

# UCSF

## UC San Francisco Previously Published Works

### Title

The active DNA-PK holoenzyme occupies a tensed state in a staggered synaptic complex

### Permalink

<https://escholarship.org/uc/item/0v65f2vz>

### Journal

Structure, 29(5)

### ISSN

1359-0278

### Authors

Hepburn, Morgan  
Saltzberg, Daniel J  
Lee, Linda  
[et al.](#)

### Publication Date

2021-05-01

### DOI

10.1016/j.str.2020.12.006

Peer reviewed



Published in final edited form as:

Structure. 2021 May 06; 29(5): 467–478.e6. doi:10.1016/j.str.2020.12.006.

## The active DNA-PK holoenzyme occupies a tensed state in a staggered synaptic complex

Morgan Hepburn<sup>1</sup>, Daniel J. Saltzberg<sup>2</sup>, Linda Lee<sup>1</sup>, Shujuan Fang<sup>1</sup>, Claire Atkinson<sup>3</sup>, Natalie C.J. Strynadka<sup>3</sup>, Andrej Sali<sup>2</sup>, Susan P. Lees-Miller<sup>1,4</sup>, David C. Schriemer<sup>\*,1,4,5,6</sup>

<sup>1</sup>Department of Biochemistry and Molecular Biology, University of Calgary, Calgary, Alberta, Canada

<sup>2</sup>Department of Bioengineering and Therapeutic Sciences, Department of Pharmaceutical Sciences, and California Institute for Quantitative Biomedical Sciences, University of California, San Francisco, CA 94158, USA

<sup>3</sup>Department of Biochemistry and Molecular Biology and High Resolution Macromolecular Electron Microscopy Facility, The University of British Columbia, Vancouver V6T 1Z3, British Columbia, Canada

<sup>4</sup>Robson DNA Science Centre, Arnie Charbonneau Cancer Institute, University of Calgary, Calgary, Alberta, Canada

<sup>5</sup>Department of Chemistry, University of Calgary, Calgary, Alberta, Canada

<sup>6</sup>Lead contact

### Summary

In the non-homologous end joining (NHEJ) of a DNA double strand break, DNA ends are bound and protected by DNA-PK, which synapses across the break to tether the broken ends and initiate repair. There is little clarity surrounding the nature of the synaptic complex and the mechanism governing the transition to repair. We report an integrative structure of the synaptic complex at a precision of 13.5Å, revealing a symmetric head-to-head arrangement with a large offset in the DNA ends and an extensive end-protection mechanism involving a previously uncharacterized plug domain. Hydrogen/deuterium exchange mass spectrometry identifies an allosteric pathway connecting DNA end-binding with the kinase domain that places DNA-PK under tension in the kinase-active state. We present a model for the transition from end-protection to repair, where the synaptic complex supports hierarchical processing of the ends and scaffold assembly, requiring displacement of the catalytic subunit and tension release through kinase activity.

---

\*Corresponding author: David C. Schriemer, dschriem@ucalgary.ca.

Present address: University of Calgary, 3330 Hospital Dr. NW, Calgary, Alberta, Canada, T2N 4N1

Author contributions

MH, DJS, and DCS conceived the study and experimental design. MH performed all experiments, with assistance from CA and NCJS on electron microscopy, SF, LL and SPLM on protein expression and isolation. MH and DCS analyzed the experimental results and co-wrote the manuscript. DJS and AS conceived the integrative approach to structure modeling. All the authors contributed to writing and/or editing the manuscript.

Declaration of interests

The authors declare no conflict of interest.

## Introduction

DNA double strand breaks (DSBs) must be repaired to prevent cell death, but also must be repaired accurately to maintain genome stability. Homologous recombination (HR) is the most error-free mechanism to repair a double strand break, but it is only active when sister chromatids are available as a template, whereas non-homologous end joining (NHEJ) is active throughout the cell cycle. NHEJ involves untemplated ligation. This seemingly simple function requires a complex multi-protein process to maintain the highest fidelity possible. Several processing factors are required, depending on the chemical and structural properties of the break, and the order in which the factors are engaged for repair has a direct impact on the process and outcomes of NHEJ (Chang et al., 2016; Conlin et al., 2017; Waters et al., 2014). DSBs, regardless of sequence (Walker et al., 2001), are quickly bound by the highly-abundant Ku70/80 heterodimer, followed by the recruitment of DNA-dependent protein kinase catalytic subunit (DNA-PKcs) to form the ultralarge holoenzyme DNA-PK (~620kDa) (Yin et al., 2017). The recruitment of a DNA-PK to each DNA end prevents resection (Weterings et al., 2003) and sets the stage for the formation of a dimeric complex (i.e. the synaptic state) that tethers the broken ends (DeFazio et al., 2002). The complex also initiates the recruitment of DNA processing factors and ultimately supports ligation of the broken ends (Liang et al., 2017).

The synaptic complex appears to contribute an organizational and gating function to NHEJ. Repair involves an iteration through different processing events that remodel the DNA ends prior to ligation (Chang et al., 2016). The entire process requires several structural factors to present the ends for ligation the instant they are sufficiently processed (Chang et al., 2016; Conlin et al., 2017; Waters et al., 2014). Specifically, the synaptic complex consists of matched holoenzymes across the break (DeFazio et al., 2002) that recruit the NHEJ factors X-ray repair cross-complementing protein 4 (XRCC4), XRCC4-like factor (XLF), paralog of XRCC4 and XLF (PAXX), and Ligase IV (LigIV) (Cottarel et al., 2013; Wang et al., 2018). In particular, an XLF-XRCC4-LigIV subcomplex (Hammel et al., 2011; Ropars et al., 2011; Sibanda et al., 2001) participates in scaffolding the break site through interactions with Ku70/80 (Costantini et al., 2007; Yano et al., 2008) and the recruitment of PAXX - minimally involving an interaction with Ku70 (Ochi et al., 2015; Tadi et al., 2016; Xing et al., 2015) - adds to stabilization (Graham et al., 2018; Wang et al., 2018).

The complexity of the preligation state raises questions regarding the overall organization of components and how processing events are regulated in such a system. Recently, compelling single-molecule FRET studies have supported the existence of long-range and short-range synaptic structures (Graham et al., 2016). The long-range structure maintains the separation of DNA ends by 100Å (Graham et al., 2016). It transitions to a short-range complex that appears to moderate at least some of the processing functions as well as ligation by LigIV (Graham et al., 2016, 2018; Stinson et al., 2019). This staging of the synaptic state argues strongly for a well-organized and hierarchical approach to repair, where LigIV activity is held in abeyance until a processing program is organized and initiated. Subsequent studies also appear to favor such a model (Stinson et al., 2019), but it is unclear how synapsis can serve an organizational role and coordinate a transition to a ligase-competent short-range

state, and indeed why the ends need to be held far apart in a long-range complex in the first place.

The core NHEJ components are involved in the transition (Graham et al., 2016). Specifically, the short-range structure consists of a well-defined ternary complex involving the XLF dimer interacting with two XRCC4-LigIV subcomplexes (Graham et al., 2018). The complex creates an environment where LigIV is held proximal to the break and poised to repair it upon sufficient end-processing (Graham et al., 2018; Stinson et al., 2019). There seems to be no explicit role for DNA-PKcs in ligation, as studies indicate that (auto)phosphorylation triggers its release prior to ligation (Hammel et al., 2010; Uematsu et al., 2007). Most models of NHEJ repair maintain the presence of DNA-PKcs up to ligation (Jette and Lees-Miller, 2015), but the recent two-stage model suggests that it is retained throughout the process, perhaps undergoing a conformational change to manage the transition (Graham et al., 2016). It is also unclear how end-protection by DNA-PK is relieved as repair proceeds and how this release is coordinated with a requirement for tethering and positioning of processing factors throughout the process.

A structural model of the long-range synaptic complex is required to understand how the system supports end-joining, but studies have proven challenging given its size and structural heterogeneity. Recently, a structure of DNA-PKcs was solved by X-ray crystallography at 4.3Å resolution, involving 89% of the sequence (Sibanda et al., 2017). It was extended to include most of Ku70/80 and a short dsDNA sequence using cryo-electron microscopy (EM), at 6.6 Å resolution (Yin et al., 2017). Together, these structures reveal a “head” feature consisting of intertwined FAT (FRAP-ATM-TRRAP), FAT-C-terminal and kinase domains assembled on a large N-terminal solenoid that forms a double ring shape. The DNA is presented in a channel formed by Ku70/80 and the solenoid, involving surprisingly few contacts with the catalytic subunit, but bending the N-terminal HEAT (Huntingtin, Elongation factor, PP2A, subunit-TOR) repeats towards the head structure (Yin et al., 2017). Attempts have been made using EM to generate structures of a synaptic state with limited results (Baretic et al., 2019; Spagnolo et al., 2006), perhaps because of the strong orientational bias of DNA-PK on sample grids. Here, we used mass spectrometry and integrative structure modeling to produce a model of the long-range synaptic complex at 13.5Å precision. This model presents the two holoenzymes in a head-to-head orientation with a considerable offset. It rationalizes key elements of the transition to a short-range repair complex, identifies a plug domain with a role in regulating the transition, and localizes unassigned structural elements (Ku80 C-terminal and PAXX). Complementary conformational analyses show a system under high tension, which must be relieved through a structural transition on the pathway to repair.

## Results

### Assembling the Synaptic Complex for Structural Mass Spectrometry

For the preparation of the synaptic complex, we used two methods. In the first, we used a DNA substrate modelled after Graham *et al.* (Graham et al., 2016), involving a 2kb stretch of DNA with an internal biotin for affinity capture. This substrate supports synapsis through the circularization of the DNA. It was used to capture purified Ku70/80 plus

DNA-PKcs (Fig. S1A–D), thus forming the holoenzyme, relying upon the high affinity of the DNA ends for the assembled trimer (West et al., 1998). Assemblies were formed with and without PAXX, a known stabilizer of core NHEJ proteins (Ochi et al., 2015) and the synaptic complex (Wang et al., 2018). The use of the biotin tag allowed for capture on streptavidin beads, followed by washing and removal of any unbound or lightly bound components. We confirmed synapsis on the unwashed state using negative-stain EM but note the presence of some possible aggregation and multiple Ku loadings (Fig. S2). The second method was based upon the use of a much shorter DNA segment. Biotinylated 100bp (blunt-end) dsDNA was used to capture Ku70/80 plus DNA-PKcs. The assembly was dispersed on the bead surface by adding excess free biotin, to model a one-ended DNA break. Protein capture was confirmed by Western blot at the expected stoichiometry (Fig. S1D). Then, holoenzyme was formed free in solution and as synapsis can only occur between holoenzymes, we extracted the holoenzyme from solution using the bead-bound holoenzyme and demonstrated enrichment (Fig. S1E). Crosslinking experiments were conducted with the 2kb construct (as the yields were greater) and we used the second method to test findings.

### Crosslinking the Synaptic Complex

Next, we crosslinked the synaptic complexes in the following forms: holoenzyme, holoenzyme with a non-hydrolyzable ATP analog (adenylyl-imidodiphosphate, AMP-PNP), holoenzyme with PAXX, and holoenzyme with AMP-PNP and PAXX. Nominally one-ended holoenzyme complexes on 100 bp DNA were used as control samples, with and without AMP-PNP. To aid in the presentation of results, Fig. 1 is provided a reference for structural features of the holoenzyme. Especially noteworthy is a long 198-residue loosely-structured segment containing the ABCDE phosphorylation sites (residues 2609-2647) that regulates the interaction of DNA-PKcs with Ku70/80 and promote DNA end-processing (Cui et al., 2005; Jiang et al., 2015; Neal et al., 2014). Both Ku70 and Ku80 have smaller structured C-termini connected to the heterodimer through short disordered regions (Walker et al., 2001; Zhang et al., 2001, 2004). Crystallographic analysis suggests that elements of the Ku80 C-terminus may interact with DNA-PKcs near the PRQ domain (Sibanda et al., 2017; Yin et al., 2017), which is another site of phosphorylation thought to regulate progression to ligation (Jiang et al., 2015).

Approximately 200-300 unique DSS crosslinks mostly between pairs of lysine residues were identified for each state and mapped on Circos plots (Grimm et al., 2015) (Fig. 2A). The maps of all states appear similar but differ in relative abundance (Fig. 2B). The set of crosslinks was then inspected for obvious trends and domain localizations, prior to integrative modeling. We consider in the first place the set of crosslinks that are satisfied by a single holoenzyme (regardless of the possibility that they may, in fact, span the two holoenzymes in a synaptic complex). On average, 42% of the crosslinks span sites on the known holoenzyme structure (Yin et al., 2017) (Fig. 3A). We observe 3 crosslinks between the Ku80 C-terminal domain and the base of DNA-PKcs, near the PQR sites around residue 1985, localizing the Ku80 C-terminal to the base. Many of the remaining crosslinks involve nominally disordered regions, primarily residues 2577-2773 of the ABCDE domain. Numerous crosslinks are formed between this region, the N-terminal HEAT repeats (the

“arm”) and the middle HEAT repeats. No crosslinks are found between this region and either the kinase or FAT domain, despite their proximity in the holoenzyme structure (Fig. 3B).

Crosslinking also indicates a nucleotide-induced conformational change in the N-terminal arm. In the DNA-free monomeric state, the “hand” at the N-terminal end is too far to crosslink to the FAT domain (~50Å) (Sibanda et al., 2017), but in the DNA-bound form of the holoenzyme, we observe a crosslink between these regions (residues 99-3196) (Fig. 3). We also observe it in the synaptic complex. These states are all in the nucleotide-free form. Interestingly, upon nucleotide binding, this crosslink is strongly reduced in intensity in the holoenzyme and it completely vanishes in the synaptic state. PAXX binding does not return it. This quantitative analysis was extended to the rest of the crosslinks (Fig. 2B). Two crosslinks stand out as unambiguously inter-holoenzyme, where the same peptide sequence is found on each side of the crosslink. The one crosslink supports a head-to-head interaction, linking residues 4085-4085 in the kinase domain. It is absent in the non-synaptic control, whereas the addition of PAXX increased its abundance considerably (Fig. 2C). Loading the kinase with AMP-PNP generated the highest abundance. In fact, AMP-PNP binding appears to drive low levels of synapsis even in the nominally non-synaptic control. This is entirely possible. The design of the control attempts to disperse DNA evenly on the beads, but the preparation likely possesses some distribution in the dispersion, allowing a subset to dimerize. Surprisingly, the addition of PAXX to the nucleotide-loaded holoenzyme increased the intensity in this crosslink relative to the nucleotide-free state, but not the nucleotide-bound state. Taken together, the similarity in the crosslinking for the two nucleotide-loaded states (Fig. 2B middle) support that one conformational state is shared between the two forms. PAXX may slightly alter the conformation of the head domain and thus reduce crosslinking locally in this region. We note that this crosslink was detected in the alternate form of the synaptic complex preparation (data not shown).

The other unambiguously inter-holoenzyme crosslink supports a base-to-base interaction, linking residues 1869-1869 in the middle HEAT repeats. The crosslink was identified in all states, including at significant levels in the non-synaptic controls (Fig. 2C). The intensity patterns are very different from the head-to-head crosslink: the intensity decreases upon adding AMP-PNP and/or PAXX. In preliminary modeling runs, we observed that the base-to-base crosslink is never satisfied. Thus, the crosslink is likely an artifact indicative of aggregation (Fig. S2) and it was not observed in the alternate synaptic complex preparation. Finally, we note that our crosslinking data localize PAXX near the base of DNA-PKcs. Crosslinks are scant but there is a link between the N-terminal of PAXX and the Ku80 C-terminal core structure (Fig. 2A). Surprisingly, no crosslinks are found to Ku70 where the PAXX tail is known to bind.

**Structure of the DNA-PK synaptic complex.**—Armed with the crosslinking data, the available high-resolution holoenzyme structure (Yin et al., 2017), the Ku70/80 termini structures (Zhang et al., 2001, 2004), the physical constraints of connectivity and excluded volume, we performed integrative structure modeling using the Integrative Modeling Platform (IMP) (Russel et al., 2012), to determine the orientation of the holoenzyme in the synaptic form (Fig. S3). We constructed separate models for each of the six crosslinked states, with each model consisting of two copies each of DNA-PKcs, Ku70 and Ku80. Each

crosslink was allowed be satisfied by either an inter-or intra-holoenzyme distance (Fig. S3). In total, 1.2 million configurations of the synaptic complex were assessed. For each dataset, the head-to-head crosslink (4085-4085) was applied to filter the ensemble post modeling, followed by structural clustering. Approximately 2000 representatives of each cluster for each dataset were combined, and clustered again to identify common states among all six crosslinked states (Fig. 4A). Two distinct clusters satisfying the head-to-head crosslink emerged, along with a more diffusive third cluster that did not. The dominant cluster of solutions represents ~30% of the total configurations and has a precision of 13.5Å. This cluster is comprised of models from every synaptic data set (highlighting the reproducibility of the structure determination from different data sets), with the dominant contributions coming from the nucleotide-loaded forms. It includes some minor contribution from the nominal control state, the nucleotide-loaded holoenzyme on 100 bp DNA. This finding is not surprising, given the observation of the head-to-head crosslink (at low levels) in the sample (Fig. 2C).

Strikingly, the structure places the holoenzymes in a side-to-side orientation, across an interaction interface that begins at the kinase and FAT domains and runs along the middle HEAT repeats (Fig. 4B). We did not explicitly include DNA strands in the modeling, but we fit the model with the holoenzyme structure (PDB 5Y3R) to orient them. This fitting places the incoming broken DNA strands in parallel but staggers their ends by ~105 Å. The symmetry of the solution is a product of the modeling and not an imposed restraint.

The smaller cluster of solutions that satisfies the head-to-head crosslink represents 12.1% of all models with a precision of 15.3Å. This solution orients one holoenzyme at a right angle to the other along its long axis, while still maintaining contact between the FAT/kinase domains of DNA-PKcs. The interaction interface does not involve any additional DNA-PKcs contacts. Rather, the Ku70/80 heterodimers contact each other extensively. We fit the holoenzyme structure as above, but here we found that the distal DNA ends (*i.e.* the ones opposite the “break”) would collide at Ku70/80 (Fig. S4A). Thus, the smaller cluster cannot represent the synaptic state. Our EM data suggest that some DNA strands could have multiple Ku heterodimers loading on a section of dsDNA, forming “beads on a string” (Fig. S2). This phenomenon has been observed previously (DeFazio et al., 2002) and explains how a subset of solutions could position two Ku heterodimers close together.

The remaining models belong to a cluster with low precision (30.5Å), placing the holoenzymes in an approximate face-to-face position, where the base of one molecule interacts with the loosely structured domain of the other (Fig. S4B). The crosslinks that support the cluster are mostly contained in this domain and in the other disordered regions. This cluster is representative of the residual aggregated states in the sample. Not surprisingly, a fraction of the models from all six crosslinked samples populate this diffusive solution set (Fig. 4). Taken together, the modeling supports a structure of the synaptic complex that brings the ends of the DNA break together in a staggered manner, separated by ~100Å.

The model also defines an unexpected “plug” domain comprised of the nominally disordered sequence (2577-2773), colored red (Fig. 4B). As noted above, crosslinks in this region

cluster in the cradle of DNA-PKcs. To extend this finding, we conducted DNA footprinting experiments on DNA-PKcs bound to various DNA constructs. The footprinting returns binding-site information at a peptide level of resolution (Fig. S5A–D). We identified peptides in the N-terminal arm as expected, but we also identified peptides that form an extended interaction surface, including the plug domain and part of the FAT domain. The extended interaction site is independent of the nature of the DNA used – blunt or overhang. When the DNA is constrained through the addition of Ku70/80, the same regions are footprinted as well as the linker between the structured regions in Ku80. While it is difficult to localize the extended binding sites with precision, the footprint is clearly relegated to one face of DNA-PKcs, consistent with the electrostatic potential on the surface of the protein (Fig. S5E). The engagement of the large disordered domain confirms its role as a plug in the center of the cradle.

**Ku70/80 C-terminal regions define a supportive base.**—We then inspected how the remaining structural elements could be placed within the model. Integrative modeling localizes the structured Ku80 C-terminal domain internal to the synaptic complex (Fig. S6A). To refine its position using interaction energies and shape complementarity, we used HADDOCK, aided by the 5 crosslinks identified between the Ku80 C-terminus and DNA-PKcs. Top-scoring clusters docked the Ku80 C-terminus directly under DNA-PKcs in the central cavity made by the middle and N-terminal HEAT repeats, readily accommodating the linker to the Ku80 core domain. This aspect of the model is well-supported by DNA footprinting (Fig. S5A–D). Interestingly, although the crosslinking data for PAXX is too sparse for precise modeling, the available data situate the structured head domains neatly in the gap between the Ku80 C-terminal domain and main Ku70/80 heterodimer (Fig. S6B). Finally, based on crosslinking alone, the Ku70 SAP domain is also difficult to locate with precision, but DNA footprinting locates it directly behind the Ku heterodimer (Fig. S6C), which is consistent with its suspected role in DNA binding to prevent inward movement of the DNA-PK complex (Hu et al., 2012; Zhang et al., 2001).

### Conformational Analysis of the Holoenzyme

The crosslinking data, particularly the dependency of the arm motion on nucleotide occupancy, suggest the regulation of long-range protein conformations through the kinase. To understand this further, we analyzed the bead-bound holoenzyme by hydrogen/deuterium exchange mass spectrometry (HX-MS), using a more sensitive version of the technology well adapted to large complexes (nanoHX-MS (Sheff et al., 2017)). This nanoHX-MS technology was essential for the project as only small amounts of DNA-PKcs could be isolated. The approach requires sub-pmol amounts per analysis, whereas conventional HX-MS systems regularly consume 10-100fold more. Excellent sequence coverage was obtained (Tables S1). Differential HX-MS experiments were then conducted (e.g., free DNA-PKcs vs. DNA-bound DNA-PKcs) to determine the conformational transitions associated with stepwise assembly of the holoenzyme. The peptides with significant changes in deuteration were mapped to the DNA-PK cryo-EM structure (PDB 5Y3R) (Yin et al., 2017) in the orientation provided in Fig. 1. Approximately 11% of the holoenzyme sequence is not resolved in available structures (Yin et al., 2017), but as MS techniques can detect these regions, they are mapped on structure as bars, scaled to their length.



**DNA flexes the arm and is constrained by a plug domain.**—We first explored the effect of DNA binding upon DNA-PKcs. The changes in deuteration are well-defined, with the most prominent change found in the N-terminal arm (Fig. 5A, Fig. S7A,B). The pattern of change is striking: a large stabilization in the first 450 residues is coupled with an even larger destabilization in residues 363-388. The stabilizations at the N-terminus agree with the biochemical evidence of binding (Meek et al., 2012) and the known DNA binding site in the holoenzyme structure (PDB 5Y3R) (Yin et al., 2017). The major difference between this structure and the DNA-PKcs structure (PDB 5LUQ) (Sibanda et al., 2017) is a rotation of the N-terminal arm toward the FAT domain, involving a “switch point” at residue 382 of DNA-PKcs in the elbow (Fig. 1) (Yin et al., 2017). The HX data show that DNA binding alone is enough to induce this arm rotation (“flexing”): the inner elbow is stabilized, and the outer elbow is destabilized (Fig. 5A). Ku70/80 is not required to induce this conformational change. Arm displacement appears to propagate to the very N-terminal end of the domain. Two patches of stabilization are seen, consistent with the end of the arm knuckling into the FAT domain. This deuteration pattern occurs independent of the type of DNA used to capture the protein (either blunt-end or overhang, Fig. S7).

Interestingly, stabilizations in the arm are not the only ones that occur upon binding to DNA. Lower-magnitude stabilizations were also detected in the middle HEAT domain, at the base of DNA-PKcs across from the DNA binding site and flanking the large disordered region (Fig. 5A, left). These additional stabilizations could represent induced conformational changes, but they may also indicate novel, secondary DNA binding sites. This interpretation is consistent with the DNA footprinting data presented above. There are other minor changes in deuteration upon DNA binding. One involves a long-range destabilization in the kinase near the activation loop. DNA binding, either indirectly through the motion of the N-terminal arm or more directly through an extended DNA binding channel, can clearly influence the kinase domain. To explore this allosteric effect further, we conducted a nanoHX-MS analysis on the entire holoenzyme.

**An allosteric pathway between DNA binding sites and the kinase domain.**—The addition of Ku70/80 induces a widespread conformational response in DNA-PKcs (Fig. 5B, Fig. S8). The changes in deuteration are so extensive that interpretation beyond the domain level is challenging. Ku binding induces conformational changes throughout the molecule, including the kinase domain ~80Å away. As expected, Ku strongly reduces deuteration at the binding interface with DNA-PKcs, which is comprised of elements of the N-terminal arm and middle HEAT repeats. Additional conformational changes are seen in the elbow of the N-terminal arm consistent with an arm flexion that is “locked in” upon Ku binding. We used a slightly longer piece of DNA (100bp) to avoid restricting the formation of the complex and to allow for possible secondary interactions with DNA (a 25bp strand would only occupy the main channel). The secondary DNA binding sites are accentuated in the HX data and supported by DNA footprinting experiments on an even longer DNA construct (Fig. S5).

The two remaining notable effects include a widespread destabilization that occurs throughout the molecule and a stabilization at the base of DNA-PKcs (Fig. 5B). For the first, the destabilizations affect all domains, even the plug. The plug becomes disordered

further in a set of helices that precede the ABCDE phosphorylation sites (Saltzberg et al., 2019b). For the second, the addition of Ku70/80 induces a set of stabilizations in the base of DNA-PKcs. These changes confirm the placement of the Ku80 C-terminus in the base region, near the PQR site (Fig. 5B) (Sibanda et al., 2017), as seen in the crosslinking data.

**Nucleotide loading primes the allosteric pathway.**—Collectively, these findings identify a long-range allosteric pathway connecting Ku70/80, the wider DNA binding site and the kinase domain. To test this hypothesis further, we added AMP-PNP to the holoenzyme and repeated the nanoHX-MS analysis. We reasoned that, if an allosteric pathway exists, the addition of nucleotide would likely activate conformational changes along a similar trajectory. As the addition of ATP induces phosphorylation and release of DNA-PKcs from the DNA (Hammel et al., 2010; Jette and Lees-Miller, 2015; Merkle et al., 2002), we used AMP-PNP to characterize the conformational state immediately prior to hydrolysis and phosphorylation. Upon nucleotide binding, a conformational response was observed that cascades from the binding site all the way to Ku70/80 (Fig. 5C, Fig. S9). As anticipated, the nucleotide induces a strong local stabilization. It propagates throughout the kinase and includes elements of the FAT domain. More distal conformational changes involve the N-terminal arm at the shoulder and elbow. The changes in Ku70/80 are remarkable, particularly in their respective C-termini. For the most part, the heterodimer is stabilized upon nucleotide binding, including the narrow bridge over the DNA and both C-termini. For Ku80, there is a patch at the extreme C-terminal end that is destabilized, which correlates with a similar destabilization in its putative binding site on DNA-PKcs, near the PQR region. These changes in stabilization identify a repositioning/weakening of the binding site in the kinase-active state. The strong stabilization of the Ku70 C-terminal SAP domain has no definite structural rationalization at this point. However, our DNA footprinting experiments identified this region as a DNA binding site (Fig. S5A–D), which is supported by previous studies (Aravind and Koonin, 2000; Hu et al., 2012; Zhang et al., 2001). Thus, the HX data suggest that the kinase can influence DNA binding directly. This linkage is more dramatically highlighted by observing the effect of nucleotide binding on the plug domain. The entire domain becomes disordered, including its DNA binding site, revealing that the plug exists in a conformationally-relaxed mode in the kinase-active state where DNA binding is weakened at least partially (Jovanovic and Dynan, 2006).

Taken together, the kinase-active DNA-PKcs is a complex conformational switch that organizes the DNA end in a well-protected cavity. The kinase domain engages a long-range allosteric axis pathway that connects the nucleotide binding pocket with all points of DNA binding, and the kinase regulates interactions with the Ku70/80 C-termini at its base, likely through a mechanism involving N-terminal arm transitions.

**PAXX engages the allosteric pathway.**—We next determined if the addition of PAXX to the holoenzyme could influence its conformational behavior. PAXX binding to nucleotide-loaded holoenzyme only induces stabilizations (Fig. 5D, Fig. S10). These stabilizations are once again distributed throughout structure, but in highly localized regions. Most notably, regions of the Ku70 core domain are stabilized, which supports an earlier study that showed the C-terminal of PAXX interacting with the Ku70 subunit (Tadi et al.,

2016). This interaction locate PAXX at the base of the holoenzyme, consistent with the crosslinking. We observe a corresponding stabilization in the Ku80 C-terminal domain and linker, and one at the base of DNA-PKcs. The interaction induces stabilization of the DNA binding sites, including elements of the N-terminal arm, the plug, and the extended channel. Once again, the nucleotide binding site is associated with these changes, likely through the conformational effects on the N-terminal arm.

## Discussion

Using integrative modeling and structural mass spectrometry methods, we determined the structure of the long-range synaptic complex and the conformational changes in the holoenzyme that could accompany a transition out of the long-range state. Our data confirm the localization of the primary DNA binding site in the center channel and highlight the influence of DNA binding on the conformational state of the N-terminal arm. This arm undergoes a major flexion at the elbow upon DNA binding. Ku70/80 supports the conformational change in the arm and completes an allosteric pathway between the kinase and the DNA-binding site. The degree of activation is remarkable. Conformational change is transmitted throughout the holoenzyme, confirming that DNA-PKcs is a large conformational switch under the control of the kinase domain. Nucleotide loading generates the active kinase, which appears to force the arm towards a conformation closer to the DNA-free form. In other words, nucleotide loading generates a tensioned state, as the arm is pushed away from the FAT domain even when DNA is bound. The tensioned state is also characterized by weakened interactions with the Ku80 C-terminus and increased disorder in the plug domain (Fig. 5E).

Our structural model of the synaptic complex is consistent with a density map determined by negative-stain EM at 33Å resolution (Spagnolo et al., 2006) and SAXS reconstructions (Hammel et al., 2010), although there are also key differences. These representations identify a two-fold axis of symmetry, with the DNA-PKcs subunits offset along the axis. However, while these orient the ends of the two DNA strands in parallel, our model positions the ends much further apart from each other (~105Å) and places the Ku heterodimers much farther away from the interface than the EM model, for example. It is also consistent with the head-to-head dimerization of other members of the phosphatidylinositol 3-kinase related kinase family, ATM (Lau et al., 2016; Wang et al., 2016; Yates et al., 2020) and ATR (Rao et al., 2018).

Our model also reveals insights into how the DNA end is presented within the holoenzyme. The “exit channel” for the DNA in each monomer is blocked by a large domain, effectively forming a space-filling plug or barrier, preventing the DNA from extruding further from the channel. It may also serve to distort the incoming DNA end, directing it towards the FAT domain, splaying DNA over a wider surface for processing functions (DeFazio et al., 2002; Jovanovic and Dynan, 2006; Leuther et al., 1999). This hypothesis would rationalize the secondary binding sites observed by HX (Fig. 5A, left), however non-specific binding of the DNA strand cannot be ruled out. We note that the upper reaches of the secondary site (Fig. S5A–D) have been suggested to form part of an RNA binding domain (Song et al., 2019). Either way, the electrostatic potential surface of DNA-PKcs is decidedly polar,

which shows that the scaffold is capable of orienting DNA across a wide contact surface on the face opposite from Ku70/80. The structure also reveals that the Ku80 C-terminal domain and linker participate in binding DNA at the base of the complex, partly filling the hole formed by the middle and N-terminal HEAT repeats. This placement defines a PAXX binding site in the same region, running from Ku70 along the base of DNA-PKcs to the Ku80 C-terminal domain. The assembly of these components places the DNA ends in a well-protected pocket, consistent with the major role of DNA-PKcs as an end protector (Jiang et al., 2015; Weterings et al., 2003). Finally, the active kinase generates the most competent state for synapsis, by reorganizing the dimerization interface (Fig. 5C). It appears that the tensioned state is preserved in the synaptic complex because the N-terminal arm does not engage the FAT domain, even with PAXX added.

Our integrative model and conformational analysis allow us to address some unanswered questions about the organization of downstream repair functions. The long distance between the DNA ends is consistent with the lack of FRET observed between labeled DNA ends (Graham et al., 2016), supporting our model of the long-range complex. Even if the ends were splayed over the secondary sites on the plug-side of the holoenzyme, the distance separating them would still exceed 100Å. Therefore, the structure of this long-range complex explains why DNA end-processing might require a synaptic state. The large separation of the DNA ends presents an opportunity for the complex to assemble the rest of the machinery prior to committing to repair. The complexity of the damage that can occur upon a double-strand break necessitates repair by a wide range of enzymes to remove the damage and allow for ligation (*e.g.*, Artemis, FEN1, PNKP, Werner protein, MRN, and DNA polymerases (Serrano-Benítez et al., 2020)); moreover, each end could require its own combination of enzymes. Separating the ends and placing them in the center of a broad surface allows for larger and more complex lesions (*e.g.*, hairpins and long overhangs) to be accommodated and for recruiting the necessary accessory enzymes such as Artemis (Wang et al., 2005). In this context, the dynamics of the plug must be significant. Phosphorylation of the ABCDE sites in the plug domain (or perhaps simply nucleotide turnover) may connect kinase activity to the factors required for engaging the extreme ends of the DNA break.

The large separation of ends has further implications for end-processing. End-joining necessitates the recruitment of XLF, XRCC4 and LigIV in a scaffolding role across the break, and the presentation of LigIV for ligation (Graham et al., 2016, 2018). The synaptic complex can nucleate their recruitment, perhaps through each holoenzyme as an intermediate, ultimately to form a supporting structure with the stoichiometry that has been recently demonstrated (Graham et al., 2018) (Fig. S11). FRET studies have shown that a measure of processing takes place in the context of a short-range complex (Stinson et al., 2019). The dimensions of this complex are uncertain, but it is difficult to see how this complex could involve DNA-PKcs. A large reorientation would be required to bring the ends close together, requiring the remodeling of a large interface ( $>10,000\text{\AA}^2$  in buried surface area). There are two possible alternative mechanisms. In one, the ends are enzymatically unblocked (if required) in the long-range complex to the point where they can exit the base of DNA-PKcs and move closer together for further processing and ligation. Here, DNA-PKcs would retain a role in positioning the downstream repair machinery until a later stage in processing. In the other, DNA-PKcs would be released

from the break immediately after the recruitment of the scaffold and the coordination of any unblocking activity, to permit further processing and ligation on the minimal scaffold. Either mechanism requires the timely exit of DNA-PKcs, which can be facilitated by tension release in the synaptic complex. Specifically, we propose that phosphorylation (or even nucleotide turnover) provides sufficient free energy in the form of released arm tension to eject DNA-PKcs from the DNA and expose the ends to a short-range complex (Jette and Lees-Miller, 2015).

The synaptic structure raises questions that need to be explored in future studies. For example, the autophosphorylation of the ABCDE sites *in cis* is readily explained through conformational transitions in the plug domain, but it is much harder to rationalize how autophosphorylation *in trans* could occur at the PQR sites (Jiang et al., 2015), given their distance from the catalytic sites. Such phosphorylation could require some relay mechanism that invokes other intermediates or perhaps it occurs during release. It is also conceivable that a conformational change in DNA-PKcs could drive closer contact for autophosphorylation. The capture of additional accessory factors (*e.g.*, XLF, XRCC4 and LigIV) and integrative structural analysis should allow us to determine how the long-range synaptic complex recruits these factors and transitions to the next stages of repair.

## STAR Methods

### Resource Availability

**Lead contact.**—Further information and requests for resources and reagents should be directed to the Lead Contact, David Schriemer (dschriem@ucalgary.ca).

**Materials Availability.**—This study did not generate new unique reagents.

**Data and Code Availability.**—The HX-MS and XL-MS data were deposited to the ProteomeXchange Consortium *via* the PRIDE (Perez-Riverol et al., 2019) partner repository with the dataset identifier PXD017931. The synaptic complex integrative models, including final structures, modeling details, and input experimental data, were deposited into the PDB-dev repository for integrative models ([www.pdb-dev.com](http://www.pdb-dev.com)) (Burley et al., 2017), accession number PDBDEV000000XX.

### Experimental Model and Subject Details

**Microbes**—*Escherichia coli* BL21 cells were cultivated in terrific broth (TB) supplemented with 100 mg/; ampicillin at 37°C.

**Cell lines**—HeLa cells were purchased directly from National Cell Culture Center as a frozen cell pellet equivalent of 150L of suspension culture.

**Coding Availability**—All scripts and data used in the modeling procedures described above are archived at [https://www.github.com/salilab/synaptic\\_complex](https://www.github.com/salilab/synaptic_complex).

## Method Details

**Protein production, DNA preparations**—Human DNA-PKcs (PRKDC, Gene ID 5591) and Ku70/80 (XRCC6, Gene ID 2547 and XRCC5, Gene ID 7520) were purified using a previously described protocol with minor modifications (Goodarzi and Lees-Miller, 2004). Briefly, 150L HeLa cell culture equivalent (National Cell Culture Center, Minneapolis MN) was lysed and pelleted. The pellet, containing mostly intact nuclei, was extracted with high salt buffer. The fraction was then applied to a series of columns to progressively enrich for DNA-PKcs and Ku70/80: DEAE Sepharose, SP Sepharose, ssDNA, Heparin Hitrap, MonoQ, and lastly MonoS. The final fractions were concentrated on a 30 kDa cutoff Vivaspin concentrator (GE Healthcare, Mississauga ON) to 2.0 mg/mL (5.2 mg for DNA-PKcs and 2.0 mg/mL (1.1 mg) for Ku70/80) and purity verified using SDS-PAGE. The phosphorylation status of DNA-PKcs was found to be minimal, based on LC-MS/MS analysis. Human PAXX (C9orf142, Accession number NM\_183241) was amplified from a HeLa cell cDNA library and cloned into pGEX6P1 vector between BamH I/Xho I sites and transformed into *E. coli* (BL21). Cells were induced with IPTG (0.2 mM) for 10 hours, lysed, clarified and extracted with Glutathione Sepharose 4B. Bound protein was eluted with glutathione (20 mM), the GST tag removed with PreScission protease (GE Healthcare) and the cleaved product polished on a HiTrap Heparin HP column. Purity was confirmed by SDS-PAGE and Western Blot.

The following DNA sequences were synthesized with the corresponding complementary strand and used in DNA footprinting and pulldown experiments:

25bp blunt and overhang (with 5' 15nt poly T): 5'-  
AAGCTTGCATGCCTGCAGGTCGACC-3'-Biotin

28bp: 5'-Biotin – GGGCGAGGGTTGTCGTCAACGCCTGAG-3'

100bp: 5'-Biotin-  
GCACAATCTTCTCGCGCAACGCGTCAGTGGGCTGATCATTA ACTATCCGCT  
GGATGACCAGGATGCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCG-3'

Internally biotinylated substrates were prepared after Graham et al., 2016 (Graham et al., 2016) with slight modifications. Briefly, pet28a was used as a template to generate a 2kb sequence with tandem single strand cut sites for insertion of an internally biotinylated single strand (/5Phos/TGAGGGATATCGAA/iBiodUK/TCCTGCAGGC). Primers LSO71F (CGGAACATTAGTGCAGGCAGC) and LSO72R (GCGTAATGGCTGGCCTGTTG) were used to generate the 2kb DNA with the insertion site. To insert the internal biotin strand, DNA was cut with Nb.BbcVI and the biotinylated strand added in 10X excess and annealed and ligated with DNA ligase 4, and the DNA purified.

**Complex formation and isolation**—With the exception of the DNA-free DNA-PKcs and the DNA footprinting experiments, all samples were immobilized on beads. Briefly, a 1:8 ratio of biotinylated DNA to biotin was bound to streptavidin agarose beads (Agarose Ultra Performance, TriLink, San Diego CA). Beads were washed in DNA binding buffer (50mM Tris (for HX) or HEPES (for XL) pH 7.5, 75mM KCl, 5mM MgCl<sub>2</sub>, 5% glycerol, and 1mM DTT added fresh) and BSA added to block nonspecific interactions. Proteins

(Ku70/80 and/or DNA-PKcs, PAXX, +/- 1mM AMP-PNP) were added and incubated. For nanoHX-MS, the beads were washed 5X with HX wash buffer (10mM Tris pH 7.5, 75mM KCl, 5mM MgCl<sub>2</sub>, 3% glycerol, 0.2μM caffeine and 1mM DTT). For crosslinking MS, the beads were washed 5X in XL wash buffer (50mM HEPES, 75mM KCl, 5mM MgCl<sub>2</sub>, 5% glycerol, 1mM DTT, +/- 1mM AMP-PNP).

### Hydrogen-Deuterium Exchange Mass Spectrometry

Proteins were analyzed under nanoHX-MS conditions first to identify peptides used in deuteration analysis. Briefly, each protein or complex was digested with rNep2 in 100mM glycine-HCl (pH 2.5) for 3 minutes at 10°C. Digests were loaded onto a 200μm x 2.5cm trap column packed with 3.6μm Jupiter C18 beads (Phenomenex, Torrance CA) washed for 2 minutes, and then separated on a 150μm x 7 cm picofrit column (New Objective, Woburn MA) packed with Jupiter C18 beads, using a 10 minute gradient of 10-40% Solvent B (97% acetonitrile, 0.23% formic acid). The separation devices were held at 2°C in a specially designed cooler (Sheff et al., 2017). Data were collected by LC-MS/MS using a nano-Ultra 2D HPLC system coupled to a 5600 QqTOF (m/z 350-1250) (Sciex, Concord ON) via a data-dependent acquisition method (top-10) in a recursive manner, until no new identifications were made (typically 3-5 runs). Peptide lists were compiled from database searches using Mascot v2.4.0 (Matrix Science, Boston MA) and PEAKS 8.5 (BSI, Waterloo ON) and combined to make the final peptide maps. Searches used a database of holoenzyme proteins, PAXX and low-level contaminating proteins, with 10ppm error in MS1, and 0.05Da error in MS2. The peptide maps were validated manually, eliminating peptides with poor intensity and unresolvable spectral overlap.

Samples were all labeled with 45% D<sub>2</sub>O in a bead slurry format. That is, 2.5μL of streptavidin bead slurry containing immobilized DNA and bound protein in HX buffer (10mM Tris pH 7.5, 75mM KCl, 5mM MgCl<sub>2</sub>, 3% glycerol, 0.2μM light caffeine, with 1mM DTT added fresh) was labeled for 5 minutes with HX buffer prepared in 90% D<sub>2</sub>O (and heavy caffeine). Samples were labeled for 5 minutes at 25°C, the optimum timepoint for differential HX-MS as determined from a previous kinetics analysis of DNA-PKcs (Sheff et al., 2017). Samples were quenched with cold rNep2 in 100mM glycine-HCl (pH 2.5) and digested for 3 minutes at 10°C. Half the sample was analyzed by LC-MS (m/z 350-1250) for deuterium analysis, using the system described above, and the remaining sample was analyzed to check for dispensing volume errors using the light and heavy caffeine signals (Sheff and Schriemer, 2014). Unbound DNA-PKcs was analyzed using the same slurry format, except all DNA binding sites were blocked with biotin. All samples were analyzed in quadruplicate. Peptide-level labeling data were analyzed using HX-DEAL in Mass Spec Studio v2 (manuscript in preparation) and normalized for back-exchange variation using a subset of peptides that showed no change in deuteration between states. Data are reported as percent corrected deuteration. Differences were mapped onto the 5Y3R holoenzyme structure in Chimera (Pettersen et al., 2004).

### DNA Footprinting

Following established protocols (Vaughan and Kao, 2015), proteins were equilibrated with DNA in binding buffer (50mM Hepes, 75mM KCl, 5mM MgCl<sub>2</sub>, 1mM DTT) and then

crosslinked with 0.1% formaldehyde for 5 minutes at room temperature. The reaction was quenched with 2M glycine and samples were digested overnight with trypsin. DNA and bound peptides were precipitated with 0.3M sodium acetate and cold ethanol in  $-80^{\circ}\text{C}$ , overnight. The precipitate was recovered with centrifugation and washed 3X with cold 75% ethanol. All supernatant was removed and the remaining ethanol evaporated, before resuspending the sample (25mM sodium acetate, pH 5.2, 200mM NaCl) and reversing the crosslinks with heating ( $70^{\circ}\text{C}$  for 1 hour). The released peptides were acidified and zip-tipped, then reconstituted in 1% formic acid and analyzed by LC-MS/MS. Peptides were detected by data-dependent acquisition (m/z 350-1200, top 12) using an Orbitrap Lumos (Thermo Scientific, San Jose CA) equipped with an EasyLC1000 and a 50cm EASY-Spray™ C18 LC column operated with a 25-minute gradient of 3-50% solvent B (98% acetonitrile, 0.1% FA). The instrument was operated at a resolution of 120000 (MS). Charge states +2-8 were selected for HCD fragmentation (35% NCE) and MS/MS measured at a resolution of 15000. Control samples were processed in an identical fashion, with the exception of formaldehyde treatment, to account for strong nonspecific interactions. Data were searched on PEAKS 8.5, with a peptide tolerance of 10ppm, and MS/MS tolerance of 0.02Da, with up to 3 missed cleavages allowed and oxidized methionine selected as variable modification (FDR of 0.1%). To identify peptides containing binding sites, feature intensities were measured in the Mass Spec Studio, and referenced to the control samples. Replicate analysis of controls and DNA-PKcs samples established a distribution that allowed us to set a minimal intensity threshold at the 98% confidence interval. Positive hits were required to exceed this threshold and demonstrate enrichment in the crosslinking samples greater than 20fold.

### Electron Microscopy

DNAPKcs (5pmol) and Ku70/80 (2.5pmol) were combined with 2kb DNA (2.5pmol) in DNA binding buffer +/- 1mM AMP-PNP. For PAXX samples, 7.5pmol of PAXX was included. Complexes were equilibrated on ice for 15 min., diluted 1:5 with DNA binding buffer and immediately 3 $\mu\text{L}$  of sample applied to a glow-discharged carbon film 400 mesh copper grid (Electron Microscopy Sciences, Hatfield PA). The sample was washed twice with water, then stained with 2% uranyl acetate. Excess stain was blotted off and the grids allowed to dry before imaging on 120kV Talos microscope with a BM-Ceta camera (Thermo Scientific).

### Crosslinking Mass Spectrometry

Complexes were isolated on DNA bound to streptavidin agarose beads (Fig. S1) to enrich for crosslinks corresponding to DNA bound complexes. Bead-bound samples were crosslinked with disuccinimidyl suberate (DSS, Thermo Scientific) at molar concentration ratios of 1:1, 2:1, 3:1 and 6:1 (crosslinker to total lysine content) for 30 minutes at  $37^{\circ}\text{C}$ , with nutation. Ammonium bicarbonate (100mM) was used to quench the reaction and then trypsin added for overnight digestion ( $37^{\circ}\text{C}$  with nutation). Peptides were recovered, acidified to 1% formic acid, zip-tipped and reconstituted in 1% formic acid for LC-MS/MS. Each state was exhaustively analyzed on two different Orbitrap platforms (Velos and Lumos) to saturate crosslinked peptide identifications. Data were collected as in DNA footprinting, except charge states 1-3 were rejected, LC gradients were extended (60-70 min) and MS/MS



resolutions were instrument specific (Velos: 7500 and Lumos: 30,000). Crosslinked peptides were identified using the CRIMP plugin in the Mass Spec Studio (Sarpe et al., 2016) with default settings. Circos plots were generated with Xvis (Grimm et al., 2015) and modified. The features for all manually-validated crosslinks were integrated in the Mass Spec Studio, normalized against the total peptide intensity and corresponding features compared across states. The natural variance in fold-change was evaluated with replicates, where ratios were recentered, fit to a normal distribution and  $\pm 3$  sigma used to establish a cut-off value. A comparison of states involved a similar recentering and features with no corresponding peptide in a given state comparison were assigned a maximal  $\log_2(\text{ratio})$  of 9. Crosslinks not selected in the data-dependent acquisition process because of sampling limitations were few, and annotated using a match-between-runs approach (Cox et al., 2014).

### **Integrative Structure Modeling of the Human Synaptic Complex**

Integrative structure determination proceeded through four stages (Schneidman-Duhovny et al., 2014; Webb et al., 2018): (1) gathering data, (2) representing subunits and translating data into spatial restraints, (3) sampling of structural components to produce an ensemble of structures that satisfy the restraints and (4) analyzing and validating the ensemble structures and data. The integrative structure modeling protocol (stages 2, 3 and 4) was scripted using the Python modeling interface (PMI), which is a library for modeling macromolecular complexes based on the open-source integrative modeling platform (IMP) version 2.11.0 (<https://integrativemodeling.org>). Analysis of ensembles was performed using the PMI\_analysis ([https://github.com/salilab/PMI\\_analysis](https://github.com/salilab/PMI_analysis)) and imp-sampcon (<https://github.com/salilab/imp-sampcon>) libraries. The specific procedures detailed below are an updated version of previously described protocols (Saltzberg et al., 2019a; Webb et al., 2018).

In stage 1, the set of data and information to be used in modeling was gathered. The cryo-EM structure 5Y3R supplied atomic coordinates for the majority of DNA-PKcs and the N-terminal domains of Ku70 and Ku80 (Yin et al., 2017) and localized the binding interfaces of Ku70 and Ku80 on DNA-PKcs. Additional structures for the C-terminal globular domains were obtained from structures 1JJR (Ku70 (Zhang et al., 2001)) and 1RW2 (Ku80 (Zhang et al., 2004)). Chemical cross-links were collected on six constructs of the synaptic complex, as described above.

In stage 2, the molecular representation of the system components was constructed, and the data translated into spatial restraints to score alternative structures. The molecular representation of the synaptic complex components must be sufficiently precise to allow for an accurate definition of spatial restraints given the data and biological interpretation of the model, yet sufficiently simple for efficient sampling of alternative models. This general guidance was applied to the representation of the synaptic complex in two ways. First, we represented the complex components with atomic models in a multi-scale fashion, using 1-10 residues per bead. Second, components of the complex for which atomic models are available were treated as rigid bodies during structural sampling, within which the distances between pairs of beads are fixed. The rigid body comprising the components of the cryo-EM structure fixes the intra-holoenzyme interaction interfaces among Ku70, Ku80,

and DNA-PKcs. All rigid bodies were consistent with the crosslinks between residues within the rigid bodies. Sequence segments missing from the atomic structures were represented by flexible strings of beads.

With this representation, input information from Stage 1 was translated into spatial restraints, which assess the parsimony between a piece of information and the structures of the components defined by the model representation. The following restraints were defined:

1. Excluded volume restraints were applied to the largest bead representation for each residue by applying a harmonic potential to the distance between sphere surfaces of close beads as described previously (Shi et al., 2014). Bead radii are determined *via* a statistical measure based on the total mass of the residues it represents (Alber et al., 2007a).
2. Sequence connectivity restraints were applied to all consecutive beads in a molecule using a harmonic upper bound on the distance between the bead centers. The center of the harmonic potential is 3.6 Å times the number of residues between the first residue of the N-terminal bead and last residue of the C-terminal bead.
3. DSS chemical crosslinks were converted into distance restraints between crosslinked residues, relying on a Bayesian scoring function (Erzberger et al., 2014) using a cross-linker arm length of 35 Å. The restraint was formulated to consider ambiguity of the cross-linked sites due to the presence of two copies of each protein in the system (Copy1 and Copy2). For each observed cross-linked residue pair, four alternative assignments were constructed: the intra-molecular pairs (Residue1::Copy1--Residue2::Copy1 and Residue1::Copy2--Residue2::Copy2) and inter-molecular pairs (Residue1::Copy1--Residue2::Copy2 and Residue1::Copy2--Residue2::Copy1). An “ambiguous” crosslink restraint was then evaluated against the model structure by multiplying the scores for all alternative restraint assignments (Shi et al., 2015). A pair of crosslinked peptides that contain the same residue were treated as unambiguously inter-molecular and for these crosslinks (4085-4085 and 4084-4085 in DNA-PKcs), only the inter-molecular assignments were considered.
4. Differential HX data that support areas of interaction between PAXX and Ku70 were used to create a restraint between the two molecules. An upper-harmonic potential with a mean of 4 Å and spring constant of 0.05 Å was used to restrain the minimum distance between the set of all PAXX residues and the set of residues in Ku70 that exhibited differential HDX (residues 413-442, 462-476 and 494-528).

In stage 3, with the scoring function and representation in hand, alternative structures of the synaptic complex were sampled. These structures were generated using Markov-Chain Monte Carlo (MC) enhanced by replica exchange, as implemented in IMP, for the set of movable objects as defined in the system representation. The MC moves included a random translation and rotation of rigid bodies with a maximum of 4 Å and 0.04 radians,

respectively, per step, and random translations for each flexible bead of a maximum of 4 Å per step. Each of the six samples were modeled as separate systems. To model each system, 30-50 independent simulations, each using 16 replicas with reduced temperature values equally distributed between 1.0 and 2.5, were initiated. Each simulation began by translating each rigid body and flexible bead by a random vector of up to 150 Å in magnitude. The positions of flexible bead coordinates were then optimized using steepest descent minimization for 500 steps, with a maximum displacement of 2.0 Å per step. Subsequently, a burn-in phase of 10,000 MC steps followed by 2M MC steps in each production run was generated. Snapshots from production runs were saved every 10,000 frames, resulting in 20,000 frames per run and ensembles of 600,000 to 1,000,000 structures produced per sample. For modeling of the synaptic complex, each independent simulation required 4-10 days on a single 16-core Intel Xeon processor with speeds between 2.3 to 2.8 GHz.

Finally, in stage 4, the hundreds of thousands of models generated for each of the six samples were analyzed for convergence and clustered. Each independent simulation was first analyzed for equilibration of score values (total score, crosslinking score, and excluded volume score). Only structures generated after equilibration of all values were considered further, reducing the population by 40-60%, resulting in 1,176,638 post-equilibrium structures over all six samples. This set of structures was then analyzed for sampling precision based on RMSD clustering using pyRMSD (Gil and Guallar, 2013), as described previously (Viswanath et al., 2017). Because of the large size of each model structure and the large number of structures, performing pairwise RMSD calculations on the entire set of models was not feasible. Instead, post-equilibration models for each of the six states were split into two independent groups (by random assignment of the 30-50 independent simulations into the A group or B group) and each group split into equal-sized subsets such that the combination of the A and B subsets contained ~2,000 models. The pairwise RMSD and sampling precision for each A and B subset pair were then calculated and models clustered at a threshold equal to the computed sampling precision. This value represents the cluster precision, defined as the average RMSD among all models in the cluster to the centroid model (i.e., the RMSF, root mean square fluctuation). Across all subsets in all six samples, the number of clusters and cluster populations were similar, indicating that the computed subset sampling precision estimates were a fair representation of the population sampling precision, which averaged ~30 Å.

To represent subsets equally within the clustered models, each subset was then clustered at 30 Å, resulting in 2-5 clusters per subset. The centroid structure from each subset cluster was extracted, resulting in 2062 model structures from the combined six samples. These combined structures were then clustered based on the pairwise RMSD of their DNA-PKcs coordinates using hierarchical linkage clustering as implemented in the Python library scikit-learn v0.20.3 (Pedregosa et al., 2011). This clustering resulted in seven sets of models. Four of these sets represented less than 10,000 models of the equilibrated set of ~1.2 million and were thus ignored. The remaining three clusters represented 399,500, 142,223 and 656,934 models and were analyzed further as described in our results. Localization densities were used to represent the final solutions, as described (Alber et al., 2007b).

## Localizing Ku70/80 termini and PAXX

Integrative modeling of the synaptic complex including PAXX was performed using the same procedure as above, except that two PAXX dimers (PDB 3WTD) (Ochi et al., 2015) were added to the system representation. Crosslinking datasets for samples including PAXX only contained a single inter-molecular crosslink to PAXX, so an additional restraint was added based on the differential HX data, as described above.

As an independent test of IMP modeling, HADDOCK docking based on crosslinks identified between Ku80 (residues 594-732) and DNA-PKcs was performed. A modified structure for the Ku80 C-terminal region (Zhang et al., 2004) (PDB 1RW2) without the variable N-terminal region in 1RW2 was used. The C-terminal 22 amino acid residues were added to the 1RW2 structure as a disordered peptide, optimized with ModLoop (Fiser and Sali, 2003), retaining the structure with the best score. The resulting refined C-terminal region was docked to the DNA-PKcs structure (PDB 5Y3R) with the HADDOCK webserver (Van Zundert et al., 2016), the crosslinks were added as unambiguous restraints, and the added C-terminal residues were considered as fully flexible.

## Quantification and Statistical Analysis

All HX-MS and XL-MS data sets were processed by Mass Spec Studio and were analyzed using the software listed in the Key Resources Table. For HX-MS, the statistical information generated for processing the data are shown in Tables S1 and S2. Significant differences in deuteration were considered significant provided the following criteria were met: passing a two-tailed t test ( $p < 0.05$ ) using pooled standard deviations from quadruplicate analyses of each state; passing a distribution analysis to guard against spectral overlap; exceeding a threshold shift value ( $\pm 2$  s.d.) based on a measurement of the shift noise and assuming its normal distribution. Selections were facilitated by the generation of volcano plots (Fig. S7–S10). For XL-MS data, the variance in fold-change was evaluated with 2 replicates: the ratios were calculated and recentered, fit to a normal distribution and  $\pm 3$  sigma of this distribution used to establish a cut-off value for fold-change. A comparison of states involved a similar recentering and features with no corresponding peptide in a given state comparison were assigned a maximal  $\log_2(\text{ratio})$  of 9 (Fig. 2).

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgements

This study was supported by NSERC (RGPIN-2017-04879 and CRDPJ 486813-15) to DCS, CFI (project #212822) to DCS, NIH (program project grant CA92584) to SPLM as well as R01 GM083960 and P41 GM109824 to AS.

## References

Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, Devos D, Suprpto A, Karni-Schmidt O, Williams R, Chait BT, et al. (2007a). Determining the architectures of macromolecular assemblies. *Nature* 450, 683–694. [PubMed: 18046405]

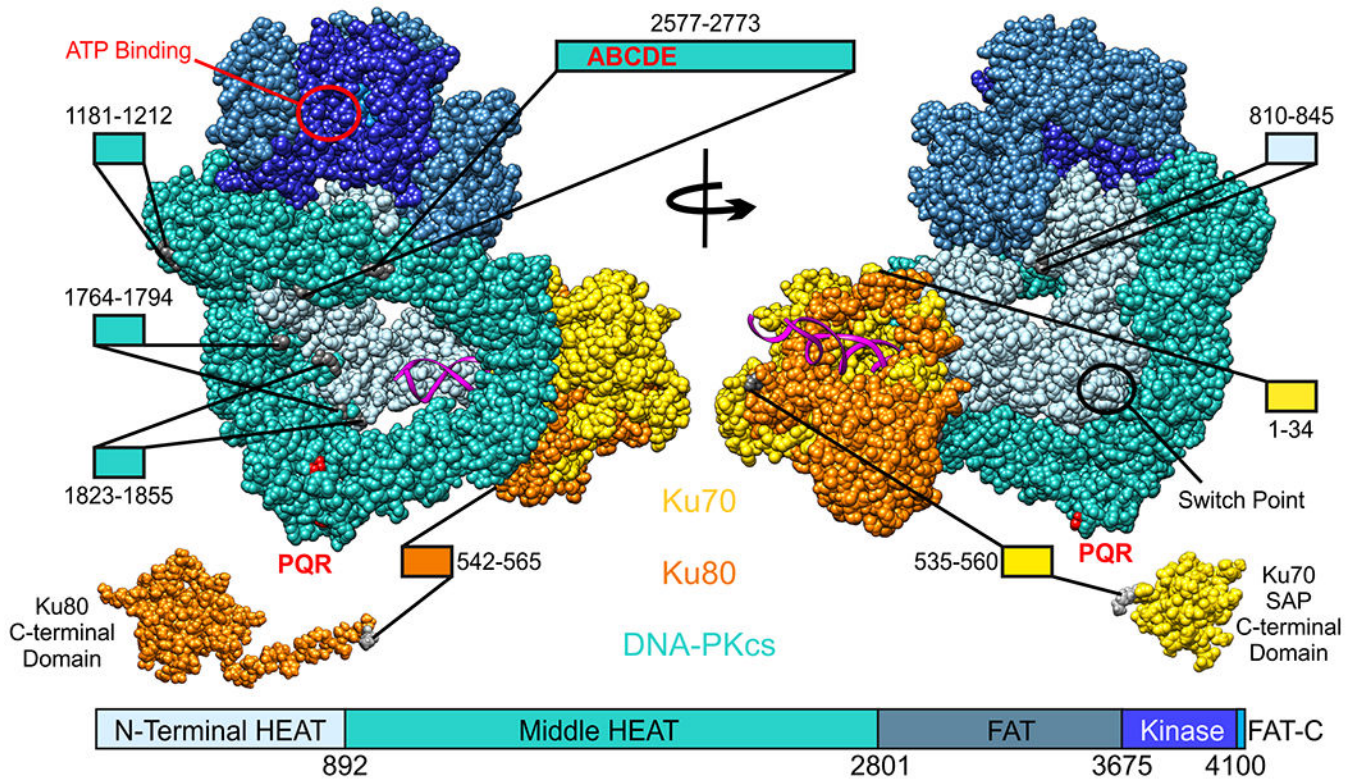
- Alber F, Dokudovskaya S, Veenhoff LM, Zhang W, Kipper J, Devos D, Suprpto A, Karni-Schmidt O, Williams R, Chait BT, et al. (2007b). The molecular architecture of the nuclear pore complex. *Nature* 450, 695–701. [PubMed: 18046406]
- Aravind L, and Koonin EV (2000). SAP - A putative DNA-binding motif involved in chromosomal organization. *Trends Biochem. Sci* 25, 112–114. [PubMed: 10694879]
- Baretic D, Maia T, Oliveira D, Niess M, Wan P, Pollard H, Johnson CM, Truman C, McCall E, Fisher D, et al. (2019). Structural insights into the critical DNA damage sensors DNA-PKcs, ATM and ATR. *Prog. Biophys. Mol. Biol* 147, 4–16. [PubMed: 31255703]
- Burley SK, Kurisu G, Markley JL, Nakamura H, Velankar S, Berman HM, Sali A, Schwede T, and Trewheella J (2017). PDB-Dev: a Prototype System for Depositing Integrative/Hybrid Structural Models. *Structure* 25, 1317–1318. [PubMed: 28877501]
- Chang HHY, Watanabe G, Gerodimos CA, Ochi T, Blundell TL, Jackson SP, and Lieber MR (2016). Different DNA end configurations dictate which NHEJ components are most important for joining efficiency. *J. Biol. Chem*
- Conlin MP, Reid DA, Small GW, Chang HH, Watanabe G, Lieber MR, Ramsden DA, and Rothenberg E (2017). DNA Ligase IV Guides End-Processing Choice during Nonhomologous End Joining. *Cell Rep.*
- Costantini S, Woodbine L, Andreoli L, Jeggo PA, and Vindigni A (2007). Interaction of the Ku heterodimer with the DNA ligase IV/Xrcc4 complex and its regulation by DNA-PK. *DNA Repair (Amst)*. 6, 712–722. [PubMed: 17241822]
- Cottarel J, Frit P, Bombarde O, Salles B, Négrel A, Bernard SS, Jeggo PA, Lieber MR, Modesti M, Calsou P, et al. (2013). A noncatalytic function of the ligation complex during nonhomologous end joining. *J. Cell Biol* 200, 173–186. [PubMed: 23337116]
- Cox J, Hein MY, Luber CA, Paron I, Nagaraj N, and Mann M (2014). Accurate proteome-wide label-free quantification by delayed normalization and maximal peptide ratio extraction, termed MaxLFQ. *Mol. Cell. Proteomics* 13, 2513–2526. [PubMed: 24942700]
- Cui X, Yu Y, Gupta S, Cho Y, Lees-Miller SP, and Meek K (2005). Autophosphorylation of DNA-Dependent Protein Kinase Regulates DNA End Processing and May Also Alter Double-Strand Break Repair Pathway Choice Autophosphorylation of DNA-Dependent Protein Kinase Regulates DNA End Processing and May Also Alter Double-Strand. *Mol. Cell. Biol* 25, 10842–10852. [PubMed: 16314509]
- DeFazio LG, Stansel RM, Griffith JD, and Chu G (2002). Synapsis of DNA ends by DNA-dependent protein kinase. *EMBO J.* 21, 3192–3200. [PubMed: 12065431]
- Erzberger JP, Stengel F, Pellarin R, Zhang S, Schaefer T, Aylett CHS, Cimerman i P, Boehringer D, Sali A, Aebersold R, et al. (2014). Molecular Architecture of the 40S·eIF1·eIF3 Translation Initiation Complex. *Cell* 158, 1123–1135. [PubMed: 25171412]
- Fiser A, and Sali A (2003). ModLoop: Automated modeling of loops in protein structures. *Bioinformatics* 19, 2500–2501. [PubMed: 14668246]
- Gil VA, and Guallar V (2013). PyRMSD: A Python package for efficient pairwise RMSD matrix calculation and handling. *Bioinformatics* 29, 2363–2364. [PubMed: 23846743]
- Goodarzi AA, and Lees-Miller SP (2004). Biochemical characterization of the ataxia-telangiectasia mutated (ATM) protein from human cells. *DNA Repair (Amst)*. 3, 753–767. [PubMed: 15177184]
- Graham TGW, Walter JC, and Loparo JJ (2016). Two-Stage Synapsis of DNA Ends during Non-homologous End Joining. *Mol. Cell*
- Graham TGW, Carney SM, Walter JC, and Loparo JJ (2018). A single XLF dimer bridges DNA ends during nonhomologous end joining. *Nat. Struct. Mol. Biol*
- Grimm M, Zimniak T, Kahraman A, and Herzog F (2015). XVis: A web server for the schematic visualization and interpretation of crosslink-derived spatial restraints. *Nucleic Acids Res.* 43, W362–W369. [PubMed: 25956653]
- Hammel M, Yu Y, Mahaney BL, Cai B, Ye R, Phipps BM, Rambo RP, Hura GL, Pelikan M, So S, et al. (2010). Ku and DNA-dependent protein kinase dynamic conformations and assembly regulate DNA binding and the initial non-homologous end joining complex. *J. Biol. Chem* 285, 1414–1423. [PubMed: 19893054]

- Hammel M, Rey M, Yu Y, Mani RS, Classen S, Liu M, Pique ME, Fang S, Mahaney BL, Weinfeld M, et al. (2011). XRCC4 protein interactions with XRCC4-like factor (XLF) create an extended grooved scaffold for DNA ligation and double strand break repair. *J. Biol. Chem*
- Hu S, Pluth JM, and Cucinotta FA (2012). Putative binding modes of Ku70-SAP domain with double strand DNA: A molecular modeling study. *J. Mol. Model* 18, 2163–2174. [PubMed: 21947447]
- Jette N, and Lees-Miller SP (2015). The DNA-dependent protein kinase: A multifunctional protein kinase with roles in DNA double strand break repair and mitosis. *Prog. Biophys. Mol. Biol*
- Jiang W, Crowe JL, Liu X, Nakajima S, Wang Y, Li C, Lee BJ, Dubois RL, Liu C, Yu X, et al. (2015). Differential Phosphorylation of DNA-PKcs Regulates the Interplay between End-Processing and End-Ligation during Nonhomologous End-Joining. *Mol. Cell* 58, 172–185. [PubMed: 25818648]
- Jovanovic M, and Dynan WS (2006). Terminal DNA structure and ATP influence binding parameters of the DNA-dependent protein kinase at an early step prior to DNA synapsis. *Nucleic Acids Res.* 34, 1112–1120. [PubMed: 16488883]
- Kosinski J, von Appen A, Ori A, Karius K, Müller CW, and Beck M (2015). Xlink analyzer: Software for analysis and visualization of cross-linking data in the context of three-dimensional structures. *J. Struct. Biol* 189, 177–183. [PubMed: 25661704]
- Lau WCY, Li Y, Liu Z, Gao Y, Zhang Q, and Huen MSY (2016). Structure of the human dimeric ATM kinase. *Cell Cycle* 15, 1117–1124. [PubMed: 27097373]
- Leuther KK, Hammarsten O, Kornberg RD, and Chu G (1999). Structure of DNA-dependent protein kinase : implications for its regulation by DNA. *EMBO J.* 18, 1114–1123. [PubMed: 10064579]
- Liang S, Esswein SR, Ochi T, Wu Q, Ascher DB, Chirgadze D, Sibanda BL, and Blundell TL (2017). Achieving selectivity in space and time with DNA double-strand-break response and repair: molecular stages and scaffolds come with strings attached. *Struct. Chem* 28, 161–171.
- Meek K, Lees-Miller SP, and Modesti M (2012). N-terminal constraint activates the catalytic subunit of the DNA-dependent protein kinase in the absence of DNA or Ku. *Nucleic Acids Res.* 40, 2964–2973. [PubMed: 22167471]
- Merkle D, Douglas P, Moorhead GBGG, Leonenko Z, Yu Y, Cramb D, Bazett-jones DP, and Lees-Miller SP (2002). The DNA-dependent protein kinase interacts with DNA to form a protein - DNA complex that is disrupted by phosphorylation. *Biochemistry* 41, 12706–12714. [PubMed: 12379113]
- Neal JA, Sugiman-Marangos S, VanderVere-Carozza P, Wagner M, Turchi J, Lees-Miller SP, Junop MS, and Meek K (2014). Unraveling the Complexities of DNA-Dependent Protein Kinase Autophosphorylation. *Mol. Cell. Biol*
- Ochi T, Blackford AN, Coates J, Jhujh S, Mehmood S, Tamura N, Travers J, Wu Q, Draviam VM, Robinson CV, et al. (2015). PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair. *Science.* 347, 185–188. [PubMed: 25574025]
- Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, Blondel M, Prettenhofer P, Weiss R, Dubourg V, et al. (2011). Scikit-learn: Machine Learning in Python. *J. Mach. Learn. Res* 12, 2825–2830.
- Perez-Riverol Y, Csordas A, Bai J, Bernal-Llinares M, Hewapathirana S, Kundu DJ, Inuganti A, Griss J, Mayer G, Eisenacher M, et al. (2019). The PRIDE database and related tools and resources in 2019: Improving support for quantification data. *Nucleic Acids Res.* 47, D442–D450. [PubMed: 30395289]
- Petterson EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, and Ferrin TE (2004). UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem* 25, 1605–1612. [PubMed: 15264254]
- Rao Q, Liu M, Tian Y, Wu Z, Hao Y, Song L, Qin Z, Ding C, Wang HW, Wang J, et al. (2018). Cryo-EM structure of human ATR-ATRIP complex. *Cell Res.* 28, 143–156. [PubMed: 29271416]
- Ropars V, Drevet P, Legrand P, Bacconnais S, Amram J, Faure G, Márquez JA, Piétrement O, Guerois R, Callebaut I, et al. (2011). Structural characterization of filaments formed by human Xrcc4-Cernunnos/XLF complex involved in nonhomologous DNA end-joining. *Proc. Natl. Acad. Sci. U. S. A* 108, 12663–12668. [PubMed: 21768349]

- Russel D, Lasker K, Webb B, Velázquez-Muriel J, Tjioe E, Schneidman-Duhovny D, Peterson B, and Sali A (2012). Putting the pieces together: Integrative modeling platform software for structure determination of macromolecular assemblies. *PLoS Biol.* 10, 1–5.
- Saltzberg D, Greenberg CH, Viswanath S, Chemmama I, Webb B, Pellarin R, Echeverria I, and Sali A (2019a). Modeling Biological Complexes Using Integrative Modeling Platform. In *Methods in Molecular Biology*, p.
- Saltzberg DJ, Hepburn M, Pilla KB, Schriemer DC, Lees-Miller SP, Blundell TL, Sali A, Bharath K, Schriemer DC, Lees-Miller SP, et al. (2019b). SSEThread: Integrative threading of the DNA-PKcs sequence based on data from chemical cross-linking and hydrogen deuterium exchange. *Prog. Biophys. Mol. Biol* 147, 92–102. [PubMed: 31570166]
- Sarpe V, Rafiei A, Hepburn M, Ostan N, Schryvers AB, and Schriemer DC (2016). High sensitivity crosslink detection coupled with integrative structure modeling in the mass spec studio. *Mol. Cell. Proteomics* 15, 3071–3080. [PubMed: 27412762]
- Schneidman-Duhovny D, Pellarin R, and Sali A (2014). Uncertainty in integrative structural modeling. *Curr. Opin. Struct. Biol* 28, 96–104. [PubMed: 25173450]
- Serrano-Benítez A, Cortés-Ledesma F, and Ruiz JF (2020). “An End to a Means”: How DNA-End Structure Shapes the Double-Strand Break Repair Process. *Front. Mol. Biosci*
- Sheff JG, and Schriemer DC (2014). Toward Standardizing Deuterium Content Reporting in Hydrogen Exchange-MS. *Anal. Chem* 86, 11962–11965. [PubMed: 25427063]
- Sheff JG, Hepburn M, Yu Y, Lees-Miller SP, and Schriemer DC (2017). Nanospray HX-MS configuration for structural interrogation of large protein systems. *Analyst* 142, 904–910. [PubMed: 28154854]
- Shi Y, Fernandez-Martinez J, Tjioe E, Pellarin R, Kim SJ, Williams R, Schneidman-Duhovny D, Sali A, Rout MP, and Chait BT (2014). Structural Characterization by Cross-linking Reveals the Detailed Architecture of a Coatmer-related Heptameric Module from the Nuclear Pore Complex. *Mol. Cell. Proteomics* 13, 2927–2943. [PubMed: 25161197]
- Shi Y, Pellarin R, Fridy PC, Fernandez-Martinez J, Thompson MK, Li Y, Wang QJ, Sali A, Rout MP, and Chait BT (2015). A strategy for dissecting the architectures of native macromolecular assemblies. *Nat. Methods* 12, 1135–1138. [PubMed: 26436480]
- Sibanda BL, Critchlow SE, Begun J, Pei XY, Jackson SP, Blundell TL, and Pellegrini L (2001). Crystal structure of an Xrc4-DNA ligase IV complex. *Nat. Struct. Biol* 8, 1015–1019. [PubMed: 11702069]
- Sibanda BL, Chirgadze DY, Ascher DB, and Blundell TL (2017). DNA-PKcs structure suggests an allosteric mechanism modulating DNA double-strand break repair. *Science* (80-. ).
- Song Z, Xie Y, Guo Z, Han Y, Guan H, Liu X, Ma T, and Zhou P (2019). Genome-wide identification of DNA-PKcs-associated RNAs by RIP-Seq. *Signal Transduct. Target. Ther* 4, 21–23. [PubMed: 31637001]
- Spagnolo L, Rivera-Calzada A, Pearl LH, and Llorca O (2006). Three-Dimensional Structure of the Human DNA-PKcs/Ku70/Ku80 Complex Assembled on DNA and Its Implications for DNA DSB Repair. *Mol. Cell*
- Stinson BM, Moreno AT, Walter JC, and Loparo JJ (2019). A Mechanism to Minimize Errors during Non-homologous End Joining. *Mol. Cell* 77, 1–12.
- Tadi SK, Tellier-Lebègue C, Nemoz C, Drevet P, Audebert S, Roy S, Meek K, Charbonnier JB, and Modesti M (2016). PAXX Is an Accessory c-NHEJ Factor that Associates with Ku70 and Has Overlapping Functions with XLF. *Cell Rep.* 17, 541–555. [PubMed: 27705800]
- Uematsu N, Weterings E, Yano KI, Morotomi-Yano K, Jakob B, Taucher-Scholz G, Mari P-OO, Van Gent DC, Chen BPCC, and Chen DJ (2007). Autophosphorylation of DNA-PKCS regulates its dynamics at DNA double-strand breaks. *J. Cell Biol* 177, 219–229. [PubMed: 17438073]
- Vaughan RC, and Kao CC (2015). Mapping Protein – RNA Interactions by RCAP , RNA-Cross-Linking and Peptide Fingerprinting. *RNA Nanotechnol. Ther. Methods Proctols, Methods Mol. Biol* 1297, 225–236.
- Viswanath S, Chemmama IE, Cimermanic P, and Sali A (2017). Assessing Exhaustiveness of Stochastic Sampling for Integrative Modeling of Macromolecular Structures. *Biophys. J* 113, 2344–2353. [PubMed: 29211988]

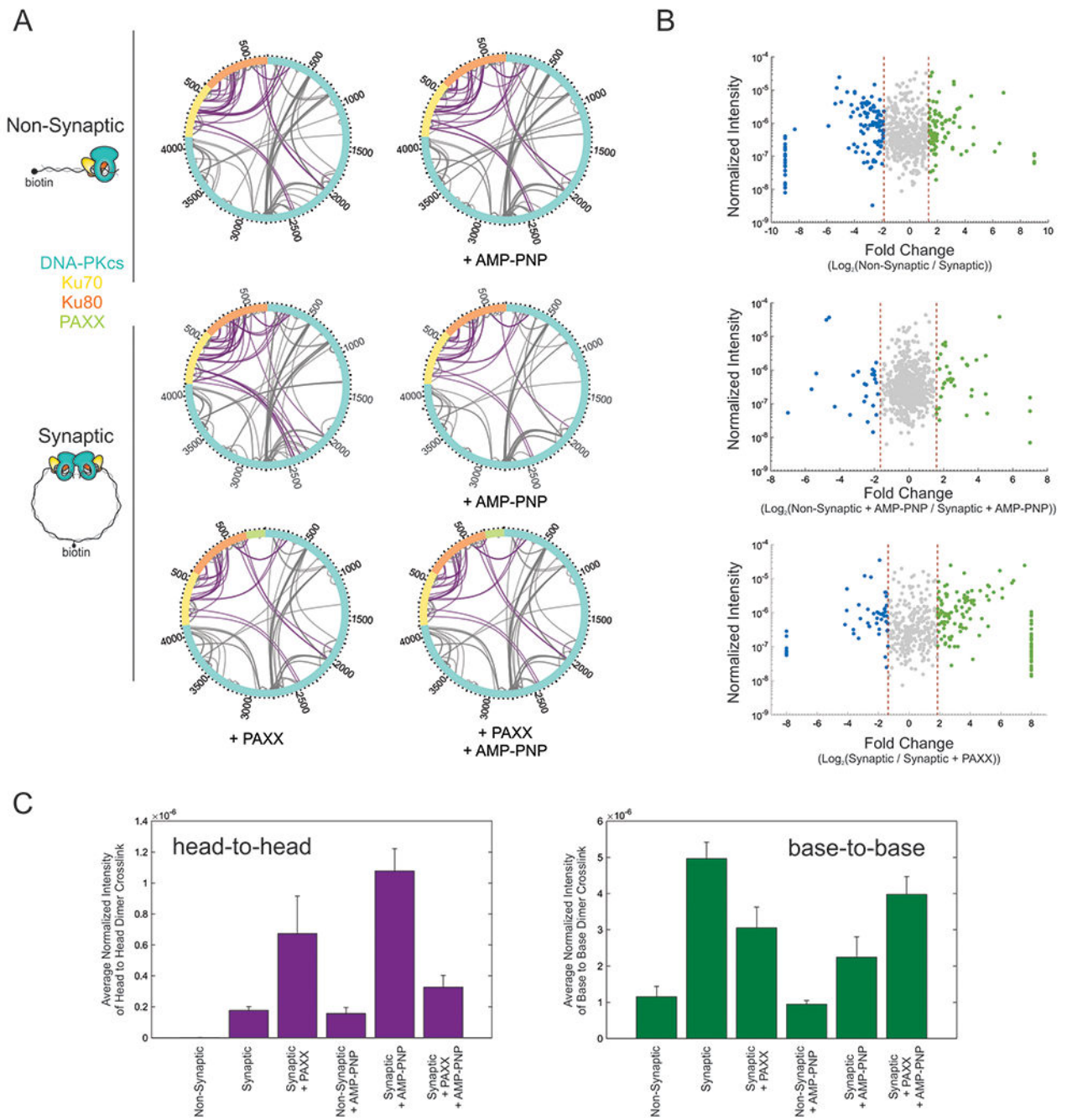
- Walker JR, Corpina RA, and Goldberg J (2001). Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature* 412, 607–614. [PubMed: 11493912]
- Wang J, Pluth JM, Cooper PK, Cowan MJ, Chen DJ, and Yannone SM (2005). Artemis deficiency confers a DNA double-strand break repair defect and Artemis phosphorylation status is altered by DNA damage and cell cycle progression. *DNA Repair (Amst)*. 4, 556–570. [PubMed: 15811628]
- Wang JL, Duboc C, Wu Q, Ochi T, Liang S, Tsutakawa SE, Lees-Miller SP, Nadal M, Tainer JA, Blundell TL, et al. (2018). Dissection of DNA double-strand-break repair using novel single-molecule forceps. *Nat. Struct. Mol. Biol*
- Wang X, Chu H, Lv M, Zhang Z, Qiu S, Liu H, Shen X, Wang W, and Cai G (2016). Structure of the intact ATM/Tel1 kinase. *Nat. Commun* 7, 1–8.
- Waters C. a, Strande NT, Pryor JM, Strom CN, Mieczkowski P, Burkhalter MD, Oh S, Qaqish BF, Moore DT, Hendrickson EA, et al. (2014). The fidelity of the ligation step determines how ends are resolved during nonhomologous end joining. *Nat. Commun* 5, 1–11.
- Webb B, Viswanath S, Bonomi M, Pellarin R, Greenberg CH, Saltzberg D, and Sali A (2018). Integrative structure modeling with the Integrative Modeling Platform. *Protein Sci*.
- West RB, Yaneva M, and Lieber MR (1998). Productive and nonproductive complexes of Ku and DNA-dependent protein kinase at DNA termini. *Mol. Cell. Biol* 18, 5908–5920. [PubMed: 9742108]
- Weterings E, Verkaik NS, Gent D.C. Van, Brüggewirth HT, Hoeijmakers JHJJ, and Van Gent DC (2003). The role of DNA dependent protein kinase in synapsis of DNA ends. *Nucleic Acids Res.* 31, 7238–7246. [PubMed: 14654699]
- Xing M, Yang M, Huo W, Feng F, Wei L, Jiang W, Ning S, Yan Z, Li W, Wang Q, et al. (2015). Interactome analysis identifies a new paralogue of XRCC4 in non-homologous end joining DNA repair pathway. *Nat. Commun* 6, 1–12.
- Yang M, Hoepfner M, Rey M, Man P, Schriemer DC. (2015) Recombinant nepenthesin II for hydrogen/deuterium exchange mass spectrometry. *Anal. Chem* 7, 6681–7.
- Yano K, Morotomi-Yano K, Wang S-Y, Uematsu N, Lee K-J, Asaithamby A, Weterings E, and Chen DJ (2008). Ku recruits XLF to DNA double-strand breaks. *EMBO Rep.* 9, 91–96. [PubMed: 18064046]
- Yates LA, Williams RM, Hailemariam S, Ayala R, Burgers P, and Zhang X (2020). Cryo-EM Structure of Nucleotide-Bound Tel1ATM Unravels the Molecular Basis of Inhibition and Structural Rationale for Disease-Associated Mutations. *Structure* 28, 96–104. [PubMed: 31740029]
- Yin X, Liu M, Tian Y, Wang J, Xu Y, Sharif H, Li Y, Dong Y, Dong L, Li W, et al. (2017). Cryo-EM structure of the human DNA-PK holoenzyme. *Cell Res.* 27, 1341–1350. [PubMed: 28840859]
- Zhang Z, Zhu L, Lin D, Chen F, Chen DJ, and Chen Y (2001). The three-dimensional structure of the C-terminal DNA-binding domain of human Ku70. *J. Biol. Chem* 276, 38231–38236. [PubMed: 11457852]
- Zhang Z, Hu W, Cano L, Lee TD, Chen DJ, and Chen Y (2004). Solution structure of the C-terminal domain of Ku80 suggests important sites for protein-protein interactions. *Structure* 12, 495–502. [PubMed: 15016365]
- Van Zundert GCP, Rodrigues JPGLM, Trellet M, Schmitz C, Kastiris PL, Karaca E, Melquiond ASJ, Van Dijk M, De Vries SJ, and Bonvin AMJJ (2016). The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol*





**Figure 1. Orientation of the DNA-PK holoenzyme.**

Based on DNA-PK (5Y3R), complete with linked Ku70 (1JJR) and Ku80 (1RW2) C-terminal structures. Boxes represent missing regions of structure greater than 30 amino acid residues. The figure presents the head, consisting of a kinase domain encased within the FAT/FAT-C domains. The middle HEATs together with the N-terminal HEATs (the arm) form loosely concentric rings that present a cavity in the base of DNA-PKcs. DNA binds under the N-terminal HEAT and protrudes into the cavity. Coloring is preserved for all HX figures, except for the head (all steel blue for clarity). See also Figure S1.



**Figure 2. Crosslinking of the DNA-PK synaptic complex.**

(A) Circos plots of crosslinks found in multiple states. Intra-protein crosslinks are shown in grey and inter-protein crosslinks in purple. (B) Comparison of crosslink abundance between states. Crosslinks enriched in the synaptic (and PAXX-bound) states are shown in blue. Crosslinks enriched in the non-synaptic (and PAXX-free) control states are shown in green. Non-significant changes are represented in grey. Redundant crosslinks (i.e. different peptides for identical linked residues) preserved for completeness. Extreme values ( $\log_2(\text{ratio})=\pm 9$ ) represent crosslinks found in only one state. (C) Quantitative evaluation of a unique head-

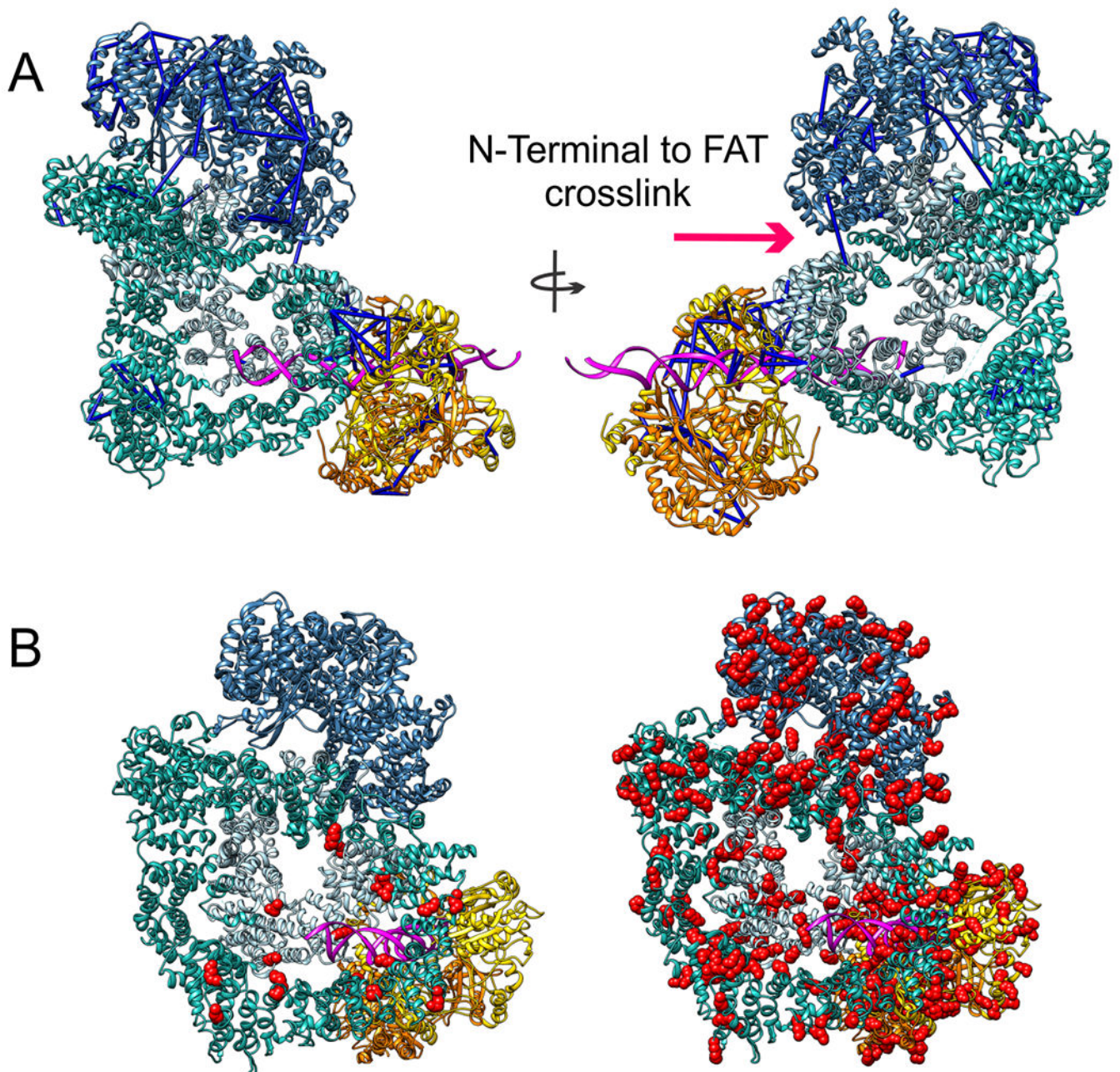
to-head (left) and a base-to-base (right) dimer crosslink. Data represent aggregated and normalized LC-MS feature intensities; error bars = SEM (n=3). See also Figures S2.

Author Manuscript

Author Manuscript

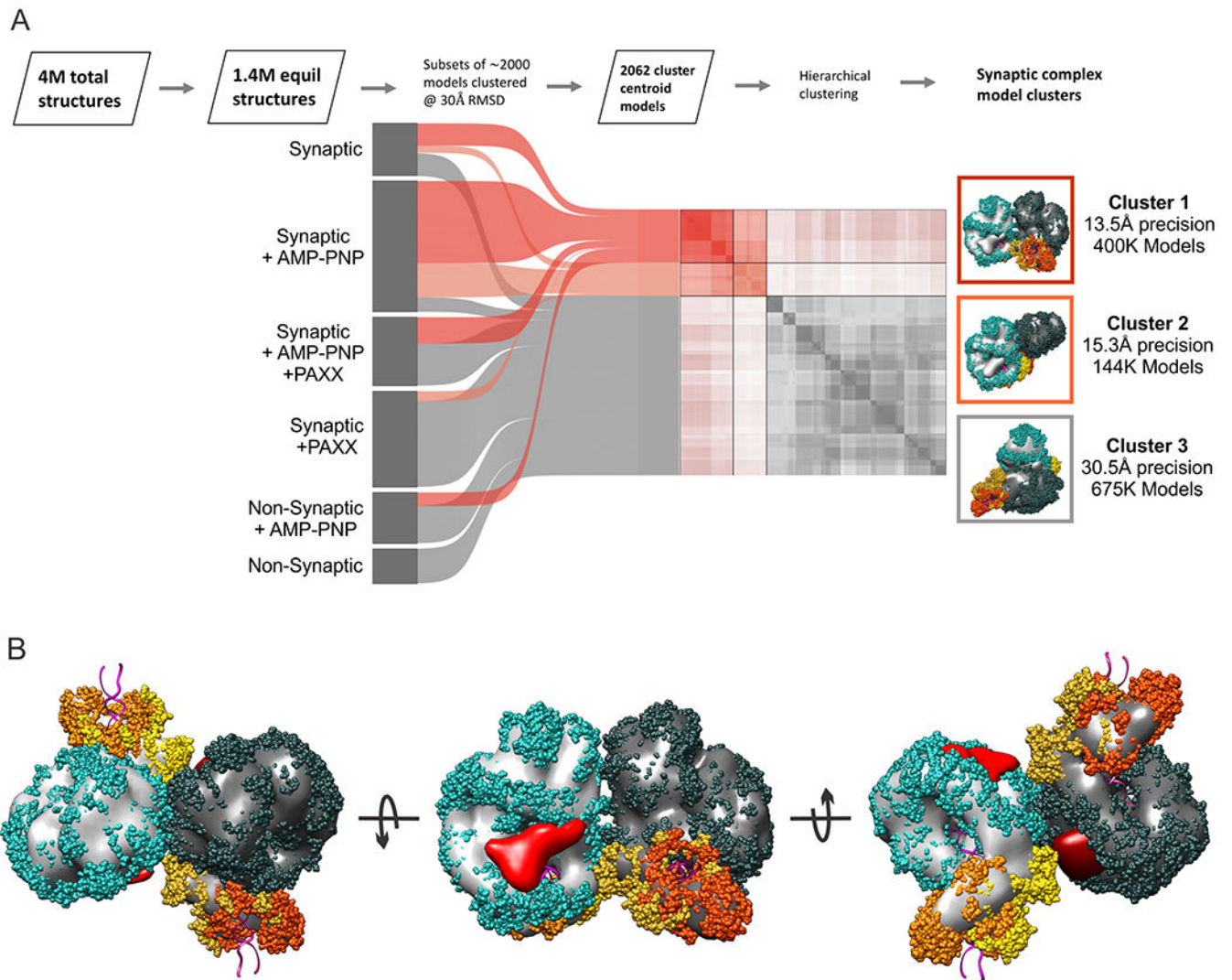
Author Manuscript

Author Manuscript



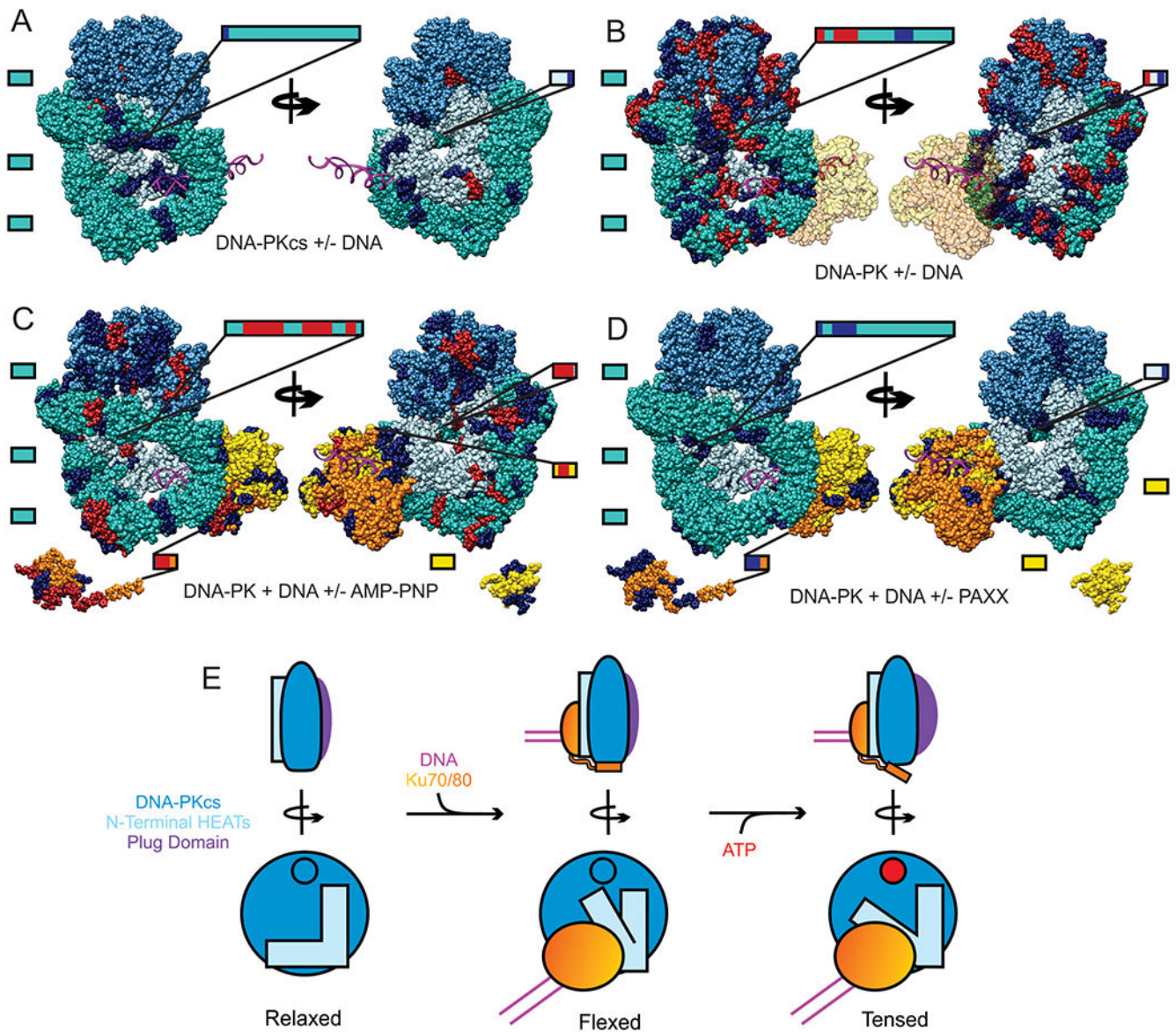
**Figure 3. A subset of detected DSS crosslinks on DNA-PK bound to DNA.**

(A) Crosslinks satisfied by the structured elements of a DNA-PK monomer from an analysis of the non-synaptic holoenzyme (nucleotide-free), mapped to the 5Y3R structure using Xlink Analyzer (Kosinski et al., 2015). Blue rods represent satisfied crosslinks (less than 35Å). (B) Left structure: crosslinking attachment points shown as red spheres for the large disordered region of DNA-PKcs (residues 2577-2774, not shown). Right structure: the availability of lysine residues for crosslinking, also shown as red spheres. Coloring of the DNA-PK structure as in Fig. 1.



**Figure 4. Integrative models of the synaptic complex.**

A) Model reduction and clustering analysis of the synaptic models from each crosslinked sample. From the set of ~4 million sampled structures of the synaptic complex, ~1.2 million remained after equilibration testing. Subsets of these models were generated for each dataset and clustered at 30Å. The resulting 2062 cluster centroids were then combined and clustered again, resulting in three major structural classes. Clusters highlighted in red are clusters which satisfy the head to head dimer crosslink. The grey cluster does not satisfy the head to head dimer crosslink. B) Detailed representation of the average structure from Cluster 1, showing the localization densities (light/dark grey surfaces) fit with the DNA-PK structures (5Y3R) where the two DNA-PKc molecules are shown in teal and dark slate grey, the Ku70s in shades of yellow, and Ku80s in shades of orange. The localization density for the plug domain (representing DNA-PKc residues 2577-2773) is highlighted in red. See also Figures S3–S6.



**Figure 5. DNA-PK is a complex conformational switch.**

Differential HX analysis of (A) DNA-PKcs upon binding to DNA, (B) DNA-PKcs upon binding to Ku70/80 and DNA, (C) DNA-PK upon binding to AMP-PNP, and (D) DNA-PK upon binding to PAXX. Peptides destabilized upon binding are shown in red. Peptides that are stabilized upon binding are shown in blue. Expanded analyses are shown in Figs. S10–13. (E) Nucleotide loading induces tension in DNA-PK. Free DNA-PKcs exists in a relaxed state, with the N-terminal arm distal from the head. Binding of DNA and Ku70/80 pushes the DNA-PKcs N-terminal into a flexed state. Loading the ATP analog places the arm in tension, where a nucleotide-induced conformational change acts against the DNA-induced flexing of the arm. The tense state destabilizes the plug domain. See also Figures S7–S11.

## KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
<b>Antibodies</b>		
Mouse monoclonal Anti-DNAPKcs	Lees-Miller Lab, University of Calgary	N/A
Rabbit/polyclonal Anti-PAXX	Abcam	Cat#Ab126353
Rabbit/polyclonal Anti-Ku80	Abcam	Cat#Ab33242
<b>Bacterial and Virus Strains</b>		
<i>E. coli</i> BL21 (DE3)	ThermoFisher	C600003
<b>Chemicals, Peptides, and Recombinant Proteins</b>		
streptavidin agarose beads	TriLink	Cat#97067-706
rNep2	(Yang et al., 2015)	N/A
XhoI	New England Biolabs	Cat#R0146
BamHI	New England Biolabs	Cat#R0136
Nb.BbvVI	New England Biolabs	Cat#R06315
PreScission protease	GE Healthcare	Cat#27-0843-01
glow-discharged carbon film 400 mesh copper grid	Electron Microscopy Sciences	Cat#CF400-CU
disuccinimidyl suberate	Thermo Scientific	Cat#21555
AMP-PNP	Sigma-Aldrich	Cat#A2647-5MG
DEAE sepharose	GE Healthcare	Cat#17-0709-01
SP sepharose	GE Healthcare	Cat#17-0729-01
Heparin Hitrap	GE Healthcare	Cat#17-5248-02
MonoQ	GE Healthcare	Cat#17-5166-01
MonoS	GE Healthcare	Cat#17-5168-01
Jupiter C18 beads	Phenomenex	Cat#04A-4053
D <sub>2</sub> O	Sigma-Aldrich	Cat#151882-25G
Caffeine (heavy and light)	Sigma-Aldrich	Cat#485365-1G, C0750
<b>Deposited Data</b>		
Cryo-EM structure of DNA-PK	(Yin et al., 2017)	PDB 5Y3R
Crystal structure of DNA-PKcs	(Sibanda et al., 2017)	PDB 5LUQ
C-terminal NMR structure of Ku80	(Zhang et al., 2004)	PDB 1RW2
C-terminal NMR structure of Ku70	(Zhang et al., 2001)	PDB 1JJR
Crystal structure of PAXX	(Ochi et al., 2015)	PDB 3WTD
Raw HX-MS and XL-MS data	This paper	PXD017931
Integrative models, structures and input data	This paper	PDBDEV000000XX
<b>Experimental Models: Cell Lines</b>		
HeLa cells	National Cell Culture Center	N/A
<b>Oligonucleotides</b>		
5Phos/TGAGGGATATCGAA/iBiodUK/TCCTGCAGGC	Integrated DNA Technologies	N/A
LSO71F primer CGGAACATTAGTGCAGGCAGC	UCDNA Synthesis Lab	N/A
LSO72R primer GCGTAATGGCTGCCTGTTG	UCDNA Synthesis Lab	N/A
AAGCTTGCATGCCTGCAGGTGCACC-3'-Biotin(forward strand)	Integrated DNA Technologies	N/A
15nt polyT overhang on: AAGCTTGCATGCCTGCAGGTGCACC-3'-Biotin (forward strand)	Integrated DNA Technologies	N/A
5'-Biotin - GGGCGAGGGTGTGTCGAACGCCTGAG-3' (forward strand)	UCDNA Synthesis Lab	N/A

REAGENT or RESOURCE	SOURCE	IDENTIFIER
5'-Biotin- GCACAATCTTCTCGCGCAACGCGTCAGTGGGCTGATCATTAACTATCCGCTGGATGACCAGGATGCCATTGCTGTGGAAGCTGCCTGCACTAATGTTCCG-3' (forward strand)	UCDNA Synthesis Lab	N/A
<b>Recombinant DNA</b>		
<i>H. Sapiens</i> PAXX	(Ochi et al., 2015)	NM_183241
pGEX6P1	GE Healthcare	Cat#GE28-9546-48
Pet28a	Novagen	Cat#69864
<b>Software and Algorithms</b>		
Mascot v2.4.0	Matrix Science	<a href="http://www.matrixscience.com">www.matrixscience.com</a>
PEAKS 8.5	BSI	<a href="http://www.bioinform.com">www.bioinform.com</a>
HX-DEAL	Schriemer lab, University of Calgary	<a href="http://www.msstudio.ca">www.msstudio.ca</a>
Chimera 1.14	UCSF	<a href="https://www.evl.ucsf.edu/chimera/download.html">https://www.evl.ucsf.edu/ chimera/download.html</a>
CRIMP	(Sarpe et al., 2016)	<a href="http://www.msstudio.ca">www.msstudio.ca</a>
Xvis	(Grimm et al., 2015)	<a href="https://xvis.genzentrum.lmu.de/">https:// xvis.genzentrum.lmu.de/</a>
Integrative Modeling Platform (IMP) 2.11.0	Sali lab, UCSF	<a href="https://integrativemodeling.org">https://integrativemodeling.org</a>
pyRMSD	(Gil and Guallar, 2013)	<a href="https://pypi.org/project/pyRMSD/">https://pypi.org/project/ pyRMSD/</a>
Python library scikit-learn v0.20.3	(Pedregosa et al., 2011).	<a href="https://scikit-learn.org/stable/whats_new/v0.20.html#version-0-20-3">https://scikit-learn.org/ stable/whats_new/ v0.20.html#version-0-20-3</a>
Modeling Scripts	This paper	<a href="https://www.github.com/salilab/synaptic_complex">https://www.github.com/ salilab/synaptic_complex</a>