Published online 22 April 2013 *Nucleic Acids Research, 2013, Vol. 41, No. 11* ***5887–5897***
doi:10.1093/nar/gkt278

Crystal structures of *Escherichia coli* exonuclease I in complex with single-stranded DNA provide insights into the mechanism of processive digestion

Sai Krishna C. Korada, Trevor D. Johns, Christopher E. Smith, Nathan D. Jones, Kimberly A. McCabe and Charles E. Bell*

Department of Molecular and Cellular Biochemistry, The Ohio State University, 1645 Neil Avenue, Columbus OH 43210, USA

Received February 21, 2013; Revised March 26, 2013; Accepted March 27, 2013

ABSTRACT

Escherichia coli **Exonuclease I (ExoI) digests single-stranded DNA (ssDNA) in the 3′-5′ direction in a highly processive manner. The crystal structure of ExoI, determined previously in the absence of DNA, revealed a C-shaped molecule with three domains that form a central positively charged groove. The active site is at the bottom of the groove, while an extended loop, proposed to encircle the DNA, crosses over the groove. Here, we present crystal structures of ExoI in complex with four different ssDNA substrates. The structures all have the ssDNA bound in essentially the predicted manner, with the 3′-end in the active site and the downstream end under the crossover loop. The central nucleotides of the DNA form a prominent bulge that contacts the SH3-like domain, while the nucleotides at the downstream end of the DNA form extensive interactions with an 'anchor' site. Seven of the complexes are similar to one another, but one has the ssDNA bound in a distinct conformation. The highest-resolution structure, determined at 1.95 Å, reveals an Mg^{2+} ion bound to the scissile phosphate in a position corresponding to Mg^{B} in related two-metal nucleases. The structures provide new insights into the mechanism of processive digestion that will be discussed.**

INTRODUCTION

Enzymes that processively digest single-stranded DNA (ssDNA) and RNA play important roles in genome maintenance and gene regulation (1,2). While much is known about the chemistry of the nuclease reaction (3), it is less clear how processive nucleases are able to use the energy from phosphodiester bond cleavage to motor along their track as they digest it. Exonuclease I from *Escherichia coli* (ExoI; M_r 54.5 kDa; 475 amino acids) is an Mg^{2+}-dependent enzyme that rapidly digests ssDNA in the 3′-5′ direction to yield 5′-mononucleotides (4). It is involved in a number of different processes related to DNA repair and recombination, most notably methyl-directed mismatch repair, where it is one of three enzymes that can digest the 3′-ended strand containing the mismatched nucleotide (5). The enzyme digests ssDNA at a rate of ~275 nucleotides per second with a high degree of processivity (6,7). ExoI forms a protein–protein interaction with *E. coli* ssDNA-binding protein (SSB), which stimulates its activity by loading it onto ssDNA substrates and unwinding secondary structures (8,9). The potency and processivity of ExoI, combined with an unusual resistance to high salt, have made it attractive for use in new biotechnology applications such as nanopore DNA sequencing (10,11).

The crystal structure of ExoI (12,13) revealed a C-shaped molecule of three domains: an N-terminal nuclease domain (residues 1–201) with homology to the proofreading domain of *E. coli* DNA polymerase I and other DnaQ superfamily enzymes (14), a central domain with a portion that resembles an SH3 domain fold (residues 202–354) and a C-terminal α-helical domain (residues 359–475). Running down the center of the molecule is a positively charged groove with the nuclease active site at the bottom end and a loop connecting the SH3-like and C-terminal domains crossing over the top. The groove is long enough to accommodate ~12–13 nucleotides of ssDNA, which is consistent with footprinting data based on quantifying the distribution of released products (6,15). Based on this structure, a model for the ExoI–ssDNA complex, and a mechanism for processivity involving topological threading of the ssDNA under the crossover loop, was proposed (12).

We have determined crystal structures of ExoI in complex with four different ssDNA substrates at resolutions ranging from 1.95 to 3.5 Å. Each structure has

*To whom correspondence should be addressed. Tel: +1 614 688 3115; Fax: +1 614 292 4118; Email: bell.489@osu.edu

two ExoI–ssDNA complexes in the asymmetric unit, giving a total of eight independently determined complexes. In all of the complexes, the ssDNA binds in essentially the predicted manner, with the 3′-end in the active site and the downstream end under the crossover loop. While seven of the complexes have the ssDNA bound in a similar conformation, one of the complexes reveals a distinct binding mode. The possible relevance of these structures to the mechanism of processivity is discussed.

MATERIALS AND METHODS

Protein expression and purification

The gene encoding ExoI was amplified from *E. coli* genomic DNA and cloned into pET14b using the NdeI and BamHI restriction sites. This vector expresses the protein with an N-terminal 6xHis tag followed by a thrombin cleavage site. The resulting plasmid was transformed into *E. coli* BL21(AI) cells (Invitrogen), grown in LB medium to OD of 0.5, induced by addition of 0.2% arabinose with shaking at 37°C for 4 h, and harvested by centrifugation for 10 min at 10 000*g*. Cells were resuspended in Buffer A (50 mM NaH_2PO_4, 300 mM NaCl, 10 mM imidazole, pH 8.0) and frozen at −80°C. Frozen cells were thawed to 4°C, incubated with 1 mg/ml lysozyme, 1 mM PMSF, 1 mM pepstatin and 1 mM leupeptin, lysed by sonication (3 × 2 min at full power using a Branson sonifier 450) and centrifuged twice for 30 min at 48 000*g*. The clarified supernatent was loaded onto a 10 ml column of nickel-charged chelating-sepharose fast flow (GE Healthcare), washed with Buffer A containing 30 mM imidazole and eluted in fractions with a gradient from 30 to 500 mM imidazole in Buffer A. Pooled fractions were incubated with 1000 U of thrombin (GE Healthcare) and dialyzed at 22°C into 150 mM NaCl, 133 mM NaH_2PO_4, pH 7.4. The cleaved mixture was loaded back onto the nickel column, and flow through fractions containing untagged ExoI was dialyzed into 20 mM Tris, pH 8.0, loaded onto a HiTrap QHP column (GE Healthcare) and eluted with a gradient from 0 to 1 M NaCl. Pooled fractions containing ExoI were dialyzed into 20 mM Tris, 1 mM dithiothreitol (DTT), pH 8.0, concentrated to 26 mg/ml using a VivaSpin 20 10K MWCO PES filtration device and stored in small aliquots at −80°C. The purified protein contains an extra N-terminal Gly-Ser-His (GSH) sequence from the expression vector. Protein concentrations were measured by OD at 280 nm using an extinction coefficient calculated from the amino acid sequence. The H181A mutant of ExoI was constructed using the QuikChange method (Stratagene), and the protein was expressed and purified as described above for wild-type ExoI.

Exonuclease reactions

Purified ExoI was incubated at 37°C with a 5′-fluorescein (FAM)-labeled 48-mer oligonucleotide that was purchased high performance liquid chromatography (HPLC)-purified from Integrated DNA Technologies (IDT), and the digestion products were monitored by denaturing polyacrylamide gel electrophoresis. Reactions (25 µl) contained 0.5 µM ExoI (WT or H181A), 1 µM 5′-FAM-oligo, 10 mM HEPES, pH 8.5, 1 mM DTT, 0–2 mM $MgCl_2$ or $CaCl_2$ and 0–10 mM EDTA as indicated. Aliquots of 10 µl were removed from each reaction at the indicated timepoints, quenched by addition of EDTA to 65 mM final concentration and loaded onto a 22% polyacrylamide gel containing 7 M urea in Tris/borate/EDTA (TBE) buffer. Bands were visualized using a UV transilluminator.

Crystallization and x-ray structure determination

Oligonucleotides used for crystallization (5′-Cy5-dT13, 5′-Cy5-dA13, dA16 and dT13) were purchased from IDT. 5′-Cy5 labeled oligonucleotides were purified by reversed phase HPLC (IDT). For crystallization by hanging drop vapor diffusion, 10 mg/ml ExoI in 20 mM Tris, 1 mM DTT, 10 mM EDTA, pH 8.0, was mixed with a 1.2 molar excess of oligonucleotide. The reservoir solution consisted of 0.9–1.5 M ammonium sulfate, 3.75–6.0% 2-propanol and 25% glycerol, and the hanging drop was prepared by mixing 2 µl of ExoI–ssDNA complex with 2 µl of reservoir solution. Crystals typically grew within 1 week. For x-ray data collection, a single crystal was mounted in a nylon loop (Hampton Research) and plunged in liquid nitrogen. For crystals of ExoI in complex with Cy5-dT13, x-ray diffraction data were collected using a Rigaku RU-H3R rotating anode generator, R-AXIS IV++ image plate detector and integrated and scaled with CrystalClear (d*Trek) software (Molecular Structure Corporation). For crystals of ExoI in complex with Cy5-dA13, dA16 and dT13, x-ray diffraction data were collected at beamline 31-ID of the Advanced Photon Source (λ = 0.97931 Å). Synchrotron data were integrated and scaled using MOSFLM and SCALA of the CCP4 suite (16). All data were further processed using TRUNCATE. Crystals with Cy5-dT13, Cy5-dA13 and dA16 belong to space group P4(3) with two molecules per asymmetric unit, and the diffraction data were twinned (twin operator—h, k, −l; twin fraction 0.20–0.33). Crystals with dT13 belong to space group P3(1)21 with two molecules per asymmetric unit with no twinning. All structures were determined by molecular replacement using the Auto Mol Rep function of CCP4 and the structure of uncomplexed ExoI as a search model (PDB code 1FXX). Structures were further refined using REFMAC5, and the models were fit using COOT (17). Refinement of the dA16, Cy5-dA13 and dT13 structures at resolutions of 3.0–3.5 Å included non-crystallographic symmetry (NCS) restraints. Amplitude-based twin refinement in REFMAC5 was used for the Cy5-dT13, Cy5-dA13 and dA16 structures. Diffraction data and refinement statistics are presented in Table 1. Solvent accessible surface area calculations were performed in CCP4 with a probe radius of 1.4 Å. Structural figures were prepared using PyMOL (The PyMOL Molecular Graphics System, Schrödinger, LLC) and COOT (17). The structure factors and coordinates for all four crystal structures have been deposited in the Protein

Table 1. Data collection and refinement statistics[a]

	Cy5-dT13	Cy5-dA13	dA16	dT13
Data collection				
Space group	P4(3)	P4(3)	P4(3)	P3(1)21
Cell dimensions				
a, *b*, *c* (Å)	91.3, 91.3, 159.7	91.3, 91.3, 162.1	91.4, 91.4, 162.5	158.3, 158.3, 151.1
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 120
Resolution (Å)	36.6–1.95 (2.02–1.95)	20.0–3.00 (3.00–3.16)	40.6–3.10 (3.10–3.27)	49.0–3.50 (3.50–3.69)
R_{merge}	0.078 (0.363)	0.065 (0.766)	0.066 (0.813)	0.100 (0.786)
Mean $I/\sigma I$	8.4 (2.0)	16.4 (2.8)	22.4 (2.9)	9.6 (2.0)
Completeness (%)	99.2 (93.9)	99.5 (100.0)	99.9 (100.0)	99.9 (100.0)
Redundancy	3.0 (1.7)	7.4 (7.7)	7.6 (7.6)	11.0 (10.1)
Refinement				
Resolution (Å)	36.6–1.95	20.0–3.00	40.6–3.10	49.1–3.50
No. reflections	89 174	25 103	22 917	26 653
$R_{work/}$ R_{free}	0.213/0.259	0.218/0.245	0.225/0.257	0.306/0.346
No. atoms				
Protein	7392	7020	7098	7431
DNA	594	548	611	420
Ligand/ion	82	45	40	10
Water	404	0	0	0
B-factors (Å^2)				
Protein	42.6	106.1	113.9	198.9
DNA	54.7	119.4	129.7	193.9
Ligand/ion	75.8	119.7	128.3	218.1
Water	41.5			
R.m.s. deviations				
Bond lengths (Å)	0.007	0.004	0.004	0.005
Bond angles (°)	1.40	0.741	0.714	0.849

[a]Each structure was refined using data collected from a single crystal. Numbers in parenthesis are calculated for the highest resolution shell only.

Data Bank. The PDB accession numbers are 4JRP (Cy5-dT13), 4JRQ (Cy5-dA13), 4JS4 (dA16) and 4JS5 (dT13).

Analytical ultracentrifugation

Sedimentation velocity experiments were performed using a Beckman XL-I analytical ultracentrifuge. Purified ExoI protein at 0.5 mg/ml, alone or mixed with an equimolar amount of a dT16 oligonucleotide, was diluted to 400 µl in 20 mM Tris, pH 8.0, 150 mM NaCl, and spun at 20°C for 4 h at 50 000 rpm. Data from interference optics were analyzed using the c(s) and c(M) model in the program Sedfit (18) to determine differential sedimentation coefficients.

RESULTS

Crystallization strategy and structural overview

We will first describe the structure of ExoI in complex with a 5′-Cy5 labeled dT13 oligonucleotide, as it was determined to significantly higher resolution (1.95 Å) than the others (Table 1). Initial attempts to crystallize ExoI in complex with various oligonucleotides used Ca^{2+} or an H181A variant to prevent nuclease activity, but this resulted in crystals that did not contain DNA, based on inspection of electron density maps. Using a gel-based assay, residual nuclease activity was observed under these conditions (Supplementary Figure S1), indicating that the ssDNA added to the crystallization mixture was likely being digested. However, addition of 10 mM EDTA with no added metal completely inhibited ExoI activity, and so crystallization screens were performed in the presence of 10 mM EDTA. The 5′-Cy5 labeled dT13 oligonucleotide was used so that crystals containing bound ssDNA would appear blue, and thus be readily distinguished from crystals without DNA. Crystallization screens yielded a single blue hit for a tetragonal crystal form grown in high salt that diffracted to 1.95 Å on a rotating anode X-ray source. Molecular replacement resulted in solutions for two monomers of ExoI in the asymmetric unit, and after initial refinement electron density maps allowed for placement of all 13 nucleotides of each Cy5-dT13 oligo, including the Cy5 groups (Figure 1).

The two ExoI-dT13 complexes in the asymmetric unit form a dimer with ~2-fold symmetry that buries 1140 Å^2 of total solvent accessible surface area through contacts involving primarily the SH3-like domains (Supplementary Figure S2). The dimer is also stabilized by interactions involving the nucleotides at the 5′-end of each bound ssDNA, including the 5′-Cy5 groups. As described below, the same dimer is formed in crystals with non-labeled oligos, indicating that the Cy5 groups are not required for dimerization. Moreover, the complex with unlabeled dT13 oligo, which crystallized in a different space group ($P3_121$ instead of $P4_3$), contained the same dimer, suggesting that the dimer is ubiquitous. Initial reports on ExoI purification indicated the possibility of a dimer (19), but gel filtration of more highly purified ExoI protein indicated that it is a monomer in solution (20). A more recent observation that the activity of ExoI at higher concentrations on a particular SSB-ssDNA substrate was cooperative, again suggested the possibility of

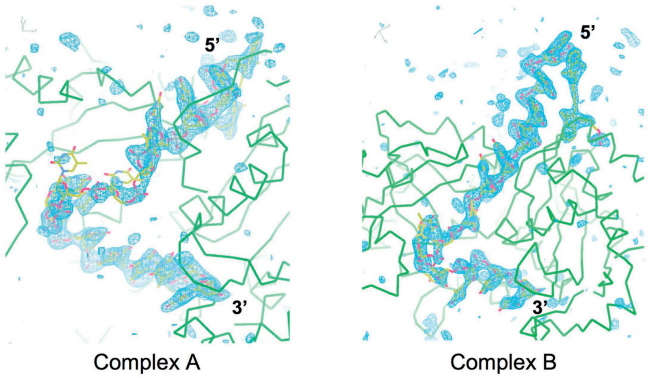

Figure 1. Electron density for the DNA in the Cy5–dT13 complex. The blue cage shows an F_o-F_c map calculated with model phases after 40 cycles of restrained refinement in REFMAC5 with the ssDNA from complex A (left) or complex B (right) omitted. Each map is contoured at +3.0 sigma and shown over the entire complex. The Cy5 groups at the 5′-end of each DNA are evident at the top of the figure. The 3′-end of each DNA binds to the active site at the bottom.

ExoI multimers, perhaps induced by DNA binding (9). To test if DNA binding induces ExoI dimerization directly, sedimentation velocity was performed on ExoI alone and in the presence of a dT16 oligonucleotide. The results clearly indicate a monomer in both cases, and hence no change in oligomeric state on DNA binding (Supplementary Figure S3). Moreover, the amino acid residues at the interface of the dimer observed in the crystal are not conserved in ExoI orthologs (Supplementary Figure S4). Based on the available evidence, we consider it unlikely that the ExoI dimer observed in the crystal is biologically relevant.

Rather remarkably, the two complexes of ExoI in the Cy5-dT13 crystal, which we will refer to as complexes A and B, have the ssDNA bound in two distinct conformations (Figure 2). Both complexes have the ssDNA threaded under the crossover loop, but in complex A, all 13 nucleotides of the ssDNA have entered the DNA-binding groove, whereas in complex B, only the first 10 nucleotides, counting from the 3′-end, have entered the groove. At the bottom of the groove, the terminal 3′-nucleotide is bound to approximately the same position within the active site of each complex. At the top of the groove, three nucleotides in each complex are bound by an extensive and equivalent set of interactions to a site on the protein that we will call the 'anchor' site. However, due to the difference in registration, in complex A, it is nucleotides 11–13 of the ssDNA that are bound to the anchor site, whereas in complex B, it is nucleotides 8–10. The region of the ssDNA between the active site and the anchor site, which we will call the 'bulge', is larger for complex A than for complex B (nucleotides 6–10 versus 5–7), but both bulge regions make a similar contact to the SH3-like domain at the left side of the DNA-binding groove, as viewed in Figure 2. The interactions at the anchor site, the active site and the bulge regions of complexes A and B, will be described in more detail below.

Based on a structural superposition (Figure 3 and Supplementary Table S1), complexes A and B differ

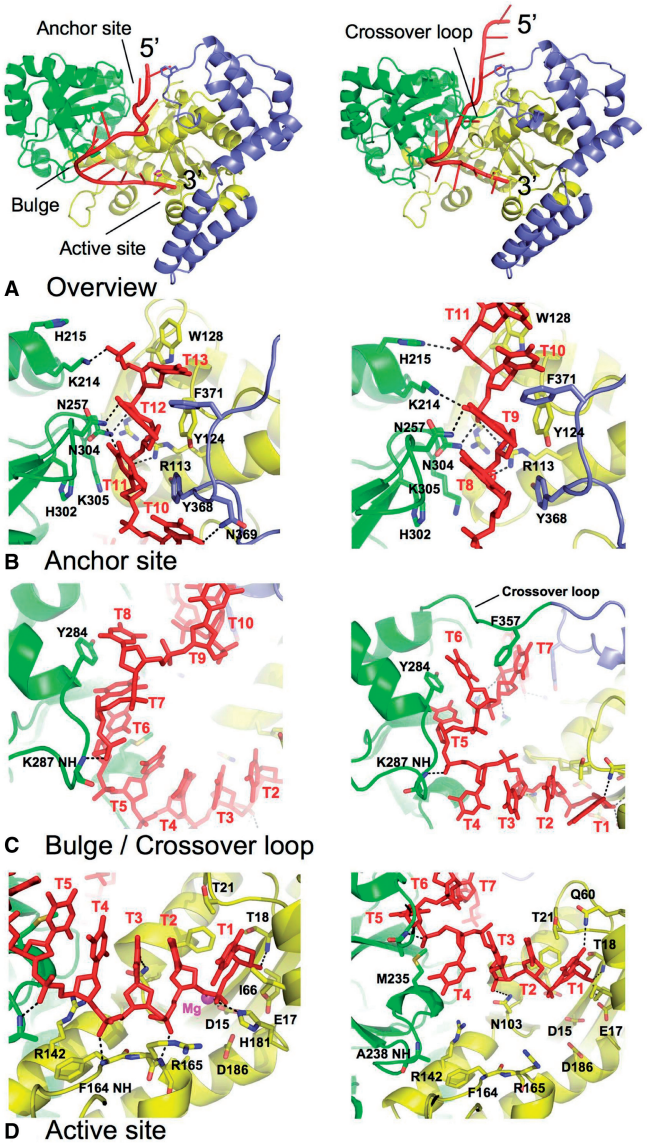

Figure 2. Two distinct complexes of ExoI bound to Cy5-dT13. (**A**) Structural overview. Complexes referred to as 'A' and 'B' in the text are shown in each panel on the left and right, respectively. The Cy5-dT13 oligo is shown in red with the 5′-Cy5 omitted. ExoI is colored by domain: exonuclease domain, yellow (residues 1–201); SH3-like domain, green (residues 202–354); C-terminal domain, blue (residues 355–475). The crossover loop shown in complex B is not present in the model for complex A due to weak electron density. (**B–D**) Close-up views of the interactions at the anchor site, bulge and active site regions of each complex, respectively. Dashed lines indicate electrostatic (<4.0 Å) and hydrogen bonding (<3.5 Å) interactions between the protein and DNA. Nucleotides of the Cy5-dT13 oligo are numbered T1–T13 starting from the 3′-end.

mainly in the conformation of the DNA, as opposed to the protein. The protein portions of the two complexes superimpose to an rmsd of 1.3 Å for all Cα atoms. The structural differences do not involve large-scale domain movements, but instead are primarily confined to a segment of the SH3-like domain (residues 268–297) that moves inward by ~4 Å in complex B to maintain its contact with the shorter bulge region (Figure 3B). There are also no

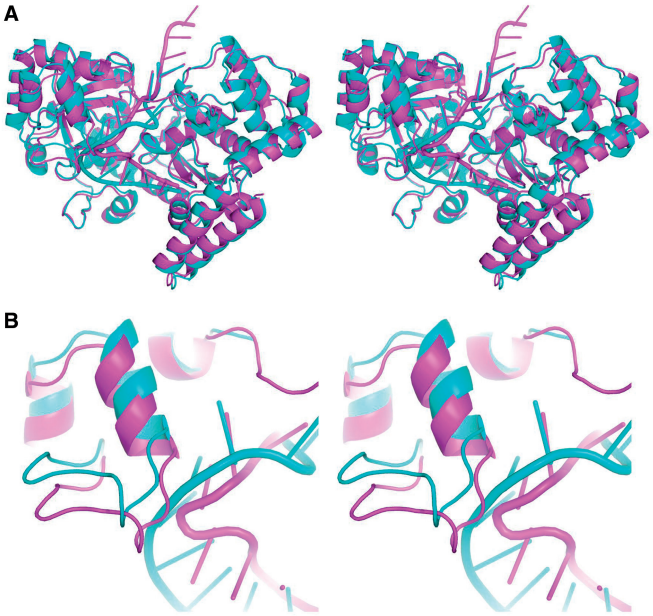

Figure 3. Overlay of complexes A and B observed in the Cy5-dT13 structure. (**A**) Complexes A (cyan) and B (magenta) were superimposed using all protein Cα atoms (rmsd 1.3 Å). Notice that all 13 nucleotides of the ssDNA are bound within the ExoI groove in complex A, whereas in complex B, only the first 10 nucleotides are bound to the groove, such that the three 5′ nucleotides extend from the surface of the complex. The 3′-nucleotide in the active site and the three nucleotides in the anchor site of each complex overlap closely, whereas those in the bulge region differ significantly. (**B**) Close-up view of the interaction with the bulge region. A portion of the SH3-like domain of complex B rotates inward by ~4 Å to maintain its contact with the shorter bulge region. The superposition is the same as in panel A, but zoomed in on the region that shows the largest difference in the conformation of the protein.

large-scale domain movements that occur in ExoI on DNA binding. Complexes A and B align to previously reported structures of unbound ExoI to rmsd values ranging from 0.9 to 1.5 Å, and 1.4 to 1.8 Å, respectively, for all Cα atoms (Supplementary Table S1). Again, the most significant differences between DNA-bound and unbound ExoI structures occur in the segment of the SH3-like domain that contacts the bulge region of the DNA. In particular, residues 277–295 of this segment adopt a particularly closed conformation in the original ExoI structure (1FXX; 12), which would clash with the bulge region of the ssDNA in both complexes. However, in other crystal structures of uncomplexed ExoI, this segment of the SH3-like domain is disordered (13,21), suggesting that it is in general a highly flexible region of the structure, as opposed to one that adopts distinct open and closed states in the presence and absence of DNA, respectively.

Anchor site

The interactions at the anchor site, which involve all three domains of the protein and three consecutive nucleotides of the ssDNA, are essentially the same for both complexes, despite the difference in registration (Figure 2B; a schematic view of ExoI–DNA interactions is provided in Supplementary Figure S5). The three nucleotides at the anchor site are bound with the phosphates buried and the bases predominantly exposed, with their Watson–Crick edges sticking out into solution. A prominent interaction is the insertion of the phenyl ring of Phe371 between the bases of nucleotides 12 and 13 of complex A, or 9 and 10 of complex B, counting from the 3′-end. This interaction appears to set an integral registry of the ssDNA relative to the protein. The aromatic rings of Trp128, Tyr124 and Tyr368 line the right side of the groove to form a hydrophobic wall that contacts the ribose groups of the three nucleotides. A number of polar residues, including Arg113, Arg134, Lys214, Asn257 and Asn304, line the bottom of the groove to form hydrogen bonds and ion pairs with the phosphates, particularly those of nucleotides 11 and 12 of complex A, or 8 and 9 of complex B. Most of the residues that contact the ssDNA at the anchor site are invariant or highly conserved in distant ExoI orthologs (Supplementary Figure S4), suggesting that these interactions are biologically relevant.

Active site

At the bottom end of each complex, the 3′-end of the ssDNA binds to the active site of the exonuclease domain (Figure 2D). The terminal 3′-nucleotide of the ssDNA overlaps closely with the dTMP bound in a previous structure of ExoI, which was formed as a digestion product (13). The terminal 3′-OH forms a close hydrogen bond (2.7 Å) with the backbone amide of Thr18. This explains why ExoI does not cleave 3′-phosphorylated ssDNA (4), as a 3′-phosphate would clash with the backbone of Thr18 to push the scissile bond out of the active site. The base of the terminal nucleotide binds to a hydrophobic pocket formed by the side chains of Thr21, Ala63 and Ile66. In complex A, the scissile phosphate of the terminal nucleotide lies near the carboxylate groups of Asp15, Glu17 and Asp186, which are the conserved metal-binding residues. Despite the presence of 10 mM EDTA in the crystallization mixture, clear electron density is observed for an active site metal ion, which lies at the position of Mg^B in related two-metal nucleases (Figure 4) (22,23). This atom is modeled as an Mg^{2+} ion, based on its distinct octahedral coordination to the scissile phosphate (O1P and O3′ atoms), the carboxylate group of Asp15 and three water molecules. No electron density is observed at the expected position for Mg^A, which is consistent with the fact that the bound DNA is not cleaved. The scissile phosphate in complex A is also near the imidazole side chain of His181, which is thought to help align and activate the attacking water molecule that would be bound to Mg^A. The interactions observed in complex A thus appear close to what would be expected for a catalytic complex, except for the absence of Mg^A and the attacking water. This conclusion is further supported by the fact that the first three nucleotides and Mg^B align well with the trinucleotide bound to the proofreading domain of *E. coli* DNA polymerase I (Figure 4B) (24). By contrast, in complex B, although the 3′-OH is bound in the same position as in complex A, there are no bound metal ions, and the scissile phosphate is rotated upward, such that it is more distant (>5.4 Å) from the active site carboxylates (Figure 4C). Complex B thus

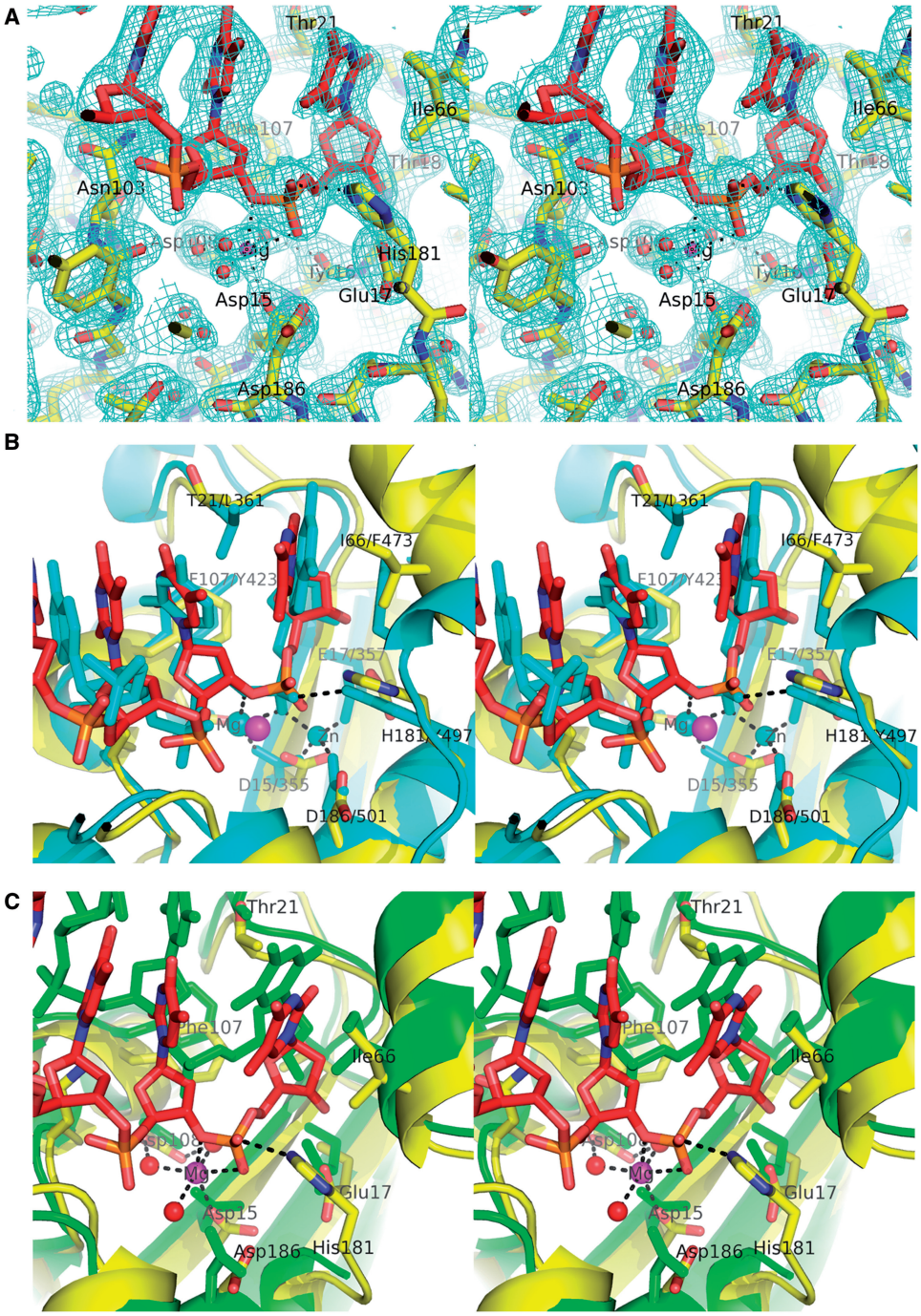

Figure 4. Electron density for a Mg^{2+} ion in the active site of complex A. (**A**) Stereo view of the $2F_{obs}$-F_{calc} electron density map for the active site of complex A. The Mg^{2+} ion (magenta) is coordinated with near octahedral geometry to the scissile phosphate (O1P and O3′ atoms), the carboxylate of Asp15 and three water molecules. This is the expected position for Mg^{B} in related two-metal enzymes. No electron density is present at the expected position for Mg^{A}. (**B**) Superposition of the active site of complex A with the structure of *E. coli* DNA polymerase I in complex with a trinucleotide (PDB code 1KFS). For ExoI, the protein is in yellow, DNA in red and Mg^{2+} magenta. For Pol I, the protein, DNA and metal ions (Mg^{2+} and Zn^{2+}) are shown in cyan. Notice that the first two nucleotides of each complex, including the scissile phosphates, overlap closely, indicating that ExoI complex A is in a near catalytic conformation, despite the absence of Mg^{A}. (**C**) Overlay of the active sites of complex A (yellow and red) and complex B (green). The two bound ssDNA molecules overlap at the 3′-OH but otherwise diverge.

appears considerably less poised for cleavage than complex A.

Moving along the ssDNA from the 3′-end in the active site, the nucleotides in each of the two complexes rapidly diverge from one another. In complex A, the first five nucleotides are in an approximate B-form conformation, with the bases predominantly exposed and stacked over one another, and the phosphates forming hydrogen bonds and electrostatic interactions with three backbone amides (Leu166, Phe164, Ala238) and two positively charged side

chains (Arg142 and Arg165). By contrast, nucleotides 2–5 of complex B follow a more irregular path, with minimal base stacking and fewer interactions with the protein (Figure 2D). Interestingly, the phosphates of nucleotides 3, 5 and 6 in complex B are pinched together to within about 4 Å of one another in the center of the bulge (Figure 2C). Such an interaction, which is clearly observed in the structure owing to the strong density of the phosphates, would be energetically unfavorable, owing to electrostatic repulsions of the phosphates. If complex B is biologically relevant, crowding of these phosphates could conceivably create a tension that could be coupled in some way to translocation of ExoI along the ssDNA to initiate further rounds of cleavage.

Bulge region

The next three nucleotides of complex A, nucleotides 6–8, make a turn at the tip of the bulge, forming a close interaction with a loop formed by residues 284–287 of the SH3-like domain (Figure 2C). A prominent interaction involves Tyr284, which wedges between the bases of nucleotides 7 and 8. A similar interaction between the SH3-like domain and the tip of the bulge occurs in complex B, where Tyr284 inserts between the bases of nucleotides 5 and 6. In complex B, this region of the SH3-like domain moves inward by ~4 Å to maintain its contact with the shorter bulge region (Figure 3B). Due to the inherent flexibility, the electron density for this region of the structure, both for the protein and for the ssDNA, is weaker than for other parts of the structure, as reflected by higher temperature factors. Tyr284 appears to be conserved as Tyr or Phe in ExoI orthologs, although the length of this loop is variable, making this region difficult to align (Supplementary Figure S4).

The ssDNA in both complexes is threaded under a stretched out 'crossover' loop that connects the SH3-like domain on one side with the helical domain on the other (Figure 2C). In complex A, the electron density did not allow for tracing of residues 355–359 of this loop, while in complex B, the electron density was slightly more clear and the entire loop is included in the model, albeit with high-temperature factors. The nucleotides of the ssDNA that bind under the crossover loop are in a rather extended conformation in both complexes, such that the bases are splayed out, making minimal interactions with the protein or with one another. A prominent interaction seen in complex B is the insertion of the phenyl ring of Phe357 of the crossover loop between the bases of nucleotides 6 and 7. A similar interaction could also occur in complex A, in this case with Phe357 inserting between bases 8 and 9, but weak electron density for this region of complex A did not permit tracing of the crossover loop. Moreover, Phe357 is not well conserved in ExoI orthologs (Supplementary Figure S4), suggesting that this interaction is not critical for function.

Crystal structures of ExoI with additional ssDNA substrates

To determine the extent to which the two complexes observed in the Cy5-dT13 structure described above are prevalent, crystal structures of ExoI in complex with three additional ssDNA substrates were determined. Complexes with unlabeled dA16 and Cy5-dA13 crystallized in the same space group as Cy5-dT13 ($P4_3$), while the complex with unlabeled dT13 crystallized in space group $P3_121$, although it still has the same dimer of ExoI present in the asymmetric unit. The three new structures were determined to lower resolution, 3.0–3.5 Å, and thus do not show the detailed protein–DNA interactions as well as in the Cy5-dT13 structure. Despite the lower resolution, electron density maps allowed for placement of all of nucleotides of each ssDNA, except for the Cy5 groups of Cy5-dA13, the two 5′-terminal nucleotide of dA16 and two nucleotides in the central bulge of dT13 (Figure 5). The

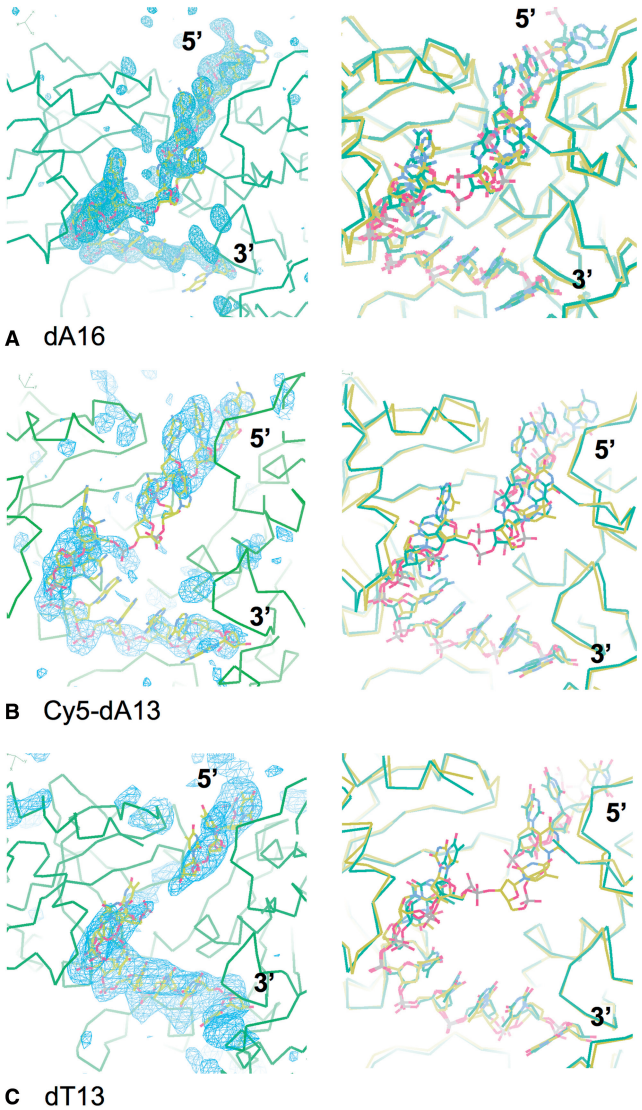

Figure 5. Structure of ExoI in complex with additional ssDNA substrates. (**A**) dA16, (**B**) Cy5-dA13, (**C**) dT13. The figure on the left of each panel shows an F_o-F_c map calculated with model phases after 40 cycles of restrained refinement in REFMAC5 with the ssDNA omitted. Each map is contoured at +3.0 sigma and shown over the entire complex. The figure on the right shows a superposition of each structure (green and blue bonds) with complex A of the Cy5-dT13 structure (yellow bonds). Notice that the ssDNA in each new structure is bound in a similar conformation as the ssDNA in complex A of Cy5-dT13.

most significant finding from the three additional structures is that all six of the new ExoI–ssDNA complexes (two complexes for each crystal structure) have the ssDNA bound in a conformation that is close to that of complex A described above, suggesting that complex B of the Cy5-dT13 crystal structure may be an outlier.

Although we report only three additional refined structures, crystals similar to those with dA16 (which are likewise similar to Cy5-dA13) could be grown with dA oligos ranging from 13–16 nucleotides in length. Preliminary analysis of a crystal with dA14 indicates that the structure is essentially identical to the complex with dA16. Likewise, crystals similar to those with unlabeled dT13 could be grown with slightly longer dT oligos (dT14–dT16), and preliminary analysis of a crystal with dT16 indicates that the structure is essentially identical to the dT13 complex. These observations indicate that the length of the ssDNA does not affect the observed binding mode: the first ~13 nucleotides of the ssDNA are bound to the ExoI groove in essentially the same conformation in all four of the structures (dT13, dT16, dA14 and dA16), such that the extra 5′-nucleotides of dT16 and dA16 simply extend out from the surface of the complex.

DISCUSSION

A hallmark of ExoI is its high degree of processivity. While the two-metal mechanism has been structurally well established for numerous Mg^{2+}-dependent nucleases (3,22,23), what is less clear is the mechanism by which enzymes such as ExoI are able to translocate along the ssDNA in a rapid and processive manner. In the original ExoI structure, encircling of the ssDNA by the crossover loop was proposed to be a key to processivity (12). Topological linkage is a recurring theme for other processive enzymes, such as λ exonuclease, PCNA and others (25). However, in an early report on processive nucleases, before any of their structures were known, a dual binding-site model was proposed (7). This model invoked the concept of an 'anchor' site on the enzyme, separate from the active site, that could allow the enzyme to remain bound to one site on the DNA as it stepped forward at the other. The crystal structures of the ExoI–ssDNA complex reported here show both features: threading of the ssDNA under the crossover loop and dual binding of the ssDNA to the active site and the anchor site. The extent to which both of these features, topological linkage and two-site binding, are necessary for processivity remains to be determined.

The crystal structures reported here have captured two distinct states of ExoI bound to ssDNA that could provide important clues to understanding the mechanism of translocation. However, it is important to consider whether both of the structures are likely to be biologically relevant, or whether one of the observed states is a structural artifact, due to interactions of the Cy5 groups or other crystal packing forces. Complex A is remarkably consistent with the biochemical studies of Brody and colleagues, who were able to make detailed predictions about the ExoI–ssDNA interactions from quantifying the distribution of products released from digestions of various oligonucleotide substrates under conditions of limiting enzyme (6,15). Based on this analysis, it was predicted that ExoI interacts with two separate sites on the ssDNA: an 'extended active site' starting at the 3′-terminal nucleotide and extending for about six nucleotides, and an 'anchor' site located at nucleotides 9–13. Comparison of these predictions with the interactions observed in complex A, as shown schematically in Supplementary Figure S5, reveals a remarkably close agreement. It was further predicted, based on analysis of oligonucleotide substrates containing abasic residues or methylphosphonates at specific positions, that while the interactions at the extended-active site require both the bases and the phosphates, those at the anchor site require only the sugar–phosphate backbone. Again, there is a reasonable agreement. While there is some contact with the bases at the anchor site, most notably the insertion of Phe371 between the bases of nucleotides 12 and 13, for the most part the bases at the anchor site are largely exposed, while the phosphates are fully buried, making extensive interactions with the protein. In addition to being consistent with the biochemical studies, complex A has the ssDNA in close to a catalytic configuration, despite the presence of only one metal ion. This is seen by the fact that the first three nucleotides in complex A overlap closely with a trinucleotide substrate bound to the proofreading domain of *E. coli* DNA Pol I in the presence of metals (24). Finally, as shown in the sequence alignment of Supplementary Figure S4, most of the key residues of ExoI that contact the DNA in complex A are predominantly conserved in distant ExoI orthologs. Thus, complex A appears to be a functionally relevant state.

Complex B is overall similar to complex A in that the 3′-end of the ssDNA is bound to the active site, the downstream end is threaded under the crossover loop and similar contacts are made at the bulge region and the anchor site. Complex B differs from complex A in that only the first 10 nucleotides from the 3′-end have entered the binding groove, such that the registry of the protein relative to the ssDNA at the anchor site is off by exactly three nucleotides. There is consequently a shorter bulge of the ssDNA in complex B, and the interactions at the active site are not as extensive and further from what would be expected for a catalytic complex. It is possible that the Cy5 label of complex B, which is clearly visualized in the structure due to its interactions at the dimer interface, could have pulled the ssDNA partially out of the binding groove to result in a structural artifact. The fact that the conformation of complex B was not seen in any of the structures with unlabeled oligos suggests that this may be the case. On the other hand, the Cy5 group apparently did not interfere with complex A, which also has its Cy5 group forming close interactions at the dimer interface, or in the structures observed with Cy5-dA13, for which the Cy5 groups are not visible in the electron density. It is also interesting to note how a region of the SH3-like domain of ExoI moves inward by ~4 Å in complex B, to maintain its contact with the shorter bulge region. It thus appears that ExoI has a built-in flexibility to accommodate different conformations of bound DNA.

While the overall conclusions from Brody *et al.* (6,15) are in general agreement with complex A, some of their observations can perhaps be more easily reconciled with complex B. For example, for all of the different oligonucleotides that were digested, only small amounts of 11-mers or longer oligonucleotides were released from the enzyme (0–6% of the total), indicating that ExoI binds nearly as tightly to the 11-mer as to longer oligonucleotides. This would suggest that any interactions of ExoI occurring beyond the 11th nucleotide are not critical for affinity. However, in complex A the main contacts at the anchor site are with nucleotides 11–13, with the phosphate of nucleotide 12 having particularly close interactions. By contrast, in complex B the shift in registry places the equivalent set of interactions at nucleotides 8–10. Significantly higher levels of 10-mers and especially 9-mers are released as end products of ExoI digestions, indicating that interactions at nucleotides 11 and 10 contribute significantly to binding. In general, the most significant effect is loss of nucleotide 10, which results in release of high levels of 9-mer (20–50% of the total). In complex A, nucleotide 10 does not make close contacts to the protein, as there are only long range (5.8 Å) electrostatic interactions to the phosphate and a weak (3.3 Å) hydrogen bond to the base. By comparison, in complex B, nucleotide 10 is one of the three that are right at the anchor site. While not all of the observations of Brody *et al.* can be reconciled with complex B in this way, in general the biochemical data tend to indicate that complex B may be relevant. One possibility is that while complex A represents a catalytic state, complex B may be representative of the end state of a processive digestion.

Based on the available structural information, one could envision two fundamentally different models for how ExoI processively digests ssDNA. In one model (Figure 6A), in which only the binding mode of complex A is relevant, ExoI translocates by one nucleotide step along the DNA after each round of cleavage, to generate a new complex A at each step. Such a model would require breakage of several close hydrogen bonds and base insertion interactions at the anchor site after each individual round of cleavage, but each translocation step could conceivably be coupled in some way to the energy released from cleavage of the terminal phosphodiester bond, estimated at −5.3 kcal/mol (26). The second model (Figure 6B) assumes that both of the observed binding modes are relevant. Starting with complex A, in which ExoI is bound to 13 nucleotides from the 3′-end, three rounds of cleavage, without displacement of the nucleotides bound to the anchor site, would generate the structure observed in complex B, in which ExoI covers 10 nucleotides from the 3′-end. Positioning of the scissile bond of the terminal nucleotide at each round of cleavage could be accommodated by shortening of the bulge region of the ssDNA, together with inward movement of the SH3-like domain. In comparing the structures of complexes A and B, it appears that such intermediate states could be easily accommodated. We note, however, that the conformation of the ssDNA in the active site of complex B is different

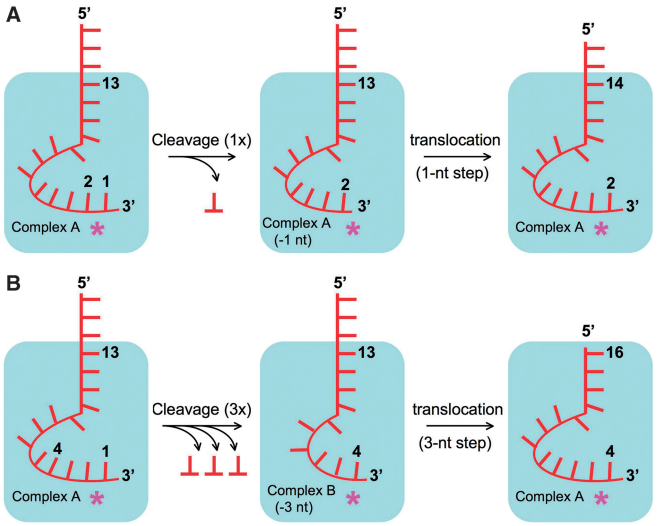

Figure 6. Cartoon depicting two possible mechanisms for ExoI processive digestion. (**A**) Single-nucleotide stepping mechanism in which translocation of ExoI along the ssDNA occurs by one nucleotide step after each individual round of cleavage. (**B**) Three-nucleotide stepping mechanism in which translocation of ExoI along the ssDNA occurs by a three-nucleotide step after every three rounds of cleavage. The first model uses only the binding mode observed in complex A, whereas the second model invokes the two distinct binding modes observed in complexes A and B.

from complex A, particularly for the third and fourth nucleotides (Figure 2D). We envision that intermediate states that are active for cleavage would more closely resemble complex A and that the conformation of complex B may occur when the ssDNA bound in the groove is too short to bind as seen in complex A. After three rounds of cleavage, the ssDNA would be too short to allow the 3′-end to fully reach the active site, and so further rounds of cleavage would require release of the nucleotides from the anchor site to generate a new complex A in which the enzyme has translocated three nucleotides forward along the DNA. In this way, ExoI would translocate along the DNA with a stepsize of three nucleotides, with three rounds of cleavage occurring between each translocation step. A possible advantage of this type of mechanism is that breakage of the interactions made by the three nucleotides at the anchor site would only need to occur once for every three rounds of cleavage, instead of after each individual step.

Intriguingly, a similar 'stepping' type of mechanism has recently been proposed to occur for Rrp44, a catalytic subunit of the yeast exosome complex, which processively digests single-stranded or structured RNA in the 3′-5′ direction (27). Single molecule FRET analysis of Rrp44 digesting a 3′-tailed 43 bp A-form RNA duplex revealed that the enzyme unwinds the duplex not in single base pair steps but rather in bursts of ∼4 bp. The data further indicated that unwinding, as opposed to the chemical cleavage steps, is rate limiting. It was surmised that as the enzyme cleaves one nucleotide at a time from the 3′-tail, it gets successively closer to the duplex region, building up tension within the protein that is elastically coupled to duplex unwinding. ExoI is fundamentally

different from Rrp44 in that it digests ssDNA instead of RNA and has a much lower capacity to unwind regions of duplex (4). Nonetheless, it is interesting to speculate that the translocation model described for ExoI in Figure 6B could conceivably generate a similar type of stepping pattern as that observed for Rrp44. It is also interesting to note that the ExoI–ssDNA complex shares some common features with Rrp44 and the related ribonuclease II of *E. coli*, for which structures have been determined in complex with 13-mer oligoribonucleotides (28,29). All three enzymes have five nucleotides at the 3′-end inserted into an active-site cleft, a turn or bulge region in which the nucleotides are more splayed out, and an 'anchor' region at which close interactions are formed with the downstream end of the substrate. In the case of RNase II, a single-step translocation mechanism was proposed, as only one binding mode of the protein on the substrate was observed (29). A single binding mode was also observed for Rrp4, although it is interesting to note that this structure differed significantly from that of RNase II in the way that the downstream portion of the DNA was bound (28).

It is also interesting to compare the two mechanisms described above for ExoI to that of λ exonuclease, an enzyme that processively digests one strand of double-stranded DNA in the 5′-3′ direction. λ Exonuclease forms a trimer with a central funnel-shaped channel for tracking along the DNA (30,31). The duplex is unwound ahead of the cleavage site by exactly 2 bp, such that the 5′-end inserts into one of the three active sites on the trimer, and the 3′-end threads through the central channel to emerge out the back. Processivity is proposed to be due to encircling of the 3′-ended strand as the trimer motors along the DNA, and translocation may be driven in part by attraction of the 5′-phosphate generated at each round of cleavage to a positively charged pocket at the end of the active-site cleft (31). Sliding of the trimer along the DNA, presumably in single base pair steps, appears to be facilitated by a rather loose set of electrostatic interactions with the DNA in the central channel, combined with the insertion of a key arginine residue into the minor groove of the downstream portion of the duplex. Because the binding of the arginine to the minor groove involves electrostatic attraction to the phosphates (32), as opposed to specific hydrogen bonds to the bases, it was proposed that the arginine could act as a sliding rudder to keep the enzyme on track as it translocates along the DNA (31). By comparison, the interactions of ExoI with the downstream portion of the ssDNA at the anchor site seem less well suited to sliding, as they involve a network of several close hydrogen bonds, as well as insertion of aromatic residues between the bases. In this regard, the three-nucleotide stepping mechanism described above for ExoI is particularly attractive.

ACCESSION NUMBERS

4JRP, 4JRQ, 4JS4, 4JS5.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Table 1, Supplementary Figures 1–5 and Supplementary Reference [33].

ACKNOWLEDGMENTS

Use of the Advanced Photon Source at Argonne National Laboratory was supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences, under Contract DE-AC02-06CH11357. Use of the Lilly Research Laboratories Collaborative Access beamline at sector 31 of the Advanced Photon Source was provided by Eli Lilly Company, which operates the facility.

FUNDING

National Science Foundation [MCB-1021066 to C.E.B.]. Funding for open access charge: National Science Foundation.

Conflict of interest statement. None declared.

REFERENCES

1. Shevelev,I.V. and Hübscher,U. (2002) The 3′–5′ exonucleases. *Nat. Rev. Mol. Cell. Biol.*, **5**, 364–376.
2. Symmons,M.F., Williams,M.G., Luisi,B.F., Jones,G.H. and Carpousis,A.J. (2002) Running rings around RNA: a superfamily of phosphate-dependent RNases. *Trends Biochem. Sci.*, **27**, 11–18.
3. Yang,W. (2011) Nucleases: diversity of structure, function and mechanism. *Q. Rev. Biophys.*, **44**, 1–93.
4. Lehman,I.R. and Nussbaum,A.L. (1964) The deoxyribonucleases of *Escherichia coli* V. On the specificity of exonuclease I (phosphodiesterase). *J. Biol. Chem.*, **239**, 2628–2636.
5. 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.
6. Brody,R.S. (1991) Nucleotide positions responsible for the processivity of the reaction of exonuclease I with oligodeoxyribonucleotides. *Biochemistry*, **30**, 7072–7080.
7. Thomas,K.R. and Olivera,B.M. (1978) Processivity of DNA exonucleases. *J. Biol. Chem.*, **253**, 424–429.
8. Sandigursky,M., Mendez,F., Bases,R.E., Matsumoto,T. and Franklin,W.A. (1996) Protein-protein interactions between the *Escherichia coli* single-stranded DNA-binding protein and exonuclease I. *Radiat. Res.*, **145**, 619–623.
9. Lu,D., Myers,A.R., George,N.P. and Keck,J.L. (2011) Mechanism of exonuclease I stimulation by the single-stranded DNA-binding protein. *Nucleic Acids Res.*, **39**, 6536–6545.
10. Hornblower,B., Coombs,A., Whitaker,R.D., Kolomeisky,A., Picone,S.J., Meller,A. and Akeson,M. (2007) Single-molecule analysis of DNA-protein complexes using nanopores. *Nat. Methods*, **4**, 315–317.
11. Clarke,J., Wu,H.C., Jayasinghe,L., Patel,A., Reid,S. and Bayley,H. (2009) Continuous base identification for single-molecule nanopore DNA sequencing. *Nat. Nanotech.*, **4**, 265–270.
12. Breyer,W.A. and Matthews,B.W. (2000) Structure of *Escherichia coli* exonuclease I suggests how processivity is achieved. *Nat. Struct. Biol.*, **277**, 1125–1128.
13. Busam,R.D. (2008) Structure of *Escherichia coli* exonuclease I in complex with thymidine 5′-monophosphate. *Acta Crystallogr. D Biol. Crystallogr.*, **64(Pt 2)**, 206–210.
14. Koonin,E. (1997) A conserved ancient domain joins the growing superfamily of 3′-5′ exonucleases. *Curr. Biol.*, **7**, R604–R606.

15. Brody,R.S., Doherty,K.G. and Zimmerman,P.D. (1986) Processivity and kinetics of the reaction of exonuclease I from *Escherichia coli* with polydeoxyribonucleotides. *J. Biol. Chem.*, **261**, 7136–7143.
16. Winn,M.D., Ballard,C.C., Cowtan,K.D., Dodson,E.J., Emsley,P., Evans,P.R., Keegan,R.M., Krissinel,E.B., Leslie,A.G.W., McCoy,A. *et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.*, **67(Pt 4)**, 235–242.
17. Emsley,P., Lohkamp,B., Scott,W.G. and Cowtan,K. (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.*, **66**, 486–501.
18. Schuck,P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys. J.*, **78**, 1606–1619.
19. Ray,R.K., Reuben,R., Molineux,I. and Gefter,M. (1974) The purification of exonuclease I from *Escherichia coli* by affinity chromatography. *J. Biol. Chem.*, **249**, 5379–5381.
20. Prasher,D.C., Conarro,L. and Kushner,S.R. (1983) Amplification and purification of exonuclease I from *Escherichi coli* K12. *J. Biol. Chem.*, **258**, 6340–6343.
21. Lu,D. and Keck,J.L. (2008) Structural basis of *Escherichia coli* single-stranded DNA-binding protein stimulation of exonuclease I. *Proc. Natl Acad. Sci. USA*, **105**, 9169–9174.
22. Beese,L.S. and Steitz,T.A. (1991) Structural basis for the 3′-5′ exonuclease activity of Escherichia coli DNA polymerase I: a two metal ion mechanism. *EMBO J.*, **10**, 25–33.
23. Yang,W., Lee,J.Y. and Nowotny,M. (2006) Making and breaking nucleic acids: Two-Mg^{2+}-ion catalysis and substrate specificity. *Mol. Cell*, **22**, 5–13.
24. Brautigam,C.A., Sun,S., Piccirilli,J.A. and Steitz,T.A. (1999) Structures of normal single-stranded DNA and deoxyribo-3′-S-phosphorothiolates bound to the 3′-5′ exonucleolytic active site of DNA polymerase I from *Escherichia coli Biochemistry*, **38**, 696–704.
25. Breyer,W.A. and Matthews,B.W. (2001) A structural basis for processivity. *Protein Sci.*, **10**, 1699–1711.
26. Dickson,K.S., Burns,C.M. and Richardson,J.P. (2000) Determination of the free energy change for repair of a DNA phosphodiester bond. *J. Biol. Chem.*, **275**, 15828–15831.
27. Lee,G., Bratkowski,M.A., Ding,F., Ke,A. and Ha,T. (2012) Elastic coupling between RNA degradation and unwinding by an exoribonuclease. *Science*, **336**, 1726–1729.
28. Lorentzen,E., Basquin,J., Tomecki,R., Dziemboski,A. and Conti,E. (2008) Structure of the active subunit of the yeast exosome core, Rrp44: diverse modes of substrate recruitment in the RNase II nuclease family. *Mol. Cell*, **29**, 717–728.
29. Frazão,C., McVey,C.E., Amblar,M., Barbas,A., Vonrhein,C., Arraiano,C.M. and Carrondo,M.A. (2006) Unravelling the dynamics of RNA degradation by ribonuclease II and its RNA-bound complex. *Nature*, **443**, 110–113.
30. Kovall,R. and Matthews,B.W. (1997) Toroidal structure of lambda-exonuclease. *Science*, **277**, 1824–1827.
31. Zhang,J., McCabe,K.A. and Bell,C.E. (2011) Crystal structures of lambda exonuclease in complex with DNA suggest an electrostatic ratchet mechanism for processivity. *Proc. Natl Acad. Sci. USA*, **108**, 11872–11877.
32. Rohs,R., West,S.M., Sosinsky,A., Liu,P., Mann,R.S. and Honig,B. (2009) The role of DNA shape in protein-DNA recognition. *Nature*, **461**, 1248–1253.
33. Papodopoulos,J.S. and Agarwala,R. (2007) COBALT: constraint-based alignment tool for multiple protein sequences. *Bioinformatics*, **23**, 1073–1079.