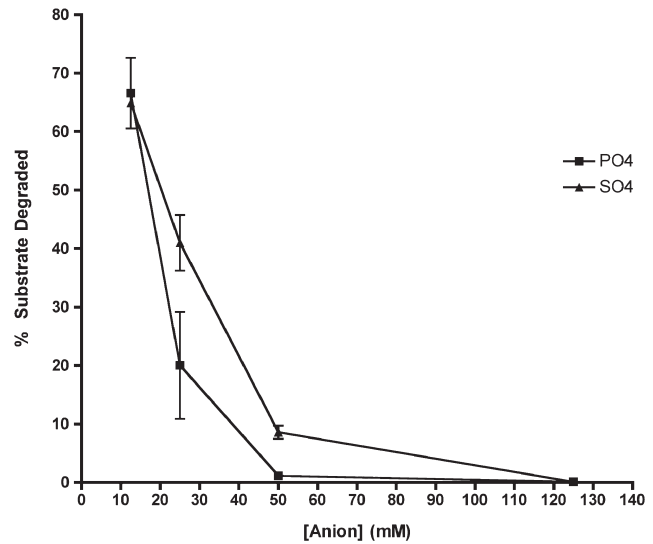
We therefore wanted to determine if the *T. maritima* homolog was also resistant to EDTA and stimulated by phosphate. Surprisingly, the enzyme was completely inhibited by the presence of EDTA (Figure 3A) and was not active in standard *E. coli* ExoVII buffer (data not shown). To determine if *T. maritima* ExoVII requires a metal cofactor, addition of several different divalent cations, including magnesium, manganese, nickel, cobalt and zinc were tested. Strong metal-dependent,



**Figure 3.** Magnesium requirement for *T. maritima* ExoVII activity. (A)  $Mg^{2+}$ ,  $Mn^{2+}$ ,  $Co^{2+}$ ,  $Zn^{2+}$  and  $Ni^{2+}$  were analyzed for their ability to stimulate enzyme-dependent DNA digestion. Percentage substrate degraded after 20 min are shown. (B) Magnesium titration shows optimal activity between 2.5 mM and 5 mM  $Mg^{2+}$ . (C) *Thermotoga maritima* ExoVII is EDTA-sensitive but the activity can be recovered by adding an excess of magnesium.

enzyme-dependent nuclease activity was only observed in the presence of magnesium (Figure 3B). Weak nuclease activity is seen when no  $Mg^{2+}$  is added, suggesting that some magnesium remains associated with the enzyme after purification. This activity can be eliminated by the addition of 1 mM EDTA but is fully recovered if sufficient magnesium is added to saturate EDTA and provide additional magnesium for the enzyme (Figure 3B). Magnesium titrations show that concentration between 2.5 and 5 mM result in optimal rates of digestion (Figure 3C).

*Escherichia coli* ExoVII activity was previously shown to be stimulated by phosphate, with an optimal concentration of 67 mM (17,18). This, coupled with the EDTA resistance of the enzyme, led us to consider the possibility that ExoVII is a phosphorolytic DNase and not a DNA hydrolase. To test this model, we examined the effect of phosphate and sulfate (an anion similar in size, shape and charge to phosphate) on ExoVII activity. Surprisingly, either phosphate or sulfate in the buffer strongly inhibited *T. maritima* ExoVII, with >50% inhibition shown at 25 mM anion (phosphate/sulfate) and almost complete inhibition by 50 mM (Figure 4). In contrast, titration of *E. coli* ExoVII with sulfate or phosphate showed stimulation similar to that previously seen with phosphate



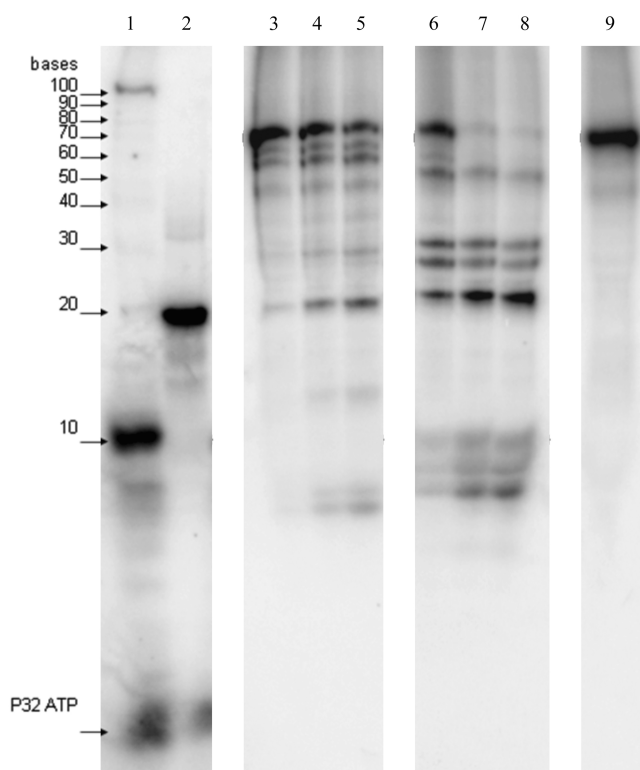
**Figure 4.** Phosphate and sulfate strongly inhibit *T. maritima* ExoVII activity. Effect of phosphate (filled square) and sulfate (filled triangle) anions on *T. maritima* ExoVII activity. It had been previously reported that *E. coli* ExoVII is strongly stimulated by the presence of phosphate in the buffer. We found that for *T. maritima* ExoVII, both phosphate and sulfate strongly inhibit the enzyme activity, with phosphate serving as a better inhibitor. Percentage substrate degraded after 20 min are shown.

(data not shown). Since *E. coli* ExoVII and *T. maritima* ExoVII share significant homology with each other and with the putative ExoVII genes in most microbial genomes, it is highly unlikely that *E. coli* ExoVII and *T. maritima* ExoVII employ different enzyme mechanisms. Furthermore, the stimulation of *E. coli* ExoVII by sulfate as well as phosphate makes it unlikely that phosphate plays a direct role in catalysis. Studies done using acetate showed no stimulation of *E. coli* ExoVII or inhibition of *T. maritima* ExoVII, suggesting that the effects of anion addition are specific to the size and shape of the phosphate and sulfate anions.

#### *Thermotoga maritima* ExoVII cleaves ssDNA to yield oligonucleotide products

*Escherichia coli* ExoVII initiates digestion from either end of a ssDNA molecule and it is this requirement for DNA ends that makes it an exonuclease. Nevertheless, the enzyme introduces endonucleolytic nicks to release oligonucleotide products of various sizes. To determine if *T. maritima* ExoVII shares these features with *E. coli* ExoVII, we analyzed the products of digestion of a synthetic DNA oligonucleotide by denaturing polyacrylamide gel electrophoresis. An example of the results from these experiments (Figure 5) shows that both *E. coli* ExoVII and *T. maritima* ExoVII release oligonucleotide products in a similar manner when digesting ssDNA. The 76-nt substrate was progressively digested to smaller oligonucleotide products over multiple turnovers. The size distribution of early and later intermediates, and of possible limit digest products formed by both enzymes, suggests that ExoVII initiates DNA digestion from both the 3'- and the 5'-ends. Although the accumulation of short





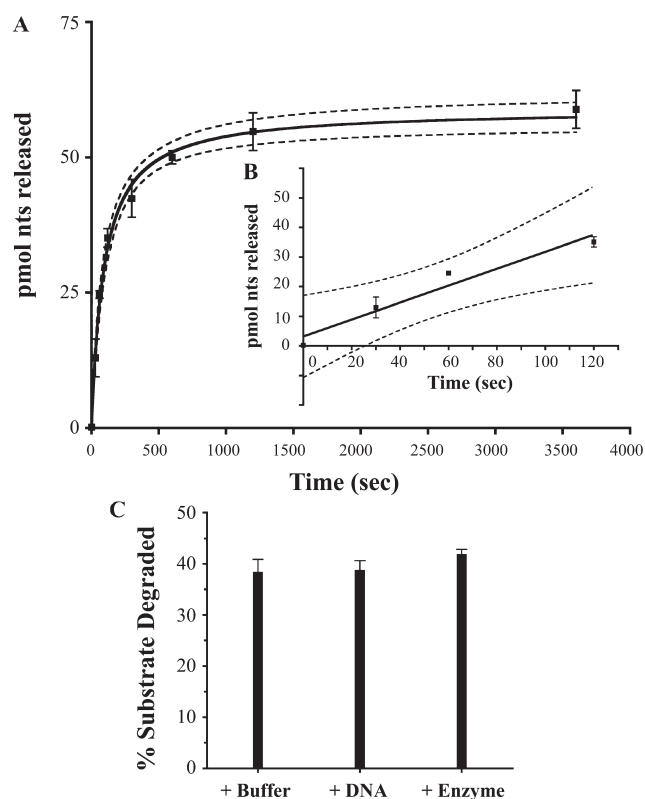
**Figure 5.** *Thermotoga maritima* ExoVII releases oligonucleotide products. A time course experiment to evaluate the length of products arising from digestion of a 5'-<sup>32</sup>P-end labeled 76 nt ssDNA substrate by *E. coli* ExoVII and *T. maritima* ExoVII with time points at 1 min, 5 min and 10 min is shown. Mobility of a 5'-<sup>32</sup>P-end labeled 10-nt ssDNA ladder is indicated with arrows. Labeled DNA species are evident in lanes 1–9. Lane 1: 100 nt, 10-nt markers and  $\gamma$ -<sup>32</sup>P ATP; lane 2: 18-nt marker; lanes 3–5: *E. coli* ExoVII digested sample after 1 min, 5 min and 10 min, respectively; lanes 6–8: *T. maritima* ExoVII digested sample after 1 min, 5 min and 10 min, respectively; lane 9: control lane with 76-nt substrate and no added nuclease. Cleavage products were oligonucleotides of different sizes.

oligonucleotide products retaining a 5'-label solely due to repeated cleavages at the 3'-end cannot be ruled out completely, this is not very likely since we also see accumulation of larger fragments  $\geq 20$  nt. Many of the initial digestion products were species above 15 nt in length, explaining why measurement of ExoVII activity using the acid solubilization protocol described above often yields a lag period with little evident activity, as the initial digestion products are too large to be soluble in trichloroacetic acid. This phenomenon complicates kinetic analysis (discussed subsequently).

A limited number of prominent cleavage species are evident and reproducible. These might reflect sequence specificity of the enzyme or limited access of the enzyme to the relatively short substrate. As the elongated structure and subunit stoichiometries of both enzymes appear quite similar, we propose that there are different DNA sequence specificities for the two enzymes.

#### *Thermotoga maritima* ExoVII kinetics

Under optimized reaction conditions, we performed time course experiments (an example is illustrated in Figure 6A

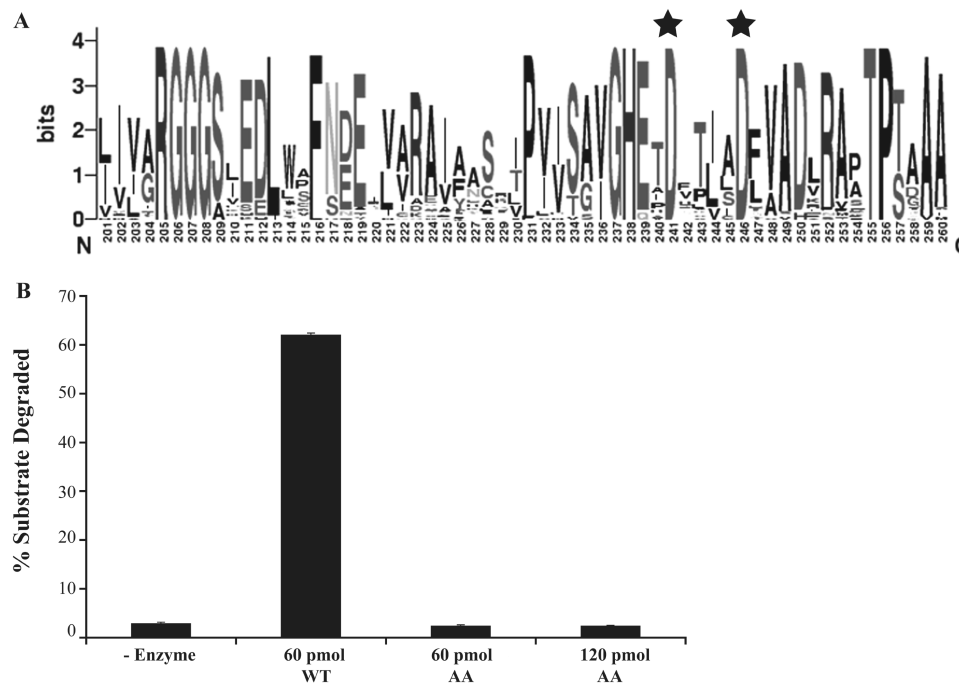


**Figure 6.** Kinetic analysis of *T. maritima* ExoVII. (A) Time-course experiment using 1.5- $\mu$ M *T. maritima* ExoVII with time points at 0 s, 30 s, 60 s, 120 s, 300 s, 600 s, 1200 s and 3600 s. (B) The initial 120 s from part (A) fit to a line. The displayed curve was fit to data averaged from experiments done in triplicate. The 95% confidence levels are shown as dashed lines. (C) To test for enzyme denaturation during the reaction, additional buffer, substrate or fresh enzyme was added after 20 min. The reaction was allowed to proceed for an additional 20 min and the total fraction of substrate digested is shown. In all cases, the same percentage of the total substrate was digested, suggesting that enzyme is still functional after the initial incubation and the reason for the limited digestion is substrate dependent.

and the inset Figure 6B) and determined that the initial rate of DNA hydrolysis by 1.5 pmol of *T. maritima* ExoVII was 286 pmol nt released from ssDNA per second. Unfortunately, determination of the apparent  $K_m$  and  $k_{cat}$  values were complicated by the unique mode of DNA hydrolysis by ExoVII. As mentioned above, *T. maritima* ExoVII binds ssDNA ends and cleaves its substrate endonucleolytically (Figure 5), releasing oligonucleotide products of various sizes, some of which contribute to initial soluble counts and some of which remain large enough to be acid precipitable. Each cleavage event creates new substrate (two new ends per cleavage event) in a manner that cannot be easily quantified, resulting in a steady increase in the number of ExoVII/DNA complexes during the course of the reaction (18).

When using the denatured 48.5-kb genome of  $\lambda$  as a substrate, a variable portion of the DNA was refractory to hydrolysis ( $\sim 40\%$  on average across all experiments). We ruled out that *T. maritima* ExoVII became inactivated during the 20-min time course (Figure 6C) and suggest that the failure to completely digest the substrate was probably a consequence of the reannealing of





**Figure 7.** Active-site determination. (A) Weblogo showing the conservation of residues at the putative active site core in the C-terminal region of XseA. Residues D235 and D240 are the two conserved aspartic acid residues (highlighted by black stars) that were mutated to alanines. (B) Activity assay comparing purified WT and D235A/D240A mutant (labeled AA) *T. maritima* ExoVII. Although the two proteins were purified using the same procedure and exhibited identical solution properties, the D235A/D240A mutant is completely inactive.

complementary ssDNA during the reaction. This might be a function of the tendency of ssDNA to adopt secondary structure, which we expect to be refractory to hydrolysis by the ExoVII family nucleases.

#### Determination of the putative active site of ExoVII

ExoVII is present in most bacterial genomes ( $\approx 90\%$ ) and in several archaea ( $\sim 30\%$ ). This wealth of sequence information was used to align all of the microbial XseA and XseB orthologs to allow the genetic drift arising during divergence of microbes from one another to erode non-conserved residues and highlight highly conserved residues. When the sequences of XseA homologs returned from a BLASTp search of all the completely sequenced microbial genomes were aligned with each other, a conserved amino acid core sequence, marked by the presence of two invariant aspartic acid residues (highlighted with black stars in Figure 7A), emerged. This core had several well-conserved acidic side chains, but the two invariant aspartic acid residues were maintained across all XseA homologs, including *E. coli* ExoVII which is EDTA-resistant (the sequence alignment of this region for 28 representative organisms is shown in Supplementary Figure S2). The presence of two acidic residues close to each other is reminiscent of the active site structures of hydrolytic nucleases in which the acidic residues serve to coordinate a divalent metal cation (37). Codons for the two residues identified, *TmD235* and *TmD240*, were mutated to alanine codons and the mutant protein was overexpressed and purified in a manner analogous to the wild-type protein. The solution properties of the mutant

enzyme derived from gel filtration and DLS analyses were identical to those seen for the wild-type protein (data not shown). The mutant however was completely inactive, even in the presence of magnesium (Figure 7C), indicating that *TmD235* and *TmD240* are critical components of the ExoVII active site.

#### DISCUSSION

Although *E. coli* ExoVII was originally characterized over 30 years ago (17–19,32,38,39), it remains an enigmatic enzyme. It is the only hydrolytic deoxyribonuclease that does not appear to require any divalent cation for activity. Some hydrolytic ribonucleases, such as RNaseA, have no metal requirement but take advantage of the 2'-OH of RNA for catalysis, a liberty not afforded to nucleases that act on DNA. Also contributing to the unusual nature of ExoVII, is its heteropentameric subunit composition, with four copies of the small XseB subunit that has no apparent activity on its own. Generically, a nuclease requires only two activities: (i) substrate recognition and (ii) substrate digestion. Bioinformatic studies strongly suggest that substrate recognition is at least partially modulated by the N-terminal OB-fold present in the large subunit. We have not been able to measure DNA binding by the purified large subunit of *T. maritima* ExoVII as it was insoluble in the absence of the small subunit, but the purified small subunit does not bind DNA while the holoenzyme binds DNA (data not shown). Our work has shown that a putative catalytic center is present in a conserved portion of the ExoVII\_Large motif (41)

towards the C-terminal end of the protein. This does not leave an obvious role for the small subunit. It is possible that this subunit assists in proper orientation of the substrate into the active site or in substrate binding. Alternatively, the small subunit may act as a processivity clamp to topologically link the holoenzyme to the DNA substrate in a manner similar to RecC, which on its own is inactive but, in the RecBC heterodimer, greatly stimulates the nuclease and helicase activity of RecB (42).

The historical description of *E. coli* ExoVII activity was that it was a processive hydrolytic ssDNA-specific nuclease that did not require divalent cations for activity and that was strongly stimulated by the presence of phosphate. The apparent lack of divalent metal ions coupled with phosphate stimulation raised the possibility that ExoVII, long believed to be a hydrolase, was actually a phosphorolase. Our finding that *T. maritima* ExoVII was strongly inhibited by phosphate raised the following possibilities:

- (i) TM1768 and TM1769 are not the genes responsible for ExoVII: although theoretically possible, this explanation is not very likely. *TM1768* and *TM1769* are adjacent to one another in the *T. maritima* genome and share a high overall percent identity with the *E. coli* orthologs (32% between large subunits and 47% between small subunits) which is even higher when the putative active site region present in the C-terminal ExoVII\_Large domain is considered (62% identity). The solution and hydrodynamic properties of *T. maritima* ExoVII are also similar to those of *E. coli* ExoVII. Like *E. coli* ExoVII, TM1768 and TM1769 form a complex that acts on linear ssDNA and releases oligonucleotide products. Both enzymes have similar pH optima and sensitivity to KCl. All together, it is extremely likely that the TM1768/TM1769 complex is the *T. maritima* ExoVII homolog.
- (ii) The *E. coli* and the *T. maritima* homologs function by two different mechanisms: these two enzymes may use different active sites, but constraints placed upon the enzymes by the substrate may cause them to share certain activities while differing in their reaction mechanism. Our data raised the possibility that *E. coli* ExoVII was a phosphorolase (no metal requirement and stimulation by phosphate), while the *T. maritima* ExoVII was a hydrolase ( $Mg^{2+}$  requirement and inhibition by phosphate). This is extremely unlikely. First, *E. coli* ExoVII was stimulated not only by phosphate but also by sulfate. This stimulation by a noncatalytic molecule weakens the phosphorolytic mechanism hypothesis. Second, the putative active site residues that we have identified in TM1768 protein (D235/D240) are maintained not only in *E. coli* XseA (D241/D246) but also in all of the XseA homologs identified, suggesting that the enzymes share a common catalytic site. Finally, it is generally accepted that chemistry is much more difficult to evolve than specificity (43). This alone makes it very unlikely that these two proteins, which share significant homology, would function by two separate mechanisms on the same substrate.
- (iii) Both are hydrolytic nucleases: we believe that ExoVII is a processive hydrolytic ssDNA nuclease with a requirement for magnesium. The conserved putative active site residues resemble a metal coordination core and the stimulation of *E. coli* ExoVII by both phosphate and sulfate weaken the phosphorolytic activity model and point to a hydrolytic mechanism that uses a divalent cation, in this case magnesium. Although our characterization of *T. maritima* ExoVII serves to distinguish it from *E. coli* ExoVII, it also aligns it with traditional hydrolytic nucleases. It remains to be determined if *E. coli* ExoVII contains a tightly bound magnesium that is resistant to chelation by EDTA. It is notable that the *E. coli* ExoVII is stimulated by the addition of magnesium (17).

We were initially concerned about the high ratio of enzyme to substrate in the reactions, but these ratios are consistent with those recommended by USB for use with their purified *E. coli* ExoVII. We believe that the reason for this is because the initial cleavages made by ExoVII release large oligonucleotides that are, themselves, susceptible to binding by additional free ExoVII. In this way, optimal activity, as defined by the rate of formation of acid-soluble products, is seen in the presence of a large excess of enzyme over initial DNA end concentrations, though we anticipate that the requirement for DNA ends for activity limits the stoichiometry to one ExoVII heteropentamer per ssDNA end. Alternatively, perhaps ExoVII requires DNA ends for loading and then passively diffuses to internal sites for endonucleolytic cleavage, thereby making the original ends available to recruit additional ExoVII complexes to the substrate.

To confirm that *T. maritima* ExoVII, like *E. coli* ExoVII, is processive, we determined the effect of adding a cold ssDNA trap at time  $t = 0$  and at  $t = 60$  s. Based on the number of DNA ends present at  $t = 0$ , we estimate that each DNA-protein complex present at  $t = 0$  releases  $\sim 2500$  nt of DNA before disassociating. This is the first report of a direct measurement of processivity by ExoVII and confirms that the enzyme is quite processive, especially when compared to other single-strand nucleases implicated in mismatch repair. For example, RecJ releases 1000 nt per binding event (14), ExoX is distributive (44) and ExoI is only weakly processive, releasing 250–300 nt per binding event (45). While the estimate of 2500 nt per binding event is very reproducible, this value for *T. maritima* ExoVII processivity may be artificially inflated. There are two reasons for this possible over-estimation. First, when the molar ratio of ExoVII to DNA ends is high, the initial products of hydrolysis generate additional substrate molecules for the enzyme, thereby generating additional DNA-protein complexes. Since the first 60 s are carried out with no trap (unlabeled substrate) added, the number of protein-DNA complexes at  $t = 60$  is likely larger than the number at  $t = 0$  s which we used to estimate processivity. Second, processivity is typically defined as the number of turnovers per binding event. In the case of a nuclease that releases individual nucleotides, this number should be the same as the number of nucleotides

released per binding event. With nucleases that release oligonucleotides this is not the case. Furthermore, for nucleases that release products of different sizes, there is no direct relationship between the number of cleavage events and the release of acid-soluble counts since some of the products are large enough to be acid precipitable (18), especially in initial samples withdrawn from reactions. Therefore, it is likely that ExoVII is actually catalyzing fewer than the apparent 2500 nicks per binding event. In either case, ExoVII is clearly able to catalyze multiple turnovers per binding event and, on average, these turnovers lead to the solubilization of ~2500 nt per binding event.

To confirm that *T. maritima* ExoVII, like *E. coli* ExoVII, functions in an endonucleolytic manner, we examined the length of initial products released by the enzyme during digestion of a synthetic ssDNA substrate. We determined that both enzymes produce oligonucleotide products. Our data are also consistent with these enzyme initiating cleavage from both ends, as was previously determined for *E. coli* ExoVII (17,18). Additional experiments with short dsDNA hairpin substrates with defined ssDNA overhangs ends confirmed this conclusion, with the caveat that the high temperatures for activating *T. maritima* ExoVII may have caused these substrates to partially denature, leading to caution in interpreting these results (data not shown).

ExoVII's prevalence in bacterial genomes makes extensive bioinformatic analysis easy. When all the XseA homologs were aligned and phylogenetically analyzed two patterns became evident. First, there is a conserved core in the C-terminal end of the protein, which is reminiscent of metal coordinating active sites in other nucleases. The region identified has the motif, RGGG(x)<sub>n</sub>GHxxDxxxxD, and is conserved among all XseA homologs. Second, the XseA homologs segregate themselves into two groups: an *E. coli*-like group and a *T. maritima*-like group, based on the amino acid composition of the C-terminal domain. We postulate that all homologs that are *E. coli*-like are likely to be EDTA-resistant, whereas all of those that are *T. maritima*-like will be EDTA-sensitive. It is likely, however, that all ExoVII homologs require divalent cations for activity since they all share the same highly conserved putative metal coordinating active site. Although previous studies of *E. coli* ExoVII showed activity in the presence of high levels of EDTA, we believe that this is not due to a novel metal-independent mechanism, but reflects the EDTA-resistant nature of the *E. coli* enzyme. The stimulation of *E. coli* ExoVII by magnesium (17) and the extremely low recovery of enzyme activity after denaturation and refolding of *E. coli* ExoVII in the presence of EDTA (32) are consistent with this view. The EDTA resistance of *E. coli* ExoVII remains a striking phenomenon and industrially useful. Because of the apparent discrepancy in the requirement for magnesium between the *E. coli* and *T. maritima* ExoVII orthologs, we analyzed the amino acid sequences for differences that could explain the EDTA sensitivity. We have identified a glutamine rich CTD that may be responsible for the differential tight binding of magnesium

and propose that this Q-rich CTD may have a role in EDTA resistance.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENTS

The CoEx\_TM1768/TM1769 construct was a generous donation from Michael DiDonato at the Joint Center for Structural Genomics (JCSG). We would like to thank Murray Deutscher and Michael A. Zundel for technical assistance and use of facilities. We also thank Bo Chen for purifying the *T. maritima* ExoVII preparation used in some of the experiments.

## FUNDING

National Institutes of Health (F31 GM070395 to A.A.L., R01-GM69972 to A.M.); the National Institutes of Health-sponsored Developmental Center for AIDS Research (to R.M.).

*Conflict of interest statement.* None declared.

## REFERENCES

- Kadyrov,F.A., Dzantiev,L., Constantin,N. and Modrich,P. (2006) Endonucleolytic function of MutLalpha in human mismatch repair. *Cell*, **126**, 297–308.
- Modrich,P. (2006) Mechanisms in eukaryotic mismatch repair. *J. Biol. Chem.*, **281**, 30305–30309.
- Narayan,S. and Roy,D. (2003) Role of APC and DNA mismatch repair genes in the development of colorectal cancers. *Mol. Cancer*, **2**, 41.
- Plotz,G., Zeuzem,S. and Raedle,J. (2006) DNA mismatch repair and Lynch syndrome. *J. Mol. Histol.*, **37**, 271–283.
- Jeter,J.M., Kohlmann,W. and Gruber,S.B. (2006) Genetics of colorectal cancer. *Oncology*, **20**, 269–276; Discussion 285–286, 288–289.
- Modrich,P. (1989) Methyl-directed DNA mismatch correction. *J. Biol. Chem.*, **264**, 6597–6600.
- Grilley,M., Holmes,J., Yashar,B. and Modrich,P. (1990) Mechanisms of DNA-mismatch correction. *Mutat. Res.*, **236**, 253–267.
- Viswanathan,M., Burdett,V., Baitinger,C., Modrich,P. and Lovett,S.T. (2001) Redundant exonuclease involvement in *Escherichia coli* methyl-directed mismatch repair. *J. Biol. Chem.*, **276**, 31053–31058.
- Cooper,D.L., Lahue,R.S. and Modrich,P. (1993) Methyl-directed mismatch repair is bidirectional. *J. Biol. Chem.*, **268**, 11823–11829.
- Ukita,T. and Ikeda,H. (1996) Role of the *recJ* gene product in UV-induced illegitimate recombination at the hotspot. *J. Bacteriol.*, **178**, 2362–2367.
- Viswanathan,M., Dower,K.W. and Lovett,S.T. (1998) Identification of a potent DNase activity associated with RNase T of *Escherichia coli*. *J. Biol. Chem.*, **273**, 35126–35131.
- Viswanathan,M., Lanjuin,A. and Lovett,S.T. (1999) Identification of RNase T as a high-copy suppressor of the UV sensitivity associated with single-strand DNA exonuclease deficiency in *Escherichia coli*. *Genetics*, **151**, 929–934.
- Zahradka,K., Simic,S., Buljubasic,M., Petranovic,M., Dermic,D. and Zahradka,D. (2006) *sbcb15* And *AsbcB* mutations activate two types of recf recombination pathways in *Escherichia coli*. *J. Bacteriol.*, **188**, 7562–7571.



14. Han, E.S., Cooper, D.L., Persky, N.S., Sutera, V.A., Whitaker, R.D., Montello, M.L. and Lovett, S.T. (2006) RecJ exonuclease: substrates, products and interaction with SSB. *Nucleic Acids Res.*, **34**, 1084–1091.
15. Burdett, V., Baitinger, C., Viswanathan, M., Lovett, S.T. and Modrich, P. (2001) *In vivo* requirement for RecJ, ExoVII, ExoI, and ExoX in methyl-directed mismatch repair. *Proc. Natl Acad. Sci. USA*, **98**, 6765–6770.
16. Hersh, M.N., Morales, L.D., Ross, K.J. and Rosenberg, S.M. (2006) Single-strand-specific exonucleases prevent frameshift mutagenesis by suppressing SOS induction and the action of DinB/DNA polymerase IV in growing cells. *J. Bacteriol.*, **188**, 2336–2342.
17. Chase, J.W. and Richardson, C.C. (1974) Exonuclease VII of *Escherichia coli*. Purification and properties. *J. Biol. Chem.*, **249**, 4545–4552.
18. Chase, J.W. and Richardson, C.C. (1974) Exonuclease VII of *Escherichia coli*. Mechanism of action. *J. Biol. Chem.*, **249**, 4553–4561.
19. Chase, J.W., Rabin, B.A., Murphy, J.B., Stone, K.L. and Williams, K.R. (1986) *Escherichia coli* exonuclease VII. Cloning and sequencing of the gene encoding the large subunit (*xseA*). *J. Biol. Chem.*, **261**, 14929–14935.
20. Nelson, K.E., Clayton, R.A., Gill, S.R., Gwinn, M.L., Dodson, R.J., Haft, D.H., Hickey, E.K., Peterson, J.D., Nelson, W.C., Ketchum, K.A. *et al.* (1999) Evidence for lateral gene transfer between Archaea and bacteria from genome sequence of *Thermotoga maritima*. *Nature*, **399**, 323–329.
21. Nelson, K.E., Eisen, J.A. and Fraser, C.M. (2001) Genome of *Thermotoga maritima* MSB8. *Methods Enzymol.*, **330**, 169–180.
22. Edgar, R.C. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. *BMC Bioinform.*, **5**, 113.
23. Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.
24. Higgins, D.G. and Sharp, P.M. (1988) CLUSTAL: a package for performing multiple sequence alignment on a microcomputer. *Gene*, **73**, 237–244.
25. Higgins, D.G., Thompson, J.D. and Gibson, T.J. (1996) Using CLUSTAL for multiple sequence alignments. *Methods Enzymol.*, **266**, 383–402.
26. Aiyar, A. (2000) The use of CLUSTAL W and CLUSTAL X for multiple sequence alignment. *Methods Mol. Biol.*, **132**, 221–241.
27. Livingstone, C.D. and Barton, G.J. (1993) Protein sequence alignments: a strategy for the hierarchical analysis of residue conservation. *Comput. Appl. Biosci.*, **9**, 745–756.
28. Pei, J. and Grishin, N.V. (2001) AL2CO: calculation of positional conservation in a protein sequence alignment. *Bioinformatics*, **17**, 700–712.
29. Pei, J., Dokholyan, N.V., Shakhnovich, E.I. and Grishin, N.V. (2003) Using protein design for homology detection and active site searches. *Proc. Natl Acad. Sci. USA*, **100**, 11361–11366.
30. Pei, J. and Grishin, N.V. (2003) Peptidase family U34 belongs to the superfamily of N-terminal nucleophile hydrolases. *Protein Sci.*, **12**, 1131–1135.
31. Pei, J. and Grishin, N.V. (2006) MUMMALS: multiple sequence alignment improved by using hidden Markov models with local structural information. *Nucleic Acids Res.*, **34**, 4364–4374.
32. Vales, L.D., Rabin, B.A. and Chase, J.W. (1982) Subunit structure of *Escherichia coli* exonuclease VII. *J. Biol. Chem.*, **257**, 8799–8805.
33. Stapleton, M.R., Norte, V.A., Read, R.C. and Green, J. (2002) Interaction of the *Salmonella typhimurium* transcription and virulence factor SlyA with target DNA and identification of members of the SlyA regulon. *J. Biol. Chem.*, **277**, 17630–17637.
34. Morelle, S., Carbonnelle, E., Matic, I. and Nassif, X. (2005) Contact with host cells induces a DNA repair system in pathogenic *Neisseriae*. *Mol. Microbiol.*, **55**, 853–861.
35. Theobald, D.L., Mitton-Fry, R.M. and Wuttke, D.S. (2003) Nucleic acid recognition by OB-fold proteins. *Annu. Rev. Biophys. Biomol. Struct.*, **32**, 115–133.
36. Yang, Z.R., Thomson, R., McNeil, P. and Esnouf, R.M. (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics*, **21**, 3369–3376.
37. Nishino, T. and Morikawa, K. (2002) Structure and function of nucleases in DNA repair: shape, grip and blade of the DNA scissors. *Oncogene*, **21**, 9022–9032.
38. Chase, J.W. and Richardson, C.C. (1977) *Escherichia coli* mutants deficient in exonuclease VII. *J. Bacteriol.*, **129**, 934–947.
39. Vales, L.D., Rabin, B.A. and Chase, J.W. (1983) Isolation and preliminary characterization of *Escherichia coli* mutants deficient in exonuclease VII. *J. Bacteriol.*, **155**, 1116–1122.
40. Marchler-Bauer, A., Anderson, J.B., Derbyshire, M.K., DeWeese-Scott, C., Gonzales, N.R., Gwadz, M., Hao, L., He, S., Hurwitz, D.I., Jackson, J.D. *et al.* (2007) CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res.*, **35**, D237–D240.
41. Phillips, R.J., Hickleton, D.C., Boehmer, P.E. and Emmerson, P.T. (1997) The RecB protein of *Escherichia coli* translocates along single-stranded DNA in the 3' to 5' direction: a proposed ratchet mechanism. *Mol. Gen. Genet.*, **254**, 319–329.
42. Petsko, G.A. (2000) Enzyme evolution. Design by necessity. *Nature*, **403**, 606–607.
43. Viswanathan, M. and Lovett, S.T. (1999) Exonuclease X of *Escherichia coli*. A novel 3'-5' DNase and Dnaq superfamily member involved in DNA repair. *J. Biol. Chem.*, **274**, 30094–30100.
44. Thomas, K.R. and Olivera, B.M. (1978) Processivity of DNA exonucleases. *J. Biol. Chem.*, **253**, 424–429.