Solution Structure of Polymerase *μ*'s BRCT Domain Reveals an Element Essential for Its Role in Nonhomologous End Joining[†]

Eugene F. DeRose[‡], **Michael W. Clarkson**[§], **Steven A. Gilmore**[‖,⊥], **Cristina J. Galban**[⊥], **Ashutosh Tripathy**[§], **Jody M. Havener**[§,‖], **Geoffrey A. Mueller**[‡], **Dale A. Ramsden**[§,‖], **Robert E. London**[‡], and **Andrew L. Lee**[*,§,‖,⊥]

Laboratory of Structural Biology, National Institute of Environmental Health Sciences, 111 T. W. Alexander Drive, Research, Triangle Park, North Carolina 27709, and Department of Biochemistry and Biophysics, Lineberger Comprehensive Cancer Center, and School of Pharmacy, University of North Carolina, Chapel Hill, North Carolina 27599

Abstract

The solution structure and dynamics of the BRCT domain from human DNA polymerase *μ*, implicated in repair of chromosome breaks by nonhomologous end joining (NHEJ), has been determined using NMR methods. BRCT domains are typically involved in protein—protein interactions between factors required for the cellular response to DNA damage. The pol *μ* BRCT domain is atypical in that, unlike other reported BRCT structures, the pol *μ* BRCT is neither part of a tandem grouping, nor does it appear to form stable homodimers. Although the sequence of the pol *μ* BRCT domain has some unique characteristics, particularly the presence of > 10% proline residues, it forms the characteristic α*β*α sandwich, in which three alpha helices are arrayed around a central four-stranded *β*-sheet. The structure of helix α1 is characterized by two solvent-exposed hydrophobic residues, F46 and L50, suggesting that this element may play a role in mediating interactions of pol *μ* with other proteins. Consistent with this argument, mutation of these residues, as well as the proximal, conserved residue R43, specifically blocked the ability of pol *μ* to efficiently work together with NHEJ factors Ku and XRCC4-ligase IV to join noncomplementary ends together in vitro. The structural, dynamic, and biochemical evidence reported here identifies a functional surface in the pol *μ* BRCT domain critical for promoting assembly and activity of the NHEJ machinery. Further, the similarity between the interaction regions of the BRCT domains of pol *μ* and TdT support the conclusion that they participate in NHEJ as alternate polymerases.

---

* Address correspondence to Dr. Andrew L. Lee, Division of Medicinal Chemistry and Natural Products, School of Pharmacy, Beard Hall - CB#7360, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599; tel: (919) 966-7821; fax: (919) 843-5150; e-mail: drewlee@unc.edu.
[‡]National Institute of Environmental Health Sciences.
[§]Department of Biochemistry and Biophysics, University of North Carolina.
[‖]Lineberger Comprehensive Cancer Center, University of North Carolina.
[⊥]School of Pharmacy, University of North Carolina.

NOTE ADDED IN PROOF

Coordinates for the pol *μ* BRCT domain determined as part of the Riken structural genomics consortium also have recently been deposited in the Protein Data Bank (2DUN). The structures are in close agreement, with the backbone rmsd over the structured region (V30–E124) of 1.27 Å.

SUPPORTING INFORMATION AVAILABLE

(1) ^{1}H–^{15}N HSQC spectrum of pol *μ* BRCT domain, (2) results from light scattering experiments to assess the monomeric/oligomeric state of the pol *μ* BRCT domain, (3) ^{1}H 1D spectra of WT, F46A, L50A, and R43A BRCT variants, and (4) an overlay of the pol *μ* BRCT domain with the C-terminal BRCT domain of XRCC1. This material is available free of charge via the Internet at http://pubs.acs.org.

DNA polymerase μ (pol μ) is a recently identified X-family polymerase that includes both a catalytic and a BRCA1 C-terminal homology (BRCT[1]) domain (1,2). The sequence of the enzyme is most similar to terminal deoxynucleotidyltransferase (TdT), with which it shares 41% sequence identity. TdT, pol μ, and pol λ (another, more distantly related pol X family member) have all been implicated in the nonhomologous end joining (NHEJ) pathway for repair of double strand breaks (DSBs) (3–7). These three polymerases provide this pathway with an apparent gradient of template strand dependence (TdT < pol μ < pol λ) (8), in keeping with the different roles NHEJ must play during V(D)J recombination (8,9) and general repair of double strand breaks (10).

All pol X members linked to NHEJ, including the sole pol X member in *S. cerevisiae,* possess an N-terminal BRCT domain required for a protein—protein interaction with the NHEJ ligase complex (XRCC4-ligase IV) (6,11). Another NHEJ factor, Ku, can then recruit both XRCC4-ligase IV and pol X members to DNA ends (5,8). The BRCT domains were determined to be essential for NHEJ activity both in vitro (3–5) and in cells (8). The three polymerases also compete with each other to form complexes with NHEJ factors in vitro (6,8) and in cells (8, 9), suggesting the three different BRCT domains interact with the same surface.

The BRCT domain is an abundant structural unit (12,13) found in more than 1000 proteins from all biological kingdoms. Structurally, it is typically an $\alpha\beta\alpha$ sandwich, characterized by four parallel β strands forming a central β-sheet and flanked by three α-helices (14–19). Nearly all of the proteins containing this motif are involved in the cell cycle checkpoint response to damaged DNA and/or more directly in DNA replication or repair.

BRCT domains most often function as protein binding domains; e.g. XRCC1, a protein involved in repair of single-stranded DNA breaks, binds DNA ligase III via its C-terminal BRCT domain (20), whereas its N-terminal BRCT domain specifically binds poly(ADP-ribose) polymerase (21). Recently, it has been shown that many BRCT domains possess a phosphoserine-specific, protein binding function (22–25). Despite the structural homology observed among BRCT domains, the modes of interaction with known ligands vary widely. For example, the DNA binding domain of p53 contacts α3 of the N-terminal BRCT domain from p53-binding protein 1, 53BP1 (14,15). Alternatively, BRCA1 tandem BRCT domains use residues from α1 and other regions to form a structural unit that specifically binds to a phosphoserine-containing sequence in the protein BACH1 (26–28). Hence, despite conservation of the three-dimensional structure of each domain, the mechanism by which BRCT domains execute their function differs significantly within the BRCT superfamily. In order to better understand the biological role of pol μ, we have investigated the structural and dynamic behavior of its BRCT domain. The resulting structure provides useful clues regarding the regions of interaction with other proteins. One surface on the pol μ BRCT domain was investigated for its functional role in complex formation with Ku and XRCC4-ligase IV, as well as in end-joining. Further, a comparison with the recently determined structure of the BRCT domain derived from TdT (PDB ID 2COE, Nagashima, et al., unpublished) supports the conclusion that the two proteins may target similar structures and thus work as alternate polymerases in the NHEJ repair complex.

---

[1]Abbreviations: NHEJ, nonhomologous end joining; DSB, double strand break; BRCT, BRCA-1 C-terminal; TdT, terminal deoxynucleotidyltransferase; NOE, nuclear Overhauser effect; CSI, chemical shift index; RDCs, residual dipolar couplings; RMSD, root-mean-square deviation.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification

The polymerase μ BRCT domain gene was subcloned into a pET-28a vector and transformed into BL21(DE3) *E. coli.* The 105 residue BRCT domain construct contained residues 21–124 of the full length protein with an additional Gly at the N-terminus. To increase expression, the first 12 codons were optimized for expression in *E. coli.* Cells were grown in M9 media with $^{15}NH_4Cl$ as the sole nitrogen source, and U-^{13}C glucose as the carbon source for doubly labeled preparations. Cultures were grown to an OD_{600} of 0.6–0.8 at 37 °C and then cold-shocked on ice for 5 min. IPTG was added to a final concentration of 400 μM to induce expression, and culture growth was continued for 5 h at 22 °C. Cells were harvested by centrifugation for 20 min at 6400g and the pellets resuspended in 40 mL of 25 mM Tris-HCl, 5 mM DTT, pH 8.0.

Cell slurries were subjected to three rounds of freeze–thaw, followed by no more than five 3-min rounds of sonication at a 30% duty cycle. Cell debris was removed by centrifugation for 30 min at 23 000g, and the supernatant was dialyzed overnight against 25 mM Tris-HCl, 5 mM DTT, pH 7.9. The lysate was then fractionated over SP-fast flow sepharose (Pharmacia), and the BRCT domain was eluted with 150 mM NaCl. The protease inhibitor AEBSF (Roche) was added to the eluted protein at the maximum recommended concentration (1 mg/mL) in order to limit proteolysis during the extended periods of data accumulation. BRCT was then purified to 99% over a G-75 superdex (GE-Amersham) column. The concentrated eluate was dialyzed overnight at 4 °C against NMR buffer (25 mM *d*-Tris-HCl, 50 mM KCl, 0.02% NaN_3, 5 mM DTT, pH 7.9). The final protein concentration was 0.5 mM.

NMR Chemical Shift Assignments

Most NMR experiments were carried out using a Varian 600 MHz INOVA spectrometer, equipped with a 5 mm Varian $^1H\{^{13}C,^{15}N\}$ Cold Probe. Backbone amide residual dipolar couplings (RDCs) were measured using a Varian 800 MHz INOVA spectrometer, also equipped with a 5 mm Varian $^1H\{^{13}C,^{15}N\}$ Cold Probe. All the NMR experiments were performed at 10 °C. Proton chemical shifts were referenced to internal DSS at 0.00 ppm, and the heteronuclear chemical shifts were referenced based on the ratio of heteronuclear and proton gyromagnetic ratios (29). The NMR data were processed with NMRPipe (30), and the spectra were assigned with NMR-View (31).

Unless stated otherwise, all NMR experiments were implemented using Varian's BioPack pulse sequences. The sequential backbone and C^β chemical shifts were assigned from a combined analysis of HNCA (32,33), HNCACB (34,35), CBCA(CO)NH (36,37), and HNCO (32,34) experiments. Several proline backbone resonances were assigned using HCA(CO)N (38) and the HACAN (39) experiments.

Most side-chain chemical shifts, including prolines, were assigned from a combined analysis of H(CCO)NH–TOCSY and (H)C(CO)NH–TOCSY experiments (40–42). Both TOCSY experiments were acquired at two different mixing times, i.e., 12.1 and 18.2 ms, using a ^{13}C spin-lock field of 9 kHz. Additional side-chain chemical shifts were assigned from a combined analysis of BioPack HCCH–TOCSY (43) and 3D ^{15}N-separated NOESY (44) experiments. The ^{13}C isotropic mixing time in the HCCH–TOCSY experiment was 12.1 ms using a ^{13}C spin-lock field of 9 kHz. Phenylalanine and tyrosine H^δ, C^δ, H^ε, and C^ε and tryptophan $H^{\delta 1}$ and $C^{\delta 1}$ chemical shifts were assigned from a combined analysis of (HB)CB(CGCD)HD, (HB)CB(CGCDCE)HE (45), and ^1H-^{13}C HSQC experiments. All chemical shift assignments have been deposited into the BMRB (http://www.bmrb.wisc.edu) as entry 7259.

All backbone amide chemical shifts, except those of T21, S42, C62, S63, S64, and C119 were assigned. The amide nitrogen, H^α, and C^α shifts of T21 were assigned from a combined analysis of HCA(CO)N and HACAN spectra. The side-chain chemical shifts of T21 were assigned from the HCCH–TOCSY spectrum. The side-chain chemical shifts of S42, S64, and C119 were assigned from CBCA(CO)NH, (H)C(CO)NH, H(CCO)NH, and HCCH–TOCSY spectra. We were not able to assign any C62 or S63 chemical shifts, possibly indicating intermediate conformational exchange for the loop containing these residues. The 105-residue BRCT domain has 10 proline residues, including 3 diproline stretches. Among the proline residues, only P96 was completely unassigned. The C^β, H^β, C^δ, and H^δ chemical shifts of P91, preceding P92, were assigned from 3D NOESY and HCCH–TOCSY data. Ninety-one percent of the proton chemical shifts were assigned for the structure calculation using CYANA.

NOE Experiment

NOE cross-peaks were assigned from a version of the CN–NOESY–HSQC experiment (46) obtained from Lewis Kay, using a 100 ms mixing time. In this NOESY experiment, the ^{13}C carrier frequency was set to 67.0 ppm to observe NOEs to both aromatic and aliphatic protons, facilitating the assignment of NOEs involving aromatic protons. The CN–NOESY–HSQC was obtained in 90% H_2O/10% D_2O to observe NOEs from amide protons.

Measurement of Residual Dipolar Coupling Constants

Backbone amide ^1H–^{15}N residual dipolar couplings (RDCs) were measured using a spin state selective HSQC experiment (47), implemented in BioPack. The alignment medium for the anisotropic phase contained C8E5 (alkyl-poly(ethylene glycol)) and *n*-octanol in a molar ratio of 0.87 and C8E5 to buffer ratio of 2.5 wt % (48) in the same buffer used to make the isotropic measurements. The alignment medium produced a 16.2 Hz ^2H quadrupolar splitting of the HDO resonance. A total of 76 RDCs were unambiguously assigned, ranging in values from −11.2 to 8.8 Hz.

Structure Calculations

Preliminary structures were generated using CYANA 2.1 and its automated NOE assignment algorithm (49,50). The new iterative option was used to improve the calibration of NOE cross-peaks in the presence of weak noise peaks. No additional manual NOEs or hydrogen bond restraints were used in the calculation. The chemical shift tolerances were set to 0.040, 0.030, and 0.45 ppm, corresponding to the indirectly detected proton, the directly detected proton, and the heteronuclear chemical shift dimensions in the 3D NOESY HSQC spectra. In CYANA, the number of initial structures in each cycle was set to 100 and the number of final structures set to 20. The number of torsion angle dynamics steps was set to 10000. A total of 147 φ and ψ dihedral angle restraints were also used in the calculation. The dihedral angle restraints were set to a range of ±2 times the standard deviation or at least ±20% about the average values predicted by TALOS (51). CSI predictions were used for angles that did not meet TALOS acceptance criteria. The angles were set to $\varphi = -70(\pm50)$ and $\psi = -50(\pm50)$ for the alpha helices and $\varphi = -140(\pm60)$ and $\psi = +130(\pm90)$ for the beta strands. Initially, CYANA assigned 1494 NOEs; 13 of these NOEs were incorrectly assigned because of degeneracies in the chemical shifts of P22/P23 and P114/P116. A second calculation was performed without the proton assignments of residues P3 and P4. In this calculation, CYANA assigned 1503 NOEs.

The final 20 CYANA structures from the second CYANA calculation were subjected to a single round of water refinement using the ARIA 2.0a/CNS 1.2 program (52–54). The CYANA upper limit NOE restraints were converted to CNS restraints using CcpNmr Format Converter (55). Three additional NOEs between beta strands 1 and 2, and 42 additional hydrogen bond restraints (two restraints per hydrogen bond, corresponding to 21 H-bonds), inferred from TALOS predictions of the locations of the α helices, were also included in the refinement. In

addition to the distance and dihedral angle restraints, 48 backbone amide RDC restraints were used in the refinement. Only RDC values for residues in regions of secondary structure were included. The error in measurement of the RDCs was estimated to be ±0.8 Hz. The molecular alignment tensor was estimated by fitting the RDCs to the NMR structures obtained from the CYANA calculation without RDC restraints (56). The initial alignment tensor had a magnitude (D_a) of 3.1 Hz and a rhombicity (R) of 0.6. Upon refinement, the alignment tensor had a magnitude of 8.7 Hz and a rhombicity of 0.6. The ten lowest energy structures and the average energy-minimized structure have been deposited into the PDB (ID 2HTF). All CYANA and ARIA restraints have also been deposited into the PDB.

Relaxation Spectra and Analysis

^{15}N relaxation spectra were collected on U-^{15}N pol μ BRCT domain at 500 and 600 MHz using standard experiments (57) with the addition of two high-power ^{1}H saturation pulses (2–4 ms) at the start of the T_1 relaxation delay to ensure ^{1}H saturation for early T_1 relaxation periods. ^{15}N T_1 data were collected with relaxation delays of 39, 109, 194, 299, 414, 544, 689, 839, and 1004 ms. ^{15}N T_2 data were collected with relaxation delays of 8,16, 24,40,63,79,95, 111, and 127 ms. Uniform uncertainties were calculated on the basis of duplicated time points (underlined above). {^{1}H}–^{15}N heteronuclear NOE spectra were obtained using a saturation period of 4.5 s, with error calculated from baseplane noise. Decay rates were fit to peak intensities using expfit2 (in-house written software).

Lipari-Szabo dynamics parameters (58) were fit to decay rates using in-house software relxn2.2 and graphical interface relaxvi. Global fitting of data from rigid residues indicated a τ_m of 10.5 ns. Using this value, order parameters were fit using the standard model-free formalism for all residues, with errors assessed on the basis of 150 Monte Carlo simulations. Model selection based on the Bayesian Information Criterion (BIC) was then applied (59). In cases where the BIC did not unambiguously identify an optimal model, preference was given to the model with the fewest parameters or to model 2 (standard model-free), except for residues which were known not to produce reasonable parameter values in model 2 fits. In these cases model 3 was applied.

Analysis of Mutant Proteins

The "quick change" (Stratagene) method was used to make mutant pol μ constructs, both in the context of the isolated BRCT domain (construct described above), as well in the context of the full length cDNA (6). Mutant BRCT domains were expressed and purified as described above. Full length constructs (wild type and BRCT substitution mutations) and the ΔBRCT (Δ21–127)-construct were expressed and initially purified on a 1 mL HiTrap Nickel column as previously described (6), except instead of subsequent purification using an S200 column, the HiTrap Nickel column eluent was directly loaded onto a 1 mL HiTrap heparin column, and bound protein eluted using a buffer containing 25 mM Tris-HCl pH 8.0, 700 mM KCl, 10% glyercol, and 1 mM DTT. An ultracel 30K (Amicon) centrifugal filtration unit was then used to concentrate and adjust the salt of heparin column eluents to 250 mM KCl. Ku and XRCC4-ligase IV complex were purified from baculovirus infected cells as previously described (60).

Intrinsic catalytic activity, as reported in Table 2, was determined using gap filling assays on a 40 bp DNA duplex with a centrally located single nucleotide gap in one strand (template C) that has been previously described (7). Reactions were performed with 5 nM of the various pol μ variants, 20 nM DNA substrate, and a buffer containing 25 mM Tris HCl pH 7.5, 50 mM NaCl, 50 mM KCl, 1 mM DTT, 5 mM MgCl$_2$, and 25 μM of each dNTP, for 10 min at 37 °C, and analyzed on a 10% denaturing polyacrylamide gel. The standard deviation from replicate assays were found to be within 25% of the mean for all six constructs.

Electrophoresis mobility shift assays (EMSA) were performed using 100 nM of a 60-bp double-stranded (ds) DNA substrate made by annealing 5′ phosphorylated DAR166 (5′-CAGCTGGGAATTCCATATGAGTACTGCAGATGCACTTGCTCGATAGAT CTAACATGAGCC-3′) to DAR167 (5′Cy5-GTAGGGCTCATGTTAGATCTATCGAGCAAGTGCATCTGCAGTACTCATATGGAA TTCCCAGCTGAG-3′). This oligonucleotide duplex was incubated with 25 nM Ku for 5 min in a buffer containing 25 mM Tris-HCl pH 7.5, 150 mM KCl, 13% (wt/vol) glycerol, 1% (wt/vol) poly(ethylene glycol) (molecular mass, greater than 8000 kD; PEG), 1 mM DTT, and 0.5 mg/mL bovine serum albumin (BSA), followed by addition of 100 ng of uncut plasmid DNA, 50 nM XRCC4-ligase IV complex, and 100 nM polymerase μ before an additional incubation for 10 min on ice. Resulting DNA–protein complexes were separated by eletrophoresis on a 3.5% polyacrylamide gel in a buffer containing 30 mM Tris-borate (pH 8.2) and 0.3 mM EDTA (1/3X TBE) and detected using a typhoon fluorescent imager (GE Biosciences).

The NHEJ assay employed a 300 bp substrate with the indicated end structures prepared as previously described (8), except substrates were fluorescently labeled by synthesis in a mix of 100 μM each dNTP supplemented with 10 μM Cy5 labeled dCTP (GE Biosciences). End-joining assays were performed by preincubating for 5 min on ice 25 nM pol μ, 25 nM Ku, 50 nM XRCC4-ligase IV, and 5 nM DNA substrate in a buffer containing 25 mM Tris pH 7.5, 75 mM NaCl, 75 mM KCl, 1 mM DTT, 30 μg/mL BSA, 3% glycerol, 0.1 mM EDTA, 12.5% PEG, and 1 μg of uncut plasmid DNA. Reactions were then supplemented with 25 μM of each dNTP and 5 mM MgCl$_2$, incubated at 37 °C for 10 min, deproteinized, and subjected to electrophoresis on a nondenaturaing 5% polyacrylamide gel. Substrate and product species were detected using a Typhoon fluorescent imager and quantified using ImageQuant total lab software (GE biosciences).

RESULTS

Assignments and Secondary Structure

The ^{1}H–^{15}N HSQC spectrum of the BRCT domain (Supporting Information, Figure S1) was unchanged for several weeks at 10 °C, allowing all of the assignment and NOESY experiments to be carried out on a single U-[^{13}C, ^{15}N]-labeled sample. The spectrum was consistent with a well-folded isolated BRCT domain, in which most of the amide resonances could be observed (Figure S1). Details of the chemical shift assignments are discussed in Experimental Procedures.

A CSI (61) analysis of the H$^\alpha$, C$^\alpha$, C$^\beta$, and C′ chemical shifts revealed four beta strands and three alpha helices (Figures 1 and 2). A TALOS analysis of the same chemical shifts identifies a fourth short helix near the C-terminus of the domain. Although the existence of the fourth helix is supported by sequential $d_{NN}(i,i+1)$ NOEs between residues R120–R122 and a $d_{\alpha N}(i,i+3)$ NOE between C119 and R122, the helix does not appear in the average energy-minimized structure but appears as a 3_{10}-helix in only a few structures of the ensemble (see below) and is not included in Figures 1–6. Since the four beta strands form a parallel β sheet, many cross-strand NOEs involving H$^\alpha$ protons were obscured by the water resonance. Consequently, no manually assigned NOEs or hydrogen bond restraints involving the beta strands were included in the structure calculation and refinement. The beta strands from the average energy-minimized structure are composed of residues A31–L34, R57–V58, H68–M71, and L99–D101. The alpha helices are composed of residues R41–K54, A76–A88, and S103–A111. The final locations of the secondary structure elements from the average energy-minimized structure of the 10 lowest energy structures are in good agreement with the TALOS predictions and are shown in Figure 1.

Solution Structure

A superposition of the ten lowest energy structures obtained after water refinement is shown in Figure 2A, and the average energy-minimized structure obtained from these 10 structures is shown in Figure 2B. The ten structures superimpose with backbone and heavy atom RMSDs of 0.70 Å and 1.08 Å, respectively, for residues 30–124. The N-terminal residues, 20–29, are disordered. A compilation of the structural statistics for the ten lowest-energy structures is shown in Table 1. The solution structure of the BRCT domain of pol μ consists of a central four-stranded parallel β sheet, flanked by helices α1 and α3 on one face of the β sheet, and helix α2 on the opposing face of the β sheet (Figure 2B). Several members of the ensemble of structures have a short 3_{10}-helix comprising residues C119–R122 that is consistent with the TALOS predictions and the sequential and intermediate NOEs (see above). It is possible that this region forms a stable 3_{10}-helix in solution; however, there were insufficient NOEs to define it in all members of the ensemble in this study. This region of the domain interacts with β4 at the edge of the β sheet via a hydrogen bond between the backbone carbonyl of H121 and the backbone amide of D101. The topology of the BRCT domain (Figures 2A and 2B) is common to other BRCT domains, such as the BRCT domain of XRCC1 (19) and the C-terminal BRCT domain of BRCA1 (62).

Several residues near the N-terminus of α1 appear slightly disordered. These residues include R41, S42, R44, and A26. Two of these residues, R41 and R44, show evidence of slow conformational exchange based on ^{15}N relaxation analysis as described below. S42 also shows evidence of conformational averaging because its amide resonance could not be assigned due to line-broadening. This conformational averaging of α1 may be important for recognition of other NHEJ proteins as discussed below. We have previously noted a similar, although more extreme effect for helix α3 in the protein θ, a subunit of *E. coli* DNA polymerase III (63). This characteristic made it difficult to determine the solution structure of θ, and the more extreme exchange broadening of resonances corresponding to this helix was ultimately dealt with by varying the polarity of the solution (63).

The domain contains a well-defined hydrophobic core. Some of the hydrophobic residues comprising the core that are conserved between pol μ, pol λ, and TdT are shown in Figure 3A. In addition, there are four salt bridges on the surface of the molecule to further stabilize the tertiary structure (Figure 3B). One salt bridge involves residues R43 on α1 and E72 on L3, a second includes R57 on β2 and E65 on L2, and the third and fourth salt bridges are between R44 on α1 and E36 on L1 and D60 on L2 (Figure 3B). There are no charged residues in the hydrophobic core, but there are two solvent-exposed hydrophobic residues on helix α1, F46 and L50, that may be involved in intermolecular interactions (Figure 3B).

Complex Formation and NHEJ Activity

R43, F46, and L50 in particular stand out as both solvent exposed in the pol μ BRCT structure (Figure 3), and invariant when comparing pol μ and TdT sequences from diverse species (Figure 1). To address their role in mediating interactions with NHEJ factors, we mutated each of these residues individually to alanine (R43A, F46A, and L50A), both in the context of the isolated domain as well as the full length protein. ^{1}H NMR spectra from purified domains with the more severe mutations (R43A, F46A, and L50A) confirmed these mutations were not sufficient to globally disrupt domain folding (Supporting Information, Figure S3). ^{1}H 1D spectra have been recently used successfully for this purpose in the case of a BRCT domain from replication factor C (16). We also generated the more conservative F46L substitution, as both pol λ and a recently reported variant TdT BRCT structure (2COE, see Discussion) possess this substitution. All four full length mutant constructs retained intrinsic synthesis activity similar to wild type pol μ and pol μ with its BRCT domain entirely deleted (ΔBRCT) (Table

2). This is consistent with prior work indicating the BRCT domain is required for interations with NHFJ core factors but not its intrinsic ability to perform synthesis.

We then employed an electrophoretic mobility shift assay to determine whether R43A, F46A, and L50A remained capable of forming a stable complex with Ku and XRCC4-ligase IV. Under the conditions used here, Ku alone forms a stable complex on DNA ends, while addition of XRCC4-ligase IV results in a less stable species of heterogeneous mobility (Figure 4B) (6). Further addition of wild type pol μ but not a pol μ truncation mutant missing its BRCT domain (ΔBRCT), generates a much more stable triple complex (Ku, XRCC4-ligase IV, and pol μ) at DNA ends. Importantly, the amount of triple complex formed with the R43A, F46A, or L50A mutants is at least 10-fold lower (Figure 4; Table 2). In contrast, F46L-promoted formation of the triple complex approximately 1/3 as well as wild type pol μ.

Pol μ is uniquely capable of promoting template dependent synthesis from ends aligned by Ku and XRCC4-ligase IV, even when ends are noncomplementary. A number of catalytic domain residues required for this activity were recently defined (8,64). We show here that this activity is also highly dependent on the ability of pol μ to form a stable complex with Ku and XRCC4-ligase IV: relative to wild type pol μ, the ability of F46L to promote joining of such ends is 3-fold lower, while activities of the more severe mutants (R43A, F46A, and L50A) are reduced by 5–10 fold (Figure 4C, Table 2). R43, F46, and L50 thus define residues in pol μ's BRCT domain that are critical for the ability of pol μ to make use of interactions with Ku and XRCC4-ligase IV in order to support template-dependent polymerization.

Conformational Heterogeneity from [15]N Relaxation

To gain further insight into BRCT structure and function, a characterization of the dynamics of pol μ BRCT was initiated using ^{15}N spin relaxation methods (Experimental Procedures). [15]N T_1, T_2, and [15]N—{[1]H} NOE were measured at 500 and 600 MHz field strengths and further analyzed using the Lipari-Szabo model-free formalism (58). The correlation time for overall tumbling of the domain, τ_m, was determined to be 10.5 ns, consistent with a monomeric BRCT domain under the conditions used (also, see Discussion for comparison to previous results). Fitting to an axially symmetric diffusion tensor yielded $D_{\parallel}/D_{\perp}$ of 1.1, indicating that tumbling anisotropy has a negligible effect on relaxation. The monomeric state was independently confirmed using static and dynamic light scattering (Supporting Information, Figure S2). The dynamic character of the backbone on a per residue basis is summarized in Figure 5. Overall, the order parameters, S^2, corresponding to rigidity of amide N–H bond vectors on a scale of 0 to 1, are approximately 0.9 in regions of secondary structure. Regions of structural flexibility are therefore restricted to "loop" regions. Lower order parameters (higher flexibility) are found in residues 90, 94, and 95 (data on 91–93 were not obtained), which reside in L4 (Figure 5). While order parameters report on fast ps-ns time scale motions, flexibility on the, μs-ms time scale can be detected from enhanced transverse relaxation rates ($R_2 = 1/T_2$) or alternatively from line broadening and loss of intensity in 2D (or nD) peaks due to chemical exchange effects. Figure 5 shows both R_2 values and peak intensities (from the 600 MHz HSQC spectrum). From this it can be seen that residues 35–41 show evidence of "slow", μs-ms motions; these residues essentially define the LI loop. From the model selection analysis of the relaxation data, model 3 (S^2, τ_e, R_{ex}) was selected for residues 35, 36, 39, 44, 47, 67, and 103, with R_{ex} values that varied from 0.4 to 3 Hz. It should also be added that residues 62–64 (in L2) could not be assigned due to line-broadening, and so this loop is also flexible or changing conformations on a slow time scale. All of these flexible residues except for 67 and 103 are in loop regions and correlate well with decreased precision in the ensemble of structures (Figure 2A).

DISCUSSION

We report here the solution structure, dynamics, and biochemical evidence for a functional surface of the BRCT domain from human polymerase μ This represents a first step toward understanding how these polymerases interact with the NHEJ machinery.

Structural Comparison to the BRCT Domain from TdT

Two other members of the pol X family of polymerases, TdT and pol λ also have N-terminal BRCT domains. Pol μ is most similar to TdT, with 41% sequence identity (2) overall and 39% sequence identity over the BRCT domain (residues V30-E124, see Figure 1). In contrast, the sequence identity with the pol μ BRCT domain is only 19% over the structured region of the domain (Figure 1). The residues that are identical between all three domains are shown in red and highlighted in yellow in Figure 1.

Recently, an NMR solution structure of the BRCT domain of human TdT has been deposited in the Protein Data Bank (PDB ID 2COE, Nagashima, et al., unpublished). The pol μ and TdT BRCT domains have the same folding topology, and the lowest energy TdT structure superimposes on the average energy-minimized pol μ structure with a backbone rmsd of 2.1 Å (Figure 6A). Overall, the structures are very similar, although α2 is shorter in the TdT domain compared to the μ domain. Since α2 is shorter in TdT, the position of L4, connecting α2 and β4, in the two domains is different: In the pol μ ensemble of structures, L4 appears more disordered (Figure 2A), and residues A71, C75, and T76 in L4 exhibit increased backbone dynamics (Figure 5). L4 is also a region of low sequence conservation (Figure 1).

The positions of many of the hydrophobic and aromatic residues in the two domains are very similar, such as interactions between Y33 and W82 and W104 and H68 (Figure 6B and corresponding residues in 6C). Both BRCT domains have exposed hydrophobic residues on α1 and α2; in pol μ the side chains of F46, L50, and V80 are exposed; in TdT the side chains of F40, L44, and L74 are exposed (Figures 6B and 6C). Note that a leucine at position 40 of the TdT structure is inconsistent with a phenylalanine found at this site in both the published human TdT sequence as well as TdT from other species (see, e.g., Figure 1), suggesting that leucine is either a rare polymorphism or a mutation that arose during cloning. The solvent exposure of hydrophobic residues F46, L50, and V80 (and corresponding residues in TdT) suggested that these residues might mediate interactions with other proteins in the NHEJ repair complex.

The electrostatic surfaces of the pol μ and TdT BRCT domains are also very similar (Figures 7A–D). Both domains have a ridge of positively charged residues on one face of the protein (compare Figures 7A and 7C), and large negative regions on the opposite faces of the proteins (compare Figures 7B and 7C). In the pol μ domain, the positive ridge is composed of residues R44, R52, R85, and R86, whereas in TdT the ridge is composed of residues R47, R46, R38, and K31. It is conceivable that this ridge may be an interaction site for phosphopeptide binding. In the pol μ domain, the ridge is partially obscured by the acidic side chains of residues E36 and D60, which form a salt bridge with R44. However, in TdT the acidic side chains of E30 and E55 are in close proximity to R34, corresponding to R44 in pol μ, and may also form salt bridges under some conditions.

Comparison to Other BRCT Domains

One of the interesting aspects of the BRCT fold is that it appears to be multifunctional. The fold accommodates direct binding of both proteins (phosphorylated and nonphosphorylated) and DNA, and it utilizes a number of different surfaces for different interactions. A variety of structures of BRCT domains have been determined (65), although the majority of these are of

tandem repeat BRCT domains, such as in BRCA1 (62), 53BP1 (14), and MDC1 (66). Tandem repeat BRCTs, in particular, have been shown to bind phospho-peptides (22–28,67). In addition, other BRCT domains appear to homo- or heterodimerize (68). The pol μ BRCT structure is atypical in that it appears to be the first reported (in the literature) BRCT domain structure that is neither part of a tandem grouping nor forms a stable dimer, as demonstrated here for μ-BRCT by ^{15}N relaxation and light scattering. (It should be mentioned that while other BRCT domains appear in the PDB as monomeric structures (see below), without additional measurements, the oligomeric state of these domains remains uncertain).

To assess the degree of structural similarity between pol μ BRCT and individual BRCTs from tandem or dimerized domains, we carried out a number of backbone superimpositions. Domains *found either in tandem pairs or as noncovalent dimers* from BRCA1, XRCC1, 53BP1, MDC1, and DNA ligase IIIα all superimposed to the energy-minimized μ-BRCT domain with RMSDs of 5 Å or greater. The sole exception was the C-terminal BRCT domain from the NMR structure of BRCA1 (1OQA), which has an rmsd of 2 Å if helix α2 is excluded from the superposition; however, this BRCT domain is structurally different because the α2 helix from this BRCT domain has only ~1 turn. An example of the different orientation of helix α2 in the μ-BRCT domain vs the C-terminal BRCT domain from XRCC1 (1CDZ) is shown in Supporting Information (Figure S4).

Next, we sought to determine the extent of structural similarity between the pol μ BRCT domain (and hence pol TdT) and BRCT domains in the protein data bank that are *not from tandem pairs or found as dimers.* The apparently monomeric BRCT domains found in the PDB include the NAD[+]-dependent DNA ligase BRCT domain from *T. thermophilus* (PDB ID: 1L7B), the second BRCT domain of epithelial transforming growth factor 2 (PDB ID: 2COU), and the BRCT domain of poly(ADP-ribose) polymerase-1 (PDB ID: 2COK), which superimpose with backbone RMSDs of 4.8, 6.3, and 11 Å, respectively. In addition, 2COU, as with 1OQA (BRCA1) and 1IMO (DNA ligase III), lacks most or all of the α2 helix. In summary, it appears that the pol μ and TDT BRCT domains have unique structural properties, as indicated by high rmsd values with other BRCT domains, and that an element of major structural difference across all BRCT domains is the α2 helix.

Relaxation/Dynamics and Monomeric State of pol μ BRCT

We have performed a standard ^{15}N relaxation analysis of the backbone dynamics of human polymerase μ. Only two BRCT domains have been previously characterized by ^{15}N relaxation: the C-terminal BRCT domain from BRCA1 (62) and the BRCT domain from DNA ligase IIIα (17). Because BRCA1 contains a tandem repeat and DNA ligase IIIα is reported to be a dimer in solution (17), this study represents the first dynamics characterization of a singly occurring (nontandem), nondimeric BRCT domain. The correlation time for tumbling, τ_m, was determined to be 10.5 ns at 10 °C. We believe that this is consistent with ^{15}N relaxation data on the dimeric BRCT domain from DNA ligase IIIα, even though that domain was reported to have a global τ_m value of 6.82 ns at 15 °C (16). In that same report (16), the average R_2 values were 19.5 s[−1] at 15 °C, consistent with a tumbling correlation time of at least 14 ns. This is reflected in the local τ_m times reported in Figure 2 of that paper, which are inconsistent with the global tumbling time cited in the text (16). These considerations lead us to conclude that the pol μ BRCT is monomeric under the conditions studied here. This has been confirmed by light scattering experiments (Supporting Information, Figure S2). For the pol μ BRCT domain, the only region found to have increased flexibility on the ps-ns time scale is L4 (Figure 5). On the other hand, slower motion on the μs-ms time scale occurs in loops LI and L2.

Functional Hot Spots in pol μ BRCT Are Proximal to Phospho-Binding Sites in Other BRCT Domains

Inspection of the solution structure and dynamics of pol μ BRCT allowed identification of the L1-α1 region as a surface containing residues that are potentially important for function. Specifically, the hydrophobic side chains F46 and L50 in α1 are conspicuously exposed to solvent and show high conservation in pol μ and TdT. In addition, R43, also in α1, shows evidence for flexibility on the μs-ms time scale, along with residues in L1 (Figure 5) that lead into α1. Because of their spatial proximity and definition of a contiguous surface (Figure 4A), residues R43, F46, and L50 were mutated to test their importance in NHEJ complex formation and activity using gel shift assays. Since all of these residues are found to be solvent-exposed in the pol μ BRCT structure, it was assumed that substitution of any of these residues with alanine would not result in significant structural changes. These general conclusions were supported by the ^{1}H spectra of the mutated proteins, which were all consistent with the expectation of a conserved structure for each of the mutants tested. The assays showed that mutation of any one of these residues to alanine in the context of full length pol μ largely abolishes pol μ's ability to form a stable complex with Ku and XRCC4-ligase IV assembled on DNA ends. This in turn strongly reduces the ability of pol μ to help Ku and XRCC4-ligase IV join ends that require prior gap filling before they can be joined. The results of these experiments demonstrate that the N-terminal region of helix α1, and possibly L1, have a pivotal role in promoting interactions with the NHEJ machinery that are critical for pol μ's biological role.

Interestingly, this general region has been implicated in other BRCT-protein interactions. In the BRCA1–BACH1 complex, S1655 and G1656, in L1 of the N-terminal BRCT domain, form hydrogen bonds to the phosphoserine (26–28,67). Nearly identical interactions are formed in the MDC1–H2AX complex (69). These residues correspond to μ-BRCT residues V35 and E36, in L1, which experience μs-ms fluctuations (Figure 5). Several lines of reasoning suggest that the details of functional interactions in μ-BRCT may be different than in the BRCA1–BACH1 or MDC1–H2AX complexes. First, μ-BRCT is not a tandem domain, although it is possible that it may form heterodimers with other BRCT domains, such as that in XRCC4. Second, L1 in μ-BRCT is composed of nearly twice as many residues as in the N-terminal BRCT domains of BRCA-1 and MDC1. Third, in the N-terminal region of α1, there are two conserved arginines (43 and 44). R44 is involved in a salt bridge to the side chain of E36, which resides in the flexible L1. These two arginines are conserved throughout species of μ and TdT polymerases (Figure 1), whereas in the N-terminal BRCA1 domain there are no positively charged side chains in this region and a glutamate at the position corresponding to R43 in μ-BRCT. It is worth mentioning, however, that in one of the BRCA1–BACH1 crystal structures (1T29), the carboxyl group of E1661 (equivalent to R43 in μ-BRCT) appears to be involved in an electrostatic interaction with the side chain of *R*3 (−3 position) in the BACH1 peptide. This is consistent with a role for the conserved arginines in μ-BRCT in peptide/protein (or even DNA) recognition.

In summary, the prominent solvent exposure of hydrophobic residues and evidence for microsecond-millisecond motions, along with sequence conservation information, has led to the biochemical identification of a group of spatially proximate residues likely to be important for protein interactions of the pol μ BRCT domain. The significance of these residues received strong support from assays performed on mutated analogs of pol μ. In addition, a structural comparison with the BRCT domain of TdT suggests that this interaction surface is well conserved, supporting the hypothesis that these polymerases can both be recruited by the NHEJ repair complex.
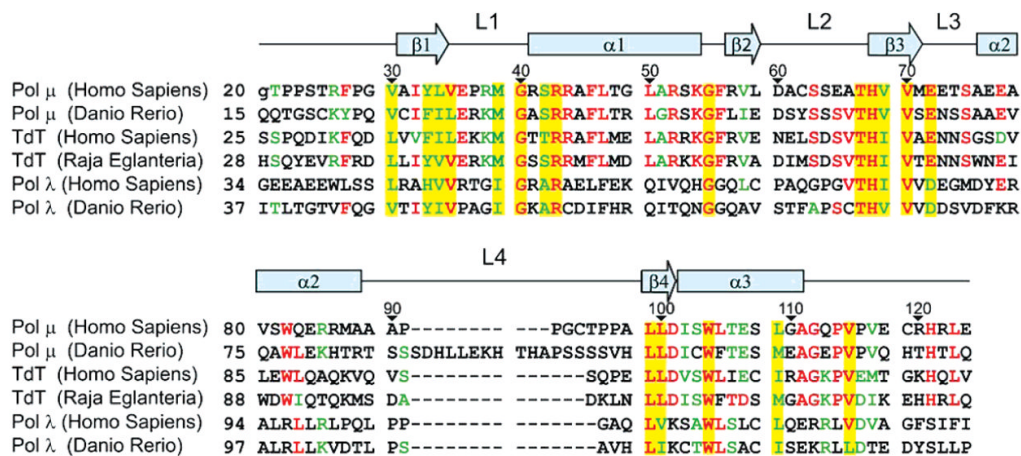
19. Zhang X, Morera S, Bates PA, Whitehead PC, Coffer AI, Hainbucher K, Nash RA, Sternberg MJ, Lindahl T, Freemont PS. Structure of an XRCC1 BRCT domain: a new protein-protein interaction module. EMBO J 1998;17:6404–6411. [PubMed: 9799248]

20. Nash RA, Caldecott KW, Barnes DE, Lindahl T. XRCC1 protein interacts with one of two distinct forms of DNA ligase III. Biochemistry 1997;36:5207–5211. [PubMed: 9136882]

21. Masson M, Niedergang C, Schreiber V, Muller S, Menissierde Murcia J, de Murcia G. XRCC1 is specifically associated with poly(ADP-ribose) polymerase and negatively regulates its activity following DNA damage. Mol Cell Biol 1998;18:3563–3571. [PubMed: 9584196]

22. Liu K, Lin FT, Ruppert JM, Lin WC. Regulation of E2F1 by BRCT domain-containing protein TopBP1. Mol Cell Biol 2003;23:3287–3304. [PubMed: 12697828]

23. Manke IA, Lowery DM, Nguyen A, Yaffe MB. BRCT repeats as phosphopeptide-binding modules involved in protein targeting. Science 2003;302:636–639. [PubMed: 14576432]

24. Rodriguez M, Yu X, Chen J, Songyang Z. Phosphopeptide binding specificities of BRCA1 COOH-terminal (BRCT) domains. J Biol Chem 2003;278:52914–52918. [PubMed: 14578343]

25. Yu X, Chini CC, He M, Mer G, Chen J. The BRCT domain is a phospho-protein binding domain. Science 2003;302:639–642. [PubMed: 14576433]

26. Clapperton JA, Manke IA, Lowery DM, Ho T, Haire LF, Yaffe MB, Smerdon SJ. Structure and mechanism of BRCA1 BRCT domain recognition of phosphorylated BACH1 with implications for cancer. Nat Struct Mol Biol 2004;11:512–518. [PubMed: 15133502]

27. Shiozaki EN, Gu L, Yan N, Shi Y. Structure of the BRCT repeats of BRCA1 bound to a BACH1 phosphopeptide: implications for signaling. Mol Cell 2004;14:405–412. [PubMed: 15125843]

28. Williams RS, Lee MS, Hau DD, Glover JN. Structural basis of phosphopeptide recognition by the BRCT domain of BRCA1. Nat Struct Mol Biol 2004;11:519–525. [PubMed: 15133503]

29. Wishart DS, Bigam CG, Yao J, Abildgaard F, Dyson HJ, Oldfield E, Markley JL, Sykes BD. 1H, 13C and 15N chemical shift referencing in biomolecular NMR. J Biomol NMR 1995;6:135–140. [PubMed: 8589602]

30. Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, Bax A. NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J Biomol NMR 1995;6:277–293. [PubMed: 8520220]

31. Johnson BA, Blevins RA. NMRView: a computer program for the visualization and analysis of NMR data. J Biomol NMR 1994;4:603–614.

32. Ikura M, Kay LE, Bax A. A novel approach for sequential assignment of 1H, 13C, and 15N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. Biochemistry 1990;29:4659–4667. [PubMed: 2372549]

33. Matsuo H, Rupee E, Li H, Wagner G. Increased sensitivity in HNCA and HN(CO)CA experiments by selective C beta decoupling. J Magn Reson B 1996;113:91–96. [PubMed: 8888595]

34. Kay LE, Xu GY, Yamazaki T. Enhanced-sensitivity triple resonance spectroscopy with minimal H$_2$O saturation. J Magn Reson A 1994;109:129–133.

35. Wittekind M, Mueller L. HNCACB, a high-sensitivity 3D NMR experiment to correlate amide-proton and nitrogen resonances with the alpha- and beta-carbon resonances in proteins. J Magn Reson B 1993;101:201–205.

36. Grzesiek S, Bax A. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. J Biomol NMR 1992;8:277–293.

37. Muhandiram DR, Kay LE. Gradient-enhanced triple-resonance three-dimensional NMR experiments with improved sensitivity. J Magn Reson B 1994;103:203–216.

38. Powers R, Gronenborn AG, Clore GM, Bax A. Three-dimensional triple-resonance NMR of 13C/15N-enriched proteins using constant-time evolution. J Magn Reson 1991;94:209–213.

39. Kanelis V, Donaldson L, Muhandiram DR, Rotin D, Forman-Kay JD, Kay LE. Sequential assignment of proline-rich regions in proteins: application to modular binding domain complexes. J Biomol NMR 2000;16:253–259. [PubMed: 10805132]

40. Grzesiek S, Anglister J, Bax A. Correlation of backbone amide and aliphatic side chain resonances in 13C/15N-enriched proteins by multiple relayed triple resonance NMR. J Magn Reson 1993;101:114–119.

41. Logan TM, Olejniczak ET, Xu RX, Fesik SW. A general method for assigning NMR spectra of denatured proteins using 3D HC(CO)NH-TOCSY triple resonance experiments. J Biomol NMR 1993;3:225–231. [PubMed: 8477187]

42. Montelione GT, Lyons BA, Emerson SD, Tashiro M. An efficient triple resonance experiment using C-13 isotropic mixing for determining sequence-specific resonance assignments of isotopically-enriched proteins. J Am Chem Soc 1992;114:10974–10975.

43. Kay LE, Xu GY, Singer AU, Muhandiram DR, Forman-Kay JD. A gradient-enhanced HCCH-TOCSY experiment for recording side-chain 1H and 13C correlations in H2O samples of proteins. J Magn Resons B 1993;101:333–337.

44. Zhang O, Kay LE, Olivier JP, Forman-Kay JD. Backbone 1H and 15N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. J Biomol NMR 1994;4:845–858. [PubMed: 7812156]

45. Yamazaki T, Forman-Kay JD, Kay LE. Two-dimensional NMR experiments for correlating ^{13}C-beta and ^{1}H-delta/epsilon chemical shifts of aromatic residues in ^{13}C-labeled proteins via scalar couplings. J Am Chem Soc 1993;115:11054–11055.

46. Pascal SM, Muhandiram DR, Yamazaki T, Forman-Kay JD, Kay LE. Simultaneous acquisition of 15N- and 13C-edited NOE spectra of proteins dissolved in H_2O. J Magn Reson B 1994;103:197–201.

47. Permi P. A spin-state-selective experiment for measuring heteronuclear one-bond and homonuclear two-bond couplings from an HSQC-type spectrum. J Biomol NMR 2002;22:27–35. [PubMed: 11885978]

48. Ruckert M, Otting G. Alignment of biological macromolecules in novel nonionic liquid crystalline media for NMR experiments. J Am Chem Soc 2000;122:7793–7797.

49. Güntert P, Mumenthaler C, Wüthrich K. Torsion angle dynamics for NMR structure calculation with the new program DYANA. J Mol Biol 1997;273:283–298. [PubMed: 9367762]

50. Herrmann T, Guntert P, Wuthrich K. Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA. J Mol Biol 2002;319:209–227. [PubMed: 12051947]

51. Cornilescu G, Delaglio F, Bax A. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. J Biomol NMR 1999;13:289–302. [PubMed: 10212987]

52. Brunger AT, Adams PD, Clore GM, DeLano WL, Gros P, Grosse-Kunstleve RW, Jiang JS, Kuszewski J, Nilges M, Pannu NS, Read RJ, Rice LM, Simonson T, Warren GL. Crystallography & NMR system: A new software suite for macromolecular structure determination. Acta Crystallogr Sect D: Biol Crystallogr 1998;54:905–921. [PubMed: 9757107]

53. Linge JP, O'Donoghue SI, Nilges M. Automated assignment of ambiguous nuclear overhauser effects with ARIA. Methods Enzymol 2001;339:71–90. [PubMed: 11462826]

54. Nilges M. Calculation of protein structures with ambiguous distance restraints. Automated assignment of ambiguous NOE crosspeaks and disulphide connectivities. J Mol Biol 1995;245:645–660. [PubMed: 7844833]

55. Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, Laue ED. The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins 2005;59:687–696. [PubMed: 15815974]

56. Yang DW, Venters RA, Mueller GA, Choy WY, Kay LE. TROSY-based HNCO pulse sequences for the measurement of 1HN-15N, 15N-13CO, 1HN-13CO, 13CO-13Calpha dipolar couplings in ^{15}N, ^{13}C, ^{2}H-labeled proteins. J Biomol NMR 1999;14:333–343.

57. Farrow NA, Muhandiram R, Singer AU, Pascal SM, Kay CM, Gish G, Shoelson SE, Pawson T, Forman-Kay JD, Kay LE. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ^{15}N NMR relaxation. Biochemistry 1994;33:5984–6003. [PubMed: 7514039]

58. Lipari G, Szabo A. Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules: 1 Theory and range of validity. J Am Chem Soc 1982;104:4546–4559.

59. d'Auvergne EJ, Gooley PR. The use of model selection in the model-free analysis of protein dynamics. J Biomol NMR 2003;25:25–39. [PubMed: 12566997]

60. Nick McElhinny SA, Snowden CM, McCarville J, Ramsden DA. Ku recruits the XRCC4-ligase IV complex to DNA ends. Mol Cell Biol 2000;20:2996–3003. [PubMed: 10757784]

61. Wishart DS, Sykes BD. The 13C chemical-shift index: a simple method for the identification of protein secondary structure using 13C chemical-shift data. J Biomol NMR 1994;4:171–180. [PubMed: 8019132]

62. Gaiser OJ, Ball LJ, Schmieder P, Leitner D, Strauss H, Wahl M, Kuhne R, Oschkinat H, Heinemann U. Solution structure, backbone dynamics, and association behavior of the C-terminal BRCT domain from the breast cancer-associated protein BRCA1. Biochemistry 2004;43:15983–15995. [PubMed: 15609993]

63. Mueller GA, Kirby TW, DeRose EF, Li D, Schaaper RM, London RE. Nuclear magnetic resonance solution structure of the Escherichia coli DNA polymerase III theta subunit. J Bacteriology 2005;187:7081–7089.

64. Moon AF, Garcia-Diaz M, Bebenek K, Davis BJ, Zhong X, Ramsden DA, Kunkel TA, Pedersen LC. Structural insight into the substrate specificity of DNA Polymerase mu. Nat Struct Mol Biol 2007;14:45–53. [PubMed: 17159995]

65. Glover JN, Williams RS, Lee MS. Interactions between BRCT repeats and phosphoproteins: tangled up in two. Trends Biochem Sci 2004;29:579–585. [PubMed: 15501676]

66. Lee MS, Edwards RA, Thede GL, Glover JN. Structure of the BRCT repeat domain of MDC1 and its specificity for the free COOH-terminal end of the gamma-H2AX histone tail. J Biol Chem 2005;280:32053–32056. [PubMed: 16049003]

67. Botuyan MV, Nomine Y, Yu X, Juranic N, Macura S, Chen J, Mer G. Structural basis of BACH1 phosphopeptide recognition by BRCA1 tandem BRCT domains. Structure 2004;12:1137–1146. [PubMed: 15242590]

68. Beernink PT, Hwang M, Ramirez M, Murphy MB, Doyle SA, Thelen MP. Specificity of protein interactions mediated by BRCT domains of the XRCC1 DNA repair protein. J Biol Chem 2005;280:30206–30213. [PubMed: 15987676]

69. Stucki M, Clapperton JA, Mohammad D, Yaffe MB, Smerdon SJ, Jackson SP. MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks. Cell 2005;123:1213–1226. [PubMed: 16377563]

70. Clore GM, Garrett DS. R-factor, free R, and complete cross-validation for dipolar coupling refinement of NMR structures. J Am Chem Soc 1999;121:9008–9012.

71. Koradi R. MOLMOL: a program for display and analysis of macromolecular structures. J Mol Graph 1996;14:51–55. [PubMed: 8744573]

72. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. PROCHECK – A program to check the stereo-chemical quality of protein structures. J Appl Crystallogr 1993;26:283–291.

73. Laskowski RA, Rullman JAC, MacArthur MW, Kaptein R, Thornton JM. AQUA and PROCHECK-NMR: Programs for checking the quality of protein structures solved by NMR. J Biomol NMR 1996;8:477–486. [PubMed: 9008363]

		L1		α1	β2	L2	L3	

```
                        β1         L1              α1           β2      L2            β3      L3    α2
                              30              40              50              60              70
Pol μ  (Homo Sapiens)   20  gTPPSTRFPG  VAIYLVEPRM  GRSRRAFLTG  LARSKGFRVL  DACSSEATHV  VMEETSAEEA
Pol μ  (Danio Rerio)    15  QQTGSCKYPQ  VCIFILERKM  GASRRAFLTR  LGRSKGFLIE  DSYSSSVTHV  VSENSSAAEV
TdT    (Homo Sapiens)   25  SSPQDIKFQD  LVVFILEKKM  GTTRRAFLME  LARRKGFRVE  NELSDSVTHI  VAENNSGSDV
TdT    (Raja Eglanteria) 28 HSQYEVRFRD  LLIYVVERKM  GSSRRMFLMD  LARKKGFRVA  DIMSDSVTHI  VTENNSWNEI
Pol λ  (Homo Sapiens)   34  GEEAEEWLSS  LRAHVVRTGI  GRARAELFEK  QIVQHGGQLC  PAQGPGVTHI  VVDEGMDYER
Pol λ  (Danio Rerio)    37  ITLTGTVFQG  VTIYIVPAGI  GKARCDIFHR  QITQNGGQAV  STFAPSCTHV  VVDDSVDFKR
```

```
                         α2              L4                     β4       α3
                                   90                      100             110             120
Pol μ  (Homo Sapiens)   80  VSWQERRMAA  AP--------  ---PGCTPPA  LLDISWLTES  LGAGQPVPVE  CRHRLE
Pol μ  (Danio Rerio)    75  QAWLEKHTRT  SSSDHLLEKH  THAPSSSSVH  LLDICWFTES  MEAGEPVPVQ  HTHTLQ
TdT    (Homo Sapiens)   85  LEWLQAQKVQ  VS--------  ------SQPE  LLDVSWLIEC  IRAGKPVEMT  GKHQLV
TdT    (Raja Eglanteria) 88 WDWIQTQKMS  DA--------  ------DKLN  LLDISWFTDS  MGAGKPVDIK  EHHRLQ
Pol λ  (Homo Sapiens)   94  ALRLLRLPQL  PP--------  -------GAQ  LVKSAWLSLC  LQERRLVDVA  GFSIFI
Pol λ  (Danio Rerio)    97  ALRLLKVDTL  PS--------  -------AVH  LIKCTWLSAC  ISEKRLLDTE  DYSLLP
```

Figure 1.
Alignment of pol X BRCT domains. Secondary structure elements (blue cartoons) in the human pol μ BRCT domain are aligned relative to its amino acid sequence and five other pol X BRCT domain sequences. Following their names and species, positions are numbered according to locations in their respective full length proteins. Positions conserved in all six members are highlighted in yellow. Residues conserved in more than four pol X members are identified in red (identical) or green (similar). Sequences were aligned using AlignX, a clustalW-based module of Vector NTI (Invitrogen). The first residue in human pol μ (G20) represents a substitution relative to the published sequence (S20).
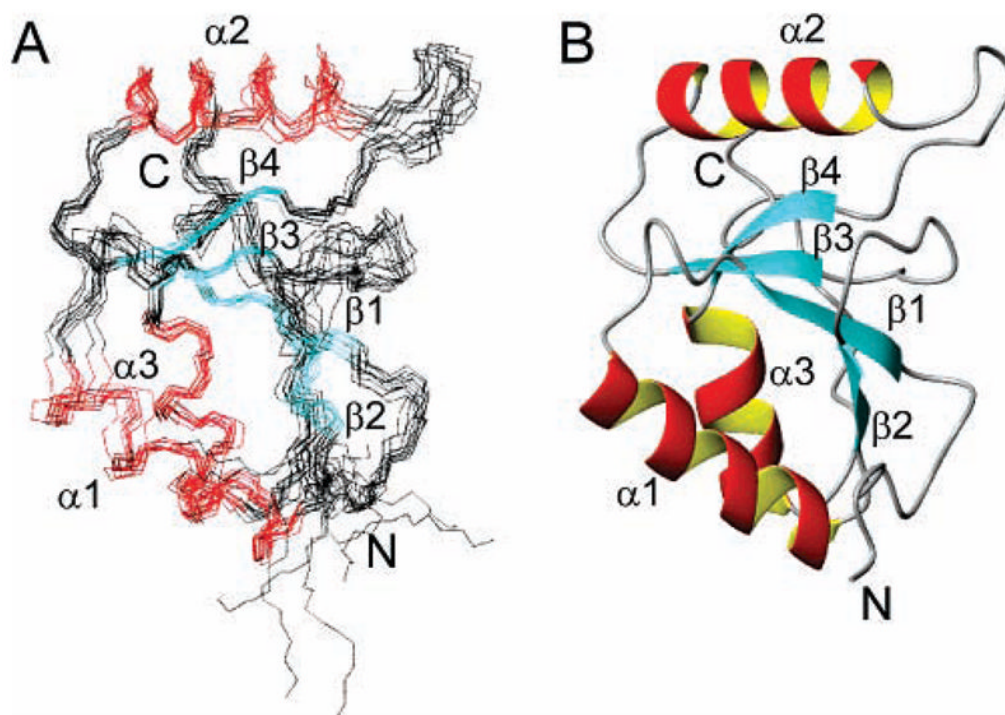
Figure 2.
Solution structure of the BRCT domain of pol μ.. (A) The 10 lowest energy structures after water refinement superimpose with a backbone rmsd of 0.70 ± 0.12 Å. The alpha helices are shown in red; the beta strands are shown in cyan. (B) The average, energy-minimized structure computed from the 10 lowest energy structures is shown. The central β sheet is shown with α1 and α3 packed against one side of the sheet and α2 packed against the opposite side of the sheet.
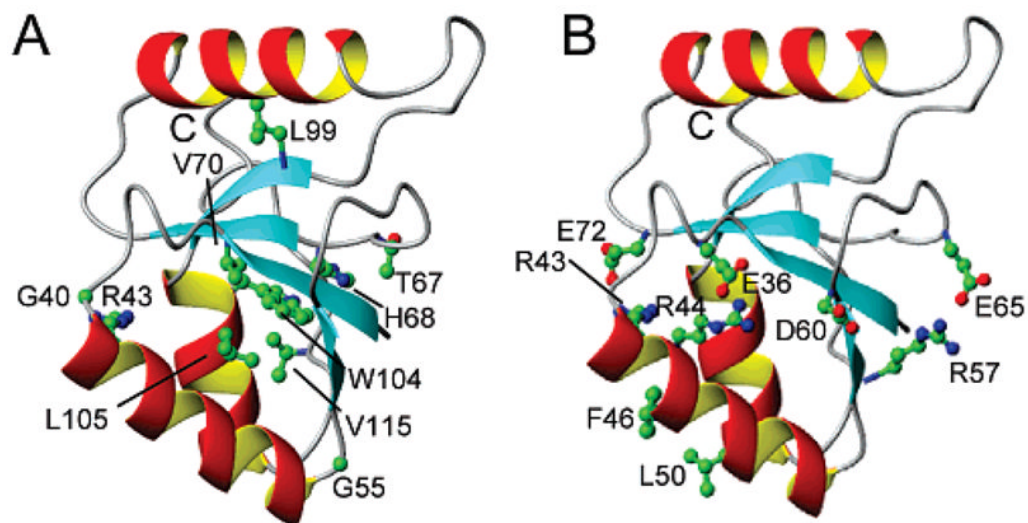
Figure 3.
Structurally important residues of the BRCT domain of pol μ Positions of fully conserved residues, salt bridges, and exposed hydrophobic residues of the BRCT domain of pol μ (A) The average, energy-minimized structure is depicted with the side chains of residues that are fully conserved between, pol μ, TdT, and pol λ. The conserved residues H68, V70, L99, W104, L105, and V115 form part of the hydrophobic core of the domain. (B) The side chains of arginine, aspartate, and glutamate residues involved in salt-bridge formation and the exposed hydrophobic residues, F46 and L50 are shown.
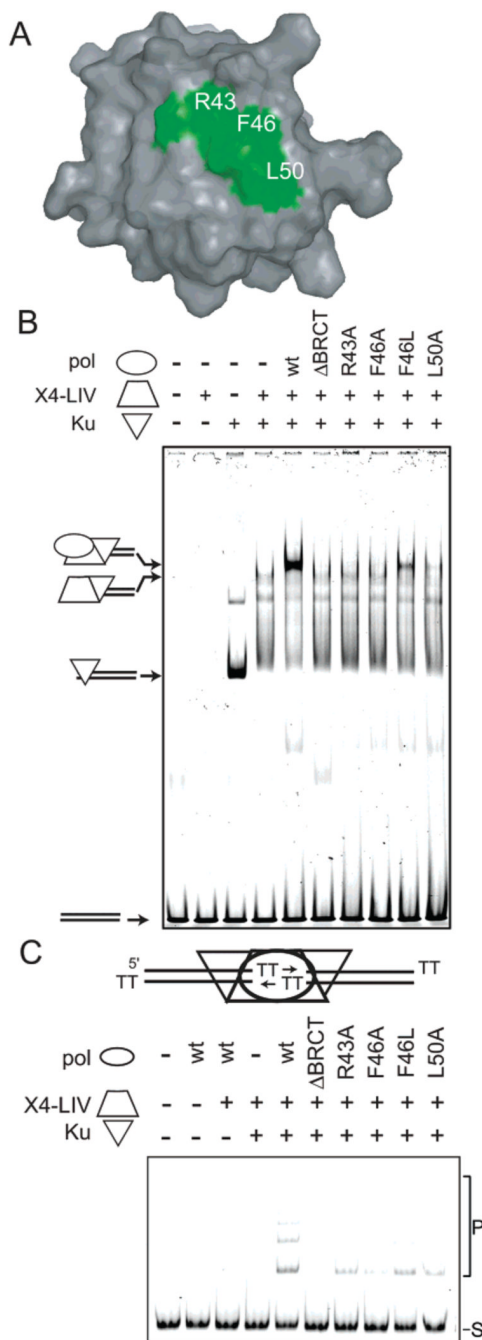
Figure 4.
Activity of pol μ BRCT domain mutants. (A) A surface representation of the pol μ BRCT domain, with the solvent exposed area of R43, F46, and L50 highlighted in green. (B) EMSA analysis was performed in the presence of a 60 bp DNA duplex, Ku, XRCC4-ligase IV (X4-LIV), and full length wild type pol μ (wt), pol μ with the BRCT domain deleted (ΔBRCT), or full length pol μ with BRCT domain substitution mutations as noted. The composition of each species of distinct mobility has previously been determined (Mahajan MCB 2002) and is noted with cartoons at the left of the panel. (C) Joining of a 300 bp substrate with two nucleotide TT 3′ overhangs was performed in the presence of Ku, XRCC4-ligase IV, and various pol μ constructs as noted. S denotes substrate, and P denotes joined concatemer products.
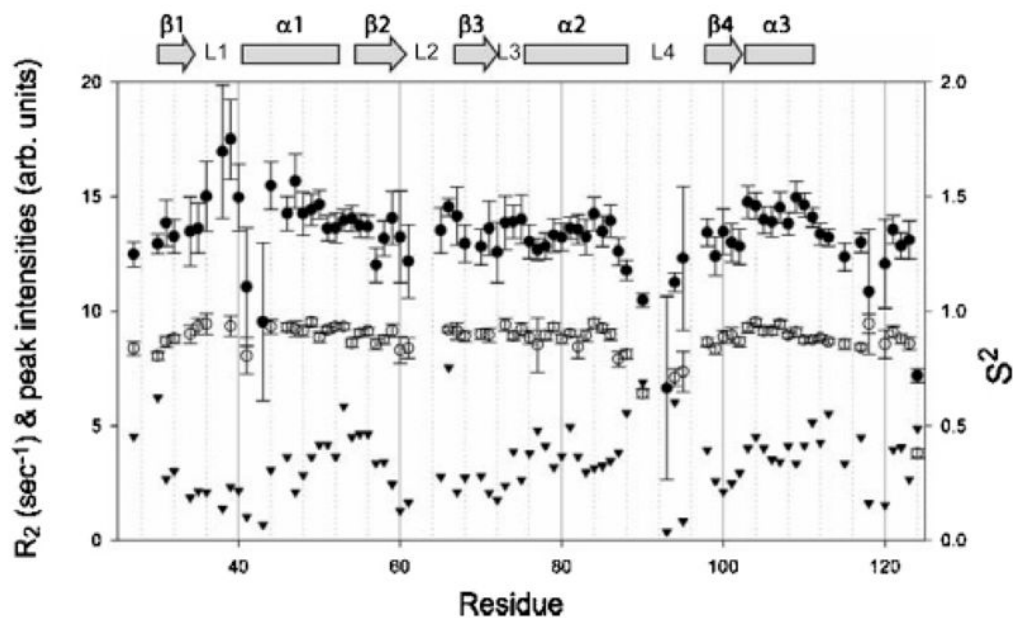
Figure 5.
Regions of flexibility in the BRCT domain of pol μ. Backbone flexibility by three different measures are shown as a function of residue number. The ^{15}N transverse relaxation rate, R_2, is shown in filled circles. Increases from the mean are indicative of "slow" motions on the μs-ms time scale. These rates are typically inversely proportional to the intensity of peaks in a ^1H–^{15}N HSQC, as shown in the filled triangles. Lipari-Szabo model-free order parameters, S^2, are shown in open circles. S^2 can take on values from 0 to 1, with 0 corresponding to the limit of complete flexibility and 1 corresponding to complete rigidity of an N–H bond vector. Order parameters report on motion on the ps-ns time scale.
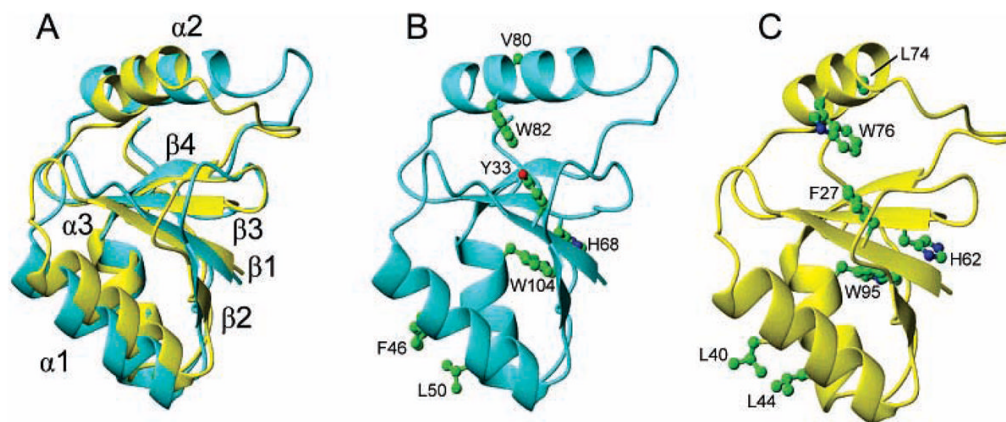
Figure 6.
Comparison of pol μ and TdT BRCT domains. (A) The average, energy-minimized structure of the BRCT domain of polμ (cyan) and the BRCT domain of TdT (yellow, PDB ID 2COE model 1) superimpose with a backbone rmsd of 2.1 Å. (B, C) The positions of the aromatic side chains in the hydrophobic core are conserved between the two domains, as well as the positions of exposed hydrophobic residues on 1 and 2.
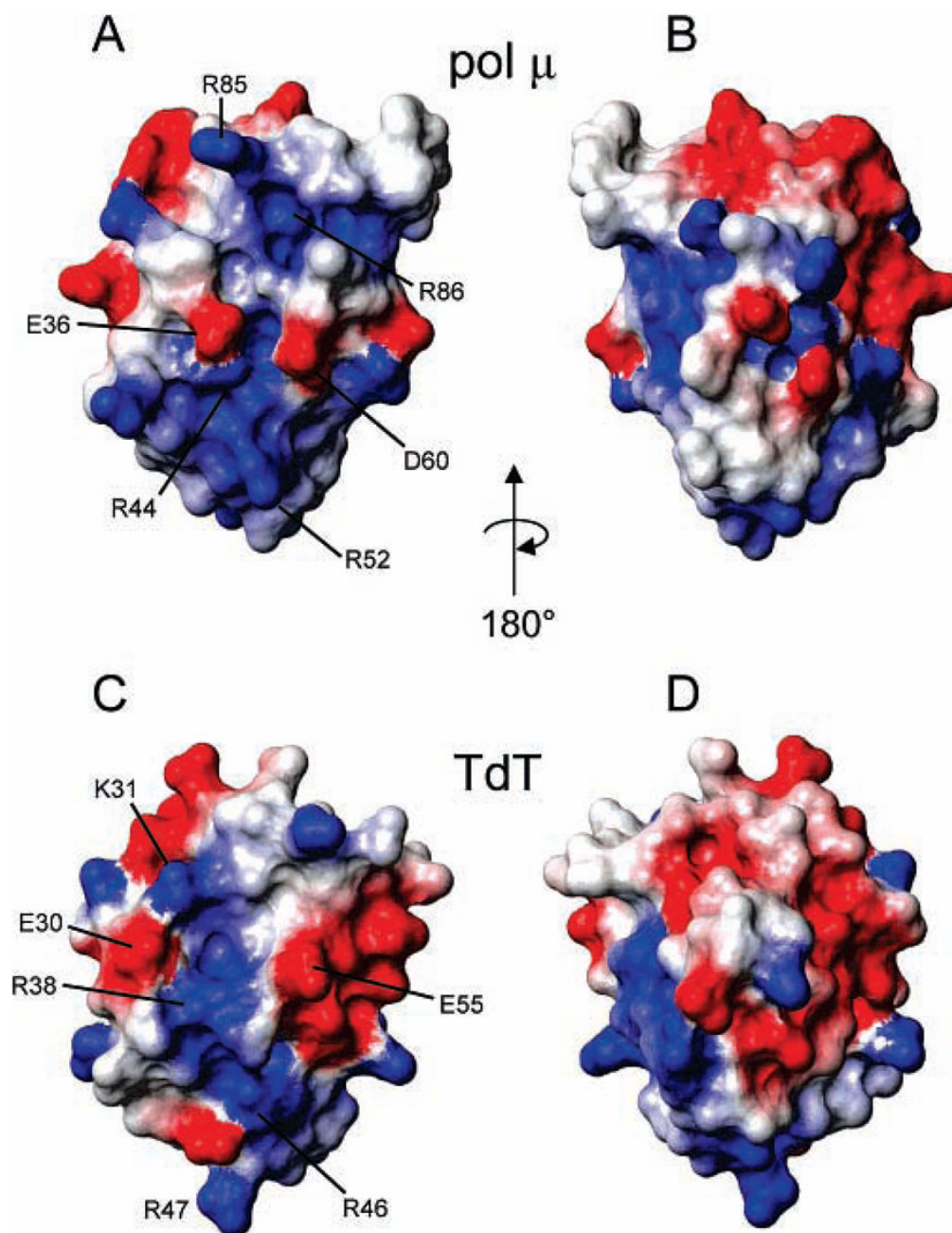
Figure 7.
Electrostatic surfaces of the BRCT domains of pol μ and TdT. (A) A ridge of positively charged residues on the surface of the BRCT domain of pol μ may be the site of phosphopeptide or DNA binding. (B) Electrostatic surface of the BRCT domain of pol μ rotated 180° about the vertical direction with respect to part A, showing a large patch of negatively charged residues. (C) As in the pol μ domain, a ridge of positively charged residues on the surface of the TdT domain may be the site of phospho-peptide or DNA binding. (D) Electrostatic surface of the BRCT domain of TdT rotated 180° about the vertical direction with respect to part C, showing a large patch of negatively charge residues.

Table 1

Statistics for the pol μ BRCT Domain Structure Ensembles

NOE distance restraints	
intraresidue	347
sequential	372
medium range ($i, i \pm 2$–4)	126
long range ($i, i > 4$)	661
total	1506
H-bond restraints (two per H-bond)	42
RDC[a]	
RMS error (Hz)	0.91 ± 0.17
R-factor[b]	7.6% ± 1.6%
Pearson's *r*	0.98 ± 0.006
ensemble rmsd(Å)	
backbone (residues 30–124)[d]	0.70 ± 0.12
heavy atoms (residues 30–124)[d]	1.08 ± 0.11
average violations per structure	
NOEs and H-bonds	1.4 ± 0.5
dihedrals	2.6 ± 0.9
rmsd (experimental restraints)[c]	
NOEs (Å)	0.0537 ± 0.0078
H-bonds (Å)	0.0126 ± 0.0078
dihedral angles (deg)	1.29 ± 0.25
rmsd (covalent geometry)[c]	
bonds (Å)	0.0042 ± 0.0001
angles (deg)	0.5910 ± 0.0195
impropers (deg)	1.9041 ± 0.1779
Ramachandran space (%)[e]	
most favored region	85.0 ± 2.8
additionally allowed	13.0 ± 2.9
generously allowed	1.4 ± 1.1
disallowed	0.5 ± 0.7

[a] The range of dipolar couplings was −11.2 to +8.8 Hz. The error in the RDCs was ±0.8 Hz.

[b] Calculated as described by Clore and Garrett (70).

[c] Output by ARIA (54), calculated by CNS (52) using the ensemble of the ten lowest energy structures.

[d] Average backbone rmsd of all structures with respect to the mean calculated with MOLMOL (71).

[e] Calculated with PROCHECK (72,73). The percentages are computed over all seven lowest energy structures.

Table 2

Activities Relative to Wild Type pol μ

	intrinsic	with Ku + XRCC4-ligase IV	
polymerase	% gap fill[a]	% complex[b]	% joining[c]
wild type	100	100	100
pol μ ΔBRCT	120	<10	1
R43A	110	<10	20
F46A	110	<10	7
F46L	120	35	34
L50A	120	<10	19

[a] Gap filling activity, relative to wild type full length pol μ (100%) determined on a gapped duplex substrate without Ku and XRCC4-ligaseIV(see Experimental Procedures).

[b] Amount of polymerase + Ku + XRCC4-ligase IV complex formed, relative to that seen with wild type full length pol μ (100%), determined by EMSA in Figure 4B.

[c] End joining activity, relative to wild type full length pol μ (100%), determined in the presence of Ku and XRCC4-ligase IV in Figure 4C.