Supporting Information

Coquelle et al. 10.1073/pnas.1112036108

SI Text

Cloning, Expression, and Purification of PNKP Constructs. The murine PNKP catalytic fragment (residues 141–522) containing an inactivating mutation (D170A, designated $PNKP^{D170A}$) was cloned with an N-terminal His-tag in a pET-16b vector. All the mutations were introduced with the QuikChange II kit (Stratagene).Proteins were overexpressed in BL21 Gold cells with 0.5 mM IPTG for 18 h at 24 °C. Cell pellets were resuspended in lysis buffer (50 mM Tris pH 8, 250 mM NaCl, 0.1% BME) complemented with protease inhibitors (PMSF, leupeptin, and pepstatin). Cells were lysed by sonication and cleared lysate was incubated with Ni-NTA beads (Qiagen) for 2 h. On-column cleavage of the His-tag was performed using 3C protease. Once eluted, the protein was further purified by size exclusion chromatography using a Superdex 75 column (Amersham) in 50 mM Tris pH 8, 150 mM NaCl, 1 mM DTT. The purified protein was finally buffer exchanged into 10 mM Tris pH 8, 150 mM KCl, 1 mM DTT, 10 mM $MgCl_2$.

Fluorescence Polarization. For the direct binding experiments, 30 nM of fluorescent ssDNA (5′-*FAM*-TCCTCp-3′, delineated *FAM*-DNA) was mixed with different protein concentrations from 20 nM to 60 μM in a reaction volume of 20 μL. The polarization was recorded with a Perkin Elmer Envision plate reader at 538 nm and the fluorescein was excited at 458 nm. The K_d values were derived from the plot of the polarization as a function of the log of protein concentration. Competition assays were carried out by increasing the 3′-phosphorylated DNA substrates with a saturated protein concentration and 30 nM *FAM*-DNA. The competing 3′-phosphorylated ssDNA substrates used in this study were PsA (5′-TCCTAp-3′), PsT (5′-TCCTTp-3′), PsC (5′-TCCTCp-3′) and PsG (5′-TCCTGp-3′) while a blunt hairpin was used as a dsDNA substrate (5′-GAGGACTCGCA**GAAT**GCGAGTCC-TCp-3′). The IC50 was obtained after fitting of the resulting sigmoidal curve. All experiments were performed in triplicate and the standard errors are represented in graphs.

Fluorescence Spectroscopy. All 2 amino-purine (2AP) solutions (ssDNA, dsDNA, and protein-DNA complexes) were prepared in buffer containing 10 mM Tris pH 8, 100 mM NaCl, and 1 mM DTT and fluorescence was recorded with DNA concentrations of 500 nM. Four different blunt-ended dsDNA substrates were used with a 3′-phosphate on one strand and a 2AP base on the complementary strand at position 2, 4, 6, and 8 (S2, S4, S6, and S8, respectively; see Fig. 3 for DNA sequences). dsDNAs were annealed by slow cooling from 95 °C. To ensure DNA was bound, a final concentration of 4 μM was used. Experimental conditions were set up to minimize protein fluorescence, but we corrected background fluorescence using a protein-dsDNA that did not contain any 2AP base. For each substrate, the fluorescence intensities were normalized; the fluorescence of the dsDNA substrate corresponding to the minimum (as the 2AP fluorescence is quenched when annealed with the complementary strand), and the fluorescence of the 2AP containing ssDNA was set as the maximum.

Structure Determination. After a brief soaking in cryoprotectant (mother solution complemented with 25% glycerol), crystals were flash frozen in liquid nitrogen. All data were collected at 100 K on diverse synchrotron beamlines (Table S1). Diffraction frames were processed with XDS (1) and scaled with XSCALE (Table S1).

Model Building and Refinement. All structures were solved using the molecular replacement technique with *PHASER* (2). The initial search model was composed of the catalytic segment from the full-length mPNKP structure (1yj5). The solution obtained with this model provided a good fit for the phosphatase domain, however the kinase domain appeared slightly tilted with respect to the phosphatase domain. We therefore repeated the calculation with two models (one for each catalytic subdomain), conserving the orientation of the phosphatase domain and only searching for the kinase domain orientation. This procedure allowed us to obtain correct orientations of the two domains (final *Z*-score of 68 and LLG of 4229 for the apo $PNKP^{D170A}$ structure for example). The refinement was conducted with PHENIX (3). Following an initial rigid body minimization, the refinement procedure was identical for all structures and consisted of refinement of atomic displacement, TLS and individual B-factors. Water molecules were added to the model using the automated water picking option from PHENIX and were checked manually for any close contacts with the protein. Model building was carried out with COOT (4). In the high-resolution apo $PNKP^{D170A}$ structure, a positive density peak in the mFo-DFc map was observed at ∼1.9 Å from the SG atom of Cys450. This distance was too close to model a water molecule and we therefore modeled an oxidized cysteine residue (CSO). After refinement, the geometry was correct and the 2mFo-Fc and mFo-DFc density were in good agreement with a CSO residue at this position. An oxidized cysteine was therefore modeled in three out of the four structures reported in this study.

Two water molecules were modeled in the vicinity of the 3′-phosphate of substrate and of the active site residue 170 (aspartate mutated into alanine in our structure). However, electron density maps clearly indicated that a solvent atom had to be added in our model. Based on multiple criteria, we modeled a Mg^{++} ion in this density. First, our crystallization buffer contains Mg^{++} ions and the position of this density corresponds to its predicted position based on analogous HAD enzyme structures. Moreover, the Mg^{++} ion is coordinated with octahedral geometry (Fig. 1*B*) and bond distances in the 2–2.2 Å range that are indicative of Mg^{++} coordination. The Mg^{++} ion was modeled in our four structures and after refinement, its occupancy ranges from 0.5 to 0.8. For $PNKP^{D170A}$ complexed to DNAs, only two nucleotides could be modeled in the density. The density was not of good enough quality to model the second base in the structure of $PNKP^{D170A}$ complexed with ssDNA 5′-TCCTTp-3′. Coordinates and structure factors of these structures were deposited in the Protein Data Bank (accession code 3u7e, 3u7f, 3u7g, 3u7h for apo $PNKP^{D170A}$ and $PNKP^{D170A}$ bound to ssDNA with a 3′-phosphorylated cytosine, adenine, and thymine, respectively).

1. Kabsch W (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *Journal of Applied Crystallography* 26:795–800.
2. Mccoy AJ, et al. (2007) Phaser crystallographic software. *Journal of Applied Crystallography* 40:658–674.
3. Adams PD, et al. (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallographica Section D-Biological Crystallography* 66:213–221.
4. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of COOT. *Acta Crystallographica Section D-Biological Crystallography* 66:486–501.

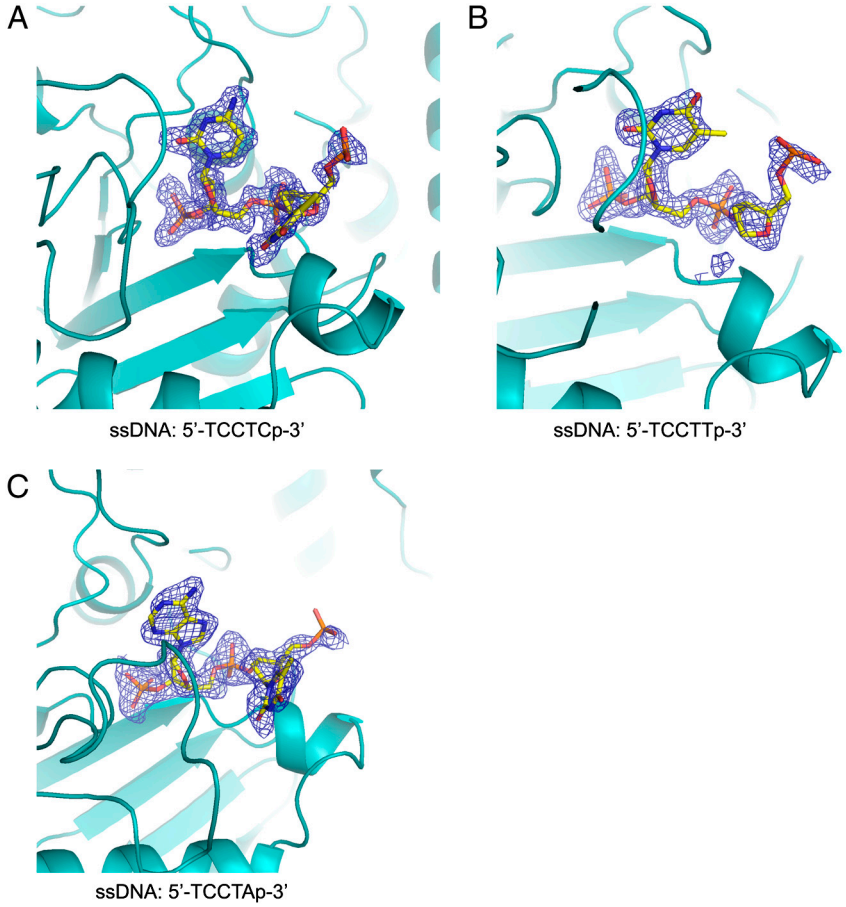

Fig. S1. Fig. S1, related to Fig. 2: 2mFo-DFc electron density map coutoured at 1σ of the different ssDNAs bound to PNKP^{D170A}.

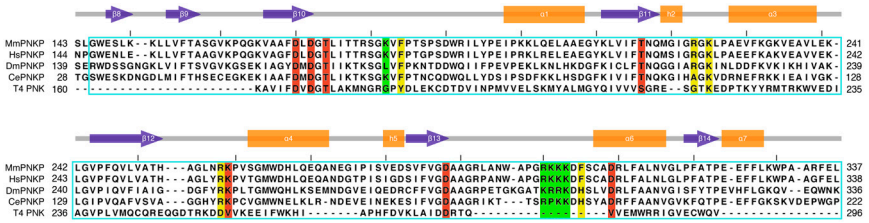

Fig. S2. Fig. S2, related to Fig. 1: Sequence comparison of phosphatase domains from mouse (MmPNKP), human (HsPNKP), fly (DmPNKP), worm (CePNKP), and T4 bacteriophage (T4 PNK) PNKP/PNK enzymes. Initial alignment of metazoan PNKP phosphatase domain (teal outline) sequences was performed and followed by structural alignment with the T4 PNK phosphatase domain. Ten residue intervals for the mouse sequence are marked (|). Secondary structure for MmPKNP is shown and α and 3_{10} helices are specified as "α"and "h", respectively. Key catalytic residues (red) and residues important for recognition and binding of single-stranded DNA substrates (yellow) and the second strand of double-stranded DNA substrates (green) have been highlighted.

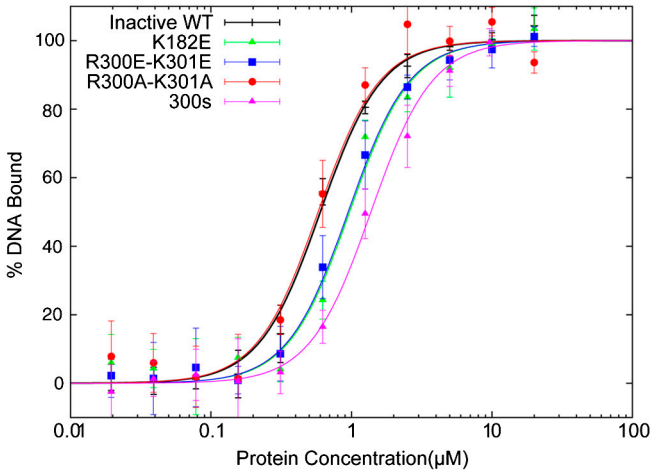

Fig. S3. Fig. S3, related to Fig. 3: Determination of binding affinity of PNKP^{D170A} and various PNKP^{D170A} mutants to *FAM*-DNA using fluorescence polarization.

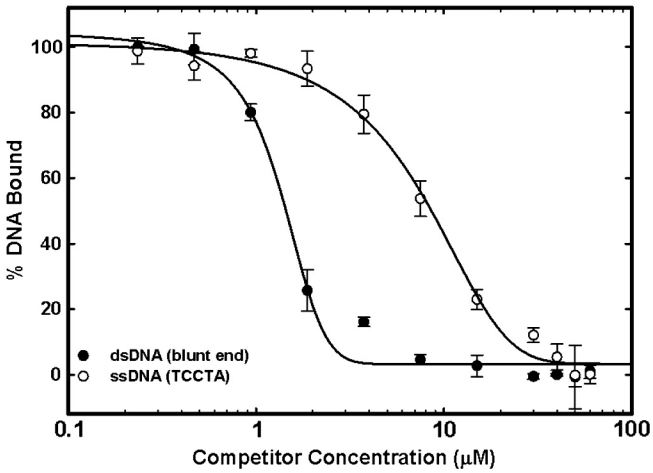

Fig. S4. Fig. S4, related to Fig. 3: Competition of 3′-phosphorylated dsDNA or ssDNAs against a PNKP^{D170A}—*FAM*-DNA complex by fluorescence polarization. The ssDNA is 5′-TCCTA-phosphate-3′, the dsDNA is a hairpin 5′-GAGGACTCGCAGAATGCGAGTCCTC-phosphate-3′. The IC50 for the dsDNA competitor is 1.4 μM, and for the ssDNA competitor, 8.6 μM.

Table S1. Data and refinement statistics

$PNKP^{D170A}$	Apo	TCCTCp	TCCTAp	TCCTTp
Synchrotron	CLS	CLS	ALS	SSRL
Beamline	08ID-1	08ID-1	12.3.1	9.2
Resolution (Å)*	46-1.7 (1.75-1.70)	46-1.8 (1.85-1.80)	40-2.1 (2.15-2.10)	31-2.0 (2.05-2.00)
Space group	$P2_1$	$P2_1$	C2	$P2_1$
Cell dimensions				
a (Å)	42.82	42.60	77.17	42.83
b (Å)	63.19	62.33	43.03	62.53
c (Å)	68.12	67.97	127.29	67.98
(°)	88.23	91.81	107.12	91.70
Completeness (%)	99.8 (99.8)	99.8 (99.5)	90.4 (82.9)	96.5 (92.7)
$\langle I/\sigma I \rangle$	13.9 (2.8)	9.3 (2.6)	10.1 (2.4)	11.3 (1.9)
Redundancy	3.8 (3.8)	3.8 (3.6)	3.2 (2.4)	3.1 (2.1)
R_{sym}	0.065 (0.534)	0.127 (0.620)	0.089 (0.443)	0.083 (0.506)
No. of subunits	1	1	1	1
R_{work}/R_{free}	0.161/0.195	0.169/0.214	0.207/0.26	0.192/0.237
No. of atoms				
Protein	3,002	2,947	2,928	2,881
DNA	0	43	45	32
Water	471	432	202	240
Mg^{++}	1	1	1	1
PO_4^{2-}	5	5	5	0
Glycerol	30	24	18	30
Overall B factor ($Å^2$)	23.6	12.2	31.0	25.3
rmsd				
Bond length (Å)	0.009	0.008	0.009	0.009
Bond angle (°)	1.16	1.17	1.04	1.00
Ramachandran†				
Preferred (%)	98.2	97.6	98.1	97.8
Allowed (%)	1.8	2.1	1.9	1.9
Disallowed (%)	0.0	0.3	0.0	0.3

*Values in parenthesis refers to highest resolution shell.

†As reported by Molprobity Server (http://molprobity.biochem.duke.edu/).