Supplementary Materials

Materials and Methods

Cloning and expression of proteins

Human $PNKP^{wt}$ and $PNKP^{E326K}$ (NP: 0099185), $XRCC4^{1-238/S237A}$ (NP: 072044), and DNA Ligase $IV^{654-911}$ (NP: 996820) were cloned into the pGEX-6P-1 (GE Healthcare) vector for expression as N-terminal GST-fusion proteins (XRCC4 and DNA Ligase IV constructs are referred here on as XRCC4 and LigIV). Mutagenesis was performed by standard PCR and QuikChange Lightning Site-Directed Mutagenesis kit (Stratagene) and verified by sequencing. Clones were transformed into *E. coli* BL21(Gold) cells (Invitrogen) and cultured individually at 37ºC in LB to OD_{600} = 0.8. Cells were induced with 0.2 mM IPTG for XRCC4 and 0.5 mM IPTG for PNKP and DNA LigIV and expressed at 18ºC for 12-16 h. Cultures were pelleted and resuspended in 150 mM KCl, 20 mM Tris pH 8.0, 1 mM EDTA, and 1 mM DTT (Buffer U). Lysis was performed by repeat passages through a cell homogenizer (Avestin Emulsiflex-C3) at 4ºC in the presence of cOmplete Protease Inhibitor tablets (Roche) at the manufacturer's suggested effective final concentration and 0.1 mg/mL PMSF. Lysate was cleared by centrifugation at 40000 g for 30 min and protein-containing supernatant was used for purification.

Purification of proteins and complexes

Wildtype and E326K GST-PNKP supernatant from cleared lysate was bound to Glutathione Sepharose 4B beads (Amersham Biosciences) and washed with Buffer U in Bio-Rad Econo-columns. The GST-tag was cleaved on-column by PreScission protease rolling overnight at 4ºC and cleaved PNKP was collected as flowthrough and wash fractions. Fractions were concentrated, diluted to 20 mM KCl, 50 mM HEPES pH 7.0, 1 mM EDTA, 1 mM DTT, and loaded onto an SP Sepharose (Amersham Biosciences) cation-exchange column. Elution was by a 0-1 M KCl, 50 mM HEPES pH 7.0, 1 mM EDTA, 1 mM DTT buffer gradient. Final PNKP purification was by Superdex 75 26/60 size-exclusion chromatography (Amersham Biosciences) with elution of protein and storage in Buffer U.

To purify XRCC4/LigIV heterodimer initial GST-LigIV cultures were grown at a 1.2:1 volume excess to GST-XRCC4 and both cultures were pelleted together prior to lysis. The GST-XRCC4/GST-LigIV pellet was lysed and centrifuged as indicated and supernatant was bound to GST beads, the GST-tag was removed, and eluted as per PNKP protocol. Fractions were concentrated, diluted to 20 mM KCl, 20 mM Tris pH 8.0, 1 mM EDTA, 1 mM DTT, and loaded onto a Q Sepharose column (Amersham Biosciences). Elution was by 0-750 mM KCl, 20 mM Tris pH 8.0, 1 mM EDTA, 1 mM DTT buffer gradient. Final purification of XRCC4/LigIV was by Superdex 200 16/60 size-exclusion chromatography (Amersham Biosciences) with elution of protein and storage in Buffer U.

pS232/pT233-phosphorylated XRCC4/LigIV (pXRCC4/LigIV) was prepared by *in vitro* CK2α phosphorylation of Q sepharose fractions followed by size-exclusion chromatography as per non-phosphorylated XRCC4/LigIV. Growth, expression, lysis, purification, and storage of XRCC4 (absent LigIV) was as per XRCC4/LigIV protocol with omission of the final size-exclusion step.

Purified recombinant PNKP and pXRCC4/LigIV were used for purification of the PNKP/pXRCC4/LigIV ternary complex. Components were quantitated by Bradford Protein Assay (Bio-Rad) and incubated at 4ºC for 1 h in a 1.2 PNKP:1 XRCC4/LigIV heterodimer molar excess. Complete complex formation was confirmed empirically with Superdex 200 10/300 (Amersham Biosciences) analytical size-exclusion chromatography as seen by a new lower elution volume peak corresponding with pXRCC4/LigIV/PNKP and a small PNKP excess peak. Large-scale preparation of the complex was by Superdex 200 16/60 (Amersham Biosciences) size-exclusion column and elution and storage of the complex was in Buffer U. The presence of XRCC4, LigIV, and PNKP in the ternary complex peak was confirmed by SDS-PAGE.

Final quantitation of the ternary complex and individual proteins for activity and binding assays was by Pierce BCA Protein Assay Kit (Thermo Scientific) using BSA and a standard.

Small-angle X-ray scattering (SAXS)

SAXS data were collected at the SYBILS beamline 12.3.1 (Advanced Light Source, Berkeley, California) using a tunable 1.0-1.5 Å wavelength and Mar165 CCD detector at a 1.5 m sample-detector distance (1,2).

The pXRCC4-LigIV complex was purified on a Superdex 200 10/300 size-exclusion column. Peak fractions were increasingly concentrated (Amicon Ultra, 30 kDa molecular weight cutoff) and quantified by Bradford Protein Assay (Bio-Rad). Concentrated samples at 0.7, 4.1 and 9.1 mg/mL were transferred to dialysis buttons and dialysed into Buffer U. Immediately prior to data collection, samples were centrifuged at 13200 rpm at 4°C for 10 min. and transferred to a 96-well plate. Post-dialysis buffer samples were used to collect data for the matching buffer subtraction. Data with scattering vectors (q) between 0.008 and 0.300 $Å^{-1}$ were collected with three consecutive exposures (6, 60 and 6 s) for each sample. PRIMUS (3) was used to scale and merge the short and long exposures of the 9.1 mg/mL sample. The two scattering curves were scaled using data points 70-130, and merged using data points 1-80 from the short exposure and 70-510 from the long exposure. Guinier analysis was performed in SCATTER [https://bl1231.als.lbl.gov/scatter/], which yielded a radius of gyration (R_g) of 46.3 ± 2.4 Å for data between $s.R_g$ 0.77-1.28, with unbiased residuals in the Guinier plot ($\log(I(q))$ vs q^2) (Supplemental Fig. 2.) Low-q data were truncated to the end of the Guinier region and the maximum particle size (D_{max}) determined and real-space $P(r)$ distance-distribution plot refined using SCATTER. DATGNOM (4) was used to output the smoothed, regularized scattering curve and the low-q data extrapolated to $I(0)$ was

added to the merged reciprocal space data to produce a scattering curve for all subsequent analysis.

The pXRCC4-LigIV-PNKP complex was purified at the SIBYLS beam line from component protein samples (pXRCC4-LigIV and PNKP). The complex was run over a Superdex 200 PC 3.2 (2.4 mL) size-exclusion column. Fractions were collected by drop (one drop per fraction) with five fractions representing each of the leading edge, peak, and tailing edge of the chromatogram trace. Data were collected analogous to the XRCC4/LigIV complex on all fractions. Fraction four from the peak and fraction one from the tail were averaged to improve the signal-to-noise of the scattering data. The resulting curve was processed in SCATTER as for the pXRCC4-LigIV data. The Guinier plot showed a linear dependence of log($I(q)$) with q^2 and a reciprocal-space R_g of 61.8 ± 1.0 Å for data between $s.R_g$ 0.77-1.30 (Supplemental Figure 2). D_{max} and real-space R_g were estimated in SCATTER to be 212 and 63.1 Å respectively.

Kratky plots for both complexes exhibited scattering data typical of partially-folded particles having an initial parabolic curve that starts to plateau at higher-q before reaching base-line (Supplemental Fig. 3). XRCC4/LigIV/PNKP flexibility was further characterized using the methodology outlined in Rambo and Tainer (5) using SCATTER. The first asymptote after the Guinier region in the Porod plot ($q^4.I(q)$ *vs* q) occurred at q ~ 0.15 $Å^{-1}$. Plots for an equivalent q-region of $q^4.I(q)$ vs q^4 (Porod-Debye), $q^3.I(q)$ vs q^3, and $q^2.I(q)$ vs q^2 (Kratky-Debye) revealed a plateau in the $q^3.I(q)$ vs q^3 plot further suggesting a partially-flexible particle for the XRCC4/LigIV/PNKP complex (Supplemental Fig. 4). Molecular masses for both the XRCC4-LigIV and PNKP-pXRCC4-LigIV complexes were estimated from the scattering data using the Q_R parameter as described (6) and implemented in SCATTER.

Ensemble modelling of pXRCC4-LigIV-PNKP was carried out with a combination of BILBOMD v1.9 (http://sibyls.als.lbl.gov/bilbomd) (7), FoXS (8) and GAJOE v2.1 (9). Crystal structures for mouse PNKP (PDB ID: 1YJ5), XRCC4-LigIV (PDB ID: 3II6), and mouse PNKP-FHA bound to cognate human XRCC4 phosphopeptide (PDB ID: 1YJM) were used to construct an initial model. Flexible polypeptide linkers were built to link the PNKP catalytic and FHA domains as well as between the XRCC4 phosphopetide and the C-terminal residue of the XRCC4 crystal structure. Three variants were modelled: two models had the PNKP FHA domain bound to either the crystallographic ‘A’ or ‘B’ chains of XRCC4; the third had PNKP FHA domains bound to both the ‘A’ and ‘B’ chains. During the BILBOMD molecular dynamic simulations, the coordinates of both chains of XRCC4 (residues 1-201) and ligase IV (residues 1:256) were fixed. The catalytic domain of PNKP (residues 137-519), and the FHA domain of PNKP (residues 1-107) with phosphorylated XRCC4 peptide bound (residues 229-235) were treated as rigid bodies, and linkers between rigid bodies as flexible polypeptide. These three models were submitted to independent BILBOMD modelling, requesting 800 conformations per 5 Å R_g bin between 40 and 95 Å. Approximately 9600 conformations derived

from each of the three starting model variants provided 28800 models in the pool for analysis by the genetic algorithms. FoXS was used to calculate theoretical scattering profiles for all models. GAJOE was used to determine minimal ensembles from the three individual model pools as well as all three pools combined using the FoXS derived scattering intensities. Default parameters for GAJOE were used with the following exceptions: number of generations was increased to 8000; number of ensembles was increased to 100; number of cycles of the genetic algorithm was increased to 200. GAJOE was repeated multiple times for each pool requesting an increasing number of models in the final ensemble for each run until adding additional models to the ensemble failed to significantly improve the ensemble fit to the experimental data.

UV laser microirradiation

Wild type human PNKP (accession number AF125807.1) was cloned into pEGFP-C2 (Clontech) to create pEGFP-C2-PNKP-wt. Vectors for expression of patient and FHA domain mutants (pEGFP-C2-PNKP-E326K and pEGFP-C2-PNKP-R35A respectively), were generated from the wt clone using a Quikchange mutagenesis kit (Stratagene). Primer sequences are available upon request. Sequences of all clones were confirmed by the University of Calgary DNA Sequencing Facility.

Asynchronously growing U2OS cells were cultured in RPMI1640 media (Life Technologies Gibco) containing 10 % Hyclone III fetal calf serum (GE Healthcare), 50 µg/mL streptomycin and 50 units/mL penicillin. Twenty-four hours prior to transfection, cells were seeded at ~50-70% confluency onto round coverslips (Fisherbrand, 12-545-102 25CIR-1) in 3 cm tissue culture dishes (Falcon, 353001) then transiently transfected with 1 µg pEGFP-C2-PNKP plasmid (WT, E326K or R35A) with Metafectene PRO (Biointex, TO40-1.0) for twenty-four hours transfection reagent according to the manufacturer's recommended conditions.

UV laser microbeam irradiation was carried out on a 355 nm CryLas FTSS355-50 UV-Laser PALM MicroBeam system as described previously (10). Briefly, 16-24 hours prior to live cell imaging, EGFP-PNKP transfected U2OS cells were pre-sensitized with 10 µM bromodeoxyuridine (BrdU, BD Pharmingen, 55091). Twenty-four hours after transfection, coverslips were placed into a metal cassette, washed 2 x with PBS (137 mM NaCL, 10 mM Na_2HPO_4, 2.7 mM KCl, and 1.8 mM KH_2PO_4) and 1 mL RPMI media lacking phenol red (Life Technologies Gibco, 11835-030) was added to the coverslips. The coverslip cassette was then placed into a chamber, pre-conditioned to 5% CO_2 and 37°C. A laser intensity of 35% energy, 15 µm/sec cutting speed, and stage focus of 66 were used in all

experiments. Live imaging settings were consistent for all experiments. Images were taken every 30 seconds for a 20-minute time period. Quantification of laser tracks was performed by subtracting the background fluorescence signal of the GFP channel prior to boxing the region of the induced laser track. The fluorescence intensity of a box of equal area adjacent to DNA damage track was subtracted to account for nuclear background cell fluorescence. This procedure was carried out on each individual laser damaged cell analyzed in order to account for changes in EGFP-PNKP expression under transient transfection conditions.

A total of 30 cells were quantified using ZenBlue ROI imaging analysis software (Zen Blue 2) for each data point. Intensities shown are a comparison of fold increase in PNKP localization to DNA induced laser tracks compared to pre-damage GFP intensities, with subtraction of background cell fluorescence. Mean intensity with standard error of the mean are shown. Statistical significance was determined using Student's t-test. P values less than 0.05 were considered significant.

Immunoprecipitation

U2OS cells were transfected with pEGFP-C2-PNKP-WT, pEGFP-C2--PNKP-R35A or pEGFP-C2-PNKP E326K at 1µg DNA/10 cm plate using Metafectene as described above. Twenty-four hours after transfection, one set of cells was irradiated at 10 Gy IR using a GammaCell 1000 Elite ^{137}Cs source (dose rate 2.9 Gy/min) and allowed to recover at 37^{o}C for 1hr. Cells were harvested and lysed in NETN buffer (150 mM NaCl, 1 mM EDTA, 50 mM Tris-HCl pH7.5 and 1%(v/v) NP-40) containing protease and phosphatase inhibitors. Samples were diluted to 10 mM Tris-HCl pH7.5, 150 mM NaCl, 0.5 mM EDTA, 0.2% NP-40 and an amount of extract equivalent to 0.5 mg total protein (determined using the BioRad Detergent Compatible Protein Assay with BSA as standard) was immunoprecipitated with 15 µl GFP-Trap beads (Chromotek) at 4^{o}C for 4 hours. Beads were washed 6 times with 1 ml of NETN buffer containing 0.25% NP-40 plus protease and phosphatase inhibitors per wash then analyzed on 10% acrylamide SDS-PAGE gels and immunoblotted for GFP (Sigma, G1544) XRCC1 (Abcam Ab47920) or XRCC4 (Serotec AHP387) as indicated. Fifty µg of whole cell extract was run on a separate gel as input. All gels were incubated with ECL reagent and exposed to X-ray film (Fuji) for an identical amount of time. Images were quantitated from film using Image Quant (GE Healthcare).

Kinase activity assay

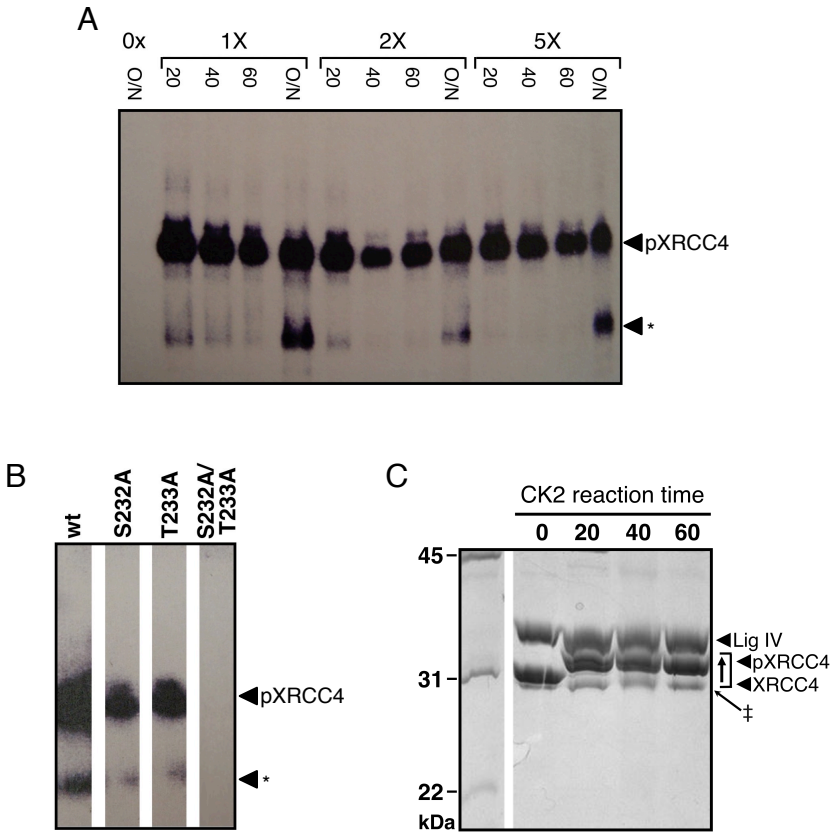
PNKP 5'-DNA kinase activity was determined as a function of ATP to ADP conversion coupled to oxidation of NADH to NAD^+ as part of the pyruvate kinase/lactate dehydrogenase (PK/LDH) ATPase assay. NADH consumption was tracked by measuring fluorescence (relative fluorescence units, RFUs) following excitation at 355 nm and emission at 460 nm. The phosphorylation activity of $PNKP^{wt}$ and $PNKP^{E326K}$ on recessed 5'-OH termini of a 50/45 nt DNA duplex substrate was tested in the presence and absence of XRCC4/LigIV and pXRCC4/LigIV. The DNA kinase substrate (5-nt 3'-overhang/recessed 5'-OH) was made by annealing oligonucleotides with the sequences 5'-P-GCTAGTGGTGGTGGGCGCCGGCGGTGTGGGCATTCGTAATCCTGACCTGA-3'-OH and 5'-OH-TCAGGATTACGAATGCCCACACCGCCGGCGCCCACCACCACTAGC-3'-OH (Integrated DNA Technologies) 95ºC for 15 min and slow cooled to room temperature. Annealed DNA was serially diluted to assess enzyme kinetics at 0, 2.5, 5, 10, 15, 20, 30, 40, 60, 80, 100 µM final substrate concentrations. Final reagent concentrations in reactions were NADH = 100 µM, phosphoenolpyruvate (PEP) = 1 µM, PK/LDH = 0.1 U/µL, and ATP = 2 mM (listed reagents from Sigma-Aldrich). Proteins in respective reactions were at final concentrations PNKP = 100 nM and XRCC4/LigIV = 1 µM. A standard curve of NADH at 0, 20, 40, 60, 80, 100, 120 µM was set up in duplicate along with the assays for conversion of initial rates from ΔNADH RFUs/sec to µmol of product/sec.

Plates containing NADH, PEP, PK/LDH, protein(s) of interest, and substrate DNA, were shaken for 5 sec and pre-incubated at 30ºC for 15 min prior to starting the reactions. Reactions for the complete DNA substrate concentration series were started simultaneously by multi-channel pipette addition of ATP dissolved in 10X kinase activity buffer (0.8 M Na-succinate pH 5.5, 0.1 $MgCl_2$, and 10 mM DTT) at a final volume of 20 µL and carried out at 30ºC. Nine replicates were completed for all experimental reactions and negative controls with XRCC4/LigIV and pXRCC4/LigIV (absent PNKP) were performed in triplicate.

Fluorescence measurements were performed in OptiPlate-384 F HB Black 384-well microplates (Perkin-Elmer) using a FLUOstar Optima plate reader and OPTIMA V1.30 R6 Reader Control software (BMG Labtech). Program parameters included 120 cycles at 5 flashes/sample and 22 sec cycles. Linear regression analysis of raw kinetics data and curve fitting for initial rate determinations was done by the program Kinetics (Coquelle, N., Glover Lab, University of Alberta). Prism 6 (GraphPad Software, Inc.) was used for Michaelis-Menten kinetics analysis: to plot initial rate vs substrate concentration graphs, for non-linear regression curve fitting, and to calculate of k_{cat} and K_m values.

1. Hura, G.L., Menon, A.L., Hammel, M., Rambo, R.P., Poole, F.L., 2nd, Tsutakawa, S.E., Jenney, F.E., Jr., Classen, S., Frankel, K.A., Hopkins, R.C. *et al.* (2009) Robust, high-throughput solution structural analyses by small angle X-ray scattering (SAXS). *Nat Methods*, **6**, 606-612.
2. Classen, S., Hura, G.L., Holton, J.M., Rambo, R.P., Rodic, I., McGuire, P.J., Dyer, K., Hammel, M., Meigs, G., Frankel, K.A. *et al.* (2013) Implementation and performance of SIBYLS: a dual endstation small-angle X-ray scattering and macromolecular crystallography beamline at the Advanced Light Source. *J. Appl. Cryst.*, **46**, 1-13.
3. Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch M.H.J. & Svergun, D.I. (2003) PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Cryst.*, **36**, 1277-1282.
4. Petoukhov, M.V., Konarev, P.V., Kikhney, A.G. and Svergun, D.I. (2007) ATSAS 2.1 - towards automated and web-supported small-angle scattering data analysis. *J. Appl. Cryst.*, **40**, S223-S228.
5. Rambo, R.P. and Tainer, J.A. (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers*, **95**, 559-571.
6. Rambo, R.P. and Tainer, J.A. (2013) Accurate assessment of mass, models and resolution by small-angle scattering. *Nature*, **496**, 477-481.
7. Pelikan, M., Hura, G.L. and Hammel, M. (2009) Structure and Flexibility within proteins as identified through small angle X-ray scattering. *General Physiology and Biophysics*, **28**, 174 - 189.
8. Schneidman-Duhovny, D., Hammel, M. and Sali, A. (2010) FoXS: a web server for rapid computation and fitting of SAXS profiles. *Nucleic Acids Res*, **38**, W540-544.
9. Bernado, P., Mylonas, E., Petoukhov, M.V., Blackledge, M. and Svergun, D.I. (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc*, **129**, 5656-5664.
10. Klement, K., Luijsterburg, M.S., Pinder, J.B., Cena, C.S., Del Nero, V., Wintersinger, C.M., Dellaire, G., van Attikum, H. and Goodarzi, A.A. (2014) Opposing ISWI- and CHD-class chromatin remodeling activities orchestrate heterochromatic DNA repair. *The Journal of Cell Biology*, **207**, 717-733.

Supplementary Figure 1



A. Phosphorylation of XRCC4-LigIV by CK2. Phosphorylation was assayed by CK2 incorporation of ^{32}P from **γ-**^{32}P-ATP into XRCC4 by SDS-PAGE. ^{32}P incorporation is at maximum within 20 min at a 1 XRCC4:0.05 CK2 molar ratio (1X). Increased CK2 concentration (2X and 5X) and reaction time to 40 and 60 mins or overnight (O/N) do not yield additional phosphorylation. Degradation of phosphorylated-XRCC4 (*) is increased in O/N reactions.

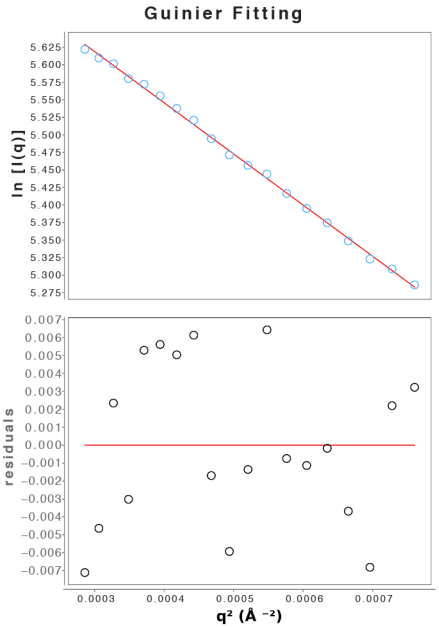
B. Phosphorylation of XRCC4/LigIV by CK2 is limited to XRCC4 C-terminal tail residues S232 and T233. The phosphorylation profile of XRCC4 by CK2 was determined by ^{32}P-incorporation (**γ-**^{32}P-ATP) in XRCC4 wt, XRCC4 S232A and XRCC4 T233A single-mutants, and the XRCC4 S232A/T233A double-mutant by SDS-PAGE. Phosphorylation is present at similar levels in both S232A and T233A single-mutants. Phosphorylation is completely absent in the XRCC4 S232A/T233A double-mutant.

C. CK2 phosphorylation of XRCC4/Lig IV causes XRCC4 mobility shift. Phos-Tag SDS-PAGE (phosphate-affinity polyacrylamide gel electrophoresis) was used to detect XRCC4 phosphorylation status. Band separation between XRCC4 and phosphorylated-XRCC4 was present within 20 min at 1X CK2. No further migration was seen with increased reaction time to 40 and 60 min. ‡ Indicates a non-phosphorylated contaminant.

Supplementary Figure 2

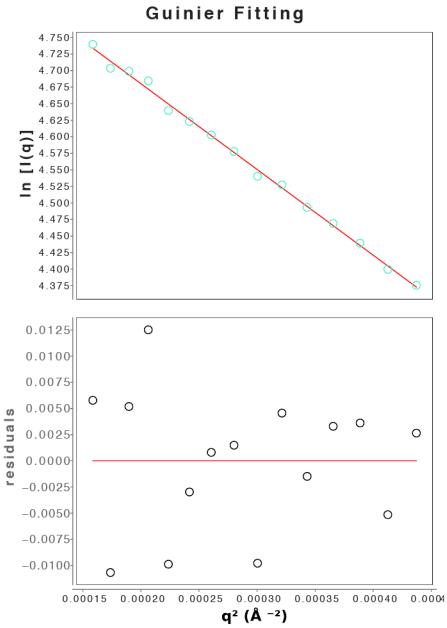
A

Guinier Plot XRCC4/Ligase IV



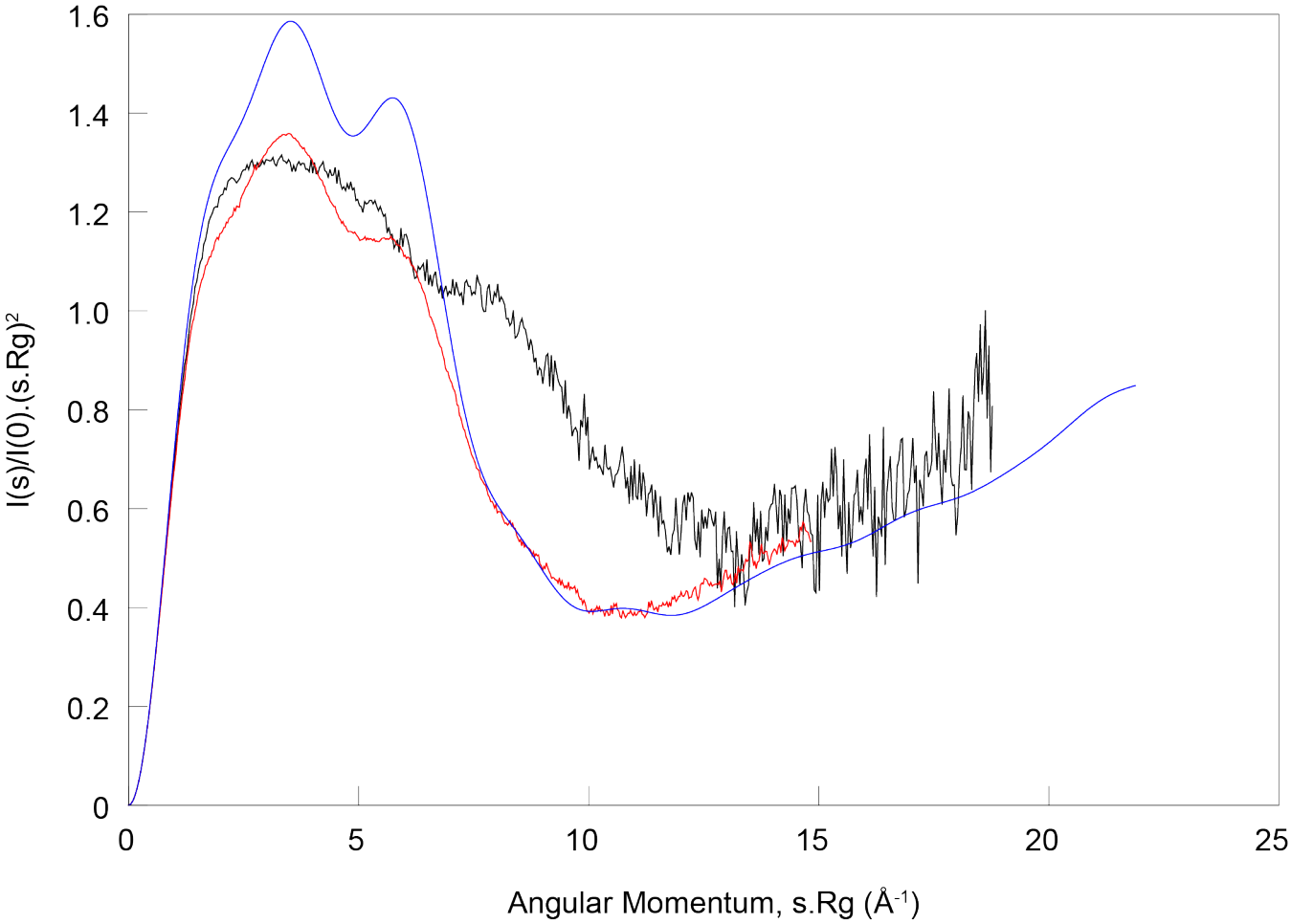
B

Guinier Plot XRCC4/Ligase IV/PNKP



Guinier analysis of SAXS data. Guinier plots are shown for XRCC4/LIGASE IV (A) and XRCC4/LIGASE IV/ PNKP (B). Deviations from the linear fit (red line) are shown in the residuals plot in the lower panels.

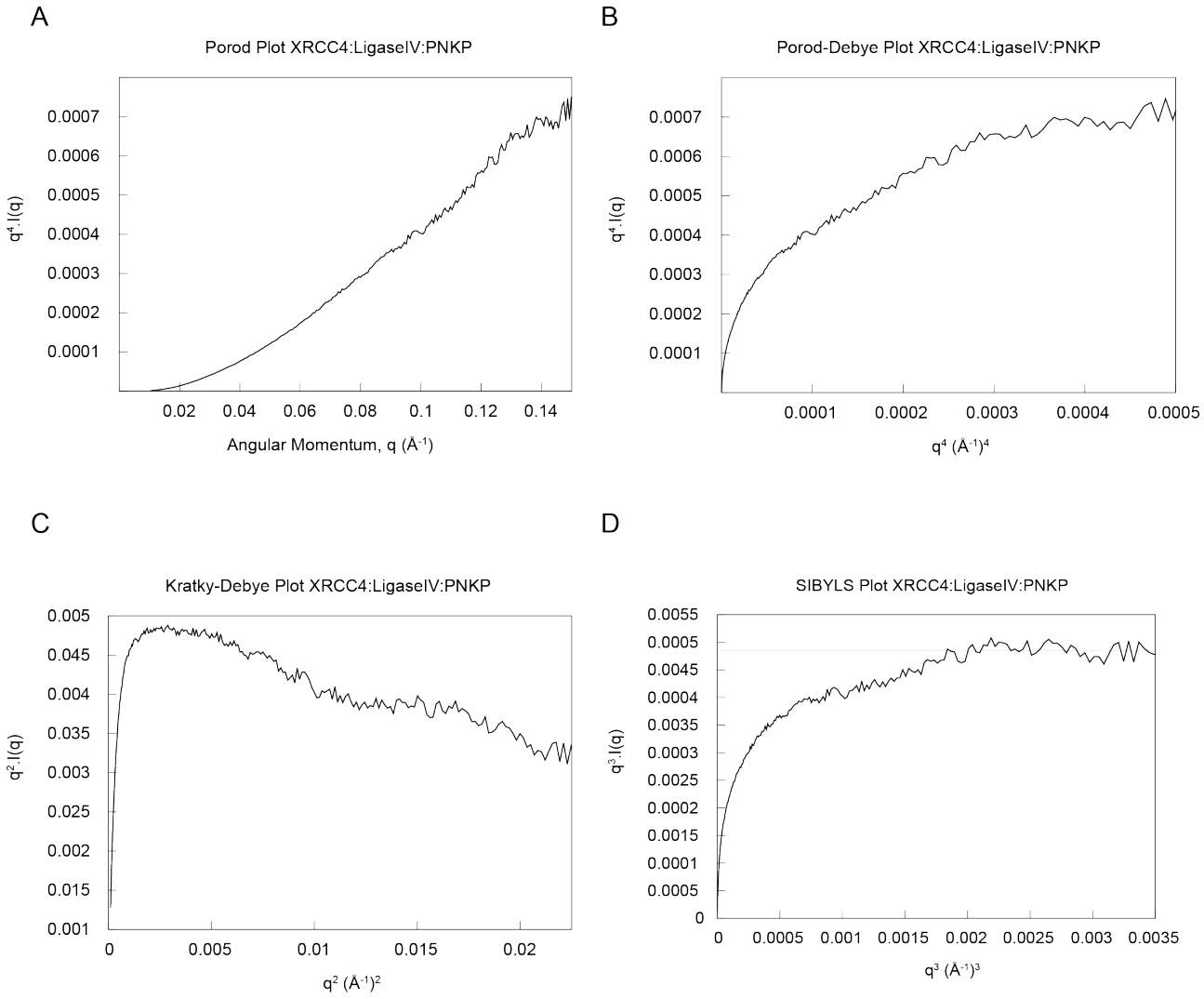
Supplementary Figure 3



Normalized Kratky plots of SAXS data. XRCC4/LIGASE IV/PNKP is in black, XRCC4/LIGASE IV is in red, and XRCC4/LIGASE IV [pdbcode:3II6] is in blue.

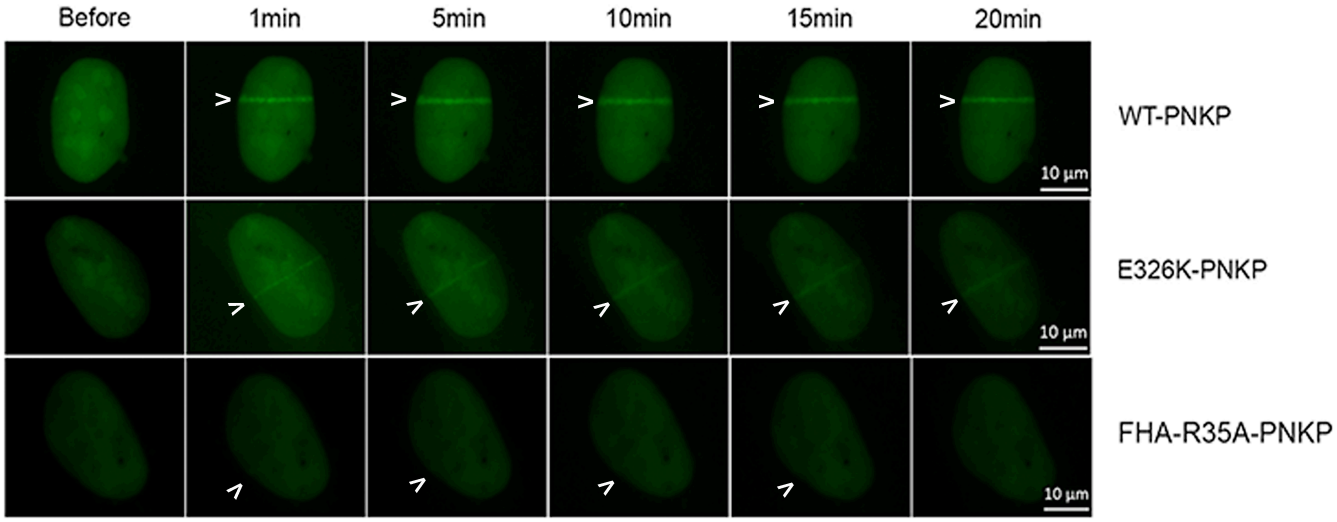
.

Supplementary Figure 4



Assessment of PNKP-pXRCC4-LigIV flexibility using the Porod-Debye law (ref. 5). (A) A Porod plot is used to define the range of q useful for subsequent Porod-Debye analysis. (B) Using the data range defined in (A), a Porod-Debye plot is used to assess particle-solvent contrast. The lack of a well-defined asymptotic plateau region suggests significantly reduced contrast compared to well-folded proteins and suggests significant conformational flexibility. (C) The Kratky-Debye plot is used to further probe conformational flexibility. Completely disordered proteins exhibit an asymptotic plateau in these plots. The fact that no plateau is obtained suggests the complex retains a significant degree of folded character. (D) The SYBYLS plot does display an asymptotic plateau (light grey line), which is a hallmark of partially folded proteins.

Supplementary Figure 5



E326K mutation abrogates recruitment of PNKP to sites of DNA damage. Shown are expanded images of nuclei shown in Figure 6B. White arrow heads indicate positions of laser tracks.