Supplemental Materials

Supplemental figure legends

Figure S1. Electrostatic Potential of hExo1, related to Figure 1. The electrostatic potential (+5 kT/e^- to −5 kT/e^-) of hExo1 catalytic domain calculated with the Adaptive Poisson-Boltzmann Solver (APBS) software (Baker et al., 2001) mapped onto the solvent-accessible molecular surface. The blue electropositive regions highlight anchor points for DNA binding and discrimination (K^+ and HLH anchors), and an acidic region (red) at the active site binds the two metal ions that position the 5'-scissile phosphate.

Figure S2. Maximum likelihood alignment of hExo1 and DNA Polymerase β helix-hairpin-helix and helix-two-turn-helix domains, related to Figure 3. Top: hExo1 is shown in green and Polβ (PDB ID 1ZQI) is blue; potassium ions are shown in similar colors (large spheres); waters for hExo1 (red) and Polβ (blue) are shown with small spheres. Bottom: coordination of potassium ion by hExo1. Six interactions are indicated with dashed lines.

Figure S3. Views of active site in pre-substrate Complexes I and II, related to Figure 4. (**A)** Complex I (D173A mutant) structure with one Ca^{2+} coordinated by conserved carboxylates (blue). Substrate-binding structural elements, Ca^{2+} (green), and N-terminus are indicated. Structural elements are colored according to scheme in Fig 1. **(B)** Close-up of complex I active site (rotated 90° from A). One Ca^{2+} ion (green) is coordinated by conserved carboxylate residues. Given the limits of resolution of these structures, interactions (dashed lines) are approximate. (**C)** Locations of residues R95 and R96

relative to substrate DNA terminal phosphate in nascent substrate Complex II. Substrate-binding structural elements, metal ions, and N-terminus are indicated as in Figure 4. Structural elements are colored according to scheme in Fig 1.

Figure S4. Representative electron density maps of Complexes I and II, related to Figure 4. (A) Anomalous difference Fourier map for Mn^{2+} superimposed on product structure (Complex III), contoured at 3σ. The map clearly confirms positions of these ions in the active site. (**B)** Electron density maps of nascent substrate complex (Complex II). A σ_A-weighted F_o-F_c electron density map shows density for the DNA substrate (3.5 σ contour). Map was generated prior to including any DNA in phase calculation.

Figure S5. Kinetic parameters of by hExo1 catalytic domain enzymatic activity with 5′-recessed substrate, related to Figure 4. (**A)** Hydrolytic reactions (Experimental Procedures) were analyzed on a 12% denaturing polyacrylamide gel containing 7M urea. Incubation of enzyme with DNA substrate results in release of deoxyguanosine 5' monophosphate in a time and substrate concentration dependent manner, as judged from the autoradiogram shown. (**B**) Initial rates of product formation (V_0) were determined by quantitation of the amount of mononucleotide as a fraction of total signal intensity in the lane, and plotted as a function of time (not shown). Error bars represent standard deviations of values from three independent experiments. V_0 values were plotted as a function of DNA substrate concentration, and were fit by non-linear least squares regression to a rectangular hyperbola to yield the indicated V_{max}, and K_m values.

(C) Effect of divalent cations on hExo1 catalytic domain activity. 5' hydrolytic activity was measured as a function of hExo1 concentration in the presence of 5 mM $MgCl_2$, $BaCl_2$, $CaCl_2$, and $MnCl_2$. Ba^{2+} and Ca^{2+} do not support hydrolysis; Ca^{2+}, Mg^{2+}, and Mn^{2+} observations are in good agreement with previously published results (Lee and Wilson, 1999).

Figure S6. Mobility of helix 4 and proposed involvement with 5' flap binding, related to Figure 4. Complex I is shown in grey; complex II is shown in orange; complex III is shown in blue. Structures were aligned using Cα atoms of the conserved β-strand core region. Movement of >3 Å between equivalent Cα atoms is observed within these complexes at helices 4 and 5; the surrounding helices and β-strands remain in a static position. Arrows are used to indicate proposed motions: rocking at the bases of the helix, as well as a twisting motion that alters the register of the helix.

Figure S7. Interactions of the hExo1 C-terminal domain and the active site, related to Figure 6. (**A**) Interactions between the active site (purple) and the C-terminal region (pink) are indicated with dashed lines; neighboring residues are labeled; metal sites are shown in gray; DNA is shown in yellow. (**B**) Surface representation of interaction view shown in (A).

Mutation	**Oligonucleotides**
Tyr32Ala	5'-CAGGTAGTAGCTGTGGATACAGCTTGCTGGCTTCACAAAGGA-3'
His36Ala	5'-GATACATATTGCTGGCTTGCCAAAGGAGCTATTGCTTGT-3'
Lys85Ala	5'-GGATGTACTTTACCTTCTGCAAAGGAAGTAGAGAGATCT-3'
Asp173Ala	5'-ATAATTACAGAGGACTCGGCTCTCCTAGCTTTTGGCTGT-3'
Arg92Ala	5'-AAGGAAGTAGAGAGATCTGCAAGAGAAAGACGACAAGCC-3'

Table S1. **Primers used in mutagenesis studies, related to Experimental Procedures.**

Data Collection					
Data Set (PDB ID)	**Native Complex I (3QE9)**	**SeMet A**	**SeMet B**	**$BaCl_2$ Complex II (3QEA)**	**$MnCl_2$ Complex III (3QEB)**
Space Group	$P2_12_12$	$P2_12_12$	$P2_12_12$	$P4_32_12$	$P4_32_12$
Cell Dimensions (Å) *abc*	a=92.50 b=95.47 c=99.78	a=92.42 b=95.70 c=100.58	a=92.23 b=95.51 c=100.30	a=b=71.83 c=181.89Å	a=b=70.41 c=182.88
Wavelength	0.9793 Å	0.9793 Å	0.9793 Å	1.2000 Å	1.0000 Å
X-ray source	APS SER-CAT 22ID	APS SER-CAT 22ID	APS SER-CAT 22ID	APS SER-CAT 22BM	APS SER-CAT 22BM
Resolution (highest bin)	50-2.5 (2.59-2.5)	50-3.3 (3.4-3.3)	50-3.0 (3.1-3.0)	50-3.1 (3.2-3.1)	50-3.0 (3.1-3.0)
R_{sym}[a]	6.9 (38.8)	9.6 (36.5)	9.8 (68.0)	6.6 (66.7)	6.1 (37.4)
I/σI	29.2 (2.2)	25.3 (2.3)	19.1 (1.7)	20.4 (3.7)	40.9 (2.9)
Completeness (%)	93.3 (62.4)	97.0 (74.1)	76.4 (25.5)	99.9 (99.9)	83.3 (53.9)
Redundancy	5.4 (3.0)	6.0 (3.2)	7.0 (5.5)	7.8 (7.8)	9.4 (5.2)
No. Reflections	28,797	25,093	13,833	16456	11,100
Experimental Phasing					
Resolution (Å)	50-2.7Å				
Isomorphous phasing power (centric/acentric)	1.895/1.598				
Anomalous phasing power (acentric)	1.085				
Figure of merit (centric/acentric)	0.224/0.127				
Refinement Statistics					
Resolution	50-2.5Å			50-3.1Å	50-3.0Å
R_{work}[b]/R_{free}[c]	21.2/23.8			21.2/24.3	22.4/26.8
No. protein atoms	5419			2716	2714
No. DNA atoms	915			452	427
RMSD bond	0.006 Å			0.005 Å	0.001 Å

lengths					
RMSD bond angles	0.843º			1.355º	0.412º
Coordinate error estimate (SigmaA)	0.33Å			0.52Å	0.62Å
Ramachandran Analysis					
Most favored (%)	98.2%			95.9%	95.6%
Allowed (%)	1.8%			4.1%	4.4%
Disallowed (%)	0%			0%	0%

[a] $R_{merge} = \Sigma_{hkl} \Sigma_i |I_i(hkl) - \langle I(hkl) \rangle| / \Sigma_{hkl} \Sigma_i I_i(hkl)$.
[b] $R_{work} = \Sigma_{hkl}(||F_o| - k|F_c||)/\Sigma_{hkl}|F_o|$ for reflections in R_{work} set.
[c] R_{free} is computed from a random selection of reflections in the asymmetric unit that was not used during refinement.

Table S2. Data collection and refinement statistics, related to Experimental Procedures.

Movie S1. Model of hExo1 DNA recognition and excision, related to Figures 4 and 5. Exo1 binds DNA at a nick or gap in the substrate; the DNA is bent and two bases fray from the 5' end in the active site; excision occurs.

Extended Experimental Procedures

Cloning of hExo1 catalytic domain- construct and mutants

Full-length hExo1 was expressed and purified as previously described (Genschel et al., 2002). Limited proteolysis of the full-length protein (1 μL of 10 μg/mL trypsin plus 50 μL of 0.6 mg/mL hExo1 at 25°C for 30 min) yielded a stable 39.5 kDa fragment. MALDI-TOF mass spectrometry identified this fragment as the hExo1 catalytic domain. The catalytic domain (residues 1-352) was cloned into the pET21+ vector using NdeI and BamHI restriction sites. The following mutations were introduced via site-directed mutagenesis according to manufacturer's protocol (QuikChange Multi Site-Directed Mutagenesis Kit, Stratagene): D173A, Y32A, K85A, H36A, and R92A (Table 1).

Expression and purification of the hExo1 catalytic domain

Wild-type and mutant hExo1 catalytic domains were expressed in *E. coli* BL21(DE3). Cells were cultured in 10 L of Luria broth supplemented with ampicillin (100 μg/mL) and were grown for 12 hours at 18°C after 0.1 mM IPTG induction. Cells were harvested via centrifugation (4,000 x g, 4°C, 30 min), resuspended in lysis buffer (100 mM HEPES-NaOH, pH 7.0; 100 mM NaCl; 5 mM DTT; 1 mM EDTA; 10% sucrose), lysed with a cell cracker and centrifuged (18,500 x g, 4°C, 30 min). The supernatant was loaded onto a 20- or 70-mL SP Sepharose Fast-Flow ion exchange column (GE Healthcare Life Sciences) pre-equilibrated with ion-exchange buffer (100 mM HEPES-NaOH, pH 7.0; 100 mM NaCl; 5 mM DTT; 1 mM EDTA; 5% glycerol) and eluted with a 10 column volume NaCl gradient (100- 500 mM). Fractions containing hExo1 were identified using SDS-PAGE and pooled. The pooled sample was diluted with buffer (100 mM HEPES-NaOH, pH 7.0; 5 mM DTT; 1 mM EDTA; 5% glycerol), loaded onto a 5- or 10- mL heparin column (HiTrap Heparin HP, GE Healthcare Life Sciences), and eluted over a 10 column volume NaCl gradient (0.1-1 M). Fractions containing hExo1 were identified using SDS-PAGE and pooled. The pooled sample was concentrated in a 10,000 MWC Centricon concentrator (Millipore) to a final volume of 1.5 mL. The sample was loaded onto a 120 mL gel filtration column (Hi-Load Superdex-200, GE Healthcare Life Sciences) and eluted in storage buffer (100 mM TRIS-HCl at pH 8.0, 100 mM NaCl, 2 mM DTT & 0.1 mM EDTA). Protein was concentrated to ~10 mg/mL, flash-frozen in liquid nitrogen, and stored at -80°C. Selenomethionine-labeled D173A mutant protein was overexpressed in *E. coli* BL21(DE3) using the metabolic repression technique (Doublie, 1997) and prepared as described above.

Preparation of Oligonucleotide Substrates for Crystallization

Oligonucleotides (Midland Certified Reagent Company, Inc.) were dissolved in TE buffer, and equimolar amounts were combined and annealed via slow-cooling using standard protocols. The 5'-recessed-end substrate was created using a 10 nucleotide (nt) top strand containing a 5' terminal phosphate (5'-(p)-TCG ACT AGC G-3') and 13 nt bottom strand (5'-CGC TAG TCG ACA C-3'). A similar 5'-recessed-end DNA substrate contained a phosphorothioate linkage between the terminal nucleotides at both ends of each oligonucleotide (top: 5'-(p)-T-thio-CGA CTA GC-thio-G-3', bottom: 5'-C-thio-GCT AGT CGA CA-thio-C-3').

Crystallization and Data Collection

hExo1 D173A (200 μM) was combined with substrate DNA (260 μM), incubated at 4°C for 30 min, and centrifuged (10,000 x g, 4°C, 2 min.). The complex (Complex I) was crystallized by sitting-drop vapor diffusion at 17°C after mixing 1 μL each of protein solution and precipitant solution (100 mM HEPES-KOH, pH 7.0; 100 mM KCl; 100 mM $CaCl_2$; 20% PEG 6000). Selenomethionine crystals were also grown in the above manner. Wild-type crystals with phosphorothioate-modified DNA were grown similarly using a final concentration of 25 mM $CaCl_2$ as an additive during substrate incubation, and an alternative precipitant solution (100 mM MES, pH 5.0; 5% PEG 6000; 3% ethylene glycol; final pH 6.0). Metals were exchanged in wild-type crystals by soaking in either 10 mM $BaCl_2$ (Complex II) or 10 mM $MnCl_2$ (Complex III) for 3 hours, and crystals were transferred stepwise into mother liquor supplemented with 35% ethylene glycol and flash cooled in liquid nitrogen.

Structural Determination and Refinement

Diffraction data were collected at The Advanced Photon Source (APS) at Argonne National Laboratory, beamlines 22-ID and 22-BM (SER-CAT), and The Advanced Light Source (ALS) at Lawrence Berkeley National Laboratory, beamline 12.3.1 (SIBYLS); experiments were conducted at 100 K. Native diffraction data were collected to 2.5 Å resolution and selenomethionine data were collected to 3.3 and 3.4 Å resolution at λ=0.9794 Å at the selenium K edge. Data were scaled in space group $P2_12_12$ using HKL2000 (Otwinowski, 1998). The structure of the hExo1 catalytic domain (D173A) DNA complex was determined by selenium-SIRAS experimental phasing (Hendrickson et al., 1990). Two Se peak data sets and a native data set were used to determine experimental phases with SHARP (Bricogne et al., 2003). The model contained two molecules in the asymmetric unit, and five selenium sites per monomer were observed. The model was built manually in COOT (Emsley and Cowtan, 2004). Initial solvent-flattened maps lacked side chain density and connectivity and were improved using partial model phase combination and B-factor sharpening in CNS (Brünger et al., 1998). PHENIX Autobuild (Adams et al., 2002) was used to improve initial model and electron density. The structure was refined using CCP4 (Winn et al., 2003) and CNS with a maximum likelihood target and phase probability distribution.

Wild-type crystals diffracted to 3.1-Å resolution and were scaled in the $P4_32_12$ space group using HKL2000. Wild-type barium derivative data were measured at λ=1.2Å, and manganese derivative data were collected at λ=1.0 Å or 1.25 Å. Molecular replacement phases were calculated using PHASER (Storoni et al., 2004) using the D173A mutant structure as a model. Initial low-resolution refinement was carried out in CNS using deformable elastic network (DEN) restraints (Schroder et al., 2007). The

hExo1 D173A structure was input as a reference model. DNA was built manually in COOT using electron-density maps from this refinement. Metal ion positions were identified using phased anomalous difference Fourier maps. Structures were refined using PHENIX and CNS with a maximum likelihood target and phase probability distribution. Final model coordinates were checked with MOLPROBITY (Davis et al., 2004).

Structure-based Sequence Alignment

The structure-based sequence alignment was generated step-wise, first aligning the structure of hExo1 and the structure of human FEN1 (PDB ID 1UL1) using the Coot LSQ function (Emsley and Cowtan, 2004). The two sequences were aligned based on residue-by-residue comparison of their secondary structural elements and residue positions. Other members of the 5'-nuclease family, human XPG and human GEN1, were then added to the alignment in a similar fashion, using tertiary-structure threading models produced by MODELLER with hExo1 and hFEN-1 structures (PDB IDs 1UL1 & 1 RXW) as reference models (Eswar et al., 2006). A ClustalX-generated (Larkin et al., 2007) sequence alignment of the 5' nuclease family and secondary-structure predictions (JPRED, PHD & PROF) for hXPG and hGEN-1 were used to align poorly ordered or divergent regions (Cole et al., 2008; Rost, 1996). It should be noted that the catalytic domain of XPG is separated into two parts (Scharer, 2008) in the primary sequence, thus only the residues predicted to be in the catalytic domain are included in this figure (1-100 and 761-1004).

hExo1 catalytic domain and mutant exonuclease activity assays

DNA substrates

Double-stranded DNA substrates for exonuclease assays had one 5' recessed terminus and one blunt end, and were prepared as follows. A 38 base synthetic oligonucleotide (IDT Inc.) (5□- GGA TCC CCG CTA GCG GGT ACC GAG CTC GAA TTC ACT GG-3□) with a 5□ phosphorylated end was radiolabeled by the kinase exchange reaction in the presence of T4 polynucleotide kinase and γ-^{32}P-ATP (Maniatis et al., 1982). Labeled DNA was separated from unincorporated radiolabel by size exclusion chromatography and annealed with a 2.5 fold molar excess of an unlabeled unphosphorylated 42 base oligonucleotide (5□- CCA GTG AAT TCG AGC TCG GTA CCC GCT AGC GGG GAT CCT CTA -3-). Annealed duplexes were purified by HPLC on a Waters GenPak Fax column as described previously (Blackwell et al., 2001), and had specific activities of 200-500 cpm/fmol. Otherwise identical, but non-radioactive DNA substrates were prepared as above, except that the radiolabeling step was omitted.

Exonuclease assays

Time courses of 5□ hydrolysis by hExo1 catalytic domain were determined in 25-μl reactions containing 0.2 nM enzyme and 6-71 nM radiolabeled DNA. Aliquots (5 μl each) were withdrawn at 15-second intervals and the reactions quenched by addition of 5-μl of 90% formamide. Samples were heated at 90 °C for 5 min prior to electrophoresis on 12% denaturing polyacrylamide gels containing 7M urea in 89 mM Tris/28.5 mM taurine, 0.5 mM EDTA. Gels were dried and visualized after exposure to a phosphorimager screen. Effects of divalent cations on 5' hydrolysis were carried out in a similar manner using 5 mM of $MgCl_2$, $BaCl_2$, $CaCl_2$, or $MnCl_2$.

Mismatch-provoked excision assays. Mismatch-provoked excision reactions were carried out in the buffer described in Experimental Procedures, and contained 24 fmol of a 6440-bp circular G-T heteroduplex (or control A-T homoduplex) with a strand break located 128 bp 5' to the mismatch. Reactions (20-μl) were assembled on ice by addition of 1 μl each of MutSα, MutLα, and RPA, diluted as described (Genschel and Modrich, 2003), to 16 μl of a solution containing all other components except ExoI. Reactions were initiated by addition of 1 μl of ExoI directly to the above solution on ice, and immediately transferred to a 37 °C water bath and incubated for 5 minutes. Samples were deproteinized by Proteinase K treatment followed by phenol extraction. Extent of excision was scored by *Nhe*I resistance assay as described (Genschel et al., 2002 JBC, 2002).

References

Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C. (2002). PHENIX: building new software for automated crystallographic structure determination. Acta Crystallogr D Biol Crystallogr *58*, 1948-1954.

Baker, N.A., Sept, D., Joseph, S., Holst, M.J., and McCammon, J.A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. Proc Natl Acad Sci U S A *98*, 10037-10041.

Blackwell, L.J., Wang, S., and Modrich, P. (2001). DNA chain length dependence of formation and dynamics of hMutSα.hMutLα.heteroduplex complexes. J Biol Chem *276*, 33233-33240.

Bricogne, G., Vonrhein, C., Flensburg, C., Schiltz, M., and Paciorek, W. (2003). Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. Acta Crystallogr D Biol Crystallogr *59*, 2023-2030.

Brünger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.-S., Kuszewski, J., Nilges, M., Pannu, N.S., *et al.* (1998). *Crystallography & NMR System:* A New Software Suite for Macromolecular Structure Determination. Acta Cryst D *54*, 905-921.

Cole, C., Barber, J.D., and Barton, G.J. (2008). The Jpred 3 secondary structure prediction server. Nucleic Acids Res *36*, W197-201.

Davis, I.W., Murray, L.W., Richardson, J.S., and Richardson, D.C. (2004). MOLPROBITY: structure validation and all-atom contact analysis for nucleic acids and their complexes. Nucleic Acids Res *32*, W615-619.

Doublie, S. (1997). Preparation of selenomethionyl proteins for phase determination. Methods Enzymol *276*, 523-530.

Emsley, P., and Cowtan, K. (2004). Coot: Model-Building Tools for Molecular Graphics. Acta Cryst D *60*, 2126-2132.

Eswar, N., Webb, B., Marti-Renom, M.A., Madhusudhan, M.S., Eramian, D., Shen, M.Y., Pieper, U., and Sali, A. (2006). Comparative protein structure modeling using Modeller. Curr Protoc Bioinformatics *Chapter 5*, Unit 5 6.

Genschel, J., Bazemore, L.R., and Modrich, P. (2002). Human exonuclease I is required for 5' and 3' mismatch repair. J Biol Chem *277*, 13302-13311.

Genschel, J., and Modrich, P. (2003). Mechanism of 5'-directed excision in human mismatch repair. Mol Cell *12*, 1077-1086.

Hendrickson, W.A., Horton, J.R., and LeMaster, D.M. (1990). Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. EMBO J *9*, 1665-1672.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., *et al.* (2007). Clustal W and Clustal X version 2.0. Bioinformatics *23*, 2947-2948.

Lee, B.I., and Wilson, D.M., 3rd (1999). The RAD2 domain of human exonuclease 1 exhibits 5' to 3' exonuclease and flap structure-specific endonuclease activities. J Biol Chem *274*, 37763-37769.

Maniatis, T., Fritsch, E.F., and Sambrook, J. (1982). Molecular Cloning (Cold Spring Harbor, Cold Spring Harbor Laboratory).

Otwinowski, Z. (1998). Denzo and Scalepackprograms: An oscillation data processing suite for macromolecular crystallography.

Rost, B. (1996). PHD: predicting one-dimensional protein structure by profile-based neural networks. Methods Enzymol *266*, 525-539.

Scharer, O.D. (2008). XPG: its products and biological roles. ADv Exp Med Biol *637*, 83-92.

Schroder, G.F., Brunger, A.T., and Levitt, M. (2007). Combining efficient conformational sampling with a deformable elastic network model facilitates structure refinement at low resolution. Structure *15*, 1630-1641.

Storoni, L.C., McCoy, A.J., and Read, R.J. (2004). Likelihood-enhanced fast rotation functions. Acta Crystallographica Section D-Biological Crystallography *60*, 432-438.

Winn, M.D., Murshudov, G.N., and Papiz, M.Z. (2003). Macromolecular TLS refinement in REFMAC at moderate resolutions. Methods Enzymol *374*, 300-321.