NORTHWESTERN UNIVERSITY

Structural Visualization of State Transition during Non-Homologous End-Joining (NHEJ)

A DISSERTATION

SUBMITTED TO THE GRADUATE SCHOOL
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

for the degree

DOCTOR OF PHILOSOPHY

Interdisciplinary Biological Sciences

By

Siyu Chen

EVANSTON, ILLINOIS

March 2022

© Copyright by Siyu Chen 2022

All Rights Reserved

ABSTRACT

DNA double-strand breaks (DSBs) are caused by either endogenous agents or exogenous ionizing radiation and chemicals. Incorrect DSB repair can lead to undesired genome rearrangements. Homologous recombination (HR) and non-homologous end-joining (NHEJ) pathways are two major DNA repair pathways that repair DSBs and maintain genome integrity. When homologous DNA is available as a template, HR uses it to direct error-free repairs, while NHEJ directly ligates the two broken DNA ends without any template. In immune cells, the V(D)J recombination pathway also uses NHEJ to assemble various immunoglobulin gene segments. While playing an important role in maintaining the stability and integrity of genomes, NHEJ is thought to be a 'dirty' DNA break repair pathway that is more error-prone compared to HR. It contributes to a large portion of chromosomal translocations in cells, making them particularly vulnerable to cancer-related mutations. It has been shown that ablation of NHEJ factors in DNA repair-deficient cell lines reduces genomic instability, and suppression of the NHEJ pathway kills cancer cells.

NHEJ is initialized by a DSB recognition step, where Ku heterodimer (Ku70/80) recognizes and binds to the broken DNA ends. Upon binding to DNA, Ku recruits the DNA-dependent protein kinase (DNA-PKcs) to the repair loci and forms the DNA-PK holoenzyme. This holoenzyme then serves as the scaffold to recruit other core NHEJ factors to the repair site, including the core NHEJ factors (DNA ligase IV, XRCC4, XLF), appendix scaffolding factors (PAXX, APLF), nucleases (Artemis), polymerases (Pol μ and Pol λ) and other processing enzymes (PNKP, Tdp1/2). While the nucleases and polymerases are only recruited when end processing is needed prior to the end ligation, the core NHEJ factors are responsible for the final ligation step of the broken DNA ends and are therefore absolutely required in NHEJ. Successful DSB repair in NHEJ relies on the efficient bridging of two DNA ends, yet the detailed molecular mechanism that lies behind this pathway remains largely unclear. This project focuses on the key intermediate complexes in the NHEJ pathway and the main goal is to structurally visualize the mechanism of how NHEJ factors orchestrate in these key complexes as molecular machinery to repair DSB DNA. To achieve this goal, in vitro reconstitution in conjunction with single-particle cryo-electron microscopy (cryo-EM) were utilized to obtain the structure of five different states in NHEJ, including the Long-range (LR) complex, PAXX bound LR complex (PAXX-LR), ATP activated PAXX-LR complex (PAXX-ATP-LR), the Short-range complex (SR) and an alternative DNA-PKcs dimeric complex. Together, the key intermediates observed with Cryo-EM provide a comprehensive understanding of how DNA- PK holoenzyme bridge the two broken DNA ends in close proximity, recruit other core NHEJ factors, transit into the ligation state upon ATP activation and dissociation of DNA-PKcs. In addition, the results shed light on how different appendix NHEJ factors can be recruited by different mechanisms at different stages. In summary, direct visualization toward the overall architecture of key nucleoprotein assemblies during NHEJ can lead to the clarification of how these essential NHEJ factors orchestrate in such a complicated, multi-step DNA repair pathway. Understanding the molecular mechanism of this dominant DSB repair pathway in eukaryotic cells can also open up new directions toward cancer drug development and rational design of optimized gene-editing tools.

ACKNOWLEDGEMENTS

As an international student pursuing his Ph.D. in a foreign country, it is simply not possible to finish the journey without the help and encouragement from the nice people around me. Although I am trying to express my gratitude here in a most formal, most ceremonial way, I understand deeply in my heart that whatever I am writing down now can not compare with what all of you have done to me, not even close.

I would like to thank my thesis advisor Dr. Yuan He, for guiding me all the way through this journey. When Yuan introduced this project to me, a fresh graduate student with only full of passion, but little experience in the field of Cryo-EM, he probably knew way better than me how significant and precious this opportunity could be. This is a project that I have been fascinated about from the very beginning, and I am glad that I did not miss this chance to not only learn how to do science, but also make solid contributions to the field. Working with Yuan is a great experience to me. He always knows how to keep good balance between when to advise on experimental details and when to leave spaces for me to explore on my own. Doing science is never a flat path, but in most cases, insightful guidance from the advisor is the best road sign to lead me through.

Also, I would like to thank everyone else in the He lab, all previous and current members. I can not remember how many times I turned to Yan Han and Ryan Marcum, former postdocs in our lab, Avinash Patel, current postdoc in our lab, and Susan Fishbain, the lab manager and all-time-ready-to-help person, and seek for all sorts of answers, discussions or even random talks. More importantly, the He lab was never a big lab during my PhD, so everyone's contribution to the warm and friendly lab atmosphere counts. I enjoyed working with every single graduate students in our lab, and would like to thank Ryan Abdella and Anna Talyzina for the fruitful collaboration on the PIC-Meditor project, and Alexis Reyes for all the discussions and Q&As about everything during my PhD. Also, I would like to express my best wishes to Alex Vogt and Jinkang Qing, the new graduate students in our lab. I am sure that you will all be successful and enjoy your time in the He lab.

I would like to thank our long-time collaborators on my thesis project, Dr. Susan Lees-Miler at University of Calgary and Dr. Alan Tomkinson at University of New Mexico. Without their help with the purified high quality protein factors, I simply can not imagine how I should even start this project. Sadly I didn't get the chance to meet any of our collaborators in person, but the advice I received during every zoom discussions and manuscript revisions are always insightful and supportive.

Also, I would like to thank my committee members, Ishwar Radhakrishnan, Jason Brickner, Reza Vafabakhsh and John Marko, for all the great advice about both my project and my career, and also Amy Rosenzweig, for mentoring my third round rotation and kindly supporting my postdoc applications with strong recommendation letters. I am so lucky to be able to work with Laura Dassama during my rotation in the Rosenzweig lab, which led to a lovely crystal that helped with the MbnH manuscript. Also, the rotation in the Vafabakhsh lab as the very first student is memorable and brought me so many special experiences and moments.

I want to thank my cohort for their accompany as well. You guys are the main reason I would like to go to the IBiS retreat every year. For every good and bad moments I experienced, it is always a relief to know that we are on this journey together, and we will make it to the other end together.

Lastly but most importantly, I would like to thank my friends and family, who were with me all the time. To my mom and dad, it is challenging to visit you during my PhD, and the COVID situation only made it worse. I totally understand how much you miss me, simply because I miss you as well. And with that, the support and encouragement from you can not be valued enough. To my wife, Xiaolin, the feeling of being together with all these ups and downs is something beyond my words can describe, but gladly, I know that both you and I share these feelings. We are still on the road, and we will be on the road together as always.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION AND BACKGROUND

1.1 DNA double strand breaks (DSB) in mammalian cells

In mammalian cells, DNA double strand breaks (DSBs), the most cytotoxic form of DNA damage, is caused by either endogenous agents such as reactive oxygen species, or exogenous ionizing radiation and chemicals(*1*). Unrepaired DSB can drive apoptosis or senescence, and incorrect DSB repair is one of the major causes of tumorigenesis, because it can potentially lead to undesired genome rearrangements such as deletions, translocations, and fusions(*2*).

1.2 Pathways responding to DSB

Several highly-regulated pathways have been evolved to repair DSB and maintain genome integrity. In mid-S phase to early G2 phase of the cell cycle, homologous DNA is available as a template for the repair reaction, thus DSBs can be repaired through homologous recombination (HR), in which the homologous DNA template directs error-free repairs(*3*). However, as the primary DSB repair pathway that remains active throughout the cell cycle(*4*), non-homologous end-joining (NHEJ) involves a direct ligation of the broken DNA ends without referring to a homologous template(*5*). NHEJ is also responsible for the assembly of gene segments in V(D)J recombination, where various immunoglobulin genes are generated by exon recombination in immune cells(*6*).

1.2.1 Homologous Recombination (HR)

HR is an evolutionarily conserved process that faithfully repair DSB when the sister chromatids or homologous chromosomes are available. The very first event that triggers HR is DNA end resection from 5' to 3'. This event is also one of the main feature of HR, and controlling this step can define a mechanism of pathway choice between HR and NHEJ(*7*). In mammalian cells, DNA end resection is initiated by the the MRN complex (Mre11, Rad50 and Nbs1) binding. When binding to the DNA end, the endonuclease in the complex, Mre11, creates short single-strand DNA overhang to recruit and activate the ataxia telangiectasia-mutated (ATM) kinase(*8*), as well as other long-range resection factors, such as Exonuclease 1 (EXO1), endonuclease DNA2 and the Bloom syndrome helicase (BLM)(*9*). ATM is an apical kinase that is believed to be the master regulator of various DSB repair pathways(*10*). It's function will be discussed more in Chapter 5.2.2, together with other phosphatidylinositol 3–kinase (PI3K)-related protein kinase (PIKK) family members. The rest steps of HR and factors involved will also be discussed in more details in Chapter 5.2.2 and will be compared with NHEJ.

1.2.2 Non-Homologous End-Joining (NHEJ)

As a predominant repair pathway in mammalian cells, NHEJ plays significant role in the maintenance of genomic stability(*11*). Mice(*12*, *13*) and human(*14*) cell lines lacking NHEJ factors are sensitive to agents that generates DSB, and have elevated levels of chromosomal translocations. Individuals with low activity of NHEJ factors tend to have higher number of unrepaired DSBs (DNA fragments) and dicentric chromosomes, thus are more susceptible to various cancers(*15*). In addition, mammalian stems cells deficient in NHEJ factors are shown to bias toward an NHEJ-independent "alternative" end-joining (A-EJ) pathway, where two DNA ends are simply ligated by joining with 6-8 bps of microhomology(*16*, *17*). It is known that A-EJ is more error-prone than

NHEJ and produces increased number of chromosome rearrangements(*18*).

Despite the guardian role of NHEJ in the stability of human genome, evidences suggest that NHEJ also contributes to the genomic rearrangements in carcinoma cells due to the error-prone nature of this DNA repair process. First, DSBs that are incompatible for direct ligation need to be processed first in NHEJ, which can be the source of small deletions or insertions. In addition, when DSBs occur on different chromosomes, oncogenic chromosome translocation could develop by fusing DNA ends from different origins(*19*). In fact, next generation sequencing results indicate that NHEJ is the major source of both pathogenic chromosomal rearrangements and complex germline chromothripsis(*20*). Meanwhile, several reports suggest that in Fanconi anemia (FA) deficient cells, defective DNA repair pathways lead to the over-use of NHEJ and thus accumulation of genetic abnormalities. Genetic knockout and chemical inhibition of NHEJ factors in such cell lines could reduce the instability of genome(*21*, *22*) (also reviewed by Bunting et al.(*18*)). Thus, small molecules that target NHEJ factors play significant roles in cancer therapies.

1.3 NHEJ general mechanism and involving factors

NHEJ is a multi-step DNA repair pathway that involves several protein complexes and processing factors (Figure1.1). It is initialized by the recognition of DSBs by DNA-PK complex, then followed by the recruitment of the remaining NHEJ factors which ultimately fuses the broken ends(*24*, *25*). When necessary, nucleases and polymerases could also be recruited to minimally process the DNA ends prior to the end ligation step(*26*). After successful ligation of DSB ends, Ku70/80 and other NHEJ factors need to be dissociated from the repaired DNA. However, as each subunit of Ku fully encircles DNA(*27*), it becomes sterically trapped on DNA and cannot leave without cleavage. Cellular and biochemical assays reveal that neddylation promotes the RNF8 dependent ubiquitination of Ku70/80, and the AAA+-type ATPase VCP/p97 targets ubi-Ku70/80

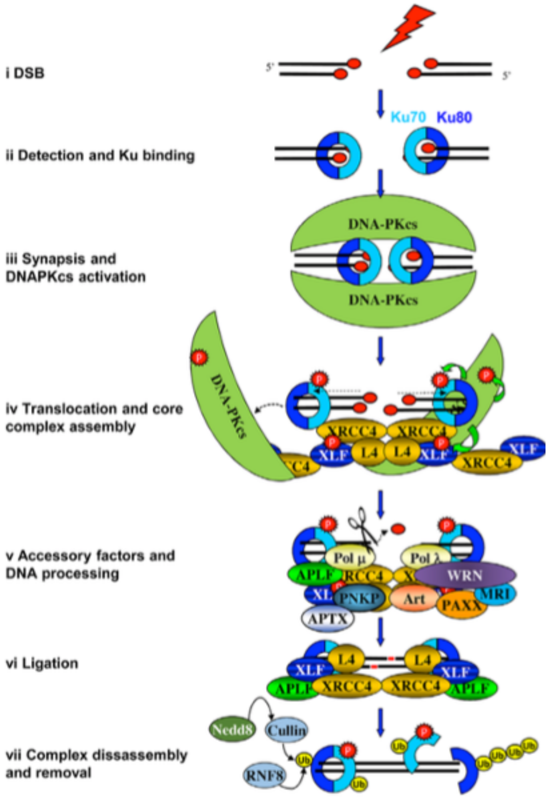


Figure 1.1: **Schematic representation of different stages in NHEJ.** Core complex (iv) contains DNA-PK holoenzyme (DNA-PK complex) and other necessary NHEJ factors. Minimal end processing is required in this example prior to ligation. Adapted from Rulten et al. 2017(*23*).

and drives its extraction from ligated dsDNA(*28–30*).

To gain better understanding of NHEJ mechanism *in vivo*, smFRET study using Xenopus egg extracts was applied to investigate this pathway recently(*31*). Two distinct synaptic stages of dsDNA are identified in the system. By supplying Ku70/80 and inactive DNA-PKcs, the DNA-PK initiation complexes tether two DNA ends to proximity in a long-range complex, and low FRET signals could be observed at this stage. However, when XLF and XRCC4-LigIV are supplied with active DNA-PKcs, the DNA ends are brought even closer, resulting in high FRET signals(*31*). Taken together, these results provide vital information about which factors contribute to different key NHEJ events. Nevertheless, atomic detail of how these factors orchestrate to ligate DSB remains unclear.

1.3.1 NHEJ core factors: DNA-PK holo-enzyme, XRCC4-XLF scaffold and LigIV

1.3.1.1 DNA-PK holo-enzyme

DNA break recognition in NHEJ requires DNA-dependent protein kinase (DNA-PK) holoenzyme, consisting of the Ku70/80 heterodimer (Figure 1.2A) and a ∼470 kDa DNA-PK catalytic subunit (DNA-PKcs, Figure 1.2B). Ku70/80 is the first factor that recognizes DSBs *in vivo*. It recruits DNA-PKcs upon binding to the end of DSB loci, and the kinase activity of DNA-PKcs remains quiescent until activated by DNA ends(*32*). The DNA-PK holoenzyme serves as the scaffold to recruit other NHEJ factors, including the necessary NHEJ factors (DNA ligase IV, XRCC4 and XLF), other scaffolding factors (PAXX), nucleases (Artemis and APLF) and polymerases (Pol µ and Pol λ)(*26*).

As one of the phosphatidylinositol 3–kinase (PI3K)-related protein kinase (PIKK) family member, DNA-PKcs is a serine-threonine kinase that phosphorylates many NHEJ factors, including DNA-PKcs itself(*33*). Known DNA-PKcs phosphorylation targets *in vitro* in NHEJ include, but not limited to Ku70/80(*34*, *35*), XRCC4(*36*, *37*), XLF(*37*, *38*), DNA Ligase IV(*39*), PNKP(*40*), APLF(*41*) and Artemis(*42*). While the biological significance of these phosphorylation events has not been fully discovered, DNA-PKcs autophosphorylation has been studied extensively and is known to be important during NHEJ. Biochemical studies suggest that DNA-PKcs undergoes autophosphorylation in a DNA damage-dependent manner, and the DNA-PKcs dependent recruitment of other NHEJ factors(*43*) is regulated by autophosphorylation(*33*). Among many post-translation modification sites identified in the online database(*44*), frequently characterized and studied DNA-PKcs autophosphorylation sites include the ABCDE cluster (T2609, S2612, T2620, S2624, T2638 and T2647 in the M-HEAT region)(*45–47*), the PQR cluster (S2056, S2023, S2029, S2041 and S2053, in the NUC194 domain of the M-HEAT region)(*48–50*), and individual sites in the head domain such as S3205 in the FAT domain(*45*) and T3950 in the kinase domain(*51*).

Both *in vitro* and *in vivo* evidences suggest that ABCDE and PQR clusters is important for NHEJ and V(D)J recombination efficiency(*47*, *52*). In addition,the DSB repair pathway choice between NHEJ and HR could also be regulated by DNA-PKcs autophosphorylation(*49*, *53*, *54*). Interestingly, DNA-PKcs undergoes autophosphorylation, loses its kinase activity(*55*) and dissociates from the NHEJ repair site(*56*). Meanwhile, S/T-to-A mutations at the clusters stabilize DNA-PKcs at DSB sites(*33*, *49*, *57*). Taken together, it is suggested that after DNA-PK complex formation, DNA-PKcs undergoes autophosphorylation and subsequently release from the repair site, potentially allowing other NHEJ proteins to access the DSB ends to facilitate processing and ligation(*58*, *59*). Although alternative mechanisms probably exist, autophosphorylation is likely to happen *in trans*, which indicates that two DNA-PKcs proteins can synapse together and phosphorylate each other(*60*). This intermolecular autophosphorylation of DNA-PKcs is required for end-ligation(*52*). However, the relationship among *in trans* phosphorylation, DNA-PK dimerization and DNA-PKcs dissociation at the molecular level remains unclear.

The crystal structure of Ku70/80 has been solved both alone and in complex of duplex DNA in the early 2000s(Figure 1.2A)(*27*). Since Ku70/80 heterodimer forms a ring-like structure that surrounds the dsDNA with a $\sim$14bp footprint, its binding affinity towards DNA is quite high, with the Kd value in the range of 0.15-0.4 nM(*70*). The DNA bound Ku acts as a hub and tool-belt to gather multiple NHEJ factors at the DSB site by direct protein-protein interactions(*71*, *72*). While many NHEJ factors interact with Ku by flexible loops and tails(*73*), DNA-PKcs and Ku-DNA form stable DNA-PK complex at the DSB ends, which activates DNA-PKcs phosphorylation activity(*74*).

The large size of DNA-PKcs makes it challenging to be over expressed and purified from either prokaryotic or eukaryotic protein expression systems. For a long period of time, DNA-PKcs can only be purified endogenously thanks to its abundance in the human cell nucleus(*75*), which is one of the main reason why decades of efforts have to be spent to push the structural study of this important kinase(*61*, *76–81*). X-ray crystallography study shows that DNA-PKcs folds into three major units: the N-terminal arm domain, the circular cradle, and the C-terminal head region harboring the kinase module (Figure 1.2B)(*61*). In the past ten years, technical breakthrough in the field of Cryo-EM brought the field of NHEJ structural biology to a new era. Comparing to crystallography, much less material is required for Cryo-EM structure determination. In addition, in-solution Cryo-EM samples greatly lifted the size limitation of target proteins and complexes, which makes the technique more suitable for large NHEJ protein-DNA complexes. The detailed discussion of Cryo-EM technical breakthrough will be in Chapter 2. With the help of cutting-edge technology, structure of DNA-PKcs(*82*), as well as the DNA-PK complex were successfully obtained in recent years. The cryo-EM model of DNA-PK complex revealed the overall architecture of how Ku70/80, DNA-PKcs and the dsDNA associate with each other (Figure 1.2C)(*62*, *83*, *84*). Extensive interaction surfaces between Ku70/80 and DNA-PKcs are observed, with Ku70/80 translocated inward along DNA(*85*) and the DSB ends protected in the cradle region of DNA-PKcs HEAT repeats. The DNA end pushes the N-terminal ‘latch’ of DNA-PKcs closer to the C-terminal head domain, activating its kinase activity.

1.3.1.2 XRCC4-XLF scaffold

After NHEJ initiation, the DNA-PK complex is expected to recruit all other necessary NHEJ factors to the DSB repair site, including XLF, XRCC4 and LigIV. Among these factors, XRCC4(X-ray cross complementing gene 4) and XLF(XRCC4-like factor, also called Cernunnos) don’t embed any enzymatic activities, and are main scaffolding factors in NHEJ.

XRCC4 was one of the early identified NHEJ factor that is a nuclear phosphoprotein and *in vitro* target of DNA-PK phosphorylation(*86*). Deficiency of XRCC4 in mice cell lines causes severe phenotypes, including growth defects, senescence, sensitivity to ionizing radiation and lack of V(D)J recombination. Defective of neurogenesis and extensive apoptotic cell death was also observed in XRCC4 deficiency mice(*87*). XLF was later identified as a XRCC4 interacting NHEJ factor(*88*) in a radiosensitive and immune-deficient patient cell line, and is required for efficient repair of DNA DSBs *in vivo*(*89*). Besides interacting with XRCC4, Ku70/80 has been shown to interact with XLF and recruit it to the NHEJ DSB repair site without the requirement of XRCC4(*90*). Thus, XLF, XRCC4 and DNA-PKcs could be recruited independently by Ku during NHEJ then assemble into large complexes(*25*). Further *in vivo* study shows that C-terminal tail of XLF, termed XLF KBM(Ku Binding Motif), binds to Ku80 and allow Ku dependent XLF recruitment and efficient NHEJ(*91*, *92*). Structure of Ku70/80 with KBM binding was later solved to illustrate the opening of Ku80 vWA domain upon KBM binding(*93*).

Crystal structures of XRCC4 (Figure 1.2D)(*63*) and XLF (Figure 1.2E)(*64*) shows that both of them form stable homodimers using their C-terminal coiled-coil stalks, with head domains pointing away from each other to form a Y-shape architecture. Since XLF and XRCC4 interacts via their head domains, it was expected, and later shown in the crystal structures that XRCC4 and XLF can interact to form repeating filaments, and the coiled-coil stalks of each XLF/XRCC4 subunit tilts for a certain degree in different filaments (Figure 1.2F)(*65*, *94–96*). In addition, in-solution models of such filaments has been characterized as well by Small-angle X-ray scattering (both with(*94*) and without(*97*) the existence of LigIV BRCT domains). While it has been proposed for a long time that DNA-PKcs, or the DNA-PK complex, can dimerize to bridge two copies of DSB ends(*61*, *98*, *99*), the XLF-XRCC4 filaments are believed to be able to wrap around naked broken DNA and bridge the ends for end-joining as well(*66*, *100*, *101*). However, several questions remain to be answered in this model. First, filament structures seen in the crystal lattices only contain the C-terminal truncated XLF and XRCC4, and adding LigIV into the system results in the disruption of DNA binding and bridging(*65*). Second, ∼1k bp long naked DNA has to be used to assemble the filament, while in eukaryotic cells, genomic DNA is mostly wrapped and protected by nucleosomes. Third, the role of DNA-PKcs in this model remains mysterious. In addition, single-molecule FRET experiments in *Xenopus laevis* egg extract indicates that only one copy of XLF dimer is enough to bridge two DSB DNA ends at the NHEJ repair site(*92*, *102*), adding another layer of complexities in different working models of NHEJ end bridging. Clearly, structural studies on this topic is needed to solve the puzzle and raise new models that are consistent with these seemingly controversial mechanisms.

1.3.1.3 LigIV and X4-L4 complex

One distinct biological function of XRCC4 in NHEJ is to complex with LigIV (Figure 1.2G)(*86*). Interaction between XRCC4 and LigIV is stable, and LigIV activity can be stimulated with XRCC4 binding(*103*). XRCC4-LigIV complex is also shown to play important roles during V(D)J recombination(*104*). Crystal structure of XRCC4-LigIV (X4L4) complex shows extensive protein-protein interaction between the coiled-coil of XRCC4 homodimer and tandem C-terminus BRCT (BRCA1 [breast cancer associated 1] C terminal) domains of LigIV (Figure 1.2H)(*68*). The coiled-coil motif of XLF is significantly shorter than XRCC4, and the helix is folded back at the C-terminus tail, which may explain why it is not able to bind LigIV in a similar manner (Figure 1.2D,E).

The catalytic domain of LigIV is on its N-terminus and is flexibly linked with tandem BRCT domains by a ∼60 aa linker (Figure 1.2G)(*68*, *69*). Like other ligases in human, LigIV catalytic region consists three domains: the N-terminal DNA-binding domain (DBD), the nucleotidyl-transferase domain (NTD) and an OB-fold domain (OBD), as shown by the crystal structure model(Figure 1.2I)(*69*). After DNA adenylation, LigIV conformation changes from the open state to the close state, with all three domains binding to substrate DNA and the NTD domain interacting with the nick. As the main ligase responsible for end ligation in NHEJ, LigIV can be stimulated by not only XRCC4, but also other NHEJ factors such as Ku(*105*), DNA-PK complex(*106*) and XLF(*89*, *107*). Hypomorphic mutations in lig4 gene result in an autosomal recessive disorder called LIG4 syndrome, which is mainly caused by NHEJ deficiency. Patients with LIG4 syndrome shows a range of phenotypes, including immunodeficiency and hypersensitivity to ionising radiation(*108*). Besides blunt end DNA, LigIV can also repair various DNA ends with short gaps or incompatible overhangs(*109–111*). However, it has been shown for a long time that unlike LigI and LigIII, individual LigIV can seal one nick on the DSB ends, and is difficult to re-adenylate after the first turn-over(*112*). This unique feature of LigIV indicates that in NHEJ, multiple factors orchestrate in complex to help performing a complete end ligation by LigIV.

1.3.2 NHEJ appendix factors: PAXX and Artemis as examples

Loss of core factors severely impact NHEJ efficiency, thus in most cases results in stronger phenotype. In contrast, disruption of appendix NHEJ factors often mildly impact cell survival rate, either because their functions in NHEJ are redundant, or because they are recruited only when certain type of DNA end processing is needed prior to end ligation(*26*).

1.3.2.1 PAXX as NHEJ scaffolding factor

As an appendix scaffolding factor in NHEJ, PAXX (PAralog of XRCC4 and XLF) are first identified by mass spectrometry interactome studies(*113*). Crystal structure of PAXX demonstrates its similarity with XLF and XRCC4(*114*), but no evidence is available so far to characterize its interaction with either of these core NHEJ scaffolding factors. Instead, interaction between C-terminus of PAXX and Ku70 has been identified in multiple biochemical studies, both in *vitro* and in *vivo*(*113–116*). While PAXX is dispensable in mice, double knock-out of PAXX and XLF results in defective DSB repair that leads to apoptosis and embryonic lethality(*113*, *115*, *116*). Thus, it is believed that the role of PAXX in NHEJ, such as Ku accumulation and stabilization, overlaps a lot with XLF. However, their molecular mechanisms could still be distinct and need further characterization.

1.3.2.2 Artemis as nuclease involved in NHEJ, especially V(D)J recombination

Another class of appendix NHEJ factors is end processing enzymes, which includes nucleases (Artemis(*117*), WRN complex(*118*)), polymerases (Pol μ, Pol λ) and damage correction enzymes (PNKP, Aprataxin(*119*)). When activated by DNA-PKcs phosphorylation, Artemis performs endonuclease activity on both 5’ and 3’ DNA overhang and can open hairpins generated by RAG1/2 complexes during V(D)J recombination(*120*). Recently published structures of Artemis(*121*) and Artemis-DNA-PK complex(*122*) reveals molecular mechanism of how Artemis can interact with DNA overhang and be regulated by DNA-PKcs phosphorylation, serving as an elegant example of how core NHEJ complexes orchestrate with appendix factors during NHEJ.

1.4 Aims of the project

Despite the progress made so far, the functional role of DNA-PKcs during factor recruitment (XRCC4, DNA ligase IV and XLF) and DNA end bridging is poorly understood. The goal of this project is to directly visualize the key stages during NHEJ using cryo-EM as the primary structural technique, in order to better understand the overall mechanism of this essential DNA repair pathway. Specifically, I have captured and solved five distinct NHEJ stages – a Long-range synaptic complex with two copies of DNA-PK complex holding two DSB ends complexed with a scaffold containing LigIV-XRCC4-XLF-XRCC4-LigIV (LR complex); the same LR complex with PAXX incorporated and interacting with Ku70 (PAXX-LR complex); the same LR-PAXX complex with ATP added showing substantial conformational changes of DNA-PKcs and Ku70/80 (PAXX-LR-ATP complex); a Short-range synaptic complex containing all necessary NHEJ factors but DNA-PKcs, with DSB ends well aligned for final ligation (SR complex); an alternative dimeric state of only two copies DNA-PKcs that can serve as a reservoir of extra DNA-PKcs copies for rapid NHEJ responses (DPKcs dimer complex). All mentioned NHEJ stages will be described in the related chapters in more detail, and finally, a comprehensive picture of NHEJ mechanism illustrated at the molecular level will be described before the last chapter.

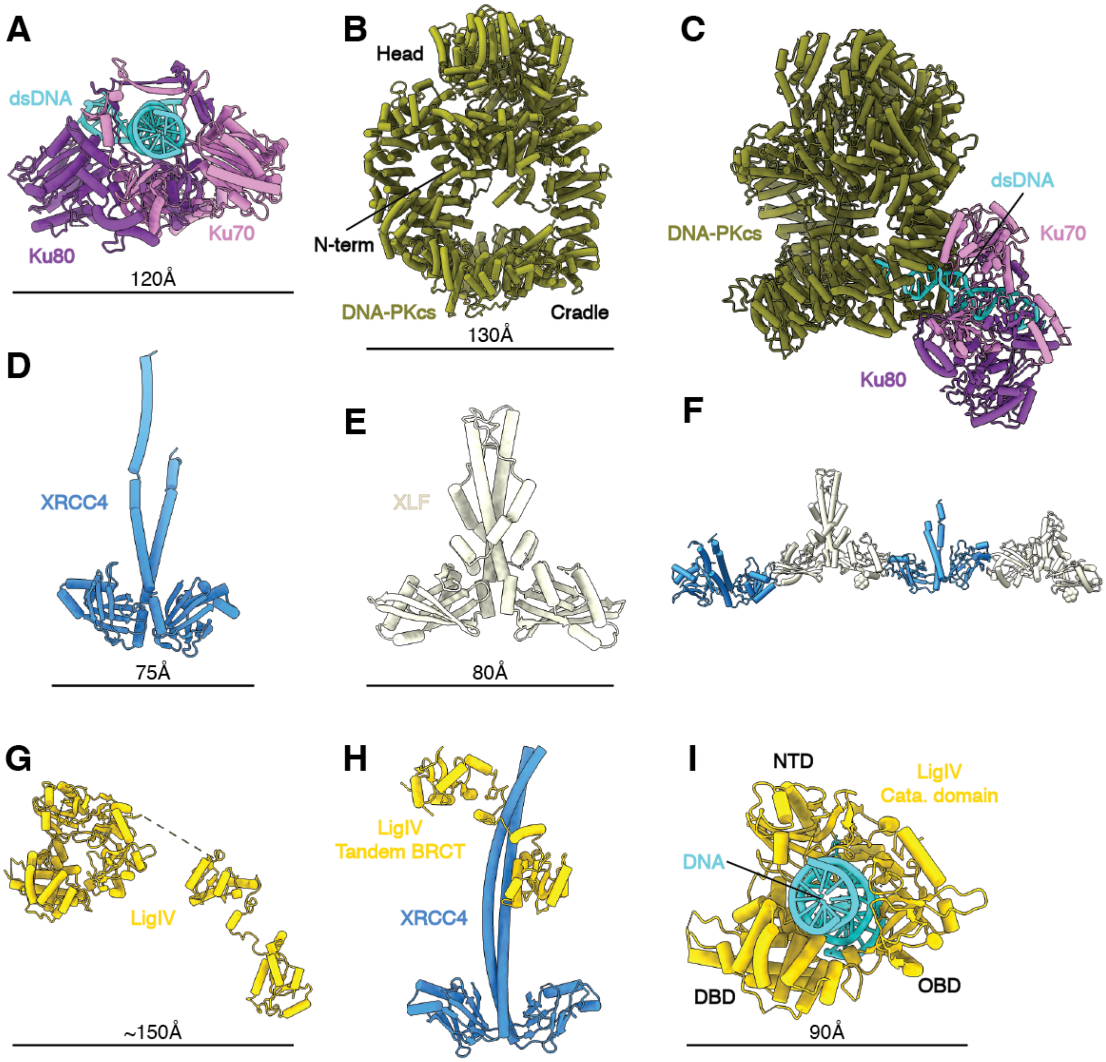


Figure 1.2: **Gallery of available NHEJ factor structures.** **A,** Ku70/80 heterodimer with dsDNA (PDB:1JEY)(*27*). **B,** DNA-PKcs (PDB:5LUQ)(*61*), showing N-terminal HEAT repeat region (N-term, 1-892), middle HEAT repeat region (Cradle, 893-2801) and Head catalytic domain (Head, 2802-4128). **C,** DNA-PK complex (PDB:6ZHA)(*62*). **D,** XRCC4 (PDB:1FU1)(*63*). **E,** XLF (PDB:2R9A)(*64*). **F,** Example of XRCC4-XLF filamentous complexes (PDB:3RWR)(*65*, *66*). **G,** Full length LigIV from the Short-range complex, showing N-terminal catalytic region flexibly linked with C-terminal tandem BRCT domains. Linker is not resolved in the complex structure (from PDB:7LSY)(*67*). **H,** X4L4(XRCC4-LigIV) complex (PDB:3II6)(*68*). **I,** Catalytic domains of LigIV in closed conformation (PDB:6BKG)(*69*), showing DBD, NTD and OBD domains interacting with nicked substrate DNA.

CHAPTER 2

GENERAL MATERIAL AND METHODS OF THE PROJECT

This chapter summarizes the general strategy of NHEJ complex structure determination, including general methods applied for DNA substrate design, complex assembly, EM sample preparation and data processing. For each NHEJ complexes mentioned in the later chapters, a detailed description of specific strategies and experimental designs will be discussed separately.

To obtain the purified stock of required protein factors (DNA-PKcs, Ku70/80, XRCC4-LigIV, XLF and other NHEJ factors), we have established a collaboration with Dr. Susan Lees-Miller at University of Calgary and Dr. Alan Tomkinson at University of New Mexico, who pioneered in discovering and characterizing key NHEJ factors.

2.1 Overall structure determination workflow

Figure 2.1 shows the overall workflow of the project. For each complex structure that was investigated, the dsDNA substrate was first designed and synthesized. Next, the DNA substrate was supplied with required protein components to assemble the target protein complex. The efficiency of complex assembly was tested using diagnostic negative staining EM, and the complex assembly protocol was adjusted to seek for higher complex reconstitution efficiency. Image data collected from the microscope contains random 2D projections of target protein complexes. After several rounds of protocol optimization, decent amount of monodispersed complex particles could be viewed under the microscope, then the initial 3D model of the complex was reconstructed from aligned 2D projections. Next, the optimized complex assembly protocol was scaled up, and the reconstituted high-concentration complex sample was subjected to cryo-EM analysis. The outcome

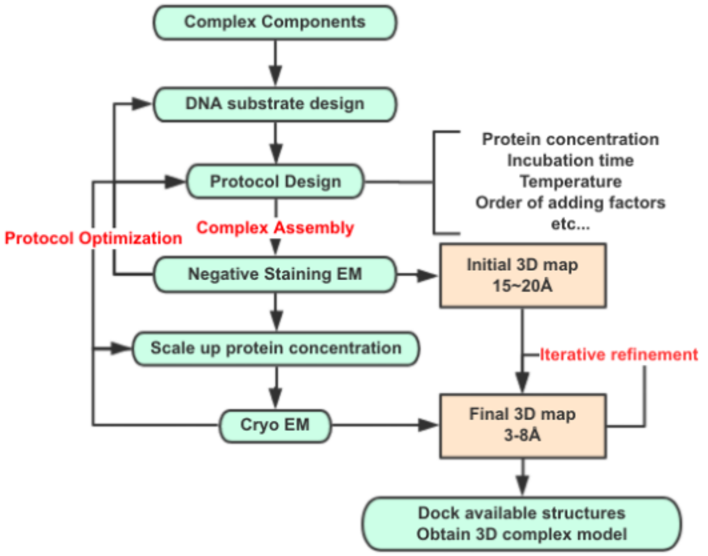


Figure 2.1: **Overall Structure Determination Workflow.**

frames of Cryo-EM was processed to reduce motion generated noise, and high resolution 2D particle projections was automatically picked and aligned similarly to reconstruct 3D structure. Finally, known structure models of complex components was docked into the refined cryo-EM volume, and the 3D model of target complex was generated and refined based on the fitting outcome.

2.2 Rationale design of DNA substrates

Figure 2.2 shows the design and sample sequences of DNA substrates for different protein complexes. Several designs were incorporated in all DNA substrates for various experimental purposes. First, a three-way Y-shape junction on one side of DNA duplex was designed to prevent Ku from binding to both ends of dsDNA(*27*). In addition, RNA-tagged biotin was attached to the end of Y-shape DNA and used to purify the fully assembled complex. During complex assembly, tagged complex was incubated with Streptavidin magnetic beads, then eluted by RNAse-H that specifically recognize and digest the DNA-RNA hybrid(*123*). Comparing to simple dsDNA, the designed

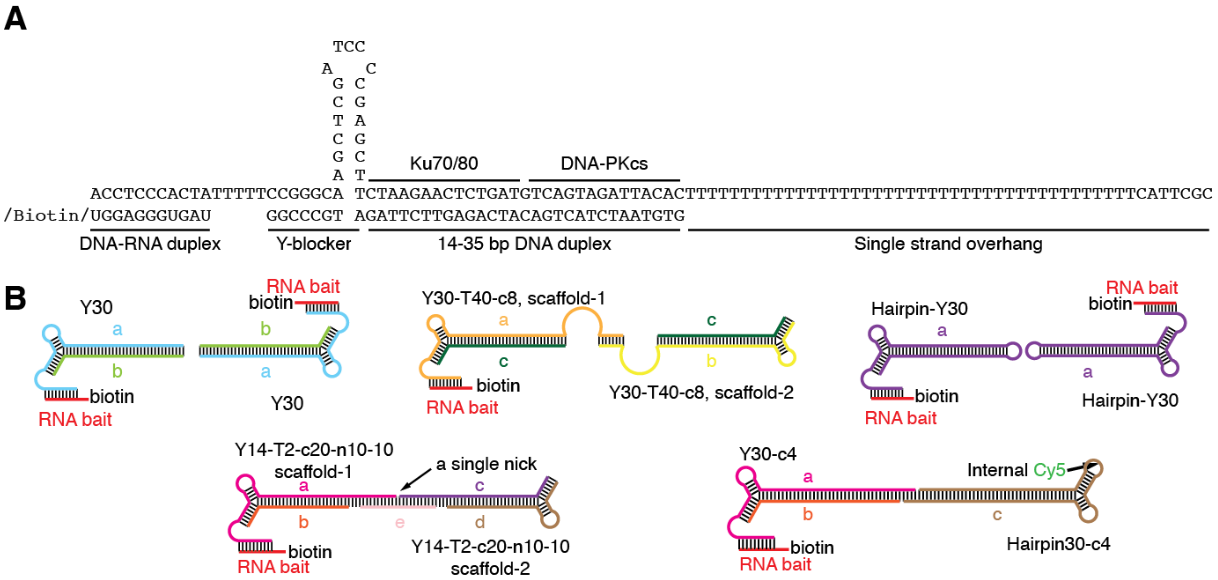


Figure 2.2: **DNA substrate designed for different NHEJ complexes. A,** An example sequence of designed DNA substrates (RNA-Y30-T40-c8). Different parts and binding footprints of Ku70/80 (14bps) and DNA-PKcs (16bps) are shown. DNA hybridization is limited at the top 8bps of DNA ends. **B,** Schematic of DNA substrate designed to assemble various NHEJ complexes. Sequences of each substrate will be listed and discussed in the corresponding chapters.

DNA substrate could greatly increase complex assembly efficiency and the extra purification step in the protocol could be used to better control the final salt condition and factor concentration of complex samples. The sequences of DNA substrates used will be listed in each chapters.

2.3 Complex assembly

To anneal the DNA substrates, oligonucleotides were purchased from Integrated DNA Technology (IDT). Lyophilized oligos were first resuspended in ultra-pure water for a final concentration of 200 µM. To prepare DNA templates used for complex assembly, complementary strands of each DNA duplex were mixed at a 1:1 molar ratio and a final concentration of 15 µM in ultra-pure water following the schematics in the corresponding chapters. They were denatured in boiling water for 5 minutes and cooled down to room temperature for 2 hours. Next, the RNA-biotin bait was annealed with DNA duplex at 45°C for 5 minutes, with a 1.5:1 molar ratio at a final concentration of 2 μM in ultra-pure water, and then cooled down to room temperature for 45 minutes.

Figure 2.3 shows example protocols for protein complex assembly. The sample assembly buffer contains 10 mM HEPES pH 7.9, 50 mM KCl, 10 mM $MgCl_2$, 2.5% glycerol, 1mM DTT and 0.05% NP-40, and the elution buffer is the same with sample buffer, except that 0.05 unit/μL RNAse-H is added. Three major steps are involved in the protocol. First, protein factors are incubated with DNA substrate at room temperature to prepare the full complex. Second, streptavidin magnetic beads are added into the sample to interact with biotin tagged DNA-RNA hybrid, and the bead will be washed with sample buffer. Finally, fully assembled complex will be eluted from the bead by the RNAse-H in the elution buffer. The eluted sample was then subjected to cross-linker treatment (Glutaraldehyde, 0.05%), and next be applied to copper grids for EM studies. Multiple steps in the protocol are optimized exclusively to assemble different target complexes. For example, to assemble the LR and SR complexes, after loading Ku70/80 on the DNA substrates, two DNA substrates with complementary single strand overhang need to be hybridized to make pseudo-dsDNA, and this DNA double strand will be used to recruit XLF and X4L4. Thus, tagged and untagged DNA substrates are incubated with Ku70/80 separately, and then combined to hybridize. Detailed adjustments of assembly protocol and rationales lying behind are discussed in the corresponding chapters.

2.4 Electron microscopy

2.4.1 Sample preparation and data collection

Comparing to other structural techniques, electron microscopy (EM) requires low amounts of protein sample. In addition, with the ability of dealing with in-solution samples that contain compo-

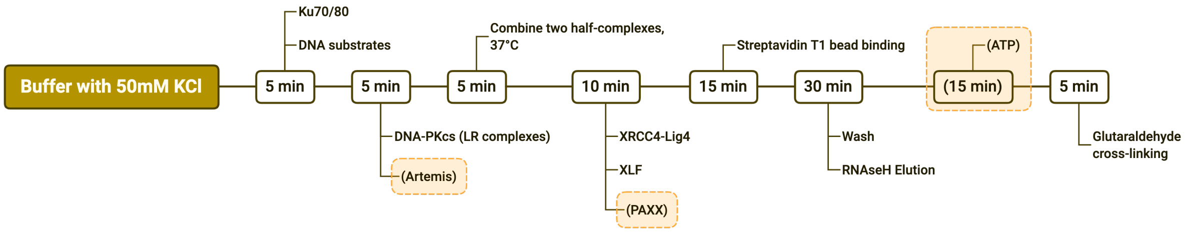


Figure 2.3: **Example of NHEJ complex assembly protocols.** Listed protocol is for negative staining sample preparation. The steps where appendix factors or ATP are added are highlighted with squared background.

sitional and conformational mixtures, EM is also a powerful technique to study functional protein complexes in different functional states(*124*). In fact, with the help of newly developed single electron detector hardware and streamlined data processing software(*125*, *126*), cryo-EM has been widely used to investigate the structures and functions of more and more molecular machines, including but not limited to proteasomes(*127*), ion channels(*128*), ribosomes(*129*) and transcription pre-initiation complexes(*130*). In this project, single-molecule cryo-EM serves as the primary technique to directly visualize the architecture of this essential DNA repair machine. At the same time, since negative staining EM, comparing to cryo-EM, requires even much less material and offers considerably fast turn-over for optimization, this method is mainly used in this project for the screening purposes.

To prepare negative staining samples, purified protein complex was applied to a carbon coated, glow-discharged mesh copper grid, in order to allow the protein to adhere to the grid surface. Next, uranyl formate was used to stain the protein on the grid, as this heavy metal compound could be well excluded from biological samples, quickly fix protein samples, and strongly scatter electrons to get high contrast imaged under the microscope(*131*). Image data was collected by JEOL 1400 transmission electron microscope operating at 120 kV at a nominal magnification of

×30,000 (3.71 Å per pixel). Leginon data collection software(*132*) was used for data collection on a Gatan 4k × 4k CCD camera using low-dose procedures (20 $e^{-}/Å^2$ exposures). Similar procedure was used for cryo-EM sample preparation, but the grid was prepared by applying a regular layer of graphene-oxide onto the holey carbon grid (C-flat or Quantifoil grids, Electron Microscopy Sciences). Cryo-EM data was collected by either JEOL 3200 or Titan Crios transmission electron microscopes.

2.4.2 Image processing and map reconstruction

Appion processing environment was used to process negative stain EM data(*133*). Automatic particle selection was performed on the micrographs with difference of Gaussians (DoG) particle picker(*134*). CTFFind4(*135*) was next used to estimate the contract transfer function (CTF) of each micrograph and flip the phases. Next, 2D classification of particles was conducted using iterative multivariate statistical analysis and multi-reference alignment analysis (MSA-MRA)(*136*), and 3D reconstruction of negative stained data was performed iteratively with dedicated image-processing software tools such as EMAN2(*137*), RELION(*138*) and Cryosparc(*139*).

Similar strategy was used to process the Cryo-EM data. First, all the movie frames were aligned using MotionCor2(*140*). After motion correction, the paragraphs were used for data pre-processing within the Appion processing environment(*133*). Particles were automatically picked by Gautomatch, and the local CTF of each particle was determined using Gctf(*141*). RELION was again used for the 3D classification and model refinement steps(*138*).

2.5 Model building

Most of the NHEJ factors have published structural models, either from X-ray crystalgraphy or Cryo-EM. To build the structure model of each complex, all published subunit structures were used to first rigidly dock into the map using Chimera(*142*) or Chimerax(*143*), then refined using Coot(*144*) and ISOLDE(*145*). To further validate our model, the deep-learning based multiple sequence alignment and structure prediction packages, such as trRosetta(*146*) and alphafold(*147*) were also used when necessary. Meanwhile, manual refinement and the Phenix real space refinement(*148*) were performed iteratively to further refine the model.

2.6 *In vitro* biochemical assays

2.6.1 DNA ligation assay

The LR and SR complexes with Cy5-labeled substrate were assembled similarly as described for EM sample preparation above, while buffer compositions before and after pull-down were modified for the reaction. Specifically, buffer with no $MgCl_2$ and 50 mM EDTA was used during complex assembly in order to prevent non-specific LigIV ligation. After the intact complex was assembled and bound to the streptavidin-coated magnetic beads, samples were washed with buffer containing 5 mM $MgCl_2$ without EDTA. 1 mM ATP was then added to the sample, and the reaction was carried out for 10 minutes at room temperature. For the control lanes, the complex assembly buffer remained unmodified, but various buffer conditions used after bead pull-down were modified accordingly. Next, the beads were boiled at 95°C for 5 minutes to elute off the DNA, and Proteinase K was added to allow incubation at 37°C for 15 minutes. 50% volume of formamide and 10 mM of EDTA were added to prepare the sample for denatured urea gel (TBE gel with 8% acrylamide (19:1)). Gel electrophoresis was performed at 250V for 40 minutes and scanned using a Sapphire Biomolecular Imager (Azure) at optimal absorbance.

2.6.2 Silver staining

To run sliver staining gel and check the complex assembly integrity, complex samples were assembled similarly as described for EM sample preparation above without adding Glutaraldehyde cross-linker. Afterward, each complex sample, together with half amount of individual input factors were diluted to 7.5 µL, then supplied with 2.5 µL of NuPAGE™ LDS Sample Buffer (4X) (Invitrogen™). Samples were boiled at 95°C for 5 minutes then loaded on to NuPAGE™ 4 to 12% Bis-Tris Mini Protein Gels (Invitrogen™). Gel electrophoresis was performed at 200V for 50 minutes with MOPS SDS Running Buffer, and subjected to silver staining. Reagents and protocol identical to the Pierce™ Silver Stain Kit (Thermo Scientific™) were used to run silver staining after electrophoresis, and the gels were developed for 3 to 5 minutes to get optimal visualization.

CHAPTER 3

STRUCTURAL BASIS OF LONG-RANGE TO SHORT-RANGE SYNAPTIC TRANSITION IN NHEJ

To understand whether DNA-PKcs is required for DNA end-bridging during NHEJ, and how core factors coordinate to bring DNA ends together, we first determined the structures of two key NHEJ complexes; a long-range (LR) synaptic complex that brings two DNA ends into proximity and a short-range (SR) synaptic complex in which the DNA ends are aligned for ligation.

3.1 Negative staining characterization of LR complex

Figure 3.1 shows the selected negative staining and Cryo-EM results during the process of protocol optimization. While co-incubation of Ku, DNA-PKcs, LigIV-XRCC4, and XLF(*31*, *149*) with duplex DNA substrates (Figure 3.1) yielded only DNA-PK complexes (Figure 3.1a-c), pre-assembly of DNA-PK on DNA ends followed by incubation with LigIV-XRCC4 and XLF generated larger complexes (Figure 3.1d-f). Thus, the optimized protocol that used Y30-T40-c8 DNA substrates as described in the next section was used to make Cryo-EM samples and collect high-resolution data for model building.

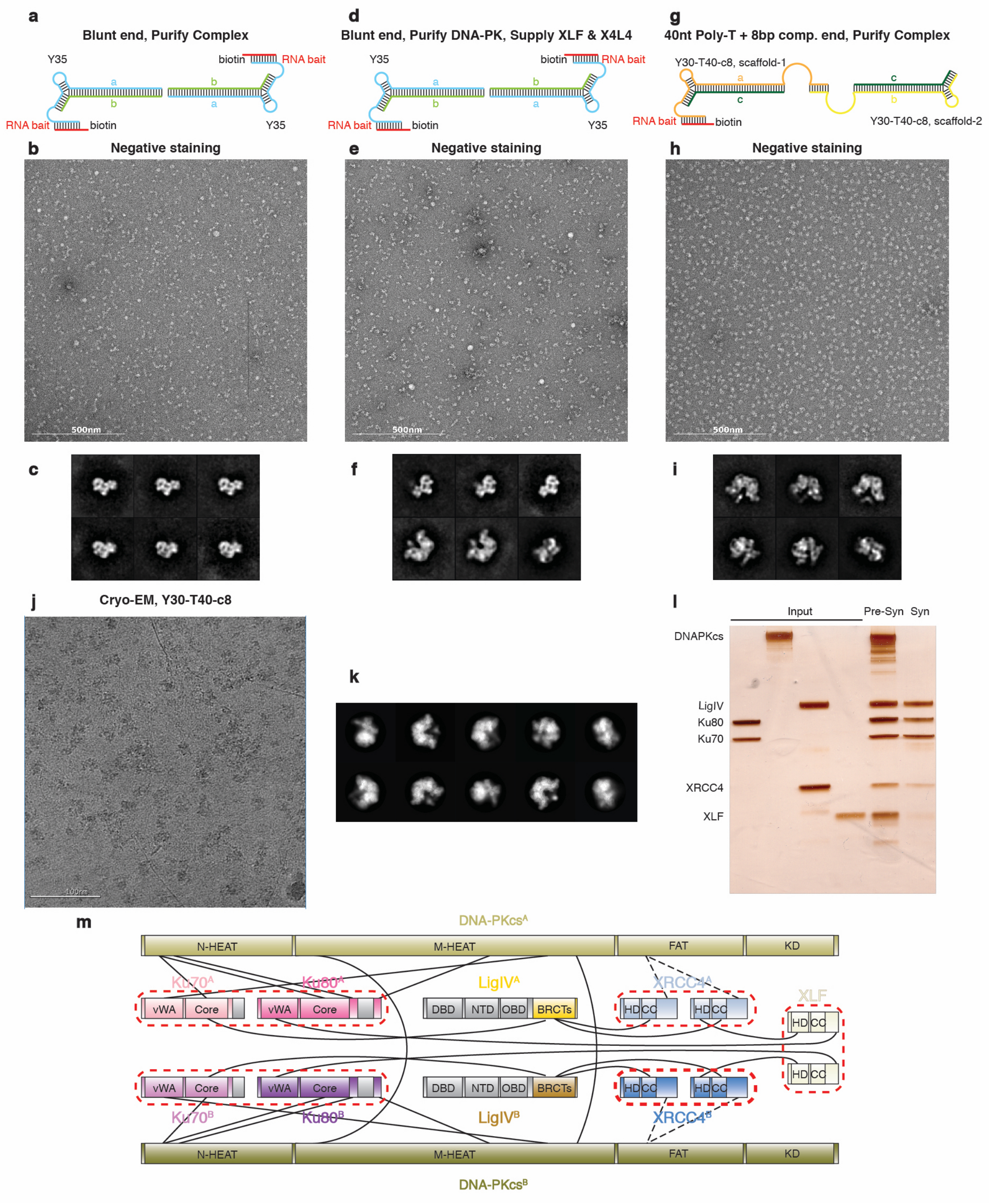


Figure 3.1: (Caption next page.)

Figure 3.1: (Previous page.) **Optimization of the LR synaptic complex assembly with various DNA substrates. a,** Schematic showing the Y35 blunt-end DNA substrate. Complex assembly was attempted (supplying DNA-PKcs, Ku, XLF and LigIV–XRCC4) before purification by RNase H elution. **b,** A representative negative-staining raw micrograph of the complex assembled as described in **a**. The raw micrograph is representative of 24 micrographs. **c,** Representative two-dimensional class averages of the complex assembled as described in **a**, showing the appearance of only the DNA-PK complex despite the addition of XLF and LigIV–XRCC4. **d–f,** Same procedure as **a–c,** showing the complex assembly with the same Y35 substrate, but adding XLF and LigIV–XRCC4 to the purified DNA-PK complex after RNase H elution. The raw micrograph is representative of 27 micrographs. In **f**, the two-dimensional class averages representing the characteristic view of scarce but existing LR complex were obtained. **g–i,** Same procedure as **a–c,** , showing the complex assembly using Y30–T40–c8 DNA substrate with 40 nt flexible poly(T) and 8 bp of complementary ends as the 3' overhang. Although the single-stranded poly(T) overhang and the 8-bp complementary region contribute to the stability of the complex, they are not observed in any part of the reconstructed density map, presumably because these single-strand DNA tethers are too flexible to be aligned with the rest of the complex. The raw micrograph is representative of 24 micrographs. The complex was assembled before RNase H elution as described in **a**. In **i**, the majority of the two-dimensional classes correspond to the LR complex. **j,** A representative cryo-EM raw micrograph (out of 17,114 micrographs in total) of the LR complex assembled with the Y30–T40–c8 DNA substrate shown in **g**. **k,** Representative two-dimensional class averages of particles (329,784 in total) contributing to the final reconstruction of the LR complex. **l,** Silver-stained SDS–PAGE (4–12% gradient, biologically replicated three times) showing the input purified subunits (Ku, DNA-PKcs, LigIV–XRCC4, and XLF) and the RNase H purified LR and SR complex for cryo-EM data collection. All representative micrographs in **b, e, h, j** are from at least three biologically replicated experiments. **m,** Protein–protein interaction network between the components of the LR complex. Major unmodelled regions are shown in grey. Well-documented hetero- or homo-dimers are grouped by red dashed lines. Alternative protein–protein interactions are depicted by black dashed lines. The globular domain within the Ku80 C-terminal region (CTR) is completely flexible in the LR complex, and we do not see evidence of the Ku80 CTR domain swap that was previously observed(*62*). The putative distance between one Ku80 CTR globular domain and the other copy of the Ku80 C-terminal helix is too far to be reached by the 18-amino acid linker within Ku80. BRCTs, tandem BRCT domains; CC, coiled-coil domain; CTD, C-terminal domain; HD, head domain; KD, kinase domain; M-HEAT, middle HEAT domain; N-HEAT, N-terminal HEAT domain; NTD, N-terminal domain; OBD, OB-fold domain.

Table 3.1: **Oligonucleotides used for the assembly of the Long-range synaptic complex**

Construct	Strand	Sequence (5’-3’)
Y30-T40-c8	a	ACCTCCCACTATTTTTCCGGGCAAGCTCGATCCCCGAGCTTCTAAGAACTCTGATGTCAGTAGATTACACTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCATTCGC
	b	CCGGGCAAGCTCGATCCCCGAGCTTCTAAGAACTCTGATGTCAGTAGATTACACTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTGCGAATGA
	c	/5phos/GTGTAATCTACTGACATCAGAGTTCTTAGATGCCCGG
	RNA bait	UAGUGGGAGGU/3BiotinTEG/
Y35	a	CGCGATGTCGTTTCGCGCCCAGCTTTCCCAGCTTGCATTGACTGGCAACGTAGCATCCTGACATCACA
	b	TGTGATGTCAGGATGCTACGTTGCCAGTCAATGCAGGGCGCG
	RNA bait	CGACAUCGCGUUU/3BiotinTEG/

3.2 Details of Cryo-EM Data Collection, Processing and Model Building

3.2.1 Sample Preparation

As described in Chapter 2, the oligos used to assemble LR complexes are listed in Table 3.1. To assemble the LR complex with the Y30-T40-c8 DNA-RNA substrate for EM studies, DNA-PK holoenzyme (Ku70/80, DNA-PKcs, dsDNA) was first assembled with two half-DNA substrates separately, then combined to assemble the full complex. Ku70/80 was first incubated with annealed DNA substrate at room temperature for 5 minutes, then diluted in sample buffer (10 mM HEPES pH 7.9, 50 mM KCl, 10 mM $MgCl_2$, 2.5% glycerol, 1mM DTT, 0.01% NP-40 for negative staining EM and 0.05% NP-40 for cryo-EM. All concentrations are final concentrations after complex assembly). DNA-PKcs was then added, and the sample was incubated for another 5 minutes. Next, two half-complexes were combined and incubated at 37°C for 5 minutes, then cooled down to room temperature in order to help the annealing of the single-strand overhang. LigIV-XRCC4 complex and XLF were then supplied and incubated for 10 minutes. Magnetic streptavidin T1 beads (Invitrogen) were equilibrated with sample buffer and added into the sample for 15 minutes to immobilize the assembled complex. Beads were then washed with sample buffer and eluted for 1 hour with 4 µL elution buffer (sample buffer with 0.05 unit/µL RNAse H, New England Biolabs).

The negative staining samples of the LR complex contained $\sim$700 nM Ku70/80 and DNA-PKcs, $\sim$2.1 µM LigIV-XRCC4, $\sim$500 nM XLF, as well as $\sim$350 nM of each half-DNA substrate.

3.2.2 Data collection and processing

400 mesh copper grid (Electron Microscopy Sciences) with continuous carbon supporting layer was used to prepare negative stain samples. Before sample deposition, the grid was plasma-cleaned for 10 seconds using a Solarus plasma cleaner (Gatan) equipped with air at 25W power. Purified complex sample (3 µL) was cross-linked for 5 minutes with 0.05% glutaraldehyde and then loaded onto the grid. The grid was incubated in a 100% humidity chamber for 10 minutes to absorb the sample particles. After the incubation, the grid was stained by four successive 35 µL drops of 2% (w/v) uranyl formate solution for 5s, 10s, 15s, and 20s, then blotted till dryness. Data was collected using a JEOL 1400 transmission electron microscope operating at 120 kV at a nominal magnification of ×30,000 (3.71 Å per pixel). Leginon data collection software(*132*) was used for data collection on a Gatan 4k × 4k CCD camera using low-dose procedures (20 $e-/Å^2$ exposures). The defocus range was -1.5 to -3 µm.

The same cross-linking procedure was used for cryo-EM sample preparation (total volume 4 µL). For the first dataset of the LR complex that was collected on C-flat grids, the grid was prepared ahead of time by floating thin carbon film onto the holey carbon grid (CF-3.5/1-4C, C-flat grid with 3.5 µm hole and 1 µm space, Electron Microscopy Sciences) and was plasma cleaned for 10s at

5 W power using the Solarus plasma cleaner (Gatan) right before sample loading. Cross-linked sample was loaded onto the grid and incubated in the 4°C, 100% humidity chamber of Vitrobot (FEI) under low illumination conditions. Grid was blotted for 4 seconds at 25 force, plunge-frozen in liquid ethane, and stored in liquid nitrogen. For the other datasets that were collected on GO grids, a thin layer of GO was applied onto either the same C-flat grids or 200 mesh Quantifoil holey carbon grids (3.5/1 copper and 2/1 copper/gold, Electron Microscopy Sciences, all yielding similar results) and the grids are prepared right before sample loading.

For all four datasets collected on the LR complex, JEM-3200FS Field Emission Electron Microscope (JEOL) operating at 200 kV was used with K2 Summit direct electron detector (Gatan) at a nominal magnification of ×30000 (pixel size 1.1 Å, counting mode). The defocus range was -2 to -4 μm. MSI-Raster2 application of the Leginon data collection software(*132*) was used for collecting 17,114 movie series in total. 30-frame exposures were taken at 0.3 s per frame (9 s total exposure time), using a dose rate of 8 e^- per pixel per second, corresponding to a total dose of 76.5e^-/Å^2 per movie series.

The initial model of LR complex was derived from the negative staining dataset. Specifically, 36,819 particles were extracted and stacked with a box size of 144 × 144 pixels (3.71Å/pix). The model was generated as described in Chapter 2.

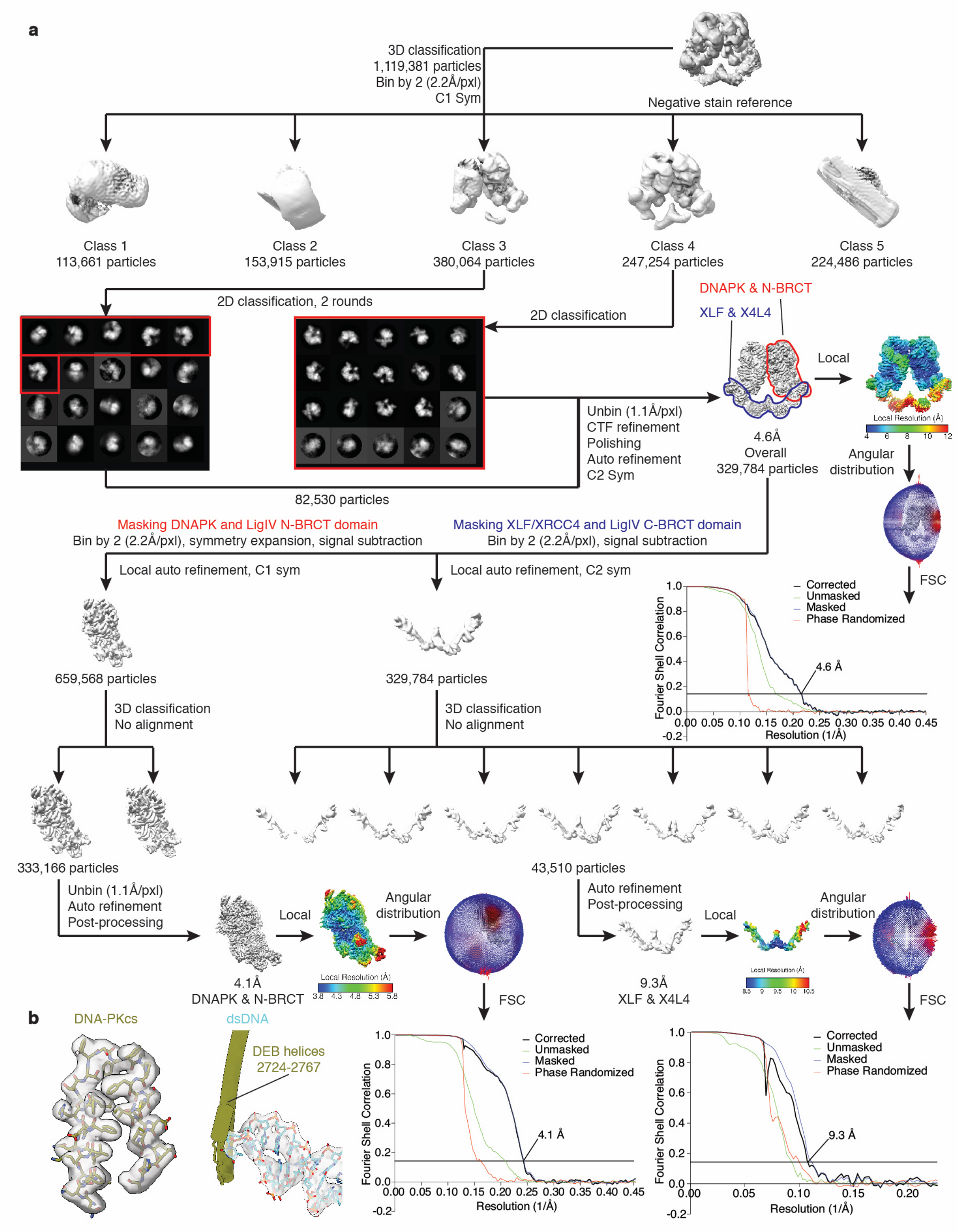


Figure 3.2: (Caption next page.)

Figure 3.2: (Previous page.) **Data-processing scheme of the LR synaptic complex sample. a,** Flow chart of the cryo-EM data processing procedure. The gold-standard Fourier shell correlation (FSC) curves (0.143 cut-off) show the final resolution of the holo-complex and each body. **b,** sample maps and fitted models of DNA-PKcs (olive) and dsDNA substrate (cyan) from the LR complex are shown at 4.1 Å resolution. Maps are shown as transparent surfaces and models are shown as sticks.

Figure 3.2 shows the data processing procedure of the LR complex. RELION 3.1(*138*) was then used for all the pre-processing, 3D classification, model refinement, post-process, and local-resolution estimation job to refine the Cryo-EM model. For the 3D reconstruction of the LR complex, 1,119,381 particles were automatically picked and supplied to an initial round of 3D classification step with the negative stain reconstruction map (low-pass filtered to 30Å) as an initial reference. Larger angular sampling interval, offset search range (pix), and offset search step (pix) were used for the first 50 iterations (15°, 10°, and 2° respectively), then followed by 190 iterations of default classification (7.5°, 10°, and 1°). Class 4 (247,254 particles) showed sharp structural features, so it was selected for further processing. Class 3 (380,064 particles) was the second-best class and was subjected to two rounds of reference-free 2D classification (20 classes, 50 iterations) to further remove bad particles. Six classes that showed sharp features and two copies of DNA-PKcs in the 2D class averages (82,530 particles) were selected and combined with class 4. Next, the selected particles were 3D auto-refined (with C2 symmetry), re-centered, and re-extracted without binning (1.1 Å per pixel, box size 432). Another round of 3D auto-refinement on this stack was performed with a soft mask applied around the whole complex, resulting in a 5.54 Å resolution reconstruction. All reported resolutions correspond to the gold-standard Fourier shell correlation (FSC) using the 0.143 criterion(*150*). Particles were then subjected to per-particle CTF refinement (per-particle defocus and per-micrograph astigmatism estimation, as well as beam tilt estimation), followed by Bayesian particle polishing(*150*). 3D auto-refinement against the shiny particles yielded a 4.85 Å resolution map, which was further improved to 4.6 Å by an additional round of CTF refinement (estimating both defocus, astigmatism, and magnification anisotropy) and Bayesian particle polishing. This map was used as the overall map for deposition.

The overall map of the LR complex contains two copies of a relatively rigid DNA-PK complex (DNA-PKcs, Ku70/80 and DNA) and a very flexible density corresponding to XRCC4-LigIV-XLF. To focus on the body of the rigid region (DNA-PKcs, Ku70/80, DNA and the first BRCT domain of LigIV) in the complex, a partial soft mask was applied to one copy of the corresponding volume, and C2 symmetry expansion was applied to align two copies of the rigid body into one dataset. Signal subtraction with the mask results in a 660,050-particle stack that contained all copies of rigid bodies. Next, the particles were binned by 2 again (2.2 Å per pixel, box size 216) and 3D auto-refined to 4.40 Å resolution. All of the 3D auto-refine steps after symmetry expansion were performed locally with an initial angular sampling interval of 3.7°. Subsequent two-class 3D classification with no alignment was performed, and the class with better features and higher resolution (333,166 particles) was selected to be unbinned (1.1 Å per pixel, box size 432) auto-refined and post-processed to 4.1 Å. Local resolution of the map was also estimated within RELION 3.1(*151*).

In order to generate a focused signal-subtracted map for the flexible body (XRCC4-LigIV BRCT-XLF), a similar strategy was applied to the same polished stack without applying symmetry expansion. The binned stack was then subjected to 7-class 3D classification without applying any alignment. The best class (43,510 particles) was further refined and post-processed to 9.3Å.

In parallel with post-processing done in RELION3.1, DeepEMhancer was applied on the refined maps as well to better correct local B-factor and yielded cleaner maps for model building and docking(*152*).

3.2.3 Model building

To build the structure model of the rigid region in the LR complex (DNA-PKcs, Ku70/80, DNA and 1st BRCT domain of LigIV), the published crystal and EM structures of DNA-PKcs (PDB 5LUQ(*61*),6ZHA(*62*)), Ku70/80 complex bound with XLF KBM (PDB 6ERH(*93*), without DNA) and the N-terminal BRCT domain of LigIV in complex with XRCC4 (PDB 3II6(*68*)) were referenced as homology models. The DNA was first built as standard B-form DNA using Coot, then fitted into the density using the interactive molecular-dynamics flexible fitting software ISOLDE. The DNA sequences were arbitrarily assigned as the poly-T overhang end of input double-strand DNA. Next, the models were manually inspected and refined to fit the model using both Coot and ISOLDE. The published cryo-EM structure model of DNA-PKcs (PDB 5W1R(*82*)) and DNA-PK holoenzyme (PDB 5Y3R(*84*), 7K0Y(*83*)), as well as the crystal structure model of DNA-bound Ku70/80 complex (PDB 1JEY(*27*)), were also referenced throughout the process. Residues 810-845, 2577-2605, and 2721-2773 in DNA-PKcs have not been modeled in any of the published structures to date and were modeled *de novo* using Coot and ISOLDE. The registers can be clearly identified as most of the densities and bulky side-chains aligned well between the map and the model. To further validate our model, the deep-learning based multiple sequence alignment module that is built-in trRosetta was used against the corresponding residues in DNA-PKcs to predict the secondary structure *de novo* and the resulted location of helices and loops aligned well with our model. Meanwhile, residues 585-601, 2902-2914, and 4008-4037 that were resolved in the previous DNA-PKcs models lacked corresponding densities in the LR complex density and were removed from the final reconstruction. DNA-PKcs kinase domain was superimposed with the structure of mTOR kinase binding to ADP (PDB 4JSV(*153*)). The location of the ADP molecule aligned well with the LR complex density, so it was directly used to dock in the map. Due to the resolution limitation, the register of residues was hard to determine unambiguously within part of the DNA-PKcs cradle domain, the DNA-PKcs FRB region, the BRCT domain, and the Ku80 vWA domain that binds to XLF peptide. For those regions, only the secondary motifs are flexibly fitted into density with minimal disruption of existing model constraints. The density that could not be incorporated into the structure of DNA-PKcs is located in the vicinity of the FAT domain (α-helix 2965-2977, 2991-3005, and 3008-3018) and is believed to be part of unstructured XRCC4 C-terminal tail. Based on the distribution of bulky side chains and XRCC4 sequence conservation, residue 270-278 is assigned to this extra density and manually built using Coot and ISOLDE (Supplementary Fig. 5). Manual refinement and the Phenix real space refinement were performed iteratively to further refine the model.

To build the model for the flexible body in the LR complex (XRCC4, XLF, and the 2nd BRCT domain of LigIV), the crystal structure of XLF-XRCC4 filament (PDB 3Q4F(*95*), with one copy of XLF dimer and two copies of XRCC4 dimer) were rigid-body docked in using UCSF Chimera as a template. Next, crystal structures of XLF (PDB 2R9A(*64*)) and XRCC4 in complex with 2nd BRCT domain of LigIV (PDB 3II6(*68*)) were aligned to the filament template, respectively. With this strategy, the interaction surfaces between XRCC4 and XLF were well-maintained, but the coiled-coil of XRCC4 and the BRCT domain of LigIV was off a little compared to the density. To better fit this part into the relative low-resolution map without disrupting the overall model architecture, ISOLDE was used to apply adaptive distance restraints between nearby atoms within this part of the model, and the helices and BRCT domains were then flexibly fitted into the density.

The complete model of the LR complex is composed of two copies of the rigid body and one copy of the flexible body. To generate such models, the above models were rigid-body docked into the composed map using UCSF Chimera, and the loop between two BRCT domains of LigIV was used to guide the linkage of the two domains, which were originally fitted into two bodies separately. The resolution of the linker loop was not good enough to perform de novo model building, but the terminal residues of the two BRCT domains were located in close proximity in the composed model. Coot was used to manually build the one missing residue (TYR765) and link them with minimal extra disruption.

For clarification, all subunits that are located on the left are named copy A, and subunits on the right are named copy B.

3.3 Architecture of the LR synaptic complex

Using the optimized order of addition approach and DNA substrates for complex assembly (Figure 3.1g-i), we determined the structure of the observed larger complex to 4.6 Å resolution using C2 symmetry (Figure 3.3a, Figure 3.1j-l, Figure 3.2). To improve resolution, symmetry expansion was applied to combine the signal from both halves of the complex. Signal subtraction followed by focused 3D classification and refinement revealed a 4.1 Å resolution reconstruction of DNA-PK complexed with the LigIV N-terminal BRCT (N-BRCT) domain. A similar focused classification resulted in a 9.3 Å reconstruction of the flexible XLF-XRCC4-LigIV sub-complex. With these density maps, we built a full model of this $\sim$1.66 MDa complex that contains two copies of DNA-PK and LigIV-XRCC4 (one LigIV complexed with an XRCC4 homodimer) on opposite sides. In accord with smFRET studies with Xenopus egg extracts(*92*, *102*), a single XLF homodimer bridges the two halves (Figure 3.3b). While the complex contains two DNA molecules, the ends are $\sim$115 Å apart (Figure 3.3c), suggesting that this LR synaptic complex corresponds to a similar complex identified by smFRET(*31*).

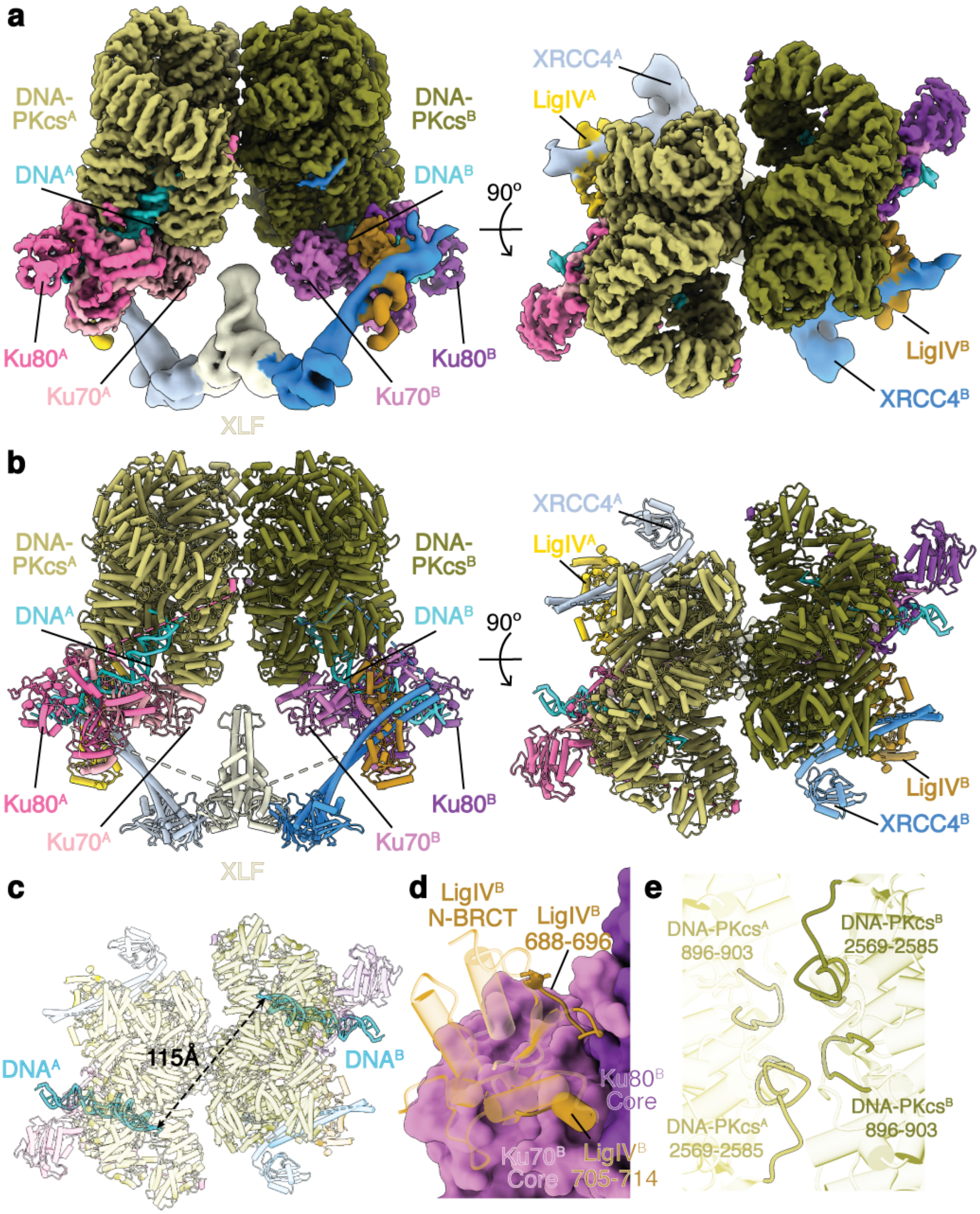


Figure 3.3: (Caption next page.)

Figure 3.3: (Previous page.) **Cryo-electron microscopy structure of the LR synaptic complex.** **a,** Front (left) and top (right) views of the cryo-electron microscopy composite map (Methods) of the LR complex assembled in the presence of ADP. **b,** Corresponding views of the structural model of the LR complex. Subunits are coloured here, and in all subsequent figures, as in **a**. **c,** An overview of the two DSBs within the LR complex. DNA elements are shown as solid ribbons and others in transparent representation. The distance between the two DSB ends is highlighted. **d,** Magnified view showing the interface between the LigIV N-BRCT (ribbon) and Ku70/80 core (surface) domains. Regions of the N-BRCT domain involved in the interaction are depicted. **e,** Magnified view showing the interface between the two copies of DNA-PKcs. For clarification, all subunits that are located on the left are named copy A, and subunits on the right are named copy B.

Multiple interaction surfaces contribute to the stabilization of the LR complex (Figure 3.1m). As in the XLF Ku-Binding Motif (KBM)-bound Ku crystal structure(*93*), the Ku80 von Willebrand (vWA) domain is rotated outward compared to the unbound state (Figure 3.4a-d). Similarly, the Ku70 vWA domain adopts an open state while engaging DNA-PKcs(*62*, *84*). While the head-to-head interaction surface between XLF and XRCC4 dimers is similar to that observed in the filamentous structure(*94*, *95*, *154*), the angle between the XLF and XRCC4 coiled-coil (CC) domains is larger (∼30°) in the LR complex compared with these structures (Figure 3.5a), suggesting a disfavored configuration to accommodate linking of the two DNA-PK complexes. While XLF-XRCC4 filaments(*94*, *95*, *154*) may reduce DNA flexibility adjacent to the DSB, the LR complex tethers the DSB ends. The XRCC4 CC domain is also bent by ∼15° more compared to the CC domain in LigIV-XRCC4 alone (Figure 3.5b), again likely accommodating linking of the two DNA-PK complexes by the LigIV-XRCC4-XLF-XRCC4-LigIV scaffold. Finally, the conformation of LigIV-XRCC4 in the LR complex is more similar to the S. cerevisiae Lig4p-Lif1p crystal structure than human LigIV-XRCC4(*155*), indicating the LigIV BRCT-XRCC4 CC interface is flexible (Figure 3.5c).

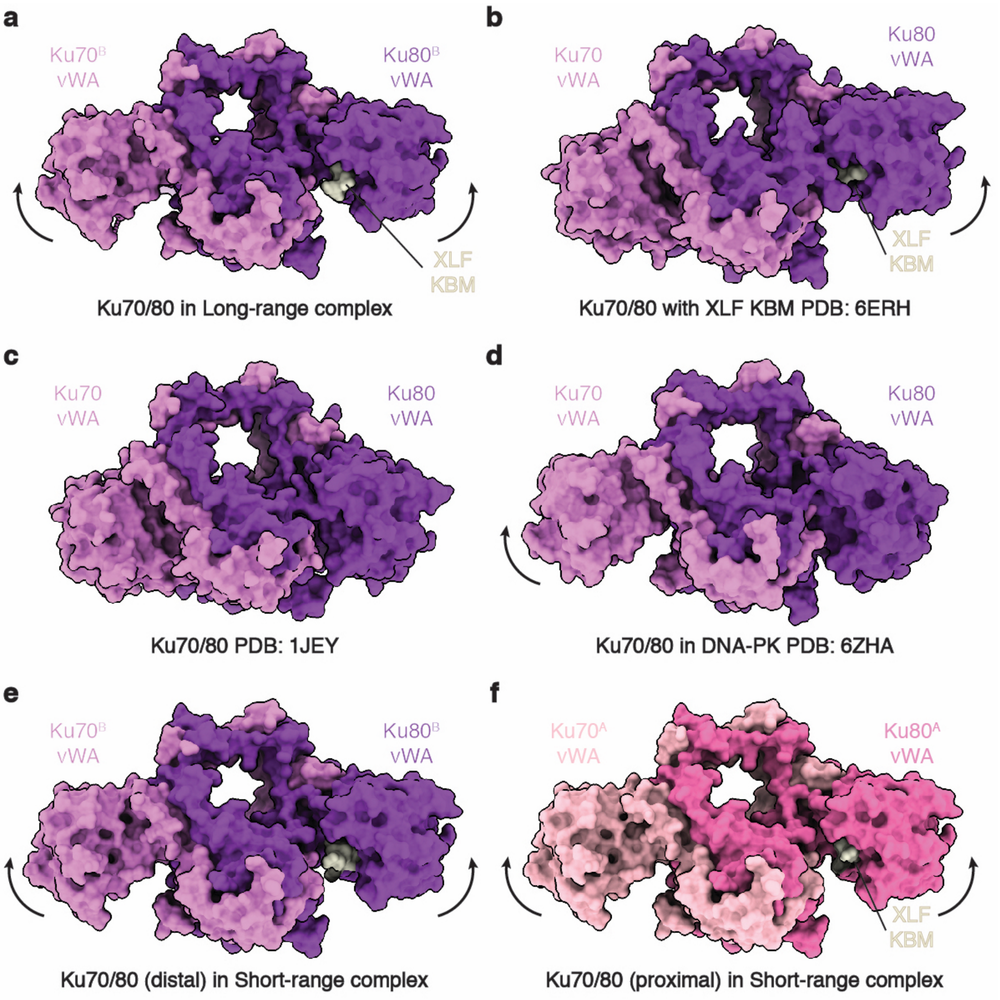


Figure 3.4: **Comparing the structure of Ku among the LR synaptic complex, the SR synaptic complex and previously published models. a,** Ku structure in the LR complex showing outward rotations of both Ku70 and Ku80 vWA domains. **b,** Crystal structure of the XLF Ku-binding motif (KBM) bound to Ku showing the outward rotation of only the Ku80 vWA domain(*93*). **c,** Crystal structure of apo Ku showing no rotation of either Ku70 or Ku80 vWA domains(*27*). **d,** Conformation of Ku shown in the cryo-EM structure of the apo DNA-PK complex(*62*). The Ku70 vWA domain is rotated outward, triggered by binding of DNA-PKcs. **e, f,** Two copies of the XLF Ku-binding motif bound to Ku in the SR complex. The conformation of both copies is the same as the one in the LR complex (**a**) despite the fact that DNA-PKcs is not present. Colour codes for Ku70 and Ku80 are the same as in Figure 3.3

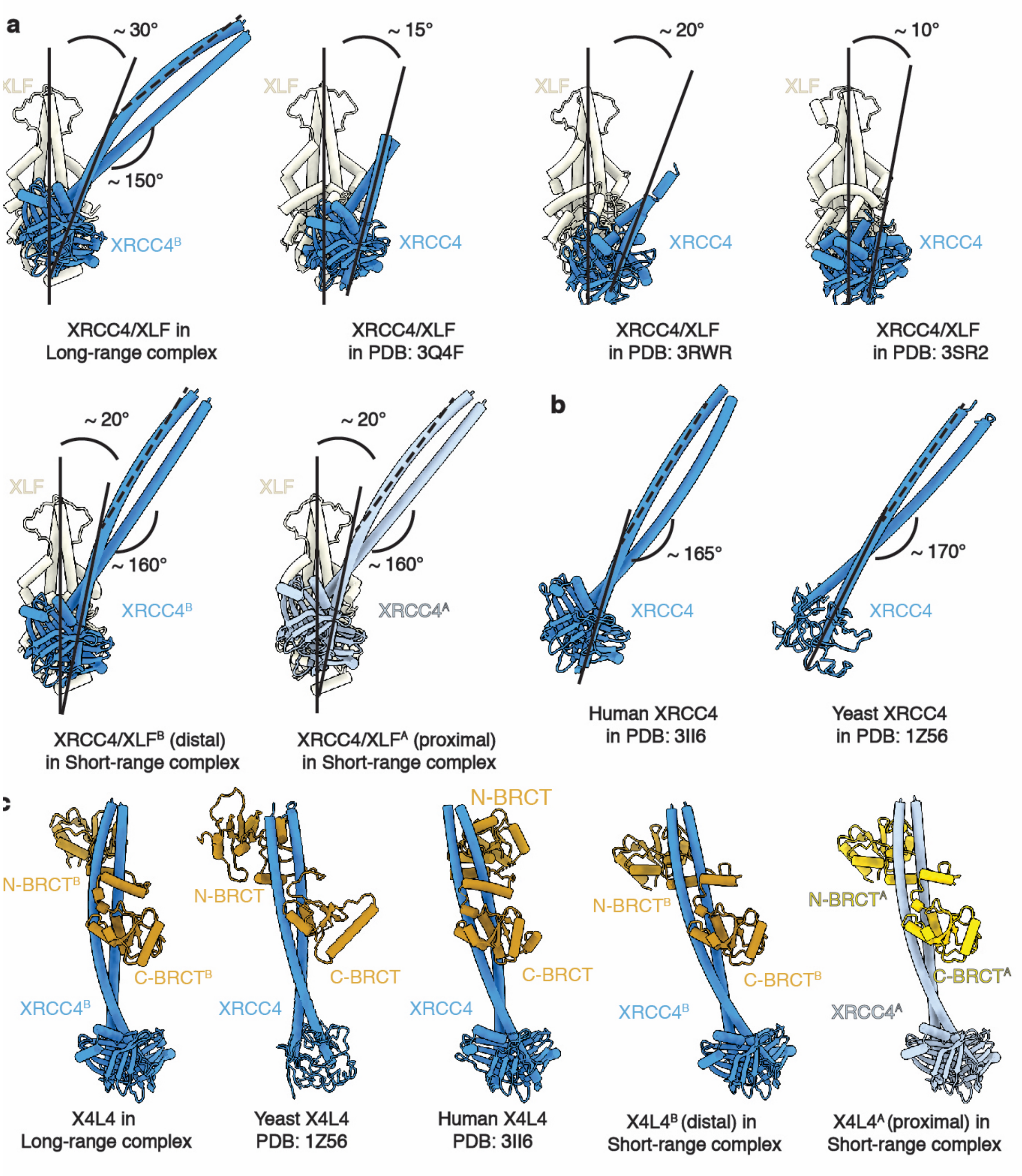


Figure 3.5: (Caption next page.)

Figure 3.5: (Previous page.) **Comparing the structure of LigIV-XRCC4-XLF scaffold among the LR synaptic complex, the SR synaptic complex, and previously published models. a,** Structure of XRCC4-XLF from the LR complex is shown in comparison with XRCC4-XLF filamentous repeat crystal structures and the ones from the SR complex (both copies). The XLF dimer is used to align all of the models shown here. Solid lines are aligned with the coiled-coils (CC) of XLF (vertical) and XRCC4 (tilted), and the angles in between are shown, respectively. Dashed lines are aligned with the C-terminal half of CC of XRCC4 when full helices are present, and the bending angles are shown as well. **b,** XRCC4 in the crystal structure of human and yeast LigIV-XRCC4 complex(*68*, *155*) are shown after aligning with XRCC4 in the LR complex shown in **a**. The bending of XRCC4 CC is more similar to the one in the SR than in the LR complex. **c,** Structure of LigIV-XRCC4 complex from the LR complex is shown in comparison with human and yeast LigIV-XRCC4 crystal structures and ones from the SR complex (both copies). Color codes for XLF, XRCC4, and LigIV BRCT domains are the same as in Figure 3.3

Three novel inter-subunit interactions were revealed within the LR complex. First, there is an interaction between conserved surfaces on N-BRCT and the Ku core region (Figure 3.3d, Figure 3.6a), predicted by biochemical pull-down experiments(*156*). Secondly, there is an extra loop within a conserved, negatively-charged groove formed by the DNA-PKcs FAT domain(*61*) (Figure 3.6c) that we attribute to the conserved, positively-charged region within XRCC4 C-terminal tails. These interactions likely stabilize contacts between the LigIV-XRCC4-XLF-XRCC4-LigIV scaffold and the DNA-PK complexes and guide XRCC4 C-terminal phosphorylation sites(*37*) to the DNA-PKcs active site. Notably, XRCC4 truncation mutation, R275X (X: stop codon), associated with both prenatal and postnatal growth failure and leukopenia(*2*), and identified in the cancer mutation database(*157*), is likely to disrupt the XRCC4-DNA-PKcs interaction and XRCC4 phosphorylation. Thirdly, we identified two conserved loops in the DNA-PKcs M-HEAT domain at the DNA-PK dimer interface (Figure 3.3e) with loop 2569-2585 interacting with the highly conserved YRPD motif(*158*) (Figure 3.6b). Both loops, together with the YRPD motif, likely serve as a molecular sensor to detect the close juxtaposition of the two DNA-PK complexes.

3.4 *In Trans* DNA-PKcs autophosphorylation

While similar overall, superimposition of DNA-PK within the LR complex with DNA-PK alone(*62*, *83*) revealed significant differences (Figure 3.7a). In the LR complex, the DNA-PKcs head module, including the FAT and the kinase domains, is rotated by ∼10° to further expose the catalytic center, a conformation that aligns better with the recent structure determined by Chen et al(*83*). The Ku core region, together with the associated DNA-PKcs N-HEAT domain in the LR complex, is also shifted by ∼4 Å, threading the DNA end further into the cradle of DNA-PKcs by ∼2 bp (Figure 3.7a,b). Moreover, a 32-amino acid DNA end-blocking (DEB) helix (residues 2736-2767) that spans the large empty space cradled by the DNA-PKcs HEAT repeats is stabilized. This helix, together with the linker loop connecting it and the DEB-appended (DEB-A) helix (residues 2724-2730), makes direct contact with the DNA end, disrupting end base-pairing (Figure 3.7b). Interestingly, the DEB helix is flanked by the unstructured ABCDE auto-phosphorylation cluster(*60*) and the conserved YRPD motif, suggesting that it coordinates DNA-PKcs dimer interactions and autophosphorylation of the ABCDE cluster (Figure 3.6d).

We compared the active sites of DNA-PK alone and within the LR complex with those of mTOR(*153*) and CDK2(*159*) (Figure 3.7c). The PIKK regulatory domain (PRD), which negatively regulates kinase activity by occupying the putative substrate-binding groove(*61*, *153*), is structured in both structures of DNA-PK alone(*62*, *83*), but mostly disordered in intrinsically active mTOR kinase(*153*), and completely disordered in the LR complex. Furthermore, the orientation between the two DNA-PKcs molecules differs substantially from the self-inhibitory conformations of the ATM and the ATR-ATRIP dimer structures(*160*, *161*) (Figure 3.8). Thus, DNA-PKcs in the LR complex is likely to be active. Intriguingly, both the ABCDE and the PQR clusters of one DNA-PKcs molecule are closer to the catalytic center of the other one (Figure 3.7d), favoring *in trans* autophosphorylation(*60*).

DNA-PKcs autophosphorylation has been hypothesized to destabilize binding to DNA ends, triggering DNA-PKcs dissociation from repair loci(*56*). In our model, the ABCDE and PQR clusters are in close proximity to other phosphorylation sites at the DNA-PKcs dimer interface, suggesting that autophosphorylation could function as a substantial and coordinated electrostatic switch (Figure 3.7d). With increased negative charge due to phosphorylation, the ABCDE cluster could compete with DNA for the DNA binding groove near the DEB helix (Figure 3.7b, Figure 3.6d), triggering the dissociation of DNA-PKcs from the DNA end and the transition from the LR to the SR synaptic state. In addition, both the ABCDE (T2609, S2612, T2620, S2624, T2638, and T2647) and the PQR (S2023, S2029, S2041, S2053, and S2056) phosphorylation sites are located within disordered loops of DNA-PKcs 2606-2720 and 1993-2084, respectively (Figure 3.7b,d). The kinase active center from the opposite side cannot reach most of the ABCDE sites unless the YRPD-Interaction (YRPD-I) loop (residue 2586-2604) is peeled off from the YRPD motif (Figure 3.7d). In turn, this conformational change potentially disrupts the DNA-PKcs-DNA-PKcs dimerization interface through loop 2569-2585 (Figure 3.3d). Similarly, some PQR sites are located too far from the trans kinase active center. PQR autophosphorylation induced changes could have a direct impact on the Ku80 CTR-DNA-PKcs interface at the bottom of the cradle (Figure 3.7d), potentially inducing the domain-swap of Ku80 observed by Chaplin et al.(*62*).

3.5 Negative staining characterization of SR complex

Figure 3.9 shows the selected negative staining and Cryo-EM results during the optimization process of the SR complex assembly protocol when DNA-PKcs is excluded. Although we were able to observe novel 2D class averages using both blunt end DNA substrate and substrate with 4nt complementary overhang (Figure 3.9b,f), neither of the corresponding Cryo-EM dataset gives high resolution reconstructions, especially the newly observed density that corresponds to LigIV catalytic domain (Figure 3.9d,h). Thus, a novel substrate design with one free complementary ssDNA that bridges a 5’ and another 3’ overhang was designed to introduce one nick in the center of the DNA, in order to ensure recruitment of single copy LigIV catalytic domain and improve sample homogeneity (Figure 3.9i-l). As expected, the rationale design of the DNA substrate improved sample quality and allow us to obtain higher resolution reconstructions with large Cryo-EM datasets.

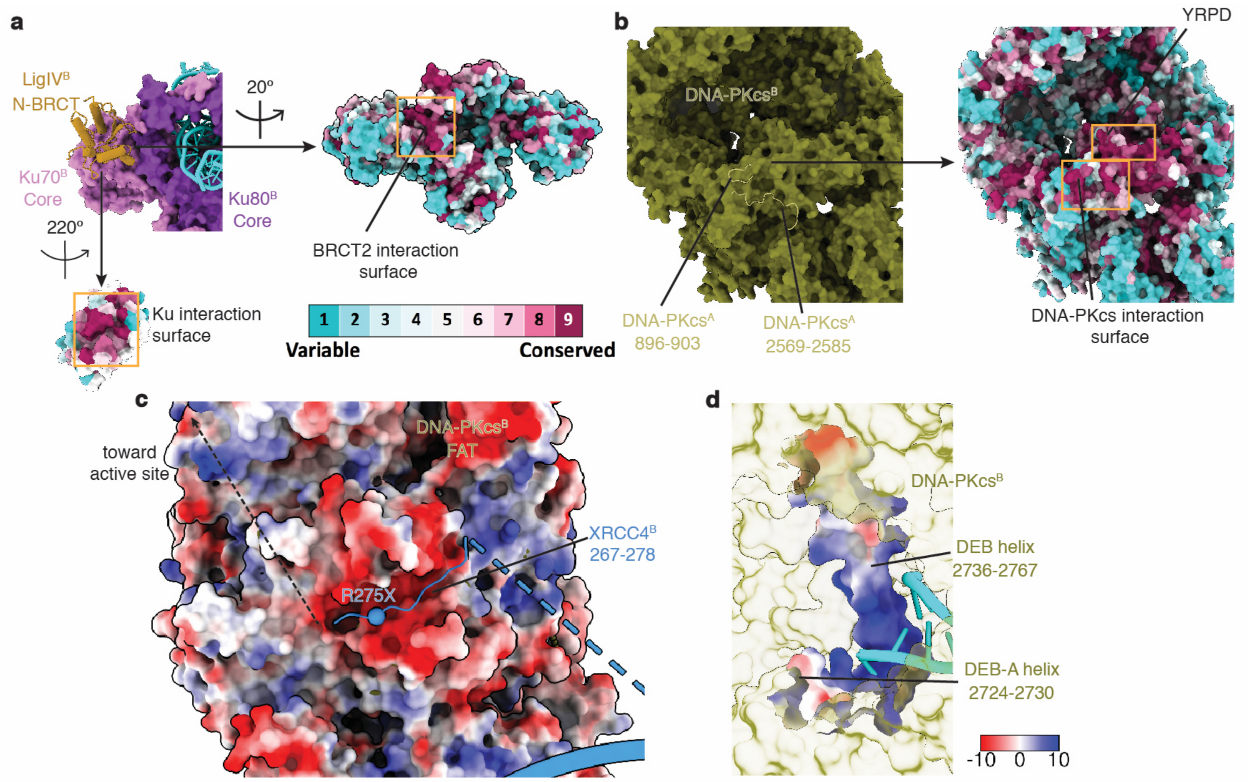


Figure 3.6: **Surface electrostatic potential and conservation of different areas in the LR synaptic complex.** **a,** Close-up view of the interaction surface between LigIV N-BRCT domain and Ku70 vWA domains colored by sequence conservation. **b,** Close-up view of the DNA-PKcs-DNA-PKcs interaction surface colored by sequence conservation. **c,** Surface electrostatic potential view of DNA-PKcs near its FAT domain, showing its negatively charged interface between XRCC4 C-terminal region (ribbon). The approximate path of the XRCC4 C-terminal peptide containing multiple phosphorylation sites is depicted. The sphere depicts the location of a cancer-associated truncation mutation that occurs at the interface. **d,** Surface electrostatic potential view of DNA-PKcs DEB and DEB-A helix. The DNA-interaction surface is positively charged. When models are not colored by either surface electrostatic potential or sequence conservation, the color codes are the same as in Figure 3.3. We cannot rule out the unlikely possibility that the stabilization of the DEB helix is due to the presence of a 3' overhang that existed in our DNA substrate design (Figure 3.1g).

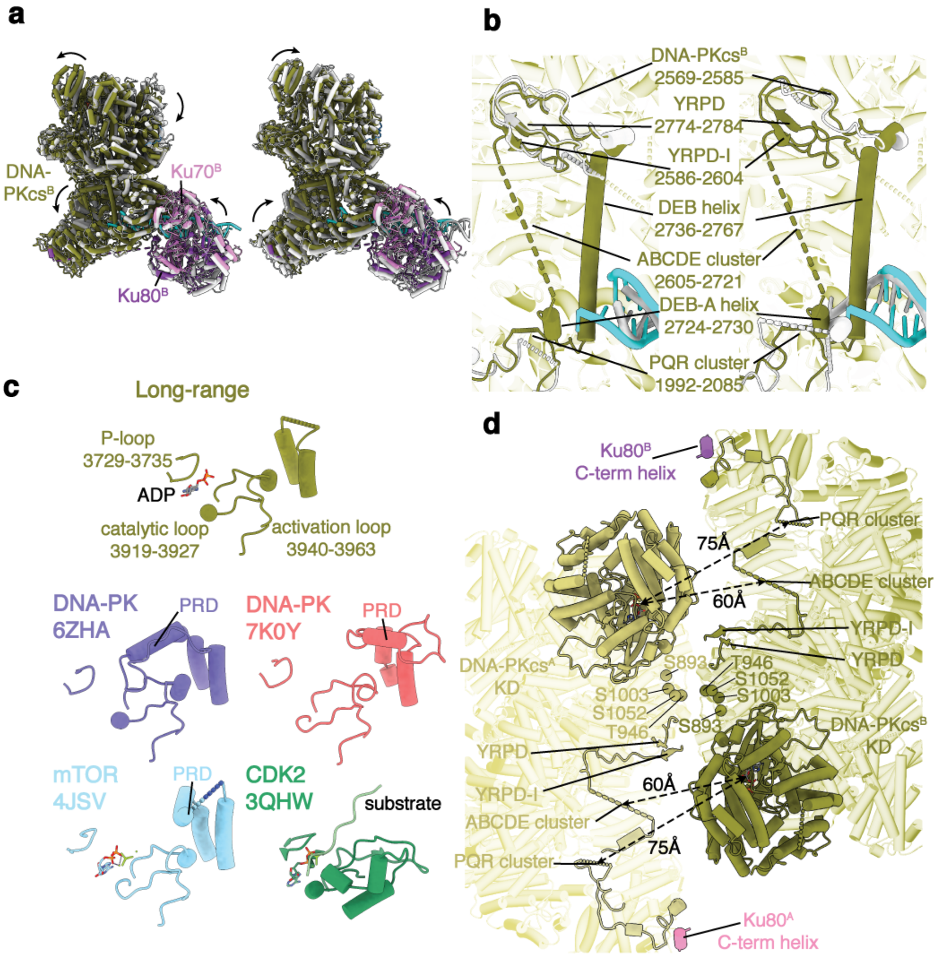


Figure 3.7: **Activation mechanism of DNA-PKcs in the LR synaptic complex.** **a,** Superimposition of the DNA-PK complex in the apo states (left, Protein Data Bank (PDB) 6ZHA(*62*), ; right, PDB 7K0Y(*83*)) with active DNA-PK derived from the LR complex (coloured). Transitions from the stand-alone states (grey) to the activated state are shown as curved arrows, indicating conformational changes induced by its assembly into the LR complex. **b,** Stabilization of a DEB helix within DNA-PKcs. The superimpositions of the loops that mediate the DNA-PKcs–DNA-PKcs interaction, as well as the DNA end, are highlighted. **c,** Comparison of the kinase active site between the LR complex (olive), apo DNA-PK (purple(*62*) and red(*83*)), mTOR (cyan(*153*)) and CDK2 (green(*159*)). **d,** Structural organization of the active sites of the DNA-PKcs kinase relative to different phosphorylation sites in the ABCDE and PQR cluster, as well as near the DNA-PK–DNA-PK and DNA-PKcs-Ku80 C-terminal helix interface. KD, kinase domain; YRPD-I, YRPD-interaction loop. ABCDE cluster, T2609, S2612, T2620, S2624, T2638 and T2647 phosphorylation sites; PQR cluster, S2023, S2029, S2041, S2053 and S2056 phosphorylation sites.

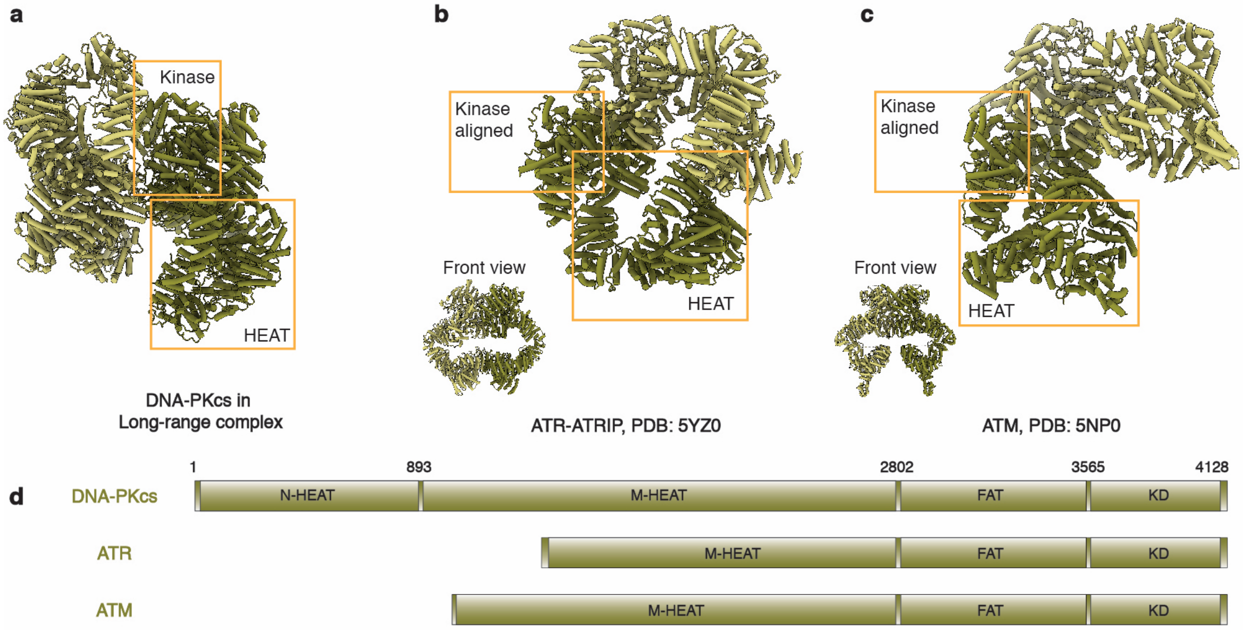


Figure 3.8: **Comparing the dimerization of DNA-PKcs in the LR synaptic complex with other PIKK family dimers. a,** Structure of the two DNA-PKcs in the LR complex. The Kinase domain (KD) is aligned with the homologous domains in **b** and **c** as an anchor point. **b, c,** Dimer of ATR-ATRIP (**b**) and ATM (**c**) showing aligned KD and corresponding N-terminus HEAT regions in the aligned copy. The symmetric-look front views are shown at the bottom left corner. Each protomer of ATR-ATRIP and ATM and colored the same as the corresponding DNA-PKcs protomer, in olive (the aligned copy) and dark khaki (the other copy). **d,** Domain organization of DNA-PKcs compared with ATR and ATM. Abbreviations are the same as in Figure 3.3.

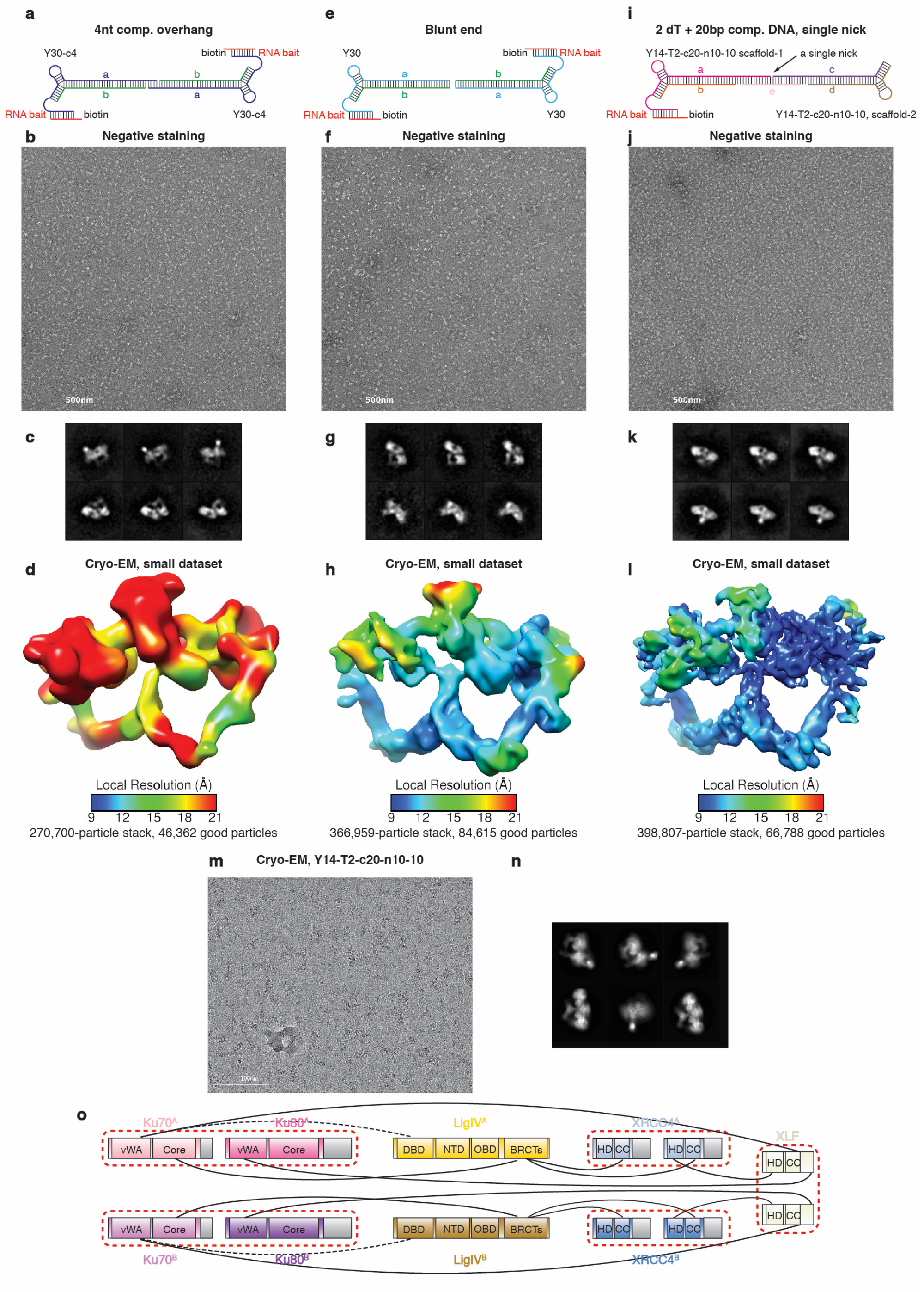


Figure 3.9: (Caption next page.)

Figure 3.9: (Previous page.) **Optimization of the SR synaptic complex assembly with various DNA substrates.** **a,** Schematic showing the Y30-c4 DNA substrate with 4nt 3' complementary overhang. The complex was assembled prior to RNAse-H elution. **b,** A representative negative staining raw micrograph of the complex assembled as described in **a**. The raw micrograph is representative of 23 micrographs. **c,** Representative two-dimensional class averages of the complex assembled as described in **a**. **d,** Cyro-EM map reconstructed from a small dataset using Y30-c4 DNA substrate shown in **a**. The map is colored by local resolution estimation. **e–h,** Same procedure as **a–d,** showing the complex assembly with the Y30 blunt end substrate. Stably assembled SR complexes on DNA substrates with either complementary or blunt ends indicates that these complexes are stable in the absence of any bridging effect from DNA. The raw micrograph is representative of 24 micrographs. **i-l,** Same procedure as **a–d,** showing the complex assembly with the Y14-T2-c20-n10-10 substrate, with one central single non-ligatable nick. The raw micrograph is representative of 24 micrographs. Strand e is added at last after mixing the two halves together with NHEJ factors. **m,** A representative cryo-EM raw micrograph (out of 32,723 total images) of the SR complex assembled with the Y14-T2-c20-n10-10 DNA substrate shown in **i**. **n,** Representative 2D class averages of particles (175,866 in total) contributing to the final reconstruction of the SR complex. All representative micrographs in **b, f, j, m** are from at least three biologically replicated experiments. **o,** Protein-protein interaction network between the components of the SR complex. Major unmodeled regions are shown in gray. Well-documented hetero- or homo-dimers are grouped by red dashed circles. Alternative protein-protein interactions are depicted by black dashed lines.

3.6 Details of Cryo-EM Data Collection, Processing and Model Building of SR complex

3.6.1 Sample Preparation

As described above, the oligos used to assemble SR complexes are listed in Table 3.2. The SR complexes were assembled using a similar protocol as described, with adjusted protein concentrations and additional incubation of combined half-complexes with single-strand complement DNA to assemble the holo-complex when necessary. The SR complex negative staining samples contained $\sim$700 nM Ku70/80, $\sim$1 µM LigIV-XRCC4 and XLF, as well as $\sim$350 nM of each half-DNA substrate. The Cryo-EM samples use half as much as the negative staining sample.

Table 3.2: **Oligonucleotides used for the assembly of the Short-range synaptic complex**

Construct	Strand	Sequence (5’-3’)
Y14-T2-c20-n10-10	a	CCGGGCCAGCTTTCCCAGCTTATATACTAAGAACTTCTGACTGTT/3ddC/
	b	GTTCTTAGTATATAGGCCCGG
	c	/5Phos/CGTTTCATTGTTGTTCTTAGTATATAGGCCCGG
	d	ACCTCCCACTATTTTTCCGGGCCAGCTTTCCCAGCTTATATACTAAGAAC
	e	CAATGAAACGGAACAGTCAG
	RNA bait	UAGUGGGAGGU/3BiotinTEG/
Y30-c4	a	ACCTCCCACTATTTTTCCGGGCAAGCTCGATCCCCGAGCTTCTAAGAACTCTGATGTCAGTAGATTACACGAT/3ddC/
	b	/5Phos/GTGTAATCTACTGACATCAGAGTTCTTAGATGCCCGG
	RNA bait	UAGUGGGAGGU/3BiotinTEG/
Y30	a	ACCTCCCACTATTTTTCCGGGCAAGCTCGATCCCCGAGCTTCTAAGAACTCTGATGTCAGTAGATTACA/3ddC/
	b	/5phos/GTGTAATCTACTGACATCAGAGTTCTTAGATGCCCGG
	RNA bait	UAGUGGGAGGU/3BiotinTEG/
Hairpin30-cy5	a	ACCTCCCACTATTTTTCCGGGCAAGCTCGATCCCCGAGCTTCTAAGAACTCTGATGTCAGCAGTCTACGTGTTC
	b	/5phos/ACGTAGACTGCTGACATCAGAGTTCTTAGATGCCCGG
	c	/5phos/ACGTAGACTGCTGACATCAGAGTTCTTAGATGCCCGGTTT/iCy5/TTCCGGGCAAGCTCGATCCCCGAGCTTCTAAGAACTCTGATGTCAGCAGTCTACGTGAAC
	RNA bait	UAGUGGGAGGU/3BiotinTEG/

3.6.2 Data collection and processing

Figure 3.10 shows the data processing procedure of the SR complex. Two datasets were collected (32,723 movie series in total) for the SR complex at The Pacific Northwest Center for Cryo-EM (PNCC) using Serial EM(*162*). Titan Krios-3 TEM (Thermo Fisher) operating at 300 kV was used with K3 direct electron detector (Gatan) at ×30000 magnification (0.83 Å/pixel, counting mode, 0.415 Å/pixel, super-resolution mode). The defocus range was -1.5 to -4 μm. 50-frame exposures were taken during 2.13s total exposure time, using a dose rate of 15 e^{-} per pixel per second, corresponding to a total dose of $\sim$46e^{-}/$Å^2$ per movie series.

A similar protocol was used to generate the SR complex initial model. 60,194 particles were extracted and stacked with a box size of 128 × 128 pixels. After that, similar protocol was used as well to process the SR complex Cryo-EM dataset. However, due to the limited resolution obtained, even after the final steps, all of the particles were binned by 2 (1.66Å/pxl) during the whole process for better alignment without further un-binning. 1,766,936 particles were picked for direct 3D classification, and 175,866 particles were selected and refined to 8.4Å after CTF refinement and polishing. This map was used as the complex map for deposition. Next, the complex was divided into three bodies for further focused refinement: 1. A C2 body similar to the LR complex low-resolution region (XRCC4-LigIV BRCT-XLF); 2. A C1 body containing distal Ku70/80, LigIV catalytic domain and their DNA footprints; 3. A C1 body containing proximal Ku70/80 and its DNA footprints. Signal subtraction and further classification and refinement improved the resolution of all three bodies. Specifically, body 1 was refined to 7.1 Å (45,934 particles), body 2 6.8 Å (57,170 particles) and body 3 6.9 Å (54,793 particles).

3.6.3 Model building

Subunits from the LR complex (no DNA) were used to flexibly fit in the density of the SR complex model. The catalytic domain model of LigIV (PDB: 6BKF(*69*)) was docked into the corresponding map region as well. Next, ISOLDE was used to flexibly refine the models with adaptive distance restraints. An intact B-form DNA was mutated into the substrate sequence and then used to flexibly fit for the entire DNA density using ISOLDE. After that, the corresponding nicks and nucleotide deletions were introduced to match the dsDNA exactly with the designed substrate. The position of the nick aligned well with the LigIV catalytic domain binding site, and the flexible single strand -TT- linkers between Ku70/80 and LigIV catalytic domain were also aligned with the worst density region that corresponds to dsDNA.

3.7 Architecture of the SR synaptic complex

Although the LR complex structure reveals how canonical NHEJ factors initially tether DNA ends, they are not oriented for joining, and the LigIV catalytic domains are not visible, presumably due to their flexibility(*41*). In the absence of DNA-PKcs, complexes were assembled on DNA substrates with either complementary or blunt ends (Figure 3.9a-h and Chapter 2). After substrate optimization for complex stability (Chapter 2, Figures 3.9i-l and Figures 3.11) and by collecting a large dataset, we were able to improve the resolution to 8.4 Å for the overall reconstruction and 6.8-7.8 Å resolution for different bodies after focused refinements, clearly revealing secondary structures of both NHEJ proteins and the DNA substrates (Figure 3.11a, Figure 3.9m-n, Figure 3.10). With this density map, we directly docked in the subunit models from the LR complex as rigid bodies, revealing a very similar overall architecture to the LR complex, except for the two DNA-PKcs molecules. In accord with smFRET studies using Xenopus egg extracts(*92*, *102*), Ku and LigIV-XRCC4 subcomplexes reside on opposite sides of the complex, with one XLF dimer bridging the two halves (Figure 3.11b). A perfectly aligned duplex DNA is visible between two Ku rings located at opposite ends of the complex, suggesting the captured species corresponds to the SR complex described previously(*31*). In our SR complex, a single LigIV catalytic domain is visible, engaging a single non-ligatable nick in the middle of the complex (Figures 3.9i).

Interactions observed in the LR complex, including XLF-XRCC4, XLF-Ku, LigIV-XRCC4, as well as the new LigIV N-BRCT-Ku interface, also occur in the SR complex and are mostly unchanged (Figures 3.4e,f, Figures 3.5c, Figure 3.9o). Interestingly, the angle between XLF and XRCC4 CCs on either end of the complex decreases to $\sim$20°, partially restoring the fully relaxed states observed in several crystal structures31-33 (Figure 3.5a). Concurrently, the gradual bending of the long XRCC4 CC is restored to 160° for the two protomers, much closer to the ground state observed in the structures of LigIV-XRCC4 alone (Figure 3.5b)(*68*).

New interactions in the SR complex include an interaction surface between the LigIV DNA binding domain (DBD) and the Ku70 vWA domain within the Ku ring on the opposite side of the complex that partially overlaps with the Ku70 interface with DNA-PKcs (Figure 3.11c, Figure 3.12a). This explains why grouping the LigIV catalytic domain with its interacting Ku improves the resolution of both parts (Figure 3.10). It also demonstrates that, when engaging a DNA nick, a single LigIV molecule contacts both Ku molecules via independent interactions with the LigIV N-BRCT and DBD, providing insights as to how Ku stimulates DNA joining by LigIV(*24*). The SR complex is further stabilized by interactions between XLF and the two Ku70 molecules (Figure 3.11d). This is consistent with the reduced ability of XLF with amino acid substitutions (L174A, R178A, and L179A) in these interfaces to stimulate NHEJ(*64*) and may provide insights into the D166R cancer mutation(*157*) associated with immunodeficiency, microcephaly, and radiosensitivity(*2*). Thus, amino acid changes within a conserved patch near the C-terminal tip of XLF

CC likely exert their effects by disrupting XLF-Ku70 interactions within the SR complex (Figure 3.12b).

3.8 Tandem ligation by two LigIV molecules

Even with substrates containing two nicks, only a single LigIV catalytic domain was visible within the SR complex (Figure 3.9). In addition, there is a severe clash of the LigIV catalytic domains as well as other discrepancies when superimposing the SR complex with itself (Figure 3.12c). The DNA nick identified by rigid-body docking of the LigIV-DNA crystal structure(*69*) is clearly off-center relative to the Ku rings on either side, rendering the Ku with which the LigIV DBD interacts distal to and the other Ku proximal to the nick (Figure 3.12d). While it has been suggested that NHEJ may only join one strand of a DSB(*112*), the positioning of two LigIV molecules with single turnover activity(*111*, *165*) within the SR complex suggests that they are in the correct orientation to sequentially join both strands of the DSB (Figure 3.12c).

To validate predictions based on the LR and SR complex structures, *in vitro* ligation assays were performed under conditions used to assemble both synaptic complexes (Chapter 2, Figure 3.13a). As expected, pre-assembly of DNA-PK prevents ligation when the other NHEJ factors are added in the absence of ATP, presumably because the transition from the LR to the SR complex is blocked. In contrast, ligation by the SR complex is ATP-independent because LigIV is pre-adenylated(*111*, *165*) (Figures 3.13b,c). Furthermore, joining of both strands occurs more frequently than joining of a single strand even in the absence of ATP and lower amounts of LigIV-XRCC4 (Figures 3.13b,c), supporting the functional relevance of the two LigIV molecules within the synaptic complexes. Notably, the flexibility of the LigIV catalytic domain(*41*) and the position of the aligned DNAs within the SR complex should allow end processing enzymes access since a significant fraction of DSBs repaired by NHEJ require processing prior to ligation(*164*, *166–168*).

3.8.1 Conclusion

We have captured several essential states during NHEJ by high-resolution cryo-EM, allowing us to propose an almost complete reaction cycle of this important DSB repair pathway. As described previously, DNA-PKcs is initially recruited by Ku to protect the DNA end from degradation and sequester the DSB for repair by NHEJ(*61*, *62*, *82*–*84*). In the LR synaptic state, two DNA-PK complexes engage through the formation of an intricate protein-protein interaction network involving the core NHEJ factors to initially tether the DNA ends. *In trans* autophosphorylation of DNA-PKcs, which triggers their simultaneous dissociation from the DSB ends due to protein-protein or protein-DNA repulsions, serves as a "checkpoint" to ensure that there are two free DNA ends in close proximity. In the SR complex, the DSB ends are aligned precisely, allowing the sealing of both strands by the two LigIV molecules on either side of the complex in a tandem manner. Our findings reveal unprecedented insights into DSB repair with implications for modulating the cellular response to radiation and chemotherapy and offer additional opportunities for targeting this and other important PIKK-dependent pathways for cancer therapy.

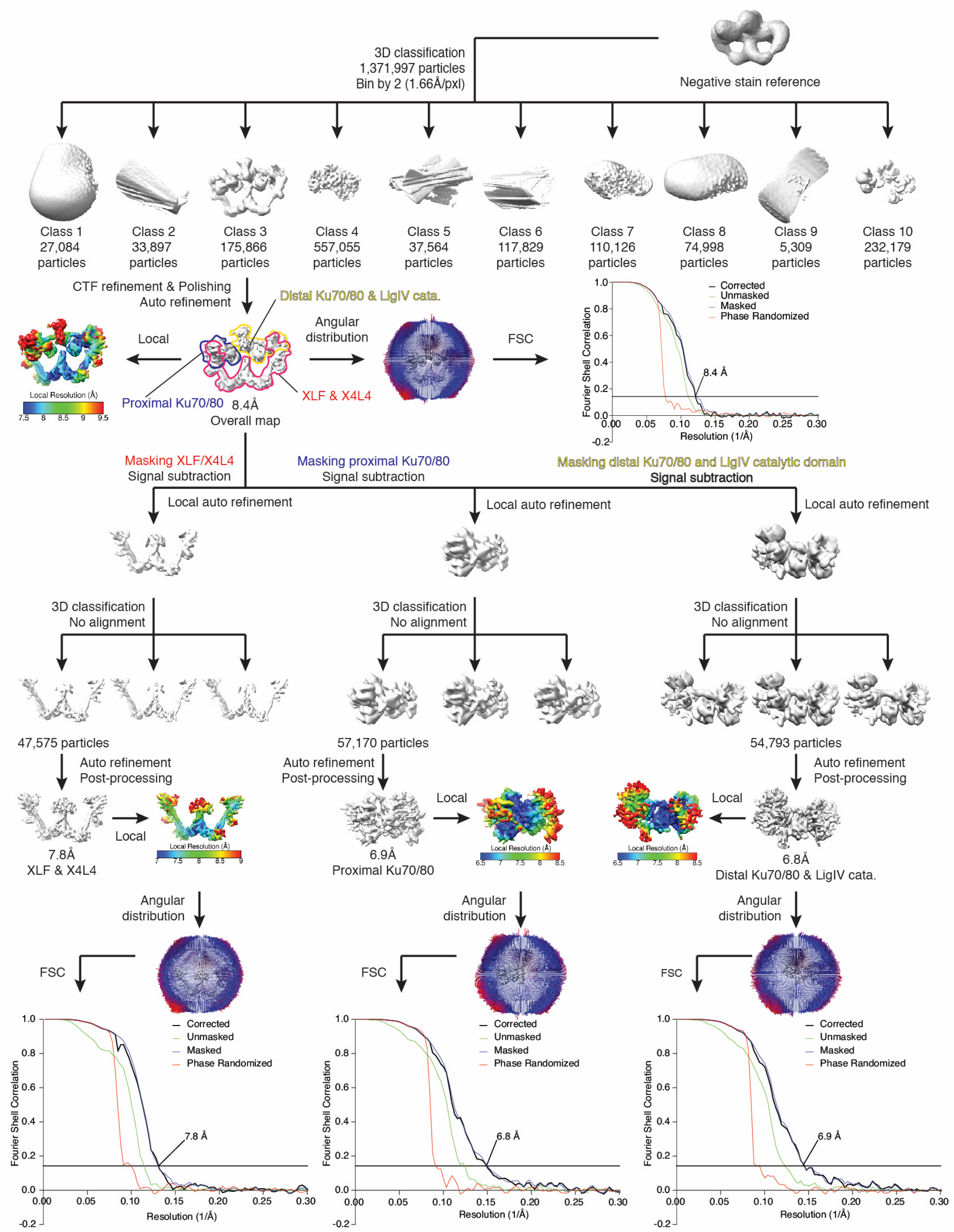


Figure 3.10: **Data-processing scheme of the SR synaptic complex sample.** Flow chart of the cryo-EM data processing procedure. The gold-standard FSC curves (0.143 cutoff) show the final resolution of the holo-complex and each body.

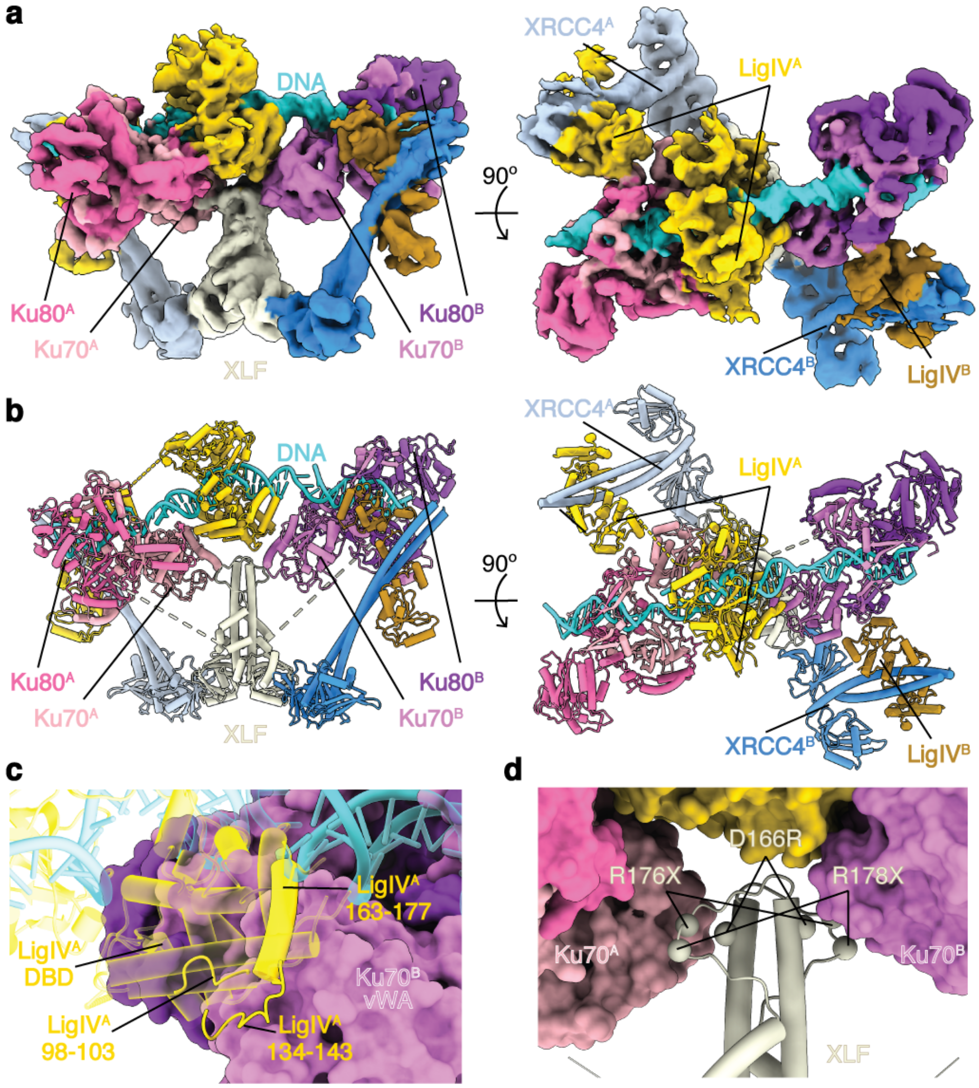


Figure 3.11: **Cryo-EM structure of the SR synaptic complex. a,** Front (left) and top (right) views of the cryo-EM composite map of the SR complex. **b,** Corresponding views of the structural model of the SR complex. **c,** Close-up view showing the interface between the LigIV DBD (ribbon) and Ku70 vWA (surface) domains. Regions of the LigIV DBD domain involved in the interaction are depicted. **d,** Close-up view showing the interface between the XLF CC (ribbon) and the two Ku70 vWA (surface) domains. Spheres depict the locations of missense or truncation mutation residues from cancer patients(*2*) that occur at the interface, although D176X and D178X truncations are also expected to impact XLF-Ku80 interaction due to the lack of XLF C-terminal KBM(*92*, *102*).

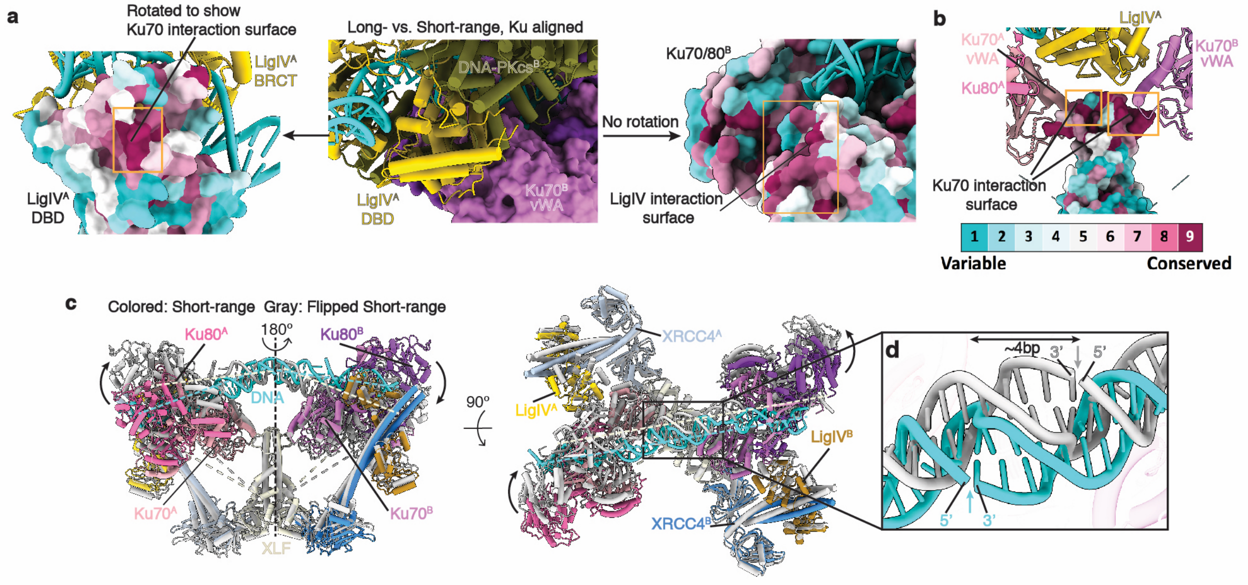


Figure 3.12: **Surface conservation of different areas in the SR synaptic complex. a,** Magnified view of the interaction surface between the LigIV DBD and Ku70 vWA domain coloured by sequence conservation. DNA-PKcs clashes with LigIV DBD when Ku is aligned between the LR and SR complex. **b,** Magnified view of the coiled-coil domain of XLF at its C-terminal tip coloured by sequence conservation. When models are not coloured by sequence conservation, the colour codes are the same as in Figure 3.11. **c,** Superimposition of two asymmetric SR complexes after a 180° flip shown in front (top) and top (bottom) views. The XLF homodimer is used for aligning the two conformers. LigIV catalytic domains are hidden for clarity purposes. The transition from the apo state to the flipped state indicates potential conformational changes during the tandem ligation. Paths of DNA are also highlighted by dashed lines. **d,** Magnified view showing the relative positions of the two off-centred nicks between the two conformers. The two preferential nick positions are separated by approximately 4 bp. Consistently, dsDNA with 4-nt 3' overhang, a major end-processing product of the NHEJ nuclease Artemis(*163*), is reported to be a favoured substrate for NHEJ(*164*). Our model suggests that dsDNA with a 4-nt 3' overhang will experience minimum DNA translocation to accommodate the two ligation steps.

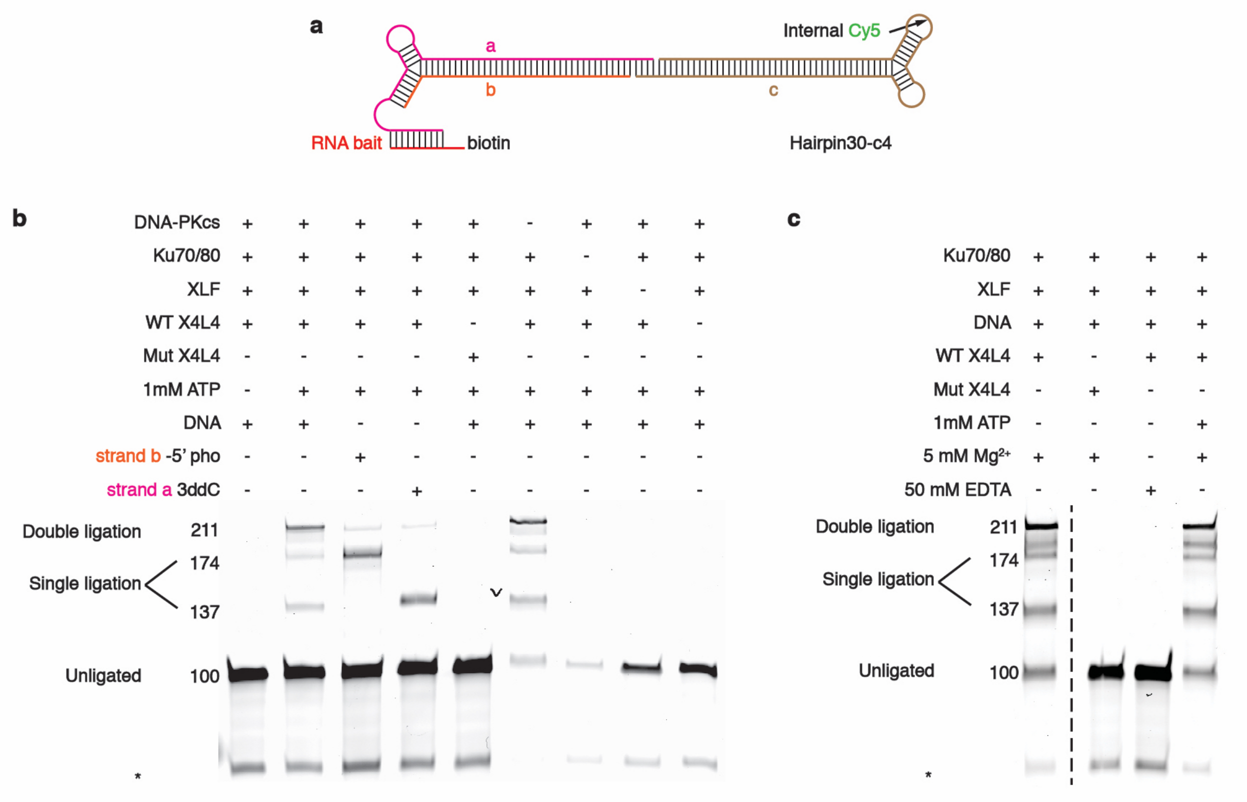


Figure 3.13: **Both LR and SR synaptic complexes are able to perform double ligation during NHEJ *in vitro*. a,** Substrate design for the ligation assay. An Internal Cy5 label is added to only the right half of the substrate to visualize the ligation products. A 4 nt 3' complementary overhang has been introduced on both sides of the substrate. **b,** Denaturing gel analysis of end-joining by the LR complex. 100nM of DNA, 200 nM of DNA-PKcs, and Ku70/80, 500nM of XLF, and 70nM of X4L4 were added, respectively. Asterisk indicates alternative secondary structure or impurity of the cy5-labeled oligo. Size of the DNA substrates and ligation products are labeled on the left (unit: bp). **c,** Denaturing gel analysis of end-joining by the SR complex. The final factor concentrations are the same as in **b**. **e, f,** Similar conditions for either of the gel have been replicated as biological replicates for two times.

CHAPTER 4

STRUCTURAL BASIS OF DNA-PKCS DISSOCIATION AND PAXX RECRUITMENT IN NHEJ

With the understanding of how core NHEJ factors (LigIV–XRCC4 and XLF) are recruited to bridge two DSB ends and undergo a Long-range (LR) to Short-range (SR) state transition triggered by *in trans* DNA-PKcs auto-phosphorylation, it remains unclear how phosphorylated DNA-PKcs dissociates from the LR complex and facilitates formation of SR complex. Furthermore, how appendix NHEJ scaffolding factors, such as PAXX(*113*), coordinates with the core NHEJ machinery during state transition is poorly understood. To shed light on these open questions, we next determined the Cryo-EM structures of three states during NHEJ: an apo-LR state with PAXX incorporated (PAXX-LR complex); an ATP induced LR state with PAXX (PAXX-ATP-LR complex), as well as a DNA-PKcs dimeric state that is pulled-down by dsDNA substrate from endogeniously expressed DNA-PKcs.

4.1 Negative staining characterization of PAXX-LR and PAXX-ATP-LR complex

To test whether PAXX stablizes the LR complex architecture, PAXX was added into the LR complex assembly protocol, and Figure 4.1 shows the selected negative staining results during the optimization of PAXX-LR assembly. As an appendix scaffolding factor of NHEJ, PAXX helped the stablilzation of the LR complex when added throughout the whole assembly process. While LR complex assembly with blunt end DNA only gives a small population of whole complex and substantical amount of free DNA-PK complexes (Figure 4.1A-C), addition of PAXX with blunt end DNA substrate produces the LR complexes much more efficiently. In addition, PAXX can still flexibly bridge two DNA-PK complexes into and intact complex without XLF and X4L4, although one copy of Ku could be flexible and well resolved in 2D class averages (Figure 4.1G-I). In conclusion, PAXX stablizes the LR complex structure obtained from various conditions.

We further added ATP into the LR complex with PAXX to obtain the PAXX-ATP-LR state. Figure 4.2 shows the selected negative staining and Cryo-EM results during the optimization process. Even with ATP addition that potentially introduces more intrinsic flexibility, PAXX still stablizes the complex and helped us to obtain stable complexes. Further more, PAXX addition results in some big complex 2D class averages that look like PAXX-ATP-LR complex, even when no XLF or XRCC4-LigIV were supplied (Figure 4.2G-I). Thanks to the stabilization effect of PAXX on complex assembly in our hand, the optimized protocol that used both Y30-T40-c8 and Y30 blunt end DNA substrates as described in the next section were used to make Cryo-EM samples and collect high-resolution data for model building.

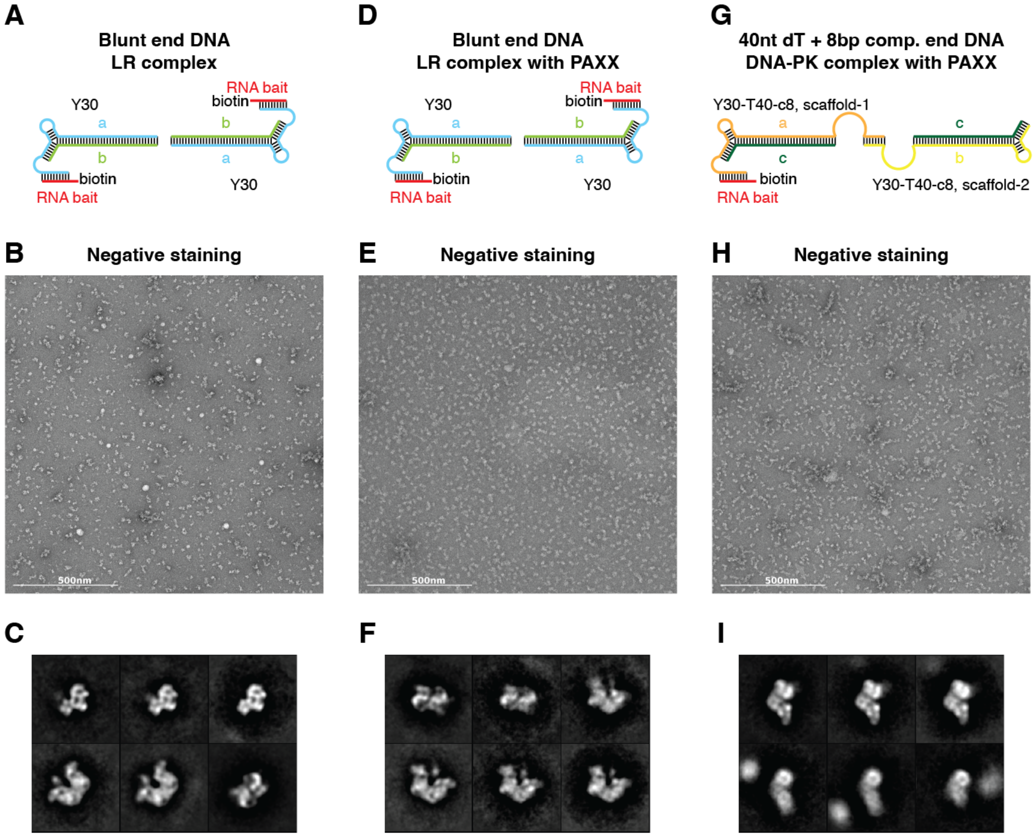


Figure 4.1: **Optimization of the PAXX-LR synaptic complex assembly with various DNA substrates.** **A-C,** Same as 3.1d-f, showing both DNA-PK complex and LR complex with no PAXX addition. **D–F,** Same procedure as **A–C,** showing the complex assembly with the blunt end substrate and PAXX addition. The raw micrograph is representative of 24 micrographs. In **F**, the two-dimensional class averages representing mostly the LR complex were obtained. **G–I,** Same procedure as **A–C** , showing the complex assembly using the Y30-T40-c8 substrate, but without adding XLF and LigIV–XRCC4. Flexible LR complex 2D class averages are observed. The raw micrograph is representative of 24 micrographs.

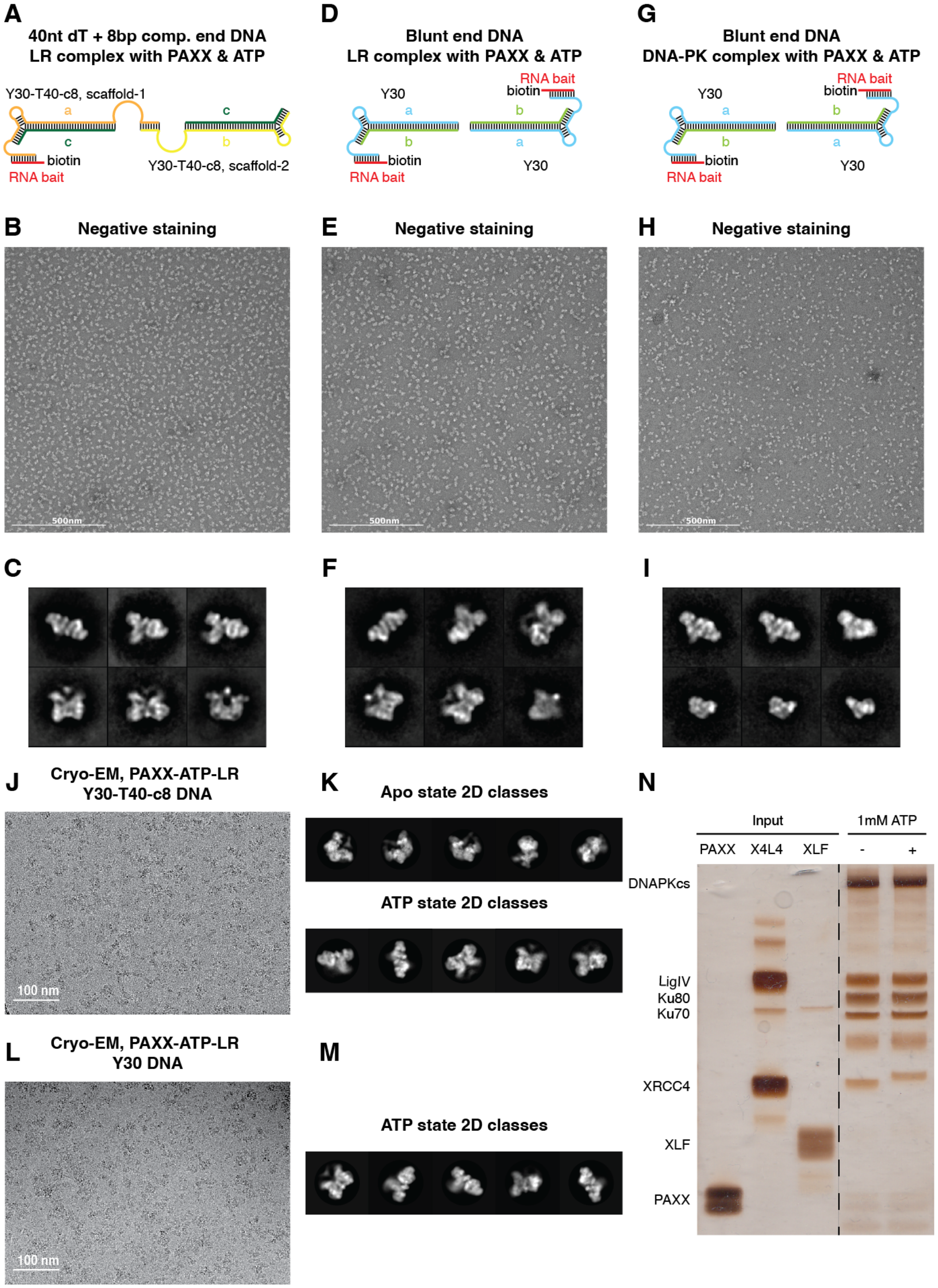


Figure 4.2: (Caption next page.)

Figure 4.2: (Previous page.) **Optimization of the PAXX-ATP-LR synaptic complex assembly with various DNA substrates. A,** Schematic showing the Y30-T40-c8 DNA substrate. PAXX was supplied during complex assembly, wash and RNase H elution. **B,** A representative negative-staining raw micrograph of the complex assembled as described in **A**. The raw micrograph is representative of 23 micrographs. **C,** Representative two-dimensional class averages of the complex assembled as described in **A**, showing characteristic view of only the both the LR-ATP complex and the apo LR complex. **D–F,** Same procedure as **A–C,** showing the complex assembly with the blunt end Y30 substrate. The raw micrograph is representative of 28 micrographs. In **F**, the two-dimensional class averages representing the characteristic view of only the ATP-LR complex were obtained. **G–I,** Same procedure as **A–C,** , showing the complex assembly using the blunt end Y30 substrate, but without adding XLF and LigIV–XRCC4. Scarce but existing LR complex 2D class averages are observed. The raw micrograph is representative of 23 micrographs. **J,** A representative cryo-EM raw micrograph (out of 11,006 micrographs in total) of the ATP-LR complex assembled with the Y30–T40–c8 DNA substrate shown in **A**. **K,** Representative two-dimensional class averages of particles (1,101,255 in total) contributing to the final reconstruction of the apo and ATP-LR complexes. **L-M,** Same procedure as **J–K,** , showing representative raw micrograph (out of 19,442 micrographs in total) and two-dimensional class averages of particles(2,684,531 in total) used to solve the ATP-LR complex assembled with the Y30 DNA substrate shown in **D**. **N,** Silver-stained SDS–PAGE (4–12% gradient, biologically replicated three times) showing some of the input purified subunits (PAXX, LigIV–XRCC4, and XLF) and the RNase H purified LR and ATP-LR complexes for cryo-EM data collection using Y30 DNA substrate shown in **D**. All representative micrographs in **B, E, H** are from at least three biologically replicated experiments.

4.2 Details of Cryo-EM data collection, processing and model building

4.2.1 Sample preparation

As described in Chapter 2, the oligos used to assemble LR complexes are listed in Table 4.1. Both the Y30-T40-c8 and Y30 DNA-RNA substrates were used to assemble PAXX-ATP-LR complexes. The overall protocol of complex assembly is similar as described in Chapter 3 with the following modifications: First, when Y30 DNA-RNA substrate was used, the steps of combining half-complexes and 37°C incubation were omitted since there is no half DNA-RNA substrates used as Y30-T40-c8. Second, PAXX was supplied together with LigIV-XRCC4 complex and XLF, followed by a 15-minute incubation at room temperature. During the wash and elution steps, 50nM

Table 4.1: **Oligonucleotides used for the assembly of the PAXX-ATP-LR complexes**

Construct	Strand	Sequence (5'-3')
Y30-T40-c8	a	ACCTCCCACTATTTTTCCGGGCAAGCTCGATCCCCGAGCTT CTAAGAACTCTGATGTCAGTAGATTACACTTTTTTTTTTTTT TTTTTTTTTTTTTTTTTTTTTTTTTTTCATTCGC
	b	CCGGGCAAGCTCGATCCCCGAGCTTCTAAGAACTCTGATG TCAGTAGATTACACTTTTTTTTTTTTTTTTTTTTTTTTTTTTT TTTTTTTTTTTGCGAATGA
	c	/5phos/GTGTAATCTACTGACATCAGAGTTCTTAGATGCCCGG
	RNA bait	UAGUGGGAGGU/3BiotinTEG/
Y30	a	ACCTCCCACTATTTTTCCGGGCAAGCTCGATCCCCGAGCTT CTAAGAACTCTGATGTCAGTAGATTACC
	b	/5phos/GTGTAATCTACTGACATCAGAGTTCTTAGATGCCCGG
	RNA bait	CGACAUCGCGUUU/3BiotinTEG/

PAXX was added throughout the process to achieve higher PAXX occupancy and further stabilize the complex. Third, to prepare the PAXX-ATP-LR complex, 1mM ATP (final concentration) was added to the sample after elution and incubated at room temperature for 30 minutes. Fourth, both the negative staining and Cryo-EM samples of the PAXX-ATP-LR complex were prepared with 0.01% NP-40 (final concentration). To protect the air-water interface of the Cro-EM sample, the NP-40 concentration was boosted up to 0.05% right before applying to the GO grids.

4.2.2 Data collection and processing

As described in Chapter 2, Negative staining and Cryo-EM data were collected and processed with a similar procedure. Two datasets were collected for the PAXX-ATP-LR complex at the Pacific Northwest Center for Cryo-EM (PNCC), one with Y30-T40-c8 substrate (pixel size 1.079Å, counting mode, super resolution 0.5395Å/pxl, 11,006 movies) and the other with Y30 substrate (pixel size 1.056Å, counting mode, super resolution 0.528Å/pxl, 19,442 movies). Titan Krios-3 TEM (Thermo Fisher) operating at 300 kV was used with K3 direct electron detector (Gatan) at

×30,000 magnification. For both datasets, the defocus range was set to -2 to -4 µm. 45-frame exposures were taken during 4s total exposure time, using a dose rate of 15 e^- per pixel per second, corresponding to a total dose of $\sim$66e^-/Å^2 per movie series.

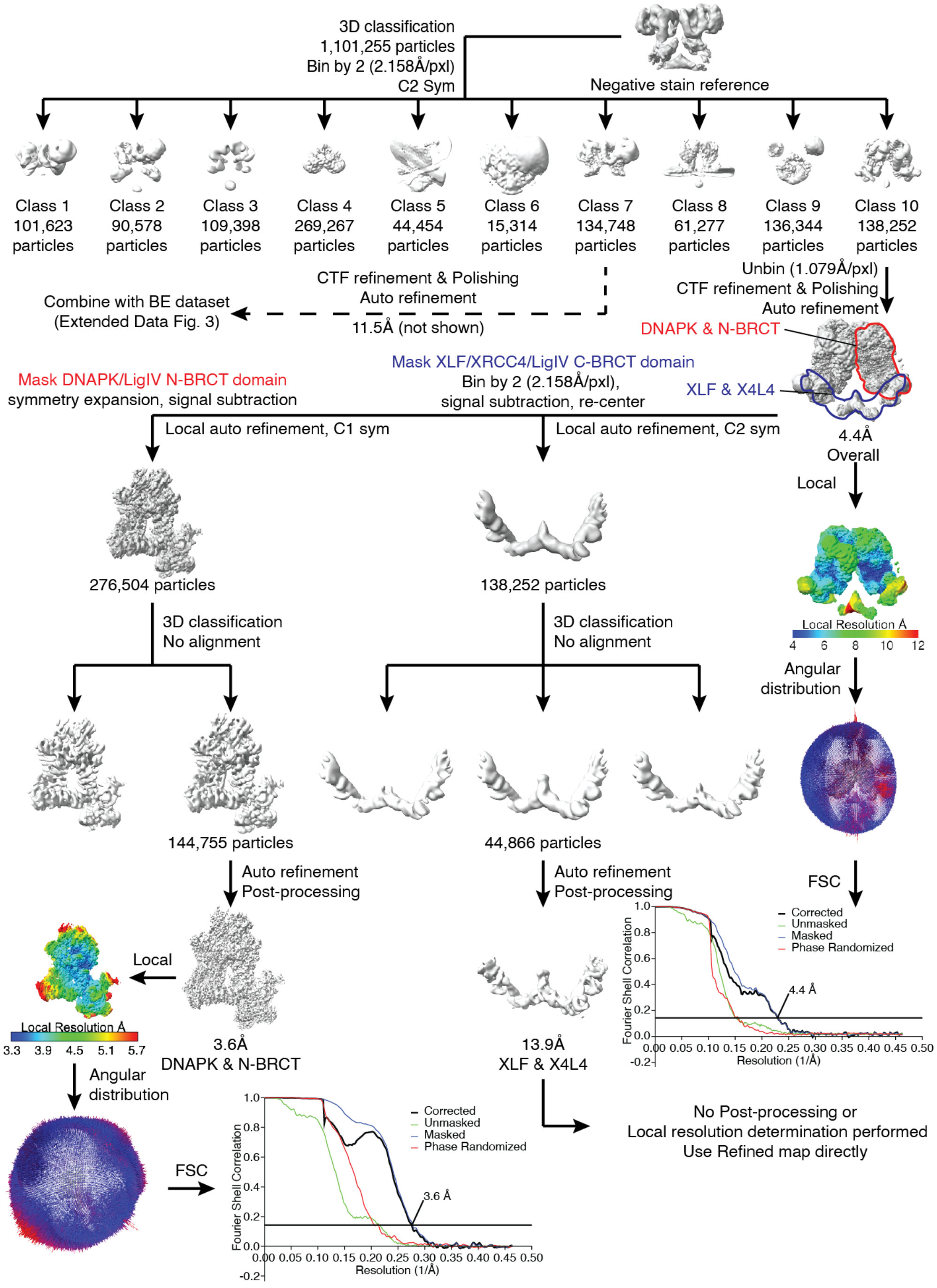


Figure 4.3: (Caption next page.)

Figure 4.3: (Previous page.) **Data-processing scheme of the PAXX-LR synaptic complex sample obtained from Y30-T40-c8 DNA substrate.** The gold-standard Fourier shell correlation (FSC) curves (0.143 cut-off) show the final resolution of the holo-complex and each body.

For the initial 3D model of PAXX-ATP-LR complex, 82,622 particles were extracted and stacked with the box size of 144 × 144 pixels (3.71Å/pix).

Figure 4.3 shows the workflow used to process the PAXX-ATP-LR dataset with Y30-T40-c8 DNA substrate. 1,101,255 particles were picked using Gautomatch, binned by 2 (2.158Å/pxl), and supplied to a 10-class 3D classification with the negative stain reconstruction map (low-pass filtered to 30Å) as an initial reference. 15° angular sampling interval, 10 pixals offset search range and 2 pixels of offset search step were used for the first 50 iterations, then followed by 120 iterations of default classification (7.5°, 5 pxl and 1 pxl). Interestingly, while class 7 (134,748 particles) showed the expected PAXX-ATP-LR state with sharp structural features, class 10 (138,252 particles) appeared to be very similar with the LR complex described in Chapter 3. Both classes were selected, auto refined with C2 symmetry and re-centered. However, class 10 was re-extracted without binning (1.079 Å per pixel, box size 480), and all further processing on class 7 were done with the bin-2 stack. After that, both stacks were supplied to another round of 3D auto-refinement with a soft mask applied around the whole complex, then subjected to per-particle CTF refinement (per-particle defocus and per-micrograph astigmatism estimation, as well as beam tilt estimation) and Bayesian particle polishing. Class 7 resulted in a reconstruction of 11.51Å resolution, likely due to its intrinsic flexibility, and class 10 provided a 4.39Å reconstruction. All reported resolutions correspond to the gold-standard Fourier shell correlation (FSC) using the 0.143 criterion. And the maps were used as the overall map for deposition.

To further improve the map resolution of class 10 (PAXX-LR complex, apo state), similar strategy of symmetry expansion, signal-subtraction and classification methods as described in Chapter

3 were used on both DNA-PKcs-Ku-LigIV BRCT domain and the XRCC4-XLF scaffold. Specifically, the rigid region (DNA-PKcs, Ku70/80, DNA and the first BRCT domain of LigIV) was masked, symmetry expanded with C2 (276,504 particles), refined locally (initial angular sampling interval of 3.7° with local search) to 3.98Å. Subsequent two-class 3D classification without alignment pooled out 144,755 good particles for auto-refinement and post-processing. The final resolution of this map reaches 3.65Å. Local resolution of the map was also estimated within RELION 3.1.3. The flexible body (XRCC4-LigIV BRCT-XLF) was binned by 2 (2.158Å/pxl), re-centered and signal subtracted without applying symmetry expansion. After a local refinement (3.7° interval with local search), the following 3-class 3D classification without alignment resulted in a cleaned stack with 44,866 particles. The stack was further refined locally (3.7° interval with local search again) and to 13.87Å.

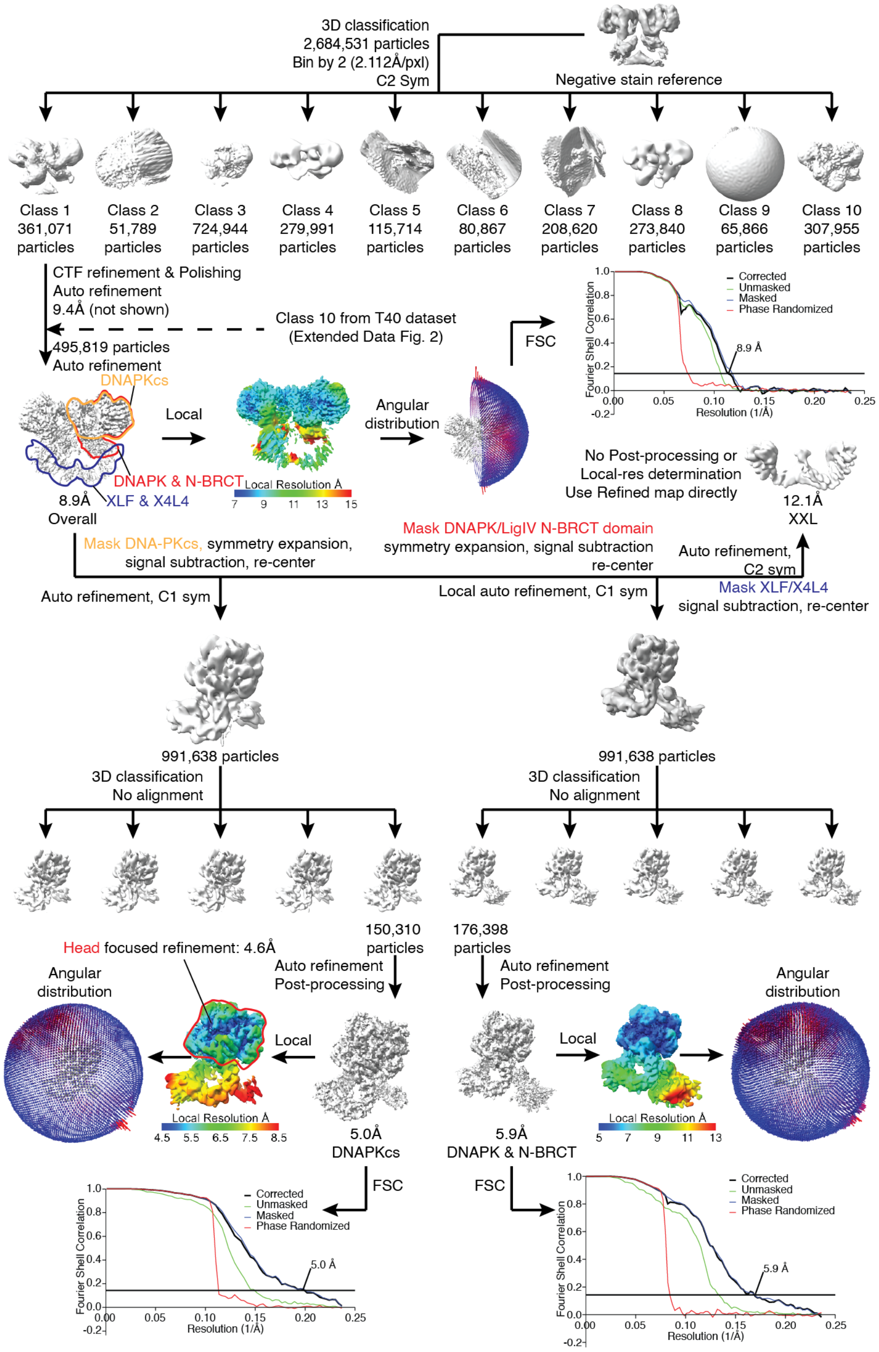


Figure 4.4: (Caption next page.)

Figure 4.4: (Previous page.) **Data-processing scheme of the PAXX-LR-ATP synaptic complex sample obtained from Y30 DNA substrate.** Class 7 from the Y30-T40-c8 stack is combined with the Y30 class1 dataset as shown. The gold-standard Fourier shell correlation (FSC) curves (0.143 cut-off) show the final resolution of the holo-complex and each body.

Figure 4.4 shows the workflow used to process the PAXX-LR-ATP dataset with Y30 blunt end DNA substrate. In total, 2,684,531 particles were picked using Gautomatch and Laplacian of Gaussian Picking built in Relion3.1.3, binned by 2 (2.112Å/pxl), and supplied to a 10-class 3D classification with the negative stain reconstruction map. Following the same procedure that produced the 11.51Å map from class 7 of the Y30-T40-c8 Cryo-EM dataset, class 1 was picked up and refined to 9.4Å. After that, the two stacks that gives PAXX-ATP-LR complex were combined and improved the whole map resolution to 8.9Å. After symmetry expansion, DNA-PKcs was masked for signal subtraction (re-centered), auto-refinement and 5-class 3D no-alignment classification. The cleaned stack with 150,310 particles was refined and post-processed to 5.0Å. Focused local refinement of the head domain region further reaches 4.6Å to help with the model building. To visualize Ku and N-terminus of DNA-PKcs, the region corresponding to DNA-PK complex and N-BRCT domain of LigIV was masked for signal subtraction (re-centered), auto-refinement and 5-class 3D no-alignment classification. After that, the cleaned stack with 176,938 particles was refined and post-processed to 5.9Å, with local resolution of Ku, DNA and N-BRCT of LigIV at around 12Å,confirming that Ku and DNA-PKcs flexibly interact in the ATP state. The XRCC4-XLF scaffold was also masked and refined to 12.1Å and was used to dock in the available scaffold model from previous works(*67*).

4.2.3 Model building

The previously described Cryo-EM model of LR complex was referenced as homology model to build the apo state of PAXX-LR complex. To build the PAXX C-terminal tail into the density, Coot was used to generate the peptide, then Coot and ISOLDE in ChimeraX were used to flexibly fit the peptide into the density. The AlphaFold2 predicted model of DNA-PKcs and PAXX was also referenced during model building. The registers can be clearly identified is most of the densities and bulky sidechains aligned well between the map and the model. Due to the resolution limitation, only the secondary motifs of the flexible regions are flexibly fitted into density with minimal disruption of existing model constraints. Manual refinement and the Phenix real space refinement were performed iteratively to further refine the model.

4.3 Architecture of the PAXX-LR complex

To study the ATP induced LR complex conformational change, as described above, we extended our established complex assembly protocol by adding ATP for 15 minutes after complex elution. Meanwhile, PAXX was added during assembly and throughout the washing and elution steps to stabilize the complex (Figure 4.2 and Methods). With the same DNA substrate as previously reported, we collected Cryo-EM data on PAXX-ATP-LR complex, resulting in an alternative state of LR complex and interestingly, a high resolution (3.63Å for the rigid, DNA-PK complex region in complex with LigIV N-terminal BRCT domain) apo-LR complex state as well after data processing (Extended data Figure 2, Figure 4.5A-D), albeit the fact that only the ATP-LR initial model was used for 3D classification. The full model we built in the apo-LR complex map has a similar overall architecture with the LR complex (Figure 3.3a,b), with RMSD value of 1.745 between the two models. Meanwhile, we were able to visualize and model in both the ATP ligand and an extra

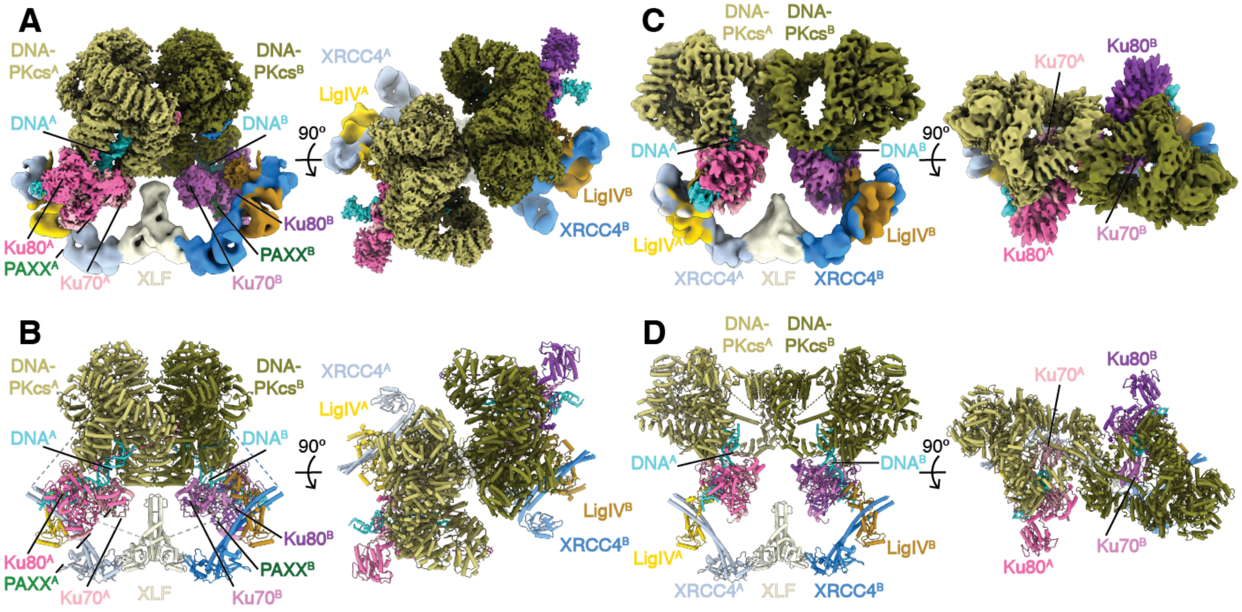


Figure 4.5: **Cryo-electron microscopy structure of the apo PAXX-LR and PAXX-ATP-LR synaptic complexes. A,** Front (left) and top (right) views of the cryo-electron microscopy composite map (Methods) of the apo PAXX-LR complex assembled in the presence of ATP. **B,** Corresponding views of the structural model of the apo PAXX-LR complex. Subunits are coloured as in previous chapters. **C,** Similar front (left) and top (right) views of the cryo-electron microscopy composite map of the ATP-PAXX-LR complex. **D,** Corresponding views of the structural model of the PAXX-ATP-LR complex.

flexible tail that corresponds to PAXX near Ku70 (4.2B).

With the high local resolution in the region corresponding to the conserved PAXX tail, we were able to unambiguously attribute the residue 180-201 into the density, with bulky side chains aligning well within the map (Figure 4.6A). The loop 184-188 forms a turn similar with the XLF Ku binding motif (KBM)(*93*), and the location of PAXX tail being inserted between Ku70 vWA domain and core region is remarkably similar with XLF KBM binding to Ku80 (Figure 4.6A,B), despite the sequence difference between PAXX tail and XLF KBM. Both the Ku70 and Ku80 pockets that PAXX and XLF tail bound are conserved, suggesting that the structural roles of PAXX and XLF tails could be conserved in evolution(Figure 4.7A). This observation is consistent with pre-

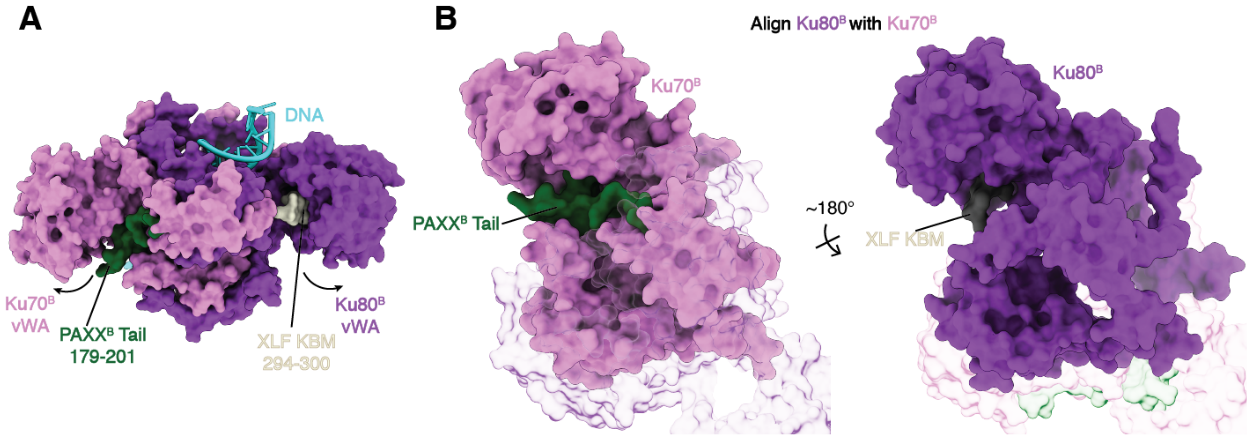


Figure 4.6: **Interaction between PAXX and Ku70 is conserved with Ku80-XLF KBM interaction. A,** Magnified view showing the interface between the C-terminal PAXX tail, XLF KBM and Ku70/80. **B,** Comparison between Ku70/PAXX subcomplex and Ku80/XLF-KBM subcomplex with Ku70 and Ku80 from the same PAXX-LR complex aligned.

vious biochemical pull-down experiments(*114*, *116*). With PAXX binding, Ku70 still adopts the open conformation similar with the reported LR, SR and DNA-PK complexes(*62*, *67*, *83*, *84*) (Figure 4.7B-D). Thus, PAXX tail stabilizes the Ku70 open conformation and its interaction with other NHEJ factors, but is not required as the XLF KBM binding to open the Ku80 vWA domain(*93*). This observation is also consistent with role of PAXX as appendix NHEJ factor, showing more severely disrupted NHEJ than XLF$^{-/-}$ cell lines when knocked down together with XLF(*115*, *169*). Although the full length PAXX was supplied within the complex assembly protocol, and is required for complex formation with blunt end DNA substrate (Figure 4.2D-F), the structured N-terminus PAXX(*114*) was not resolved within the complex. It is possible that PAXX dimer links two copies of Ku70 in a similar manner with XLF linking Ku80, but is not able to interact with XRCC4 to form a stable scaffold.

4.4 Architecture of the PAXX-ATP-LR complex

Stable LR complex can form without the complementary DNA overhang with PAXX present, as confirmed by negative staining and Cryo-EM 2D classification (Figure 4.2D-F,L-M). Notably, even without XLF and XRCC4-LigIV, PAXX is able to flexibly link two copies of DNA-PK and show similar PAXX-ATP-LR 2D class averages at lower efficiency(Figure 4.2G-I). Thus, we were able to use blunt end DNA as substrate to assemble the PAXX-ATP-LR complex and collected Cryo-EM data. Interestingly, no apo-LR state was observed anymore in either negative staining or Cryo-EM reconstructions with more biologically relative blunt end DNA substrate, and the overall architectures of the PAXX-ATP-LR state obtained from overhang and blunt end DNA substrates are very similar. Thus, we combined the two PAXX-ATP-LR complex stacks and improved the resolution of the reconstruction to overall 8.9Å. Similar signal subtraction and focused refinement strategy was performed afterward, resulted in a 5.0Å map of DNA-PKcs (4.6Å with the C-terminal head domain) and 5.9Å map of DNA-PK complex (Figure 4.5B). Due to the intrinsic flexibility of the complex, the XLF-XRCC4-LigIV scaffold could only be resolved at near atomic resolution ($\sim$12Å).

4.5 Autophosphorylation dependent rearrangement and dissociation of DNA-PKcs

The DNA-PKcs in the PAXX-ATP-LR complex shows substantial conformational change comparing to the apo PAXX-LR state. A rotation of $\sim$60° of its main body (without its N-terminus HEAT repeat regions, aa 1-500) alternates its dimeric interaction surfaces from the forehead region near the terminus of ABCDE loop to the NUC194 domain closer to the PQR loop (Figure 4.8A). The DEB helix that was structured and blocking the DNA end cannot be resolved in the ATP state anymore, and the ABCDE loop undergoes significant reassembly and covers the forehead region used for dimeric interaction in the PAXX-LR state (Figure 4.5D)(*67*, *122*). The novel DNA-PKcs dimeric conformation blocks the bindings surface on DNA-PKcs that interacts with C-terminal helix of Ku70/80 as reported in DNA-PK standing-alone complexes(*62*, *83*) and apo LR complex(*67*), which can potentially destablize interactions between DNA-PKcs and Ku, promoting the DNA-PKcs dissociation upon ATP activation(Figure 4.8B). Meanwhile, Interaction between DNA-PKcs and Ku-DNA seems to be less stable comparing to the apo state, resulting to limited resolution of Ku70/80. With the focus-refined map, we were able to observe similar interaction surfaces between Ku and N-terminus of DNA-PKcs, but the rest of DNA-PKcs rotates out and forms an opened cradle space between middle HEAT and N-terminal NEAT of DNA-PKcs (Figure 4.8C). It is possible that DNA end could be escaped in this opened space, and potentially recruits other NHEJ factors to process the ends that are not protected by DNA-PKcs in the apo state anymore.

Besides relative rearrangement between DNA-PKcs and Ku, we also observed significant rotations of the entire DNA-PK complex relative to the XLF-X4L4 scaffold within the LR complex. With the XLF-X4L4 scaffold aligned, Ku/DNA is rotated upward as well for ∼30° from PAXX-LR state to PAXX-ATP-LR state, with the anchoring point likely located between the tandem BRCT domains of LigIV (Figure 4.8D). In coordinate with the DNA-PKcs upward rotation, DNA-PKcs rotates out in a total of ∼90° to dimerize using its lower HEAT repeats, making the kinase more prone to complete dissociation. To transfer from PAXX-ATP-LR state to the SR state, Ku rotates back with the DNA substrate to align the end(Figure 4.8E), and the DSB end is shown to be mostly exposed within the PAXX-ATP-LR state. However, the DSB end observed in the PAXX-ATP-LR complex does not locate in the trajectory between the apo LR and the SR complex. It instead forms a triangle with similar distances separated from both the LR and SR DSB ends (Figure 4.8F). It is possible that once end ligation failed in the SR stage, the SR complex can directly rearrange back

Table 4.2: **Oligonucleotides used for the assembly of the DNA-PKcs dimer complex**

Construct	Strand	Sequence (5'-3')
Y25	a	CGCGATGTCGTTTCGCGCCCAGCTTTCCCAGCTGGTTGAC CTACAGGCACGCAAGGTT
	b	AACCTTGCGTGCCTGTAGGTCAACCGGGCGCG
	RNA bait	CGACAUCGCGUUU/3BiotinTEG/

to the original apo LR state by recruiting another two copies of DNA-PKcs back, and the non-ligatable DSB ends can have another opportunity to be processed in either the LR, the ATP-LR or the SR states.

4.6 Negative staining characterization of a DNA-PKcs alternative dimeric state

We were also able to capture and solve another state containing dimeric DNA-PKcs when Y-25 DNA substrate (Table 4.2) was used to directly pull-down DNA-PKcs without other NHEJ factors. Similar dimer was also observed when DNA-PKcs and DNA were in excess comparing to other factors. Interestingly, the reconstruction contains two copies of DNA-PKcs, adopting a novel interaction surface that is different from previously published DNA-PK dimeric states (Figure 4.9C,E)(*62*, *67*, *170*).

4.7 Cryo-EM characterization of the DNA-PKcs alternative dimeric state

4.7.1 Sample preparation

As described in Chapter 2, the oligos used to assemble LR complexes are listed in Table 4.2. The complex assembly protocol is as described in Chapter 3. The DNA-PKcs dimer complex negative staining samples contained ∼350 nM DNA-PKcs and the Y25 DNA substrate. The Cryo-EM sample used five fold as much as the negative staining sample.

4.7.2 Data collection and processing

The DNA-PKcs dimer complex data was collected on C-flat grids (CF-3.5/1-4C, Electron Microscopy Sciences) with amorphous thin carbon film floated on ahead of time.

For the Cryo-EM dataset, JEM-3200FS Field Emission Electron Microscope (JEOL) with K2 Summit direct electron detector (Gatan) was used at 200 kV and a nominal magnification of ×30,000 (pixel size 1.09 Å, counting mode). The defocus range was -2 to -4 µm. MSI-Raster2 application of Leginon was used to collect 1,159 movie series in total. 30-frame exposures were taken at 0.3 s per frame (9 s total exposure time), using a dose rate of 8 e^- per pixel per second, corresponding to a total dose of 76.5e^-/$Å^2$ per movie series.

Figure 4.10 shows the data processing procedure of the DNA-PKcs dimer complex. The DNA-PKcs dimer class appeared when DNA-PKcs map was used as the initial model during preliminary data processing. Thus, the 3D model of dimer was directly low-pass filtered and used as initial model for C2 3D classification. A protocol similar as described in Chapter 3 was used to process the Cryo-EM dataset, with all of the particles binned by 2 during the data processing (2.18Å per pixel) for better alignment without further un-binning. Totally 306,459 particles were picked for direct 6-class 3D classification, and C2 symmetry was forced to capture most particles corresponding to the dimeric DNA-PKcs complex. Finally, 36,379 particles were selected and refined to 8.0Å after CTF refinement and polishing. About 30% of particles goes to the final reconstruction of DNA-PK monomer complex when C1 symmetry was applied during initial 3D classification. Resulting in a DNA-PK map with 7.3Å final resolution (data not shown). Further attempts of signal subtraction and focused refinement on DNA-PKcs monomer or smaller regions failed to improve the local resolution.

4.7.3 Model building

DNA-PKcs within the Cryo-EM model of LR complex (Figure 3.3) was referenced as homology model to build the alternative DNA-PKcs dimer complex. Rigid fitting of the model into the map results in clashes areas at the dimeric interaction surface. Due to the resolution limitation, ISOLDE with distance restraints was applied to flexibly fit the secondary structures into the density. To completely resolve the clashing contacts between two copies of DNA-PKcs, Coot and ISOLDE was also used to manually build the backbones at the dimerization surfaces. When one copy of DNA-PKcs model was refined substantially, it was duplicated to rigidly fit into the other half of the Cryo-EM map to ensure no clashing in between. This process was performed for several iterations, then the refined model was supplied to Namdinator(*171*) and Phenix real space refinement for further refinement.

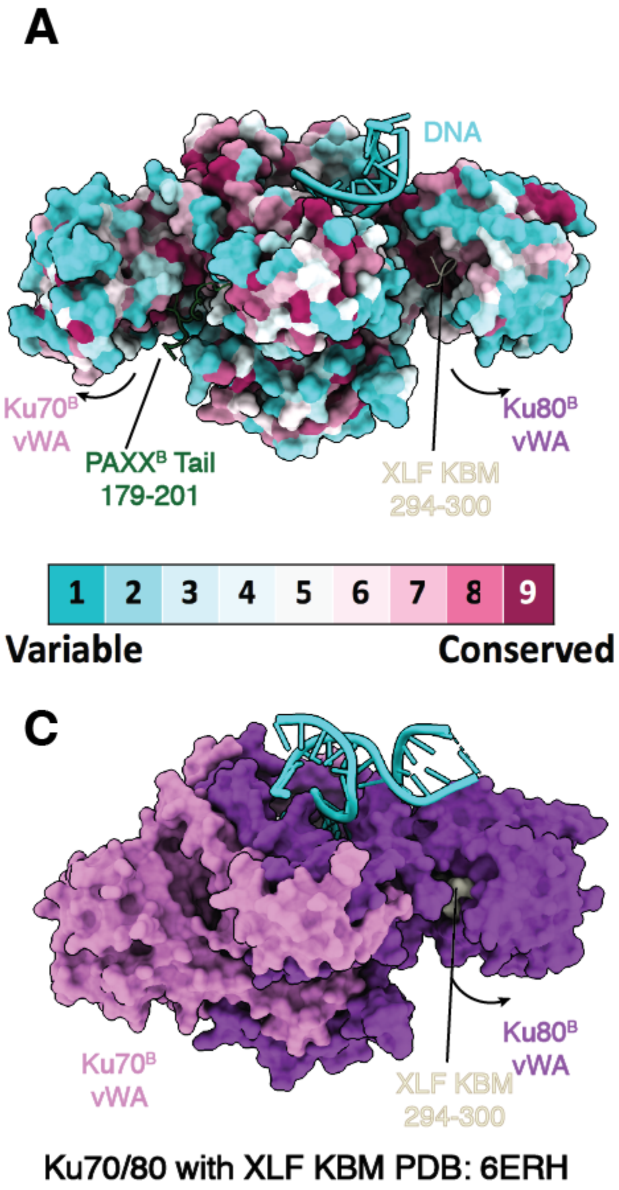
4.8 Architecture of the DNA-PKcs alternative dimeric state

Although limited by particles numbers and overall resolutions, the dimeric DNA-PKcs state observed in our negative staining and Cryo-EM datasets (Figure 4.11A,B) uses the N-terminus HEAT repeats (1-200), the central cradle domain (2300-2500) as well as the FAT regions in the head unit (2820-2850, 3000-3100) to make extensive dimerization surfaces (Figure 4.12). Among these interaction patches, the N-terminus dimerization surface buries most of the surface area ($900Å^2$), and is mutually exclusive with the DNA-PKcs-Ku70/80 interactions within the DNA-PK complexes obtained in both the LR state and the ATP-LR state (Figure 4.13A-C), despite the fact that DNA-PK complex ungergoes significant conformational changes between the two states (Figure 4.13D-F). This observation indicates that the dimeric state of DNA-PKcs we obtained functions as a standing-alone units, and is more likely to be inactive without Ku binding. Upon dimerization, both the N-terminal HEAT and middle HEAT regions of DNA-PKcs that are involved with dimerization are pushed inwardly, resulting in a smaller space within the cradle area(Figure 4.13D,E). This movement is opposite from the opening of the DNA-PKcs cradle space observed in the transition from LR to ATP-LR states(Figure 4.13F).

4.8.1 Possible roles of dimeric DNA-PKcs in NHEJ

Based on the structural observations described above, it is possible that the dimeric state of DNA-PKcs exists in the nucleus, and functions as a reservoir to store extra copies of DNA-PKcs. Recruiting two copies of DNA-PKcs by Ku bound DNA can potentially increase the factor recruitment efficiency, and can facilitate the bridging of free DNA ends as well. In addition, the dimeric DNA-PKcs can also form when both copies are completely dissociated from the repair site, and can be recycled later on to perform next round of NHEJ DNA repair when necessary. If the dimeric state of DNA-PKcs is favored, it could shift the equilibrium toward the formation of both dimeric DNA-PKcs and SR complex from the LR-ATP state. If this is the case, a condition with both dimeric DNA-PKcs and SR complex formed from the LR-ATP state should exist, and will be interesting to be characterized by following-up structural studies. Interestingly, DNA is not observed in the reconstruction, although the DNA-PKcs dimer complex has to be pull-down by DNA in the assembly protocol. It is possible that this dimeric DNA-PKcs state can be flexibly linked to DNA and the rest of NHEJ factors, for example, via flexible tails of Ku70/80.

4.8.2 Conclusion – Multiple Steps During LR to SR State Transition in NHEJ



Ku70/80 PDB: 1JEY

Ku70/80 with XLF KBM PDB: 6ERH

Ku70/80 in LR complex PDB: 7LT3

Figure 4.7: **Comparing the structure of Ku with PAXX and XLF tail inserted among the the PAXX-LR complex and previously published models. A,** Close-up view of PAXX and XLF bound Ku colored by sequence conservation. **B,** Crystal structure of apo Ku showing no rotation of either Ku70 or Ku80 vWA domains(*27*). **C,** Crystal structure of the XLF Ku-binding motif (KBM) bound to Ku showing the outward rotation of only the Ku80 vWA domain(*93*). **D,** Ku structure in the LR complex showing outward rotations of both Ku70 and Ku80 vWA domains(*67*).

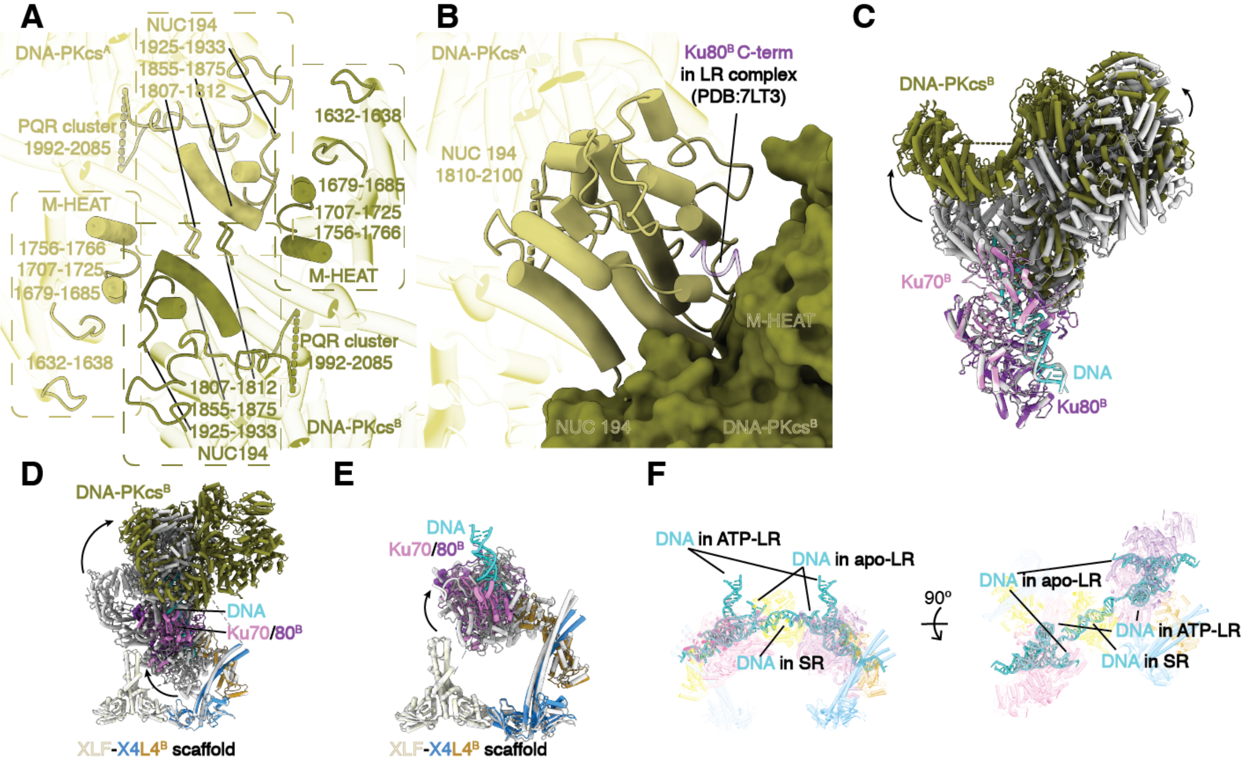


Figure 4.8: **Novel interaction between two copies of DNA-PKcs triggered by ATP induced DNA-PK conformational changes in the LR complex.** **A,** Magnified view showing the novel interface between the middle HEAT region of DNA-PKcs. Secondary motifs involved in the NUC194 and surrounding middle HEAT regions are highlighted. **B,** Superimposition of the DNA-PK complex near NUC194 domain in the apo PAXX-LR and PAXX-ATP-LR complexes. Surface binding to Ku80 C-terminal helix in the apo state (Copy B, surface) is blocked by another copy of DNA-PKcs in the PAXX-ATP-LR state (Copy A, ribbon). **C,** Superimposition of the DNA-PK complex in the apo PAXX-LR (light gray) and PAXX-ATP-LR (colored) complexes with Ku70/80 and DNA aligned. **D,** Superimposition of the half apo PAXX-LR (light gray) and PAXX-ATP-LR (colored) complexes with XLF-X4L4 scaffold aligned. **E,** Superimposition of the half SR (light gray) and PAXX-ATP-LR (colored) complexes with XLF-X4Lr scaffold aligned. **F,** An overview of the two DSBs within the apo PAXX-LR complex, the PAXX-ATP-LR complex and the SR complex with XLF-X4L4 scaffold aligned. DNA elements are shown as solid ribbons and others in transparent representation. Front view (left) and up view (right) highlights the DSB ends at each state.

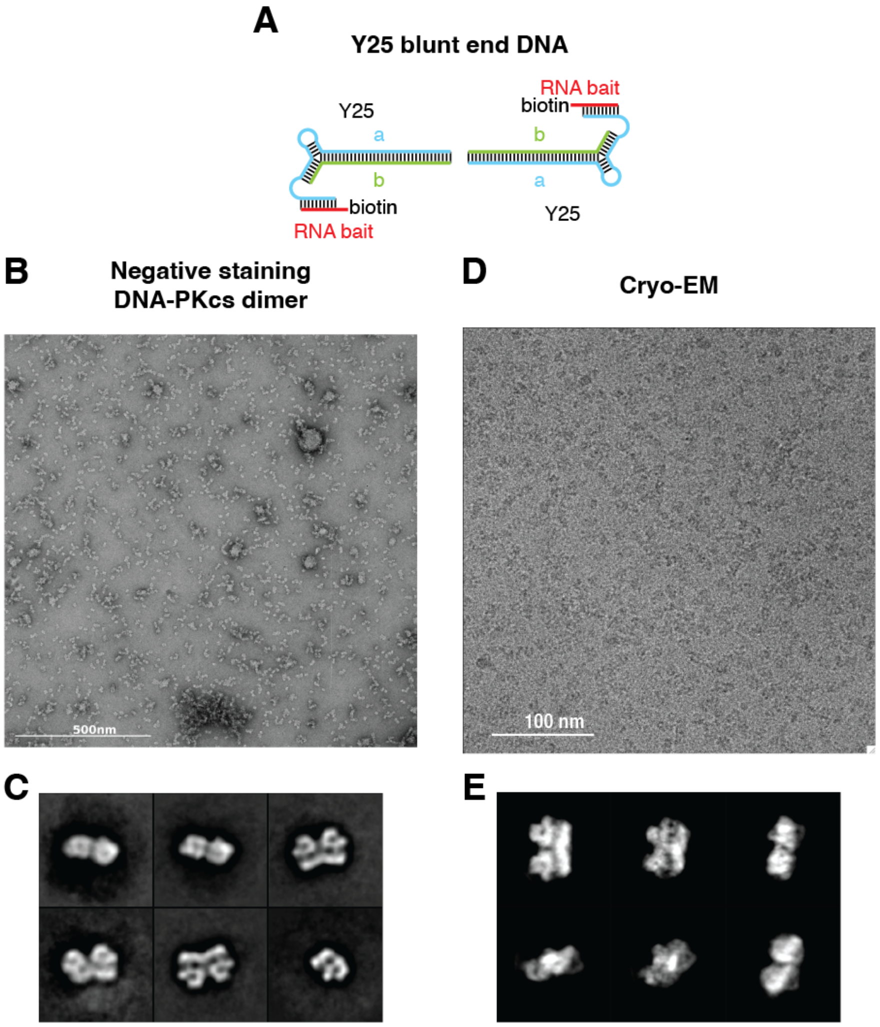


Figure 4.9: **Optimization of the DNA-PKcs alternative dimeric complex assembly with DNA.** **A,** Schematic showing the Y25 blunt-end DNA substrate. **B,** A representative negative-staining raw micrograph of the complex assembled using the LR complex formation protocol (Chapter 3). The raw micrograph is representative of 69 micrographs. **C,** Representative two-dimensional class averages of the complex assembled, showing the appearance of the alternative dimeric DNA-PKcs complex. **D,** A representative cryo-EM raw micrograph (out of 1,159 micrographs in total) of the dimeric DNA-PKcs complex assembled with the Y25 DNA substrate shown in **A**. **E,** Representative two-dimensional class averages of particles (306,459 in total) contributing to the final reconstruction of the dimeric DNA-PKcs complex.

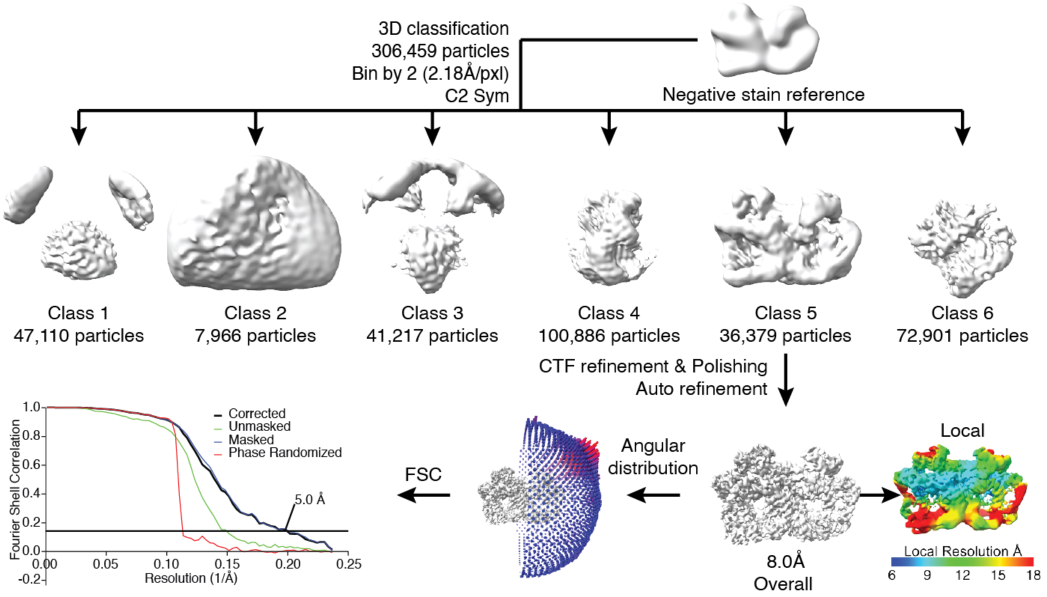


Figure 4.10: **Data-processing scheme of the DNA-PKcs dimer complex sample.** Flow chart of the cryo-EM data processing procedure. The gold-standard FSC curves (0.143 cutoff) show the final resolution of the dimer.

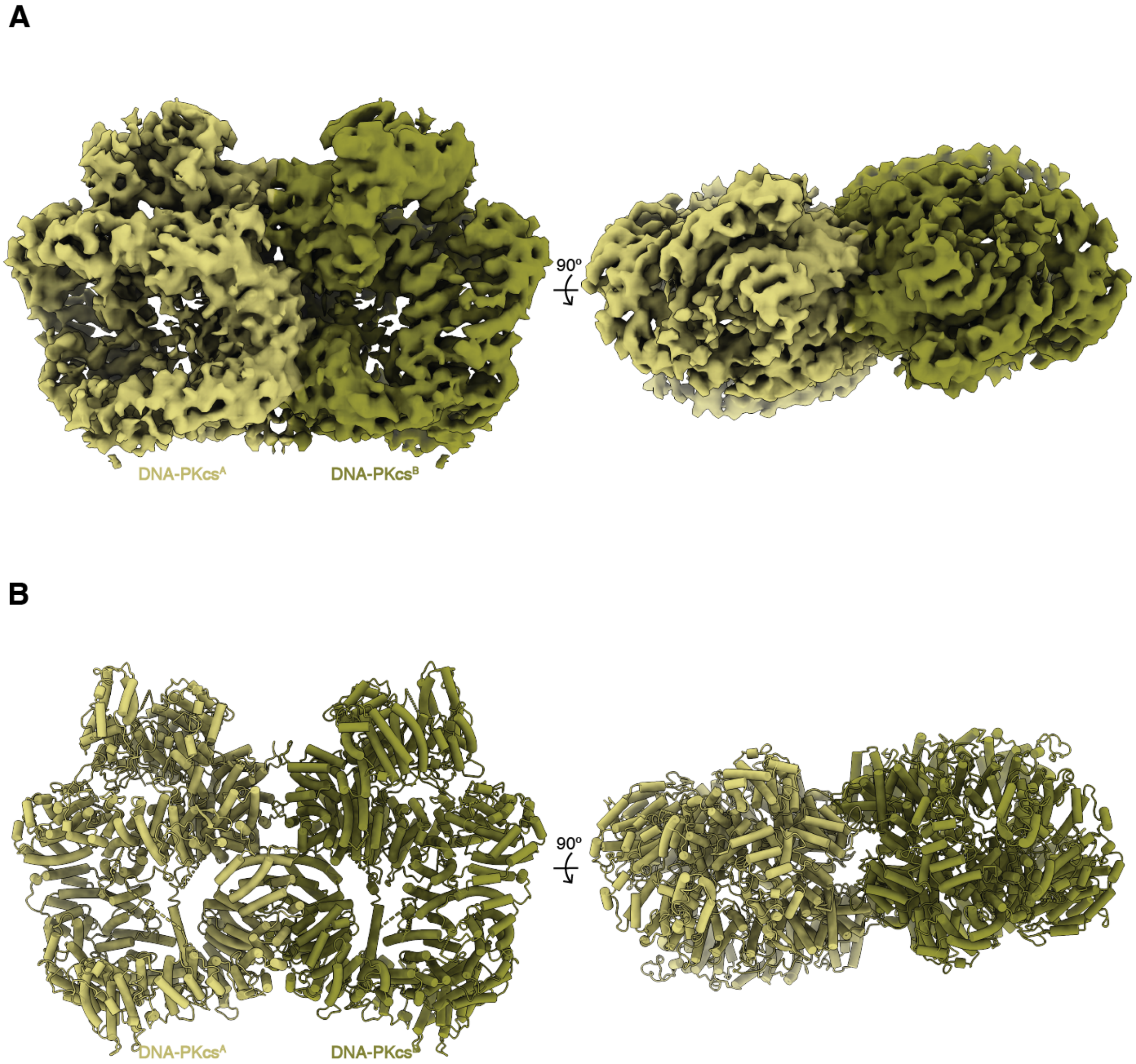


Figure 4.11: **Cryo-electron microscopy structure of the alternative DNA-PKcs dimeric complex. A,** Front (left) and top (right) views of the cryo-electron microscopy map of the DNA-PKcs dimeric complex. **B,** Corresponding views of the structural model of the DNA-PKcs dimeric complex. Subunits are coloured as in previous chapters.

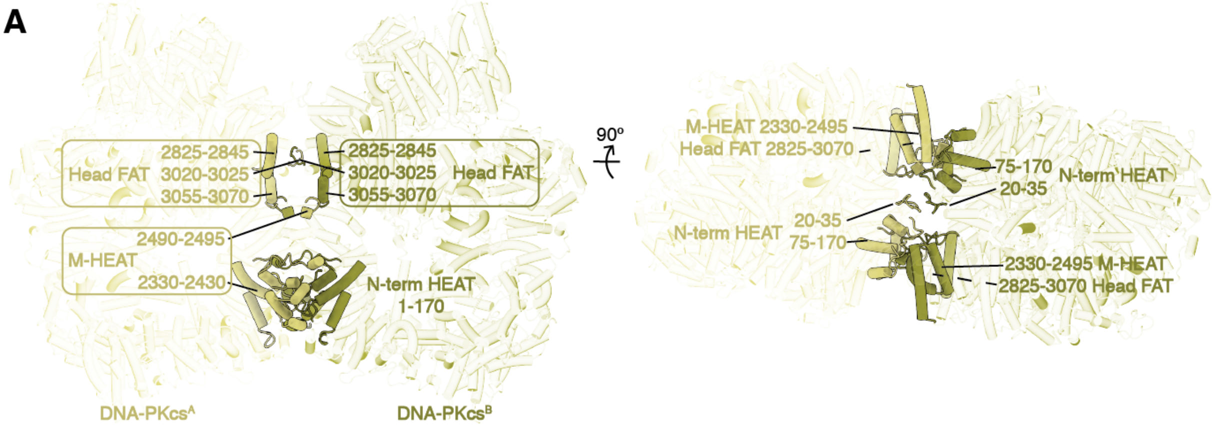


Figure 4.12: **Dimerization surface of DNA-PKcs in the alternative dimeric state. A,** Front (left) and bottom (right) views of the DNA-PKcs dimeric complex interaction surfaces. secondary motifs and loops involved in the N-terminal HEAT region (N-HEAT), middle HEAT region (M-HEAT) and head domain (Head FAT) are highlighted.

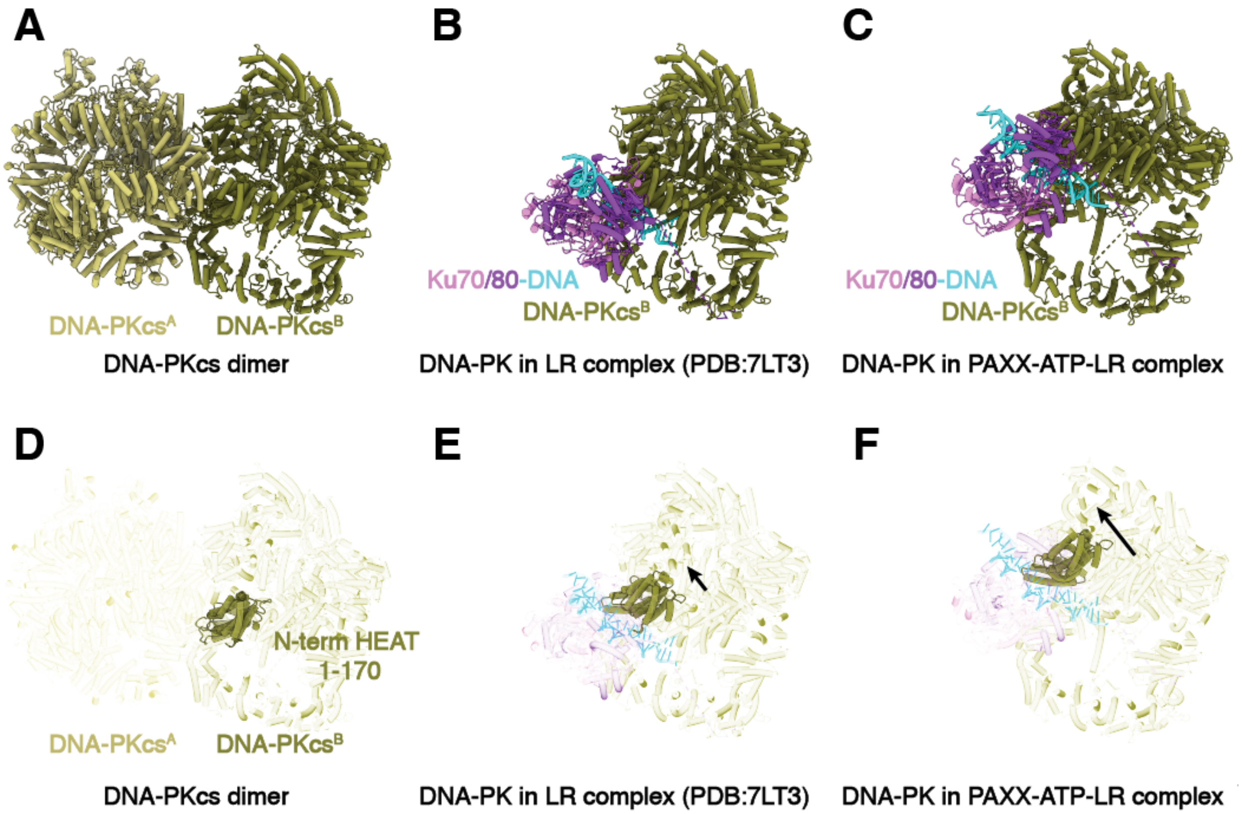


Figure 4.13: **Dimerization of DNA-PKcs is mutually exclusive with DNA-PK complex. A,** DNA-PKcs dimeric complex model. The model is rotated up for 45° to visualize the dimerization surface from the bottom. **B,** DNA-PK complex in the LR complex model. **C,** DNA-PK complex in the PAXX-ATP-LR complex model. DNA-PKcs in **B,C** are aligned with **A,** for comparison. **D-F,** Same view as **A-C** , with N-terminal HEAT repeats highlighted for comparison. The rest of the models are in transparent.

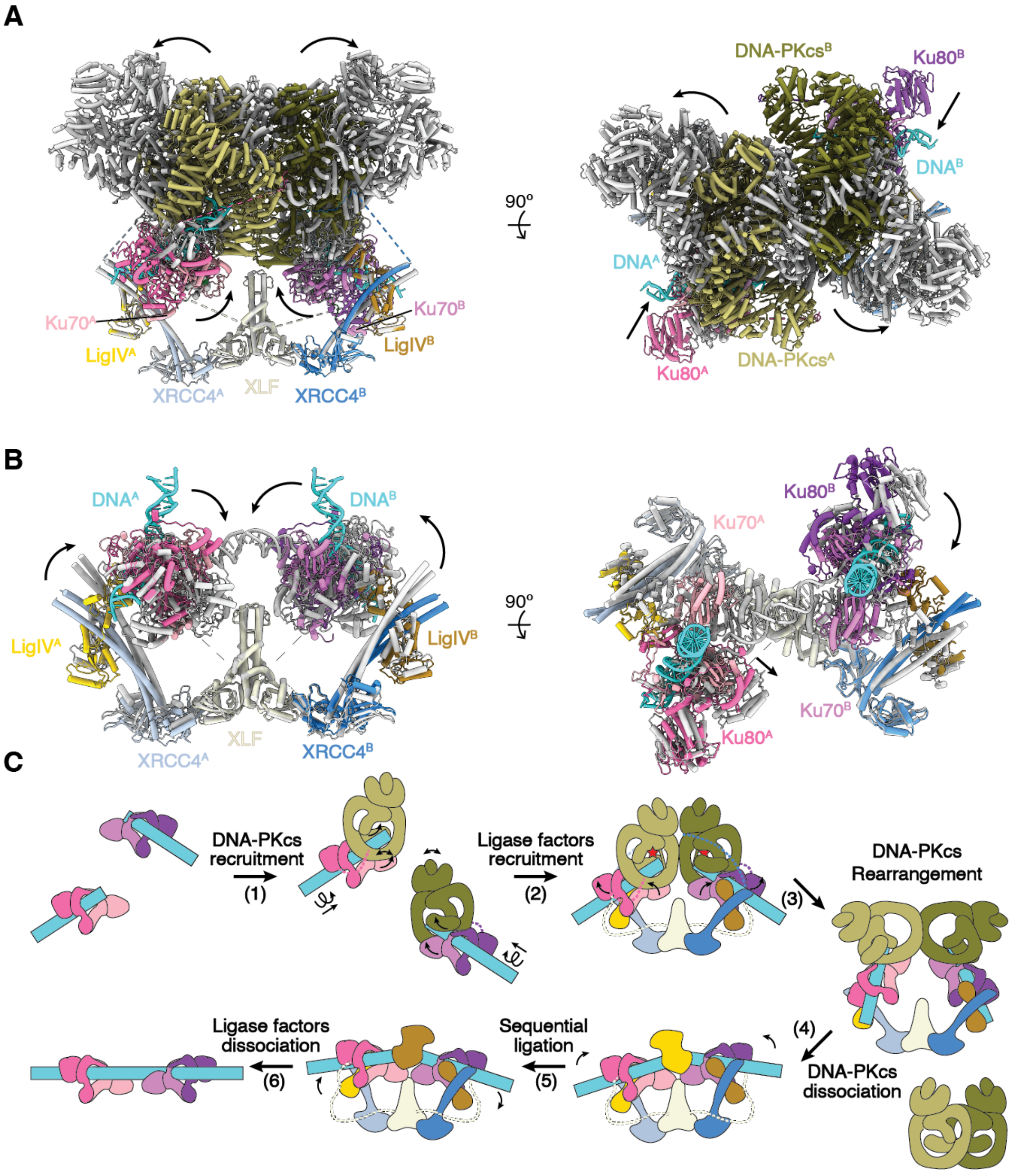


Figure 4.14: (Caption next page.)

Figure 4.14: (Previous page.) **Structural transition from the LR to the SR synaptic complex. A,** Superimposition of the LR (coloured) and LR-ATP (grey) complexes shown in front (left) and top (right) views. The XLF homodimer is used to align the two conformers. The transition from the LR to the ATP-LR synaptic state indicates potential conformational changes induced by DNA-PKcs activation and auto-phosphorylation. **B,** Superimposition of the LR-ATP (coloured) and SR (grey) complexes shown in front (left) and top (right) views. The XLF homodimer is used to align the two conformers. The catalytic domains of DNA-PKcs and LigIV are hidden for clarity. The transition from the LR-ATP to the SR synaptic state indicates potential conformational changes induced by dissociation of DNA-PKcs and association of the LigIV catalytic domain. **C,** Model of structural transitions during NHEJ. After DSB detection by Ku70/80, DNA-PKcs is recruited to form a DNA-PK complex by inward translocation on DNA (1). A LigIV–XRCC4–XLF–XRCC4–LigIV scaffold assembles the two DNA-PK complexes into the LR complex, positioning the major autophosphorylation clusters near the kinase active centres *in trans* (2). After autophosphorylation and of DNA-PKcs, the complex transitions into the LR-ATP complex, promoting the dissociation of DNA-PKcs and exposing DNA ends for processing and ligation (3). Next, DNA-PKcs dissociate and the complex transitions into the SR complex, enabling the recognition and sealing of one nick by one LigIV catalytic domain (4). The complex next undergoes a conformational change and possibly DNA translocation to enable the recognition and sealing of the other nick by the second LigIV catalytic domain from the opposite side (5). The ligase factors then dissociate and the DSB is repaired (6). During the whole process, the alternative DNA-PKcs dimer can serve as the reservoir to provide two copies of DNA-PKcs rapidly.

To extrapolate the conformational changes during the transition from LR to SR synapsis, we superimposed the XLF dimers from the three synaptic states (Figure 4.14A,B). While the approximate positions of individual NHEJ factors remain the same relative to each other during the transition, each of the Ku-XRCC4-LigIV subcomplexes rotate to different agnles in order to align the DSB ends with the rotation pivot points in the middle of each XRCC4 coiled-coil. A similar discrepancy is observed when comparing the crystal structures of the human LigIV-XRCC4 complex and its yeast ortholog (Lif1p/Lig4p)(*155*) (Figure 3.5b), indicating that the bending angle of the intrinsically twistable XRCC4 CC domain can adjust to accommodate different states. Interestingly, the $\sim$16 bp DNA footprint of DNA-PKcs that would be exposed by DNA-PKcs dissociation from the LR complex is precisely the length required to align the two DSB ends in the SR complex

(Figure 4.14A), suggesting that the HEAT cradle region of DNA-PKcs serves as a “ruler” by holding the appropriate length of DNA within the DNA-PK sub-complex for subsequent alignment in the SR complex(*172*). Finally, newly observed interfaces within the SR complex, including XLF-Ku70 and LigIV DBD-Ku70 interfaces, also likely contribute to the transition from the LR complex (Figure 3.11c,d).

In summary, based on the NHEJ reaction cycle we proposed in the last Chapter, we further captured three more essential states during NHEJ by high-resolution cryo-EM. As appendix NHEJ scaffolding factor, PAXX stabilizes both the LR state and LR-ATP state structurally, ensuring the intactness of the whole complex when subunits undergo significant conformational changes. With novel DNA-PKcs dimerization surfaces obtained in both the LR-ATP state and the DNA-PKcs dimeric state, the multi-step mechanism of state transition from the LR state to the SR state in NHEJ can be comprehensively illustrated (Figure 4.14C). Besides the function of DNA-PKcs discussed in Chapter 3, DNA-PKcs could also work as a ’DNA ruler’ during state transition, since the length of the DNA released from the HEAT cradles of the DNA-PKcs molecules is appropriate for subsequent alignment within the SR complex. Finally, a concerted conformational change is triggered by the restoring force within the LigIV-XRCC4-XLF-XRCC4-LigIV scaffold as well as protein-protein interactions involving XLF-Ku70 and LigIV DBD-Ku70,

CHAPTER 5

CONCLUSION AND FUTURE WORK

5.1 Summary of key findings and significance

In summary, the high-resolution structures of multiple NHEJ complexes obtained in this study provide unprecedented insight into the mechanism of NHEJ. With these new information, biological roles of DNA-PKcs in DNA end synapsis, especially how DSBs are detected, protected, bridged then aligned for ligation are well explained at atomistic level. Further more, the Long-range complex captured for the first time all of the core NHEJ factors structurally, which provides a critical foundation to move the field forward from focusing on individual pieces of NHEJ machinery to understanding the whole system in a comprehensive manner. Meanwhile, this works raises up new questions and directions in the field. A number of new interaction surfaces observed in different stages are worth to be further characterized both *in vitro* and *in vivo*. Certain patient mutations and truncations on NHEJ factors can now be re-visited as well with the context of different complexes, and be probed further using biochemical and cellular methods.

5.2 Opportunities for future research

5.2.1 Recruitment of other appendix factors during NHEJ at different stages

While represented in details at different stages, the core NHEJ machinery consists only five NHEJ factors, and is only capable of repairing DSBs that need no further end processing. As mentioned in Chapter 1, many appendix NHEJ factors have to be recruited when end processing is necessary, and such event will likely occur with the context of various NHEJ stages. As proof of principle, PAXX has shown to be interacting with Ku70 in the LR complex to stabilize complex structure as described in Chapter 4, and Artemis has shown recently to complex with DNA-PK complex when activated by DNA-PK phosphorylation(*122*). Complex structures solved in this study will be fundamental toward the understanding of appendix factor recruitment during NHEJ. Notably, it is possible that different appendix factors could be recruited at different stages during NHEJ. For example, Pol λ and Tdp1, polymerases that are responsible for gap fillings in NHEJ, is observed to function after DNA end alignment, thus more likely to be recruited with the SR complex(*164*). It is important to understand how different mechanisms can be applied to recruit required NHEJ appendix factors in various scenarios. Furthermore, by solving complexes structures containing different NHEJ appendix factors, how these factors work together to create compatible and ligatable DSB ends prior to LigIV binding could be well characterized.

5.2.2 Common role of DNA damage response(DDR) PIKKs in their DSB repair pathways

DNA-PKcs is one of the three PIKK family proteins that has principal roles of activating DNA damage response (DDR). All three kinases, DNA-PKcs, ATM and ATR, share similar domain organizations and 3D structural architectures (Figure 3.8d). With the recently published complex structures containing dimeric DNA-PKcs(*62*, *67*, *170*), all three DDR PIKKs have been observed to form dimers structurally(*160*, *161*, *173*) (Figure 3.8a-c). Meanwhile, DNA-PKcs, ATM and ATR share similar mechanisms of recruitment to their targeted DNA damage sites(*74*). Comparable with the Ku dependent recruitment of DNA-PKcs to DSB sites, the C-terminus of NBS-1(*74*), one subunit of the Mre11-Rad50-Nbs1 (MRN) complex(*174*), recruits ATM to DSB cites to initiate the homologous recombination (HR) pathway. Two tumor suppressor proteins, p53-binding protein 1 (53BP1) and breast cancer type 1 (BRCA1), were found to promote this process. Notably, the tandem breast cancer carboxy-terminal (BRCT) repeats of 53BP1 establishes direct interaction with Rad50 in MRN complex, which promotes ATM phosphorylation activity and is required for ATM dependent DNA repair(*175*). Similar interaction between Ku and LigIV tandem BRCT domains was observed as well in the LR and SR complexes(*67*). Both Mre11 in the MRN complex and another important HR factor, CtIP(*176*), are required endonucleases for DNA resection in HR, and are both regulated by ATM phosphorylation(*177*). In addition, Exonuclease 1 (EXO1), endonuclease DNA2 and the Bloom syndrome helicase (BLM) are also involved to perform long range resection(*178*). After DNA resection, the exposed ssDNA overhang will be sensed by replication protein A (RPA), and RPA will accumulate at the exposed ssDNA to recruit ATR/ATRIP by direct interaction(*179*), again similar with Ku recruiting DNA-PKcs and MRN recruiting ATM (Figure 5.1).

The RPA-ssDNA filament can be formed at the stalled replication forks as well, initiating the same ATR/ATRIP dependent DDR pathway. Thus, it is believed that in the S/G2 phase, ATM dependent HR and ATR dependent DDR pathways often function at the same DNA damage site, playing overlapping yet distinct roles(*180*). The ATR activity is stimulated by other independently recruited factors as well. For example, the Rad9-Hus1-Rad1 (9-1-1) clamp recognize the 5'-ended ssDNA-dsDNA junction and recruits topoisomerase II binding protein 1 (TOPBP1)(*181*, *182*) that interacts with ATR/ATRIP(*183*). Another activator, Ewing tumor-associated antigen 1 (ETAA1), interacts directly with RPA and ATR/ATRIP(*184–186*).

DNA-PKcs is by far the only factor that has been shown to make stable and synaptic complexes with other factors functioning in its DNA repair pathway(*67*, *170*). Based on the structure of both LR and SR complex, it is proposed that in the LR complex, synaptic DNA-PK complexes hold two DSB DNA ends in proximity, which serves as the major checking point prior to the LR to SR state transition. When two DNA ends are on site together with other required NHEJ factors, in trans autophosphorylation of DNA-PKcs(*60*) triggers its dissociation from the repair site, which lead

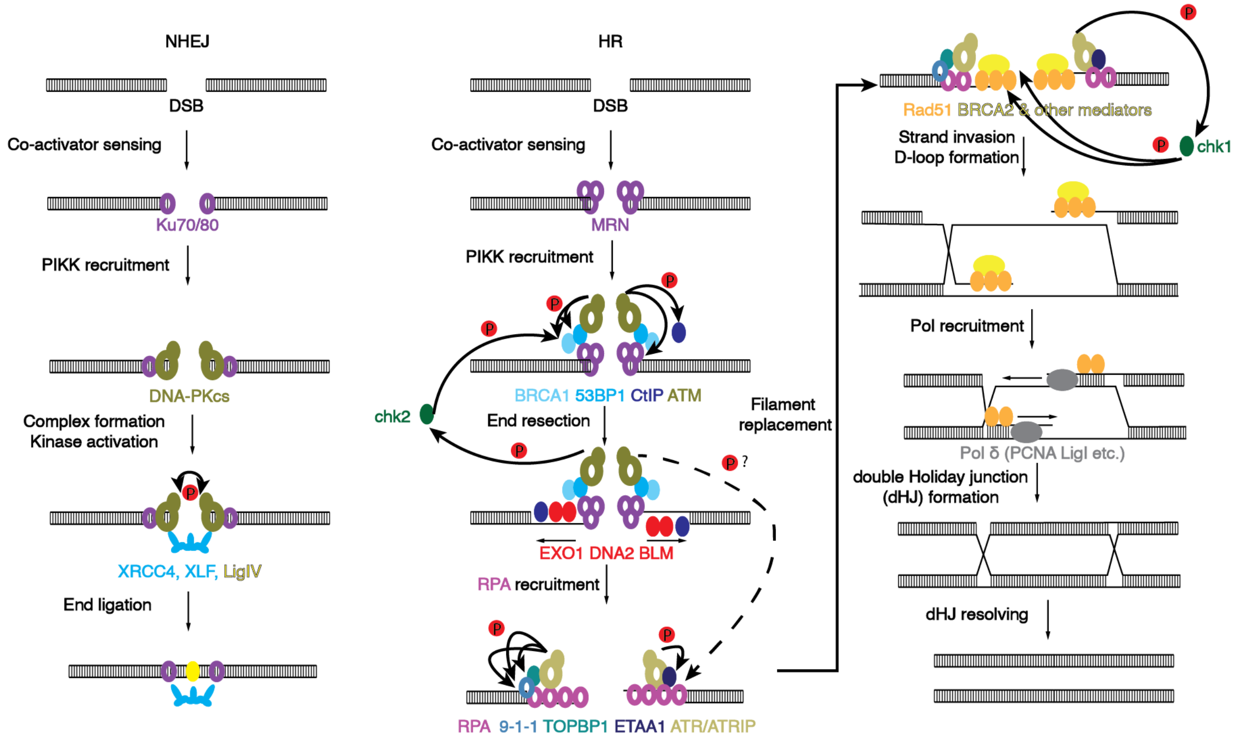


Figure 5.1: **Comparison between the function of PIKK family member DDR kinases in their corresponding DSB repair pathways.** Different factors are depicted with different colors. Potential phosphorylation events are highlighted with curved arrows and red 'P' labels.

NHEJ to its next step(*67*). Despite the structural and evolutional similarities among three DDR PIKKs, however, how ATM or ATR/ATRIP can form synaptic complexes with DNA and other related DDR factors are poorly understood. Like DNA-PKcs, all three DDR PIKKs undergo autophosphorylation as well, with targeted S/T-Q sites clustered mostly on their HEAT domains(*187*), but the biological roles of ATM and ATR autophosphorylation are more controversial and less clear as well.

As apical kinases that regulate DDR mainly by phosphorylation, both ATM and ATR phosphorylate many substrates in response to DSB agents. To ensure efficient and specific phosphorylation at the DNA repair site, it is natural to speculate that similar with DNA-PKcs, ATM and ATR can form a series of key intermediates on the DNA substrate with other DDR factors, and the autophosphorylation activity of the two kinases also serve as the checking point to regulate its dissociation from the repair site when all required factors have been recruited and phosphorylated (Figure 5.1). In NHEJ, the dimeric states of DNA-PKcs are important for DNA end bridging, which could be achieved with the existence of template DNA in both HR and replication stress response. However, a stable synaptic protein-DNA complex would still be beneficial prior to strand invasion and D-loop formation (more likely formed by ATM), or during the second invasion event on the opposite side of D-loop (more likely formed by ATR/ATRIP, Figure 5.1). During the formation of extensive duplex DNA, the ATM/ATR synaptic complexes become less important to the stability of DNA synapses, and the dissociation of such complexes can be triggered by either the removal of their recruiting factors (MRN complex and RPA), or their auto-phosphorylation.

Besides forming DNA synapses structurally, it is also beneficial functionally in the kinase signal cascade to form ATM-DNA or ATR/ATRIP-DNA complexes at the repair site. Multiple copies of ATM or ATR can either allow inter-molecular interactions to regulate its kinase activity(*160*, *161*), or to rapidly amplify the signal cascade by creating positive feedback loops in the phosphorylation network(*10*, *188*), often via phosphorylation of their activator proteins mentioned above. ATM and ATR also phosphorylate many kinases, and two well-characterized examples are in the ATM-chk2 and ATR-chk1 pathways, which primarily coordinate with each other to regulate DDR(*180*). Proteomic analysis identifies hundreds of ATM and ATR phosphorylation targets involved in DDR pathways, but the biological roles of most connections are unknown(*189*). Based on the proposed model above, RPA functions latest in HR among all the well-studied ATM phosphorylation targets. However, *in vivo* results suggest that ATR and DNA-PKcs phosphorylation on RPA is dominant(*190*). Thus, it is more likely that ATM autophosphorylation and dissociation happen before the RPA recruitment, and ATR, after being recruited and activated upon RPA binding, will undergo autophosphorylation and dissociation before Rad51 replaces RPA and wraps around ssDNA (Figure 5.1). Clearly, further investigation and characterization of such events during DDR are expected.

Comparison of the relative orientations of DNA-PKcs, ATM and ATR-ATRIP in the published dimer structures shows substantially different dimerization surfaces, indicating that the observed ATM and ATR-ATRIP dimers may play different roles in their DDR pathways than the DNA-PKcs dimer in NHEJ, respectively. Besides the possibility that ATM and ATR may also form large complexes with their activators and substrates at DNA damage sites, it is possible as well, that DNA-PKcs may have another structurally stabled state that doesn't interacts extensively with any other NHEJ factors. This DNA-PKcs-only dimeric stage, if exists, is more likely to be inactive and serve similarly as the ATM and ATR/ATRIP dimers to be the reservoir of extra kinases that can quickly respond to various DNA damages.

5.2.3 NHEJ in CRISPR based gene-editing strategies

Visualization of NHEJ DNA repair complexes provides new insights toward rational design and engineering of this system, which will lead to the development of novel research tools. One promising example will be the combination of NHEJ and CRIPSR/Cas9 system as a new generation gene editing tool. NHEJ pathway has been utilized years ago to couple with CRISPR/Cas9 system and quantitatively detect the CRISPR/Cas9 directed on- and off-target mutations(*191*, *192*). In addition, programmed CRISPR/Cas9 system generates DSBs downstream of the target site for its genome editing function, so its gene manipulating efficiency is strongly affected by the fast, error-prone NHEJ DSB repair activity. Thus, structure-based rational engineering on NHEJ factors can be coupled with the usage of CRISPR/Cas9 and other genome engineering tools to improve gene editing efficiency(*193*), introduce frame-shift gene knock-off mutations(*194*) and expand its usage in a wide range of organisms(*195*).

5.2.4 NHEJ in cancer drug development

Understanding the molecular basis of NHEJ mechanism will provide the foundation for potential therapeutic application. It is known that suppression of NHEJ activity in tumor cells potentiates their sensitivity to radiation therapy(*196*), and small molecule inhibitors toward key repair factors have been designed to offer new therapeutic strategies toward cancer treatments(*197*, *198*). Some small molecule DNA-PKcs inhibitors are currently in clinical trials(*199*). With detailed structural elucidation of multiple NHEJ stages, advanced NHEJ inhibitors can be designed to target not only specific factors and their allosteric conformations, but also novel protein-protein and protein-DNA interaction surfaces that are crucial during the state transition processes. The more small molecule inhibitors researchers could discover with distinct targets and designing rationales, the more likely that precision oncology and personalized cancer immunotherapy could be available in more and more cases. There is no doubt that comprehensive understanding of NHEJ DNA repair machinery will shed light on more applications of this system in the field of cancer research, nanoscience and synthetic biology.

REFERENCES

(*1*) Aplan, P. D. (2006). Causes of oncogenic chromosomal translocation. *Trends Genet 22*, 46–55.

(*2*) Zhao, B., Rothenberg, E., Ramsden, D. A., and Lieber, M. R. (2020). The molecular basis and disease relevance of non-homologous DNA end joining. *Nat Rev Mol Cell Biol 21*, 765–781.

(*3*) Moynahan, M. E., and Jasin, M. (2010). Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis. *Nat Rev Mol Cell Biol 11*, 196–207.

(*4*) Guirouilh-Barbat, J., Huck, S., Bertrand, P., Pirzio, L., Desmaze, C., Sabatier, L., and Lopez, B. S. (2004). Impact of the KU80 pathway on NHEJ-induced genome rearrangements in mammalian cells. *Mol Cell 14*, 611–23.

(*5*) Critchlow, S. E., and Jackson, S. P. (1998). DNA-end-joining: from yeast to man. *Trends in Biochemical Sciences 23*, 394–398.

(*6*) Malu, S., Malshetty, V., Francis, D., and Cortes, P. (2012). Role of non-homologous end joining in V(D)J recombination. *Immunol Res 54*, 233–46.

(*7*) Wright, W. D., Shah, S. S., and Heyer, W. D. (2018). Homologous recombination and the repair of DNA double-strand breaks. *J Biol Chem 293*, 10524–10535.

(*8*) Jazayeri, A., Balestrini, A., Garner, E., Haber, J. E., and Costanzo, V. (2008). Mre11-Rad50-Nbs1-dependent processing of DNA breaks generates oligonucleotides that stimulate ATM activity. *EMBO J 27*, 1953–62.

(*9*) Tatin, X., Muggiolu, G., Sauvaigo, S., and Breton, J. (2021). Evaluation of DNA double-strand break repair capacity in human cells: Critical overview of current functional methods. *Mutat Res Rev Mutat Res 788*, 108388.

(*10*) Shiloh, Y., and Ziv, Y. (2013). The ATM protein kinase: regulating the cellular response to genotoxic stress, and more. *Nat Rev Mol Cell Biol 14*, 197–210.

(*11*) Sishc, B. J., and Davis, A. J. (2017). The Role of the Core Non-Homologous End Joining Factors in Carcinogenesis and Cancer. *Cancers (Basel) 9*.

(*12*) Karanjawala, Z. E., Grawunder, U., Hsieh, C. L., and Lieber, M. R. (1999). The nonhomologous DNA end joining pathway is important for chromosome stability in primary fibroblasts. *Current Biology 9*, 1501–1504.

(*13*) Ferguson, D. O., Sekiguchi, J. M., Chang, S., Frank, K. M., Gao, Y., DePinho, R. A., and Alt, F. W. (2000). The nonhomologous end-joining pathway of DNA repair is required for genomic stability and the suppression of translocations. *Proc Natl Acad Sci U S A 97*, 6630–3.

(*14*) Brunet, E., Simsek, D., Tomishima, M., DeKelver, R., Choi, V. M., Gregory, P., Urnov, F., Weinstock, D. M., and Jasin, M. (2009). Chromosomal translocations induced at specified loci in human stem cells. *Proc Natl Acad Sci U S A 106*, 10620–5.

(*15*) Sakata, K., Someya, M., Matsumoto, Y., and Hareyama, M. (2007). Ability to repair DNA double-strand breaks related to cancer susceptibility and radiosensitivity. *Radiat Med 25*, 433–8.

(*16*) Boboila, C., Yan, C., Wesemann, D. R., Jankovic, M., Wang, J. H., Manis, J., Nussenzweig, A., Nussenzweig, M., and Alt, F. W. (2010). Alternative end-joining catalyzes class switch recombination in the absence of both Ku70 and DNA ligase 4. *J Exp Med 207*, 417–27.

(*17*) Simsek, D., and Jasin, M. (2010). Alternative end-joining is suppressed by the canonical NHEJ component Xrcc4-ligase IV during chromosomal translocation formation. *Nat Struct Mol Biol 17*, 410–6.

(*18*) Bunting, S. F., and Nussenzweig, A. (2013). End-joining, translocations and cancer. *Nat Rev Cancer 13*, 443–54.

(*19*) Weinstock, D. M., Elliott, B., and Jasin, M. (2006). A model of oncogenic rearrangements: differences between chromosomal translocation mechanisms and simple double-strand break repair. *Blood 107*, 777–80.

(*20*) Chiang, C. et al. (2012). Complex reorganization and predominant non-homologous repair following chromosomal breakage in karyotypically balanced germline rearrangements and transgenic integration. *Nat Genet 44*, 390–7, S1.

(*21*) Adamo, A., Collis, S. J., Adelman, C. A., Silva, N., Horejsi, Z., Ward, J. D., Martinez-Perez, E., Boulton, S. J., and La Volpe, A. (2010). Preventing nonhomologous end joining suppresses DNA repair defects of Fanconi anemia. *Mol Cell 39*, 25–35.

(*22*) Pace, P., Mosedale, G., Hodskinson, M. R., Rosado, I. V., Sivasubramaniam, M., and Patel, K. J. (2010). Ku70 corrupts DNA repair in the absence of the Fanconi anemia pathway. *Science 329*, 219–23.

(*23*) Rulten, S. L., and Grundy, G. J. (2017). Non-homologous end joining: Common interaction sites and exchange of multiple factors in the DNA repair process. *Bioessays 39*.

(*24*) Mari, P. O., Florea, B. I., Persengiev, S. P., Verkaik, N. S., Bruggenwirth, H. T., Modesti, M., Giglia-Mari, G., Bezstarosti, K., Demmers, J. A., Luider, T. M., Houtsmuller, A. B., and van Gent, D. C. (2006). Dynamic assembly of end-joining complexes requires interaction between Ku70/80 and XRCC4. *Proc Natl Acad Sci U S A 103*, 18597–602.

(*25*) Yano, K., and Chen, D. J. (2008). Live cell imaging of XLF and XRCC4 reveals a novel view of protein assembly in the non-homologous end-joining pathway. *Cell Cycle 7*, 1321–5.

(*26*) Stinson, B. M., and Loparo, J. J. (2021). Repair of DNA Double-Strand Breaks by the Nonhomologous End Joining Pathway. *Annu Rev Biochem 90*, 137–164.

(*27*) Walker, J. R., Corpina, R. A., and Goldberg, J. (2001). Structure of the Ku heterodimer bound to DNA and its implications for double-strand break repair. *Nature 412*, 607–614.

(*28*) Brown, J. S., Lukashchuk, N., Sczaniecka-Clift, M., Britton, S., le Sage, C., Calsou, P., Beli, P., Galanty, Y., and Jackson, S. P. (2015). Neddylation promotes ubiquitylation and release of Ku from DNA-damage sites. *Cell Rep 11*, 704–14.

(*29*) Feng, L., and Chen, J. (2012). The E3 ligase RNF8 regulates KU80 removal and NHEJ repair. *Nat Struct Mol Biol 19*, 201–6.

(*30*) van den Boom, J., Wolf, M., Weimann, L., Schulze, N., Li, F., Kaschani, F., Riemer, A., Zierhut, C., Kaiser, M., Iliakis, G., Funabiki, H., and Meyer, H. (2016). VCP/p97 Extracts Sterically Trapped Ku70/80 Rings from DNA in Double-Strand Break Repair. *Mol Cell 64*, 189–198.

(*31*) Graham, T. G., Walter, J. C., and Loparo, J. J. (2016). Two-Stage Synapsis of DNA Ends during Non-homologous End Joining. *Mol Cell 61*, 850–8.

(*32*) Lees-Miller, S. P., Chen, Y. R., and Anderson, C. W. (1990). Human cells contain a DNA-activated protein kinase that phosphorylates simian virus 40 T antigen, mouse p53, and the human Ku autoantigen. *Mol Cell Biol 10*, 6472–81.

(*33*) Jette, N., and Lees-Miller, S. P. (2015). The DNA-dependent protein kinase: A multifunctional protein kinase with roles in DNA double strand break repair and mitosis. *Prog Biophys Mol Biol 117*, 194–205.

(*34*) Chan, D. W., Ye, R., Veillette, C. J., and Lees-Miller, S. P. (1999). DNA-dependent protein kinase phosphorylation sites in Ku 70/80 heterodimer. *Biochemistry 38*, 1819–28.

(*35*) Douglas, P., Gupta, S., Morrice, N., Meek, K., and Lees-Miller, S. P. (2005). DNA-PK-dependent phosphorylation of Ku70/80 is not required for non-homologous end joining. *DNA Repair (Amst) 4*, 1006–18.

(*36*) Yu, Y. P., Wang, W., Ding, Q., Ye, R. Q., Chen, D., Merkle, D., Schriemer, D., Meek, K., and Lees-Miller, S. P. (2003). DNA-PK phosphorylation sites in XRCC4 are not required for survival after radiation or for V(D)J recombination. *DNA Repair 2*, 1239–1252.

(*37*) Normanno, D., Negrel, A., de Melo, A. J., Betzi, S., Meek, K., and Modesti, M. (2017). Mutational phospho-mimicry reveals a regulatory role for the XRCC4 and XLF C-terminal tails in modulating DNA bridging during classical non-homologous end joining. *Elife 6*.

(*38*) Yu, Y., Mahaney, B. L., Yano, K., Ye, R., Fang, S., Douglas, P., Chen, D. J., and Lees-Miller, S. P. (2008). DNA-PK and ATM phosphorylation sites in XLF/Cernunnos are not required for repair of DNA double strand breaks. *DNA Repair (Amst) 7*, 1680–92.

(*39*) Wang, Y. G., Nnakwe, C., Lane, W. S., Modesti, M., and Frank, K. M. (2004). Phosphorylation and regulation of DNA ligase IV stability by DNA-dependent protein kinase. *J Biol Chem 279*, 37282–90.

(*40*) Zolner, A. E., Abdou, I., Ye, R., Mani, R. S., Fanta, M., Yu, Y., Douglas, P., Tahbaz, N., Fang, S., Dobbs, T., Wang, C., Morrice, N., Hendzel, M. J., Weinfeld, M., and Lees-Miller, S. P. (2011). Phosphorylation of polynucleotide kinase/ phosphatase by DNA-dependent protein kinase and ataxia-telangiectasia mutated regulates its association with sites of DNA damage. *Nucleic Acids Res 39*, 9224–37.

(*41*) Hammel, M., Yu, Y., Radhakrishnan, S. K., Chokshi, C., Tsai, M. S., Matsumoto, Y., Kuzdovich, M., Remesh, S. G., Fang, S., Tomkinson, A. E., Lees-Miller, S. P., and Tainer, J. A. (2016). An Intrinsically Disordered APLF Links Ku, DNA-PKcs, and XRCC4-DNA Ligase IV in an Extended Flexible Non-homologous End Joining Complex. *J Biol Chem 291*, 26987–27006.

(*42*) Goodarzi, A. A., Yu, Y., Riballo, E., Douglas, P., Walker, S. A., Ye, R., Harer, C., Marchetti, C., Morrice, N., Jeggo, P. A., and Lees-Miller, S. P. (2006). DNA-PK autophosphorylation facilitates Artemis endonuclease activity. *EMBO J 25*, 3880–9.

(*43*) Kurimasa, A., Kumano, S., Boubnov, N. V., Story, M. D., Tung, C. S., Peterson, S. R., and Chen, D. J. (1999). Requirement for the kinase activity of human DNA-dependent protein kinase catalytic subunit in DNA strand break rejoining. *Mol Cell Biol 19*, 3877–84.

(*44*) Hornbeck, P. V., Zhang, B., Murray, B., Kornhauser, J. M., Latham, V., and Skrzypek, E. (2015). PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res 43*, D512–20.

(*45*) Douglas, P., Sapkota, G. P., Morrice, N., Yu, Y., Goodarzi, A. A., Merkle, D., Meek, K., Alessi, D. R., and Lees-Miller, S. P. (2002). Identification of in vitro and in vivo phosphorylation sites in the catalytic subunit of the DNA-dependent protein kinase. *Biochem J 368*, 243–51.

(*46*) Chan, D. W., Chen, B. P., Prithivirajsingh, S., Kurimasa, A., Story, M. D., Qin, J., and Chen, D. J. (2002). Autophosphorylation of the DNA-dependent protein kinase catalytic subunit is required for rejoining of DNA double-strand breaks. *Genes Dev 16*, 2333–8.

(*47*) Ding, Q., Reddy, Y. V., Wang, W., Woods, T., Douglas, P., Ramsden, D. A., Lees-Miller, S. P., and Meek, K. (2003). Autophosphorylation of the catalytic subunit of the DNA-dependent protein kinase is required for efficient end processing during DNA double-strand break repair. *Mol Cell Biol 23*, 5836–48.

(*48*) Chen, B. P., Chan, D. W., Kobayashi, J., Burma, S., Asaithamby, A., Morotomi-Yano, K., Botvinick, E., Qin, J., and Chen, D. J. (2005). Cell cycle dependence of DNA-dependent protein kinase phosphorylation in response to DNA double strand breaks. *J Biol Chem 280*, 14709–15.

(*49*) Cui, X., Yu, Y., Gupta, S., Cho, Y. M., Lees-Miller, S. P., and Meek, K. (2005). Autophosphorylation of DNA-dependent protein kinase regulates DNA end processing and may also alter double-strand break repair pathway choice. *Mol Cell Biol 25*, 10842–52.

(*50*) Neal, J. A., and Meek, K. (2019). Deciphering phenotypic variance in different models of DNA-PKcs deficiency. *DNA Repair (Amst) 73*, 7–16.

(*51*) Douglas, P., Cui, X., Block, W. D., Yu, Y., Gupta, S., Ding, Q., Ye, R., Morrice, N., Lees-Miller, S. P., and Meek, K. (2007). The DNA-dependent protein kinase catalytic subunit is phosphorylated in vivo on threonine 3950, a highly conserved amino acid in the protein kinase domain. *Mol Cell Biol 27*, 1581–91.

(*52*) Jiang, W., Crowe, J. L., Liu, X., Nakajima, S., Wang, Y., Li, C., Lee, B. J., Dubois, R. L., Liu, C., Yu, X., Lan, L., and Zha, S. (2015). Differential phosphorylation of DNA-PKcs regulates the interplay between end-processing and end-ligation during nonhomologous end-joining. *Mol Cell 58*, 172–85.

(*53*) Convery, E., Shin, E. K., Ding, Q., Wang, W., Douglas, P., Davis, L. S., Nickoloff, J. A., Lees-Miller, S. P., and Meek, K. (2005). Inhibition of homologous recombination by variants of the catalytic subunit of the DNA-dependent protein kinase (DNA-PKcs). *Proc Natl Acad Sci U S A 102*, 1345–50.

(*54*) Neal, J. A., Dang, V., Douglas, P., Wold, M. S., Lees-Miller, S. P., and Meek, K. (2011). Inhibition of homologous recombination by DNA-dependent protein kinase requires kinase activity, is titratable, and is modulated by autophosphorylation. *Mol Cell Biol 31*, 1719–33.

(*55*) Chan, D. W., and Lees-Miller, S. P. (1996). The DNA-dependent protein kinase is inactivated by autophosphorylation of the catalytic subunit. *J Biol Chem 271*, 8936–41.

(*56*) Uematsu, N., Weterings, E., Yano, K., Morotomi-Yano, K., Jakob, B., Taucher-Scholz, G., Mari, P. O., van Gent, D. C., Chen, B. P., and Chen, D. J. (2007). Autophosphorylation of DNA-PKCS regulates its dynamics at DNA double-strand breaks. *J Cell Biol 177*, 219–29.

(*57*) Hammel, M., Yu, Y., Mahaney, B. L., Cai, B., Ye, R., Phipps, B. M., Rambo, R. P., Hura, G. L., Pelikan, M., So, S., Abolfath, R. M., Chen, D. J., Lees-Miller, S. P., and Tainer, J. A. (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*, 1414–23.

(*58*) Dobbs, T. A., Tainer, J. A., and Lees-Miller, S. P. (2010). A structural model for regulation of NHEJ by DNA-PKcs autophosphorylation. *DNA Repair (Amst) 9*, 1307–14.

(*59*) Chen, S., Lees-Miller, J. P., He, Y., and Lees-Miller, S. P. (2021). Structural insights into the role of DNA-PK as a master regulator in NHEJ. *Genome Instab Dis 2*, 195–210.

(*60*) Meek, K., Douglas, P., Cui, X., Ding, Q., and Lees-Miller, S. P. (2007). trans Autophosphorylation at DNA-dependent protein kinase's two major autophosphorylation site clusters facilitates end processing but not end joining. *Mol Cell Biol 27*, 3881–90.

(*61*) Sibanda, B. L., Chirgadze, D. Y., Ascher, D. B., and Blundell, T. L. (2017). DNA-PKcs structure suggests an allosteric mechanism modulating DNA double-strand break repair. *Science 355*, 520–524.

(*62*) Chaplin, A. K., Hardwick, S. W., Liang, S., Kefala Stavridi, A., Hnizda, A., Cooper, L. R., De Oliveira, T. M., Chirgadze, D. Y., and Blundell, T. L. (2021). Dimers of DNA-PK create a stage for DNA double-strand break repair. *Nat Struct Mol Biol 28*, 13–19.

(*63*) Junop, M. S., Modesti, M., Guarne, A., Ghirlando, R., Gellert, M., and Yang, W. (2000). Crystal structure of the Xrcc4 DNA repair protein and implications for end joining. *EMBO J 19*, 5962–70.

(*64*) Andres, S. N., Modesti, M., Tsai, C. J., Chu, G., and Junop, M. S. (2007). Crystal structure of human XLF: a twist in nonhomologous DNA end-joining. *Mol Cell 28*, 1093–101.

(*65*) Andres, S. N., Vergnes, A., Ristic, D., Wyman, C., Modesti, M., and Junop, M. (2012). A human XRCC4-XLF complex bridges DNA. *Nucleic Acids Res 40*, 1868–78.

(*66*) Ochi, T., Wu, Q., and Blundell, T. L. (2014). The spatial organization of non-homologous end joining: from bridging to end joining. *DNA Repair (Amst) 17*, 98–109.

(*67*) Chen, S., Lee, L., Naila, T., Fishbain, S., Wang, A., Tomkinson, A. E., Lees-Miller, S. P., and He, Y. (2021). Structural basis of long-range to short-range synaptic transition in NHEJ. *Nature 593*, 294–298.

(*68*) Wu, P. Y., Frit, P., Meesala, S., Dauvillier, S., Modesti, M., Andres, S. N., Huang, Y., Sekiguchi, J., Calsou, P., Salles, B., and Junop, M. S. (2009). Structural and functional interaction between the human DNA repair proteins DNA ligase IV and XRCC4. *Mol Cell Biol 29*, 3163–72.

(*69*) Kaminski, A. M., Tumbale, P. P., Schellenberg, M. J., Williams, R. S., Williams, J. G., Kunkel, T. A., Pedersen, L. C., and Bebenek, K. (2018). Structures of DNA-bound human ligase IV catalytic core reveal insights into substrate binding and catalysis. *Nat Commun 9*, 2642.

(*70*) Dynan, W. S., and Yoo, S. (1998). Interaction of Ku protein and DNA-dependent protein kinase catalytic subunit with nucleic acids. *Nucleic Acids Res 26*, 1551–9.

(*71*) Frit, P., Ropars, V., Modesti, M., Charbonnier, J. B., and Calsou, P. (2019). Plugged into the Ku-DNA hub: The NHEJ network. *Prog Biophys Mol Biol 147*, 62–76.

(*72*) Grundy, G. J., Moulding, H. A., Caldecott, K. W., and Rulten, S. L. (2014). One ring to bring them all–the role of Ku in mammalian non-homologous end joining. *DNA Repair (Amst) 17*, 30–8.

(*73*) Kim, K., Min, J., Kirby, T. W., Gabel, S. A., Pedersen, L. C., and London, R. E. (2020). Ligand binding characteristics of the Ku80 von Willebrand domain. *DNA Repair (Amst) 85*, 102739.

(*74*) Falck, J., Coates, J., and Jackson, S. P. (2005). Conserved modes of recruitment of ATM, ATR and DNA-PKcs to sites of DNA damage. *Nature 434*, 605–11.

(*75*) Goodarzi, A. A., and Lees-Miller, S. P. (2004). Biochemical characterization of the ataxia-telangiectasia mutated (ATM) protein from human cells. *DNA Repair (Amst) 3*, 753–67.

(*76*) Boskovic, J., Rivera-Calzada, A., Maman, J. D., Chacon, P., Willison, K. R., Pearl, L. H., and Llorca, O. (2003). Visualization of DNA-induced conformational changes in the DNA repair kinase DNA-PKcs. *EMBO J 22*, 5875–82.

(*77*) Rivera-Calzada, A., Maman, J. D., Spagnolo, L., Pearl, L. H., and Llorca, O. (2005). Three-dimensional structure and regulation of the DNA-dependent protein kinase catalytic subunit (DNA-PKcs). *Structure 13*, 243–55.

(*78*) Williams, D. R., Lee, K. J., Shi, J., Chen, D. J., and Stewart, P. L. (2008). Cryo-EM structure of the DNA-dependent protein kinase catalytic subunit at subnanometer resolution reveals alpha helices and insight into DNA binding. *Structure 16*, 468–77.

(*79*) Sibanda, B. L., Chirgadze, D. Y., and Blundell, T. L. (2010). Crystal structure of DNA-PKcs reveals a large open-ring cradle comprised of HEAT repeats. *Nature 463*, 118–21.

(*80*) Morris, E. P., Rivera-Calzada, A., da Fonseca, P. C., Llorca, O., Pearl, L. H., and Spagnolo, L. (2011). Evidence for a remodelling of DNA-PK upon autophosphorylation from electron microscopy studies. *Nucleic Acids Res 39*, 5757–67.

(*81*) Villarreal, S. A., and Stewart, P. L. (2014). CryoEM and image sorting for flexible protein/DNA complexes. *J Struct Biol 187*, 76–83.

(*82*) Sharif, H., Li, Y., Dong, Y., Dong, L., Wang, W. L., Mao, Y., and Wu, H. (2017). Cryo-EM structure of the DNA-PK holoenzyme. *Proc Natl Acad Sci U S A 114*, 7367–7372.

(*83*) Chen, X., Xu, X., Chen, Y., Cheung, J. C., Wang, H., Jiang, J., de Val, N., Fox, T., Gellert, M., and Yang, W. (2021). Structure of an activated DNA-PK and its implications for NHEJ. *Mol Cell 81*, 801–810 e3.

(*84*) Yin, X., Liu, M., Tian, Y., Wang, J., and Xu, Y. (2017). Cryo-EM structure of human DNA-PK holoenzyme. *Cell Res 27*, 1341–1350.

(*85*) Yoo, S., and Dynan, W. S. (1999). Geometry of a complex formed by double strand break repair proteins at a single DNA end: recruitment of DNA-PKcs induces inward translocation of Ku protein. *Nucleic Acids Research 27*, 4679–4686.

(*86*) Critchlow, S. E., Bowater, R. P., and Jackson, S. P. (1997). Mammalian DNA double-strand break repair protein XRCC4 interacts with DNA ligase IV. *Current Biology 7*, 588–598.

(*87*) Gao, Y. et al. (1998). A Critical Role for DNA End-Joining Proteins in Both Lymphogenesis and Neurogenesis. *Cell 95*, 891–902.

(*88*) Callebaut, I., Malivert, L., Fischer, A., Mornon, J. P., Revy, P., and de Villartay, J. P. (2006). Cernunnos interacts with the XRCC4 x DNA-ligase IV complex and is homologous to the yeast nonhomologous end-joining factor Nej1. *J Biol Chem 281*, 13857–60.

(*89*) Ahnesorg, P., Smith, P., and Jackson, S. P. (2006). XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining. *Cell 124*, 301–13.

(*90*) Yano, K., Morotomi-Yano, K., Wang, S. Y., Uematsu, N., Lee, K. J., Asaithamby, A., Weterings, E., and Chen, D. J. (2008). Ku recruits XLF to DNA double-strand breaks. *EMBO Rep 9*, 91–6.

(*91*) Yano, K., Morotomi-Yano, K., Lee, K. J., and Chen, D. J. (2011). Functional significance of the interaction with Ku in DNA double-strand break recognition of XLF. *FEBS Lett 585*, 841–6.

(*92*) Carney, S. M., Moreno, A. T., Piatt, S. C., Cisneros-Aguirre, M., Lopezcolorado, F. W., Stark, J. M., and Loparo, J. J. (2020). XLF acts as a flexible connector during non-homologous end joining. *Elife 9*.

(*93*) Nemoz, C. et al. (2018). XLF and APLF bind Ku80 at two remote sites to ensure DNA repair by non-homologous end joining. *Nat Struct Mol Biol 25*, 971–980.

(*94*) Hammel, M., Rey, M., Yu, Y., Mani, R. S., Classen, S., Liu, M., Pique, M. E., Fang, S., Mahaney, B. L., Weinfeld, M., Schriemer, D. C., Lees-Miller, S. P., and Tainer, J. A. (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 286*, 32638–50.

(*95*) Ropars, V., Drevet, P., Legrand, P., Baconnais, S., Amram, J., Faure, G., Marquez, J. A., Pietrement, O., Guerois, R., Callebaut, I., Le Cam, E., Revy, P., de Villartay, J. P., and Charbonnier, J. B. (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*, 12663–8.

(*96*) Roy, S., Andres, S. N., Vergnes, A., Neal, J. A., Xu, Y., Yu, Y., Lees-Miller, S. P., Junop, M., Modesti, M., and Meek, K. (2012). XRCC4's interaction with XLF is required for coding (but not signal) end joining. *Nucleic Acids Res 40*, 1684–94.

(*97*) Hammel, M., Yu, Y., Fang, S., Lees-Miller, S. P., and Tainer, J. A. (2010). XLF regulates filament architecture of the XRCC4.ligase IV complex. *Structure 18*, 1431–42.

(*98*) DeFazio, L. G., Stansel, R. M., Griffith, J. D., and Chu, G. (2002). Synapsis of DNA ends by DNA-dependent protein kinase. *Embo Journal 21*, 3192–3200.

(*99*) Spagnolo, L., Rivera-Calzada, A., Pearl, L. H., 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 22*, 511–9.

(*100*) Mahaney, B. L., Hammel, M., Meek, K., Tainer, J. A., and Lees-Miller, S. P. (2013). XRCC4 and XLF form long helical protein filaments suitable for DNA end protection and alignment to facilitate DNA double strand break repair. *Biochem Cell Biol 91*, 31–41.

(*101*) Reid, D. A., Keegan, S., Leo-Macias, A., Watanabe, G., Strande, N. T., Chang, H. H., Oksuz, B. A., Fenyo, D., Lieber, M. R., Ramsden, D. A., and Rothenberg, E. (2015). Organization and dynamics of the nonhomologous end-joining machinery during DNA double-strand break repair. *Proc Natl Acad Sci U S A 112*, E2575–84.

(*102*) Graham, T. G. W., Carney, S. M., Walter, J. C., and Loparo, J. J. (2018). A single XLF dimer bridges DNA ends during nonhomologous end joining. *Nat Struct Mol Biol 25*, 877–884.

(*103*) Grawunder, U., Wilm, M., Wu, X., Kulesza, P., Wilson, T. E., Mann, M., and Lieber, M. R. (1997). Activity of DNA ligase IV stimulated by complex formation with XRCC4 protein in mammalian cells. *Nature 388*, 492–5.

(*104*) Grawunder, U., Zimmer, D., Fugmann, S., Schwarz, K., and Lieber, M. R. (1998). DNA ligase IV is essential for V(D)J recombination and DNA double-strand break repair in human precursor lymphocytes. *Molecular Cell 2*, 477–484.

(*105*) Li, F., Zhang, Y., Bai, J., Greenberg, M. M., Xi, Z., and Zhou, C. (2017). 5-Formylcytosine Yields DNA-Protein Cross-Links in Nucleosome Core Particles. *J Am Chem Soc 139*, 10617–10620.

(*106*) Chen, L., Trujillo, K., Sung, P., and Tomkinson, A. E. (2000). Interactions of the DNA ligase IV-XRCC4 complex with DNA ends and the DNA-dependent protein kinase. *J Biol Chem 275*, 26196–205.

(*107*) Riballo, E., Woodbine, L., Stiff, T., Walker, S. A., Goodarzi, A. A., and Jeggo, P. A. (2009). XLF-Cernunnos promotes DNA ligase IV-XRCC4 re-adenylation following ligation. *Nucleic Acids Res 37*, 482–92.

(*108*) Altmann, T., and Gennery, A. R. (2016). DNA ligase IV syndrome; a review. *Orphanet J Rare Dis 11*, 137.

(*109*) Gu, J., Lu, H., Tippin, B., Shimazaki, N., Goodman, M. F., and Lieber, M. R. (2007). XRCC4:DNA ligase IV can ligate incompatible DNA ends and can ligate across gaps. *EMBO J 26*, 1010–23.

(*110*) Gu, J., Lu, H., Tsai, A. G., Schwarz, K., and Lieber, M. R. (2007). Single-stranded DNA ligation and XLF-stimulated incompatible DNA end ligation by the XRCC4-DNA ligase IV complex: influence of terminal DNA sequence. *Nucleic Acids Res 35*, 5755–62.

(*111*) Wang, Y., Lamarche, B. J., and Tsai, M. D. (2007). Human DNA ligase IV and the ligase IV/XRCC4 complex: analysis of nick ligation fidelity. *Biochemistry 46*, 4962–76.

(*112*) Chen, X., Ballin, J. D., Della-Maria, J., Tsai, M. S., White, E. J., Tomkinson, A. E., and Wilson, G. M. (2009). Distinct kinetics of human DNA ligases I, IIIalpha, IIIbeta, and IV reveal direct DNA sensing ability and differential physiological functions in DNA repair. *DNA Repair (Amst) 8*, 961–8.

(*113*) Xing, M. et al. (2015). Interactome analysis identifies a new paralogue of XRCC4 in non-homologous end joining DNA repair pathway. *Nat Commun 6*, 6233.

(*114*) Ochi, T., Blackford, A. N., Coates, J., Jhujh, S., Mehmood, S., Tamura, N., Travers, J., Wu, Q., Draviam, V. M., Robinson, C. V., Blundell, T. L., and Jackson, S. P. (2015). DNA repair. PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair. *Science 347*, 185–188.

(*115*) Liu, X., Shao, Z., Jiang, W., Lee, B. J., and Zha, S. (2017). PAXX promotes KU accumulation at DNA breaks and is essential for end-joining in XLF-deficient mice. *Nat Commun 8*, 13816.

(*116*) Tadi, S. K., Tellier-Lebegue, C., Nemoz, C., Drevet, P., Audebert, S., Roy, S., Meek, K., Charbonnier, J. B., 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.

(*117*) Kurosawa, A., and Adachi, N. (2010). Functions and regulation of Artemis: a goddess in the maintenance of genome integrity. *J Radiat Res 51*, 503–9.

(*118*) Cooper, M. P., Machwe, A., Orren, D. K., Brosh, R. M., Ramsden, D., and Bohr, V. A. (2000). Ku complex interacts with and stimulates the Werner protein. *Genes & Development 14*, 907–912.

(*119*) Ahel, I., Rass, U., El-Khamisy, S. F., Katyal, S., Clements, P. M., McKinnon, P. J., Caldecott, K. W., and West, S. C. (2006). The neurodegenerative disease protein aprataxin resolves abortive DNA ligation intermediates. *Nature 443*, 713–6.

(*120*) Ma, Y., Pannicke, U., Schwarz, K., and Lieber, M. R. (2002). Hairpin opening and overhang processing by an Artemis/DNA-dependent protein kinase complex in nonhomologous end joining and V(D)J recombination. *Cell 108*, 781–94.

(*121*) Yosaatmadja, Y., Baddock, H. T., Newman, J. A., Bielinski, M., Gavard, A. E., Mukhopadhyay, S. M. M., Dannerfjord, A. A., Schofield, C. J., McHugh, P. J., and Gileadi, O. (2021). Structural and mechanistic insights into the Artemis endonuclease and strategies for its inhibition. *Nucleic Acids Res 49*, 9310–9326.

(*122*) Liu, L., Chen, X., Li, J., Wang, H., Buehl, C. J., Goff, N. J., Meek, K., Yang, W., and Gellert, M. (2021). Autophosphorylation transforms DNA-PK from protecting to processing DNA ends. *Mol Cell.*

(*123*) Han, Y., Reyes, A. A., Malik, S., and He, Y. (2020). Cryo-EM structure of SWI/SNF complex bound to a nucleosome. *Nature 579*, 452–455.

(*124*) Nogales, E., and Scheres, S. H. (2015). Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity. *Mol Cell 58*, 677–89.

(*125*) Bai, X. C., McMullan, G., and Scheres, S. H. (2015). How cryo-EM is revolutionizing structural biology. *Trends Biochem Sci 40*, 49–57.

(*126*) Kuhlbrandt, W. (2014). Cryo-EM enters a new era. *Elife 3*, e03678.

(*127*) Li, X., Mooney, P., Zheng, S., Booth, C. R., Braunfeld, M. B., Gubbens, S., Agard, D. A., and Cheng, Y. (2013). Electron counting and beam-induced motion correction enable near-atomic-resolution single-particle cryo-EM. *Nat Methods 10*, 584–90.

(*128*) Liao, M., Cao, E., Julius, D., and Cheng, Y. (2013). Structure of the TRPV1 ion channel determined by electron cryo-microscopy. *Nature 504*, 107–12.

(*129*) Fischer, N., Neumann, P., Konevega, A. L., Bock, L. V., Ficner, R., Rodnina, M. V., and Stark, H. (2015). Structure of the E. coli ribosome-EF-Tu complex at ¡3 A resolution by Cs-corrected cryo-EM. *Nature 520*, 567–70.

(*130*) Abdella, R., Talyzina, A., Chen, S., Inouye, C. J., Tjian, R., and He, Y. (2021). Structure of the human Mediator-bound transcription preinitiation complex. *Science 372*, 52–56.

(*131*) Ohi, M., Li, Y., Cheng, Y., and Walz, T. (2004). Negative Staining and Image Classification - Powerful Tools in Modern Electron Microscopy. *Biol Proced Online 6*, 23–34.

(*132*) Suloway, C., Pulokas, J., Fellmann, D., Cheng, A., Guerra, F., Quispe, J., Stagg, S., Potter, C. S., and Carragher, B. (2005). Automated molecular microscopy: the new Leginon system. *J Struct Biol 151*, 41–60.

(*133*) Lander, G. C., Stagg, S. M., Voss, N. R., Cheng, A., Fellmann, D., Pulokas, J., Yoshioka, C., Irving, C., Mulder, A., Lau, P. W., Lyumkis, D., Potter, C. S., and Carragher, B. (2009). Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J Struct Biol 166*, 95–102.

(*134*) Voss, N. R., Yoshioka, C. K., Radermacher, M., Potter, C. S., and Carragher, B. (2009). DoG Picker and TiltPicker: software tools to facilitate particle selection in single particle electron microscopy. *J Struct Biol 166*, 205–13.

(*135*) Mindell, J. A., and Grigorieff, N. (2003). Accurate determination of local defocus and specimen tilt in electron microscopy. *Journal of Structural Biology 142*, 334–347.

(*136*) van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R., and Schatz, M. (1996). A new generation of the IMAGIC image processing system. *J Struct Biol 116*, 17–24.

(*137*) Tang, G., Peng, L., Baldwin, P. R., Mann, D. S., Jiang, W., Rees, I., and Ludtke, S. J. (2007). EMAN2: an extensible image processing suite for electron microscopy. *J Struct Biol 157*, 38–46.

(*138*) Zivanov, J., Nakane, T., Forsberg, B. O., Kimanius, D., Hagen, W. J., Lindahl, E., and Scheres, S. H. (2018). New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife 7*.

(*139*) Punjani, A., Rubinstein, J. L., Fleet, D. J., and Brubaker, M. A. (2017). cryoSPARC: algorithms for rapid unsupervised cryo-EM structure determination. *Nat Methods 14*, 290–296.

(*140*) Zheng, S. Q., Palovcak, E., Armache, J. P., Verba, K. A., Cheng, Y., and Agard, D. A. (2017). MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat Methods 14*, 331–332.

(*141*) Zhang, K. (2016). Gctf: Real-time CTF determination and correction. *J Struct Biol 193*, 1–12.

(*142*) Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., and Ferrin, T. E. (2004). UCSF Chimera–a visualization system for exploratory research and analysis. *J Comput Chem 25*, 1605–12.

(*143*) Pettersen, E. F., Goddard, T. D., Huang, C. C., Meng, E. C., Couch, G. S., Croll, T. I., Morris, J. H., and Ferrin, T. E. (2021). UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci 30*, 70–82.

(*144*) Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr D Biol Crystallogr 66*, 486–501.

(*145*) Croll, T. I. (2018). ISOLDE: a physically realistic environment for model building into low-resolution electron-density maps. *Acta Crystallogr D Struct Biol 74*, 519–530.

(*146*) Yang, J., Anishchenko, I., Park, H., Peng, Z., Ovchinnikov, S., and Baker, D. (2020). Improved protein structure prediction using predicted interresidue orientations. *Proc Natl Acad Sci U S A 117*, 1496–1503.

(*147*) Jumper, J. et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature 596*, 583–589.

(*148*) Afonine, P. V., Poon, B. K., Read, R. J., Sobolev, O. V., Terwilliger, T. C., Urzhumtsev, A., and Adams, P. D. (2018). Real-space refinement in PHENIX for cryo-EM and crystallography. *Acta Crystallogr D Struct Biol 74*, 531–544.

(*149*) Zhao, B., Watanabe, G., Morten, M. J., Reid, D. A., Rothenberg, E., and Lieber, M. R. (2019). The essential elements for the noncovalent association of two DNA ends during NHEJ synapsis. *Nat Commun 10*, 3588.

(*150*) Scheres, S. H., and Chen, S. (2012). Prevention of overfitting in cryo-EM structure determination. *Nat Methods 9*, 853–4.

(*151*) Kucukelbir, A., Sigworth, F. J., and Tagare, H. D. (2014). Quantifying the local resolution of cryo-EM density maps. *Nat Methods 11*, 63–5.

(*152*) Sanchez-Garcia, R., Gomez-Blanco, J., Cuervo, A., Carazo, J. M., Sorzano, C. O. S., and Vargas, J. (2021). DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun Biol 4*, 874.

(*153*) Yang, H., Rudge, D. G., Koos, J. D., Vaidialingam, B., Yang, H. J., and Pavletich, N. P. (2013). mTOR kinase structure, mechanism and regulation. *Nature 497*, 217–23.

(*154*) Andres, S. N., and Junop, M. S. (2011). Crystallization and preliminary X-ray diffraction analysis of the human XRCC4-XLF complex. *Acta Crystallogr Sect F Struct Biol Cryst Commun 67*, 1399–402.

(*155*) Dore, A. S., Furnham, N., Davies, O. R., Sibanda, B. L., Chirgadze, D. Y., Jackson, S. P., Pellegrini, L., and Blundell, T. L. (2006). Structure of an Xrcc4-DNA ligase IV yeast ortholog complex reveals a novel BRCT interaction mode. *DNA Repair (Amst) 5*, 362–8.

(*156*) Costantini, S., Woodbine, L., Andreoli, L., Jeggo, P. A., 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*, 712–22.

(*157*) Tate, J. G. et al. (2019). COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Res 47*, D941–D947.

(*158*) Lees-Miller, J. P., Cobban, A., Katsonis, P., Bacolla, A., Tsutakawa, S. E., Hammel, M., Meek, K., Anderson, D. W., Lichtarge, O., Tainer, J. A., and Lees-Miller, S. P. (2021). Uncovering DNA-PKcs ancient phylogeny, unique sequence motifs and insights for human disease. *Prog Biophys Mol Biol 163*, 87–108.

(*159*) Bao, Z. Q., Jacobsen, D. M., and Young, M. A. (2011). Briefly bound to activate: transient binding of a second catalytic magnesium activates the structure and dynamics of CDK2 kinase for catalysis. *Structure 19*, 675–90.

(*160*) Baretic, D., Pollard, H. K., Fisher, D. I., Johnson, C. M., Santhanam, B., Truman, C. M., Kouba, T., Fersht, A. R., Phillips, C., and Williams, R. L. (2017). Structures of closed and open conformations of dimeric human ATM. *Sci Adv 3*, e1700933.

(*161*) Rao, Q., Liu, M., Tian, Y., Wu, Z., Hao, Y., Song, L., Qin, Z., Ding, C., Wang, H. W., Wang, J., and Xu, Y. (2018). Cryo-EM structure of human ATR-ATRIP complex. *Cell Res 28*, 143–156.

(*162*) Mastronarde, D. N. (2005). Automated electron microscope tomography using robust prediction of specimen movements. *J Struct Biol 152*, 36–51.

(*163*) Chang, H. H., Watanabe, G., and Lieber, M. R. (2015). Unifying the DNA end-processing roles of the artemis nuclease: Ku-dependent artemis resection at blunt DNA ends. *J Biol Chem 290*, 24036–50.

(*164*) Stinson, B. M., Moreno, A. T., Walter, J. C., and Loparo, J. J. (2020). A Mechanism to Minimize Errors during Non-homologous End Joining. *Mol Cell 77*, 1080–1091 e8.

(*165*) Riballo, E., Doherty, A. J., Dai, Y., Stiff, T., Oettinger, M. A., Jeggo, P. A., and Kysela, B. (2001). Cellular and biochemical impact of a mutation in DNA ligase IV conferring clinical radiosensitivity. *J Biol Chem 276*, 31124–32.

(*166*) Chang, H. H. Y., Watanabe, G., Gerodimos, C. A., Ochi, T., Blundell, T. L., Jackson, S. P., and Lieber, M. R. (2016). Different DNA End Configurations Dictate Which NHEJ Components Are Most Important for Joining Efficiency. *J Biol Chem 291*, 24377–24389.

(*167*) Tsai, C. J., Kim, S. A., and Chu, G. (2007). Cernunnos/XLF promotes the ligation of mismatched and noncohesive DNA ends. *Proc Natl Acad Sci U S A 104*, 7851–6.

(*168*) Gerodimos, C. A., Chang, H. H. Y., Watanabe, G., and Lieber, M. R. (2017). Effects of DNA end configuration on XRCC4-DNA ligase IV and its stimulation of Artemis activity. *J Biol Chem 292*, 13914–13924.

(*169*) Hung, P. J., Chen, B. R., George, R., Liberman, C., Morales, A. J., Colon-Ortiz, P., Tyler, J. K., Sleckman, B. P., and Bredemeyer, A. L. (2017). Deficiency of XLF and PAXX prevents DNA double-strand break repair by non-homologous end joining in lymphocytes. *Cell Cycle 16*, 286–295.

(*170*) Chaplin, A. K., Hardwick, S. W., Stavridi, A. K., Buehl, C. J., Goff, N. J., Ropars, V., Liang, S., De Oliveira, T. M., Chirgadze, D. Y., Meek, K., Charbonnier, J. B., and Blundell, T. L. (2021). Cryo-EM of NHEJ supercomplexes provides insights into DNA repair. *Mol Cell 81*, 3400–3409 e3.

(*171*) Kidmose, R. T., Juhl, J., Nissen, P., Boesen, T., Karlsen, J. L., and Pedersen, B. P. (2019). Namdinator - automatic molecular dynamics flexible fitting of structural models into cryo-EM and crystallography experimental maps. *IUCrJ 6*, 526–531.

(*172*) Kysela, B., Doherty, A. J., Chovanec, M., Stiff, T., Ameer-Beg, S. M., Vojnovic, B., Girard, P. M., and Jeggo, P. A. (2003). Ku stimulation of DNA ligase IV-dependent ligation requires inward movement along the DNA molecule. *J Biol Chem 278*, 22466–74.

(*173*) Lau, W. C., Li, Y., Liu, Z., Gao, Y., Zhang, Q., and Huen, M. S. (2016). Structure of the human dimeric ATM kinase. *Cell Cycle 15*, 1117–24.

(*174*) Lee, J. H., and Paull, T. T. (2005). ATM activation by DNA double-strand breaks through the Mre11-Rad50-Nbs1 complex. *Science 308*, 551–4.

(*175*) Lee, J. H., Goodarzi, A. A., Jeggo, P. A., and Paull, T. T. (2010). 53BP1 promotes ATM activity through direct interactions with the MRN complex. *EMBO J 29*, 574–85.

(*176*) Sartori, A. A., Lukas, C., Coates, J., Mistrik, M., Fu, S., Bartek, J., Baer, R., Lukas, J., and Jackson, S. P. (2007). Human CtIP promotes DNA end resection. *Nature 450*, 509–14.

(*177*) You, Z., Shi, L. Z., Zhu, Q., Wu, P., Zhang, Y. W., Basilio, A., Tonnu, N., Verma, I. M., Berns, M. W., and Hunter, T. (2009). CtIP links DNA double-strand break sensing to resection. *Mol Cell 36*, 954–69.

(*178*) Wittkopp, P. J., Haerum, B. K., and Clark, A. G. (2004). Evolutionary changes in cis and trans gene regulation. *Nature 430*, 85–8.

(*179*) Cortez, D., Guntuku, S., Qin, J., and Elledge, S. J. (2001). ATR and ATRIP: partners in checkpoint signaling. *Science 294*, 1713–6.

(*180*) Smith, J., Tho, L. M., Xu, N., and Gillespie, D. A. (2010). The ATM-Chk2 and ATR-Chk1 pathways in DNA damage signaling and cancer. *Adv Cancer Res 108*, 73–112.

(*181*) Delacroix, S., Wagner, J. M., Kobayashi, M., Yamamoto, K., and Karnitz, L. M. (2007). The Rad9-Hus1-Rad1 (9-1-1) clamp activates checkpoint signaling via TopBP1. *Genes Dev 21*, 1472–7.

(*182*) Lee, J., Kumagai, A., and Dunphy, W. G. (2007). The Rad9-Hus1-Rad1 checkpoint clamp regulates interaction of TopBP1 with ATR. *J Biol Chem 282*, 28036–44.

(*183*) Kumagai, A., Lee, J., Yoo, H. Y., and Dunphy, W. G. (2006). TopBP1 activates the ATR-ATRIP complex. *Cell 124*, 943–55.

(*184*) Bass, T. E., Luzwick, J. W., Kavanaugh, G., Carroll, C., Dungrawala, H., Glick, G. G., Feldkamp, M. D., Putney, R., Chazin, W. J., and Cortez, D. (2016). ETAA1 acts at stalled replication forks to maintain genome integrity. *Nat Cell Biol 18*, 1185–1195.

(*185*) Haahr, P., Hoffmann, S., Tollenaere, M. A., Ho, T., Toledo, L. I., Mann, M., Bekker-Jensen, S., Raschle, M., and Mailand, N. (2016). Activation of the ATR kinase by the RPA-binding protein ETAA1. *Nat Cell Biol 18*, 1196–1207.

(*186*) Lee, Y. C., Zhou, Q., Chen, J., and Yuan, J. (2016). RPA-Binding Protein ETAA1 Is an ATR Activator Involved in DNA Replication Stress Response. *Curr Biol 26*, 3257–3268.

(*187*) Blackford, A. N., and Jackson, S. P. (2017). ATM, ATR, and DNA-PK: The Trinity at the Heart of the DNA Damage Response. *Mol Cell 66*, 801–817.

(*188*) Saldivar, J. C., Cortez, D., and Cimprich, K. A. (2017). The essential kinase ATR: ensuring faithful duplication of a challenging genome. *Nat Rev Mol Cell Biol 18*, 622–636.

(*189*) Meek, K. (2020). Activation of DNA-PK by hairpinned DNA ends reveals a stepwise mechanism of kinase activation. *Nucleic Acids Res 48*, 9098–9108.

(*190*) Liu, S., Opiyo, S. O., Manthey, K., Glanzer, J. G., Ashley, A. K., Amerin, C., Troksa, K., Shrivastav, M., Nickoloff, J. A., and Oakley, G. G. (2012). Distinct roles for DNA-PK, ATM and ATR in RPA phosphorylation and checkpoint activation in response to replication stress. *Nucleic Acids Res 40*, 10780–94.

(*191*) Hendel, A., Fine, E. J., Bao, G., and Porteus, M. H. (2015). Quantifying on- and off-target genome editing. *Trends Biotechnol 33*, 132–40.

(*192*) Zischewski, J., Fischer, R., and Bortesi, L. (2017). Detection of on-target and off-target mutations generated by CRISPR/Cas9 and other sequence-specific nucleases. *Biotechnol Adv 35*, 95–104.

(*193*) Lin, C., Li, H., Hao, M., Xiong, D., Luo, Y., Huang, C., Yuan, Q., Zhang, J., and Xia, N. (2016). Increasing the Efficiency of CRISPR/Cas9-mediated Precise Genome Editing of HSV-1 Virus in Human Cells. *Sci Rep 6*, 34531.

(*194*) Ghezraoui, H., Piganeau, M., Renouf, B., Renaud, J. B., Sallmyr, A., Ruis, B., Oh, S., Tomkinson, A. E., Hendrickson, E. A., Giovannangeli, C., Jasin, M., and Brunet, E. (2014). Chromosomal translocations in human cells are generated by canonical nonhomologous end-joining. *Mol Cell 55*, 829–842.

(*195*) Su, T., Liu, F., Gu, P., Jin, H., Chang, Y., Wang, Q., Liang, Q., and Qi, Q. (2016). A CRISPR-Cas9 Assisted Non-Homologous End-Joining Strategy for One-step Engineering of Bacterial Genome. *Sci Rep 6*, 37895.

(*196*) Dynan, W. S., Takeda, Y., and Li, S. (2006). Modifying the function of DNA repair nanomachines for therapeutic benefit. *Nanomedicine 2*, 74–81.

(*197*) Davidson, D., Coulombe, Y., Martinez-Marignac, V. L., Amrein, L., Grenier, J., Hodkinson, K., Masson, J. Y., Aloyz, R., and Panasci, L. (2012). Irinotecan and DNA-PKcs inhibitors synergize in killing of colon cancer cells. *Invest New Drugs 30*, 1248–56.

(*198*) Srivastava, M., Nambiar, M., Sharma, S., Karki, S. S., Goldsmith, G., Hegde, M., Kumar, S., Pandey, M., Singh, R. K., Ray, P., Natarajan, R., Kelkar, M., De, A., Choudhary, B., and Raghavan, S. C. (2012). An inhibitor of nonhomologous end-joining abrogates double-strand break repair and impedes cancer progression. *Cell 151*, 1474–87.

(*199*) Cleary, J. M., Aguirre, A. J., Shapiro, G. I., and D'Andrea, A. D. (2020). Biomarker-Guided Development of DNA Repair Inhibitors. *Mol Cell 78*, 1070–1085.

STRUCTURAL VISUALIZATION OF STATE TRANSITION DURING NON-HOMOLOGOUS END-JOINING (NHEJ)

Approved by:

Ishwar Radhakrishnan
Department of Molecular Biosciences
Northwestern University

Jason Brickner
Department of Molecular Biosciences
Northwestern University

John Marko
Department of Molecular Biosciences
Northwestern University

Date Approved: February 16, 2022

VITA

SIYU CHEN

EDUCATION

Ph. D. 2022 Northwestern University *Interdisciplinary Biological Sciences*

B. Sc. 2014 Peking University *Life Sciences*

Dual-degree B. Sc. 2014 Peking University *Economics*

PUBLICATION

S. Chen, J. P. Lees-Miller, Y. He, S. Lees-Miller, “Structural insights into the role of DNA-PK as a master regulator in NHEJ”, *Genome Instab. Dis.* (2021).

S. Chen, L. Lee, T. Naila, S. Fishbain, A. Wang, A. Tomkinson, S. Lees-Miller, Y. He, “Structural basis of long-range to short-range synaptic transition in NHEJ”, *Nature* (2021).

R. Abdella, A. Talyzina, **S. Chen**, C. J. Inouye, R. Tjian, Y. He, “Structure of the human Mediator-bound transcription pre-initiation complex”, *Science* (2021).

G. Kenney, L. Dassama, A. Manesis, M. Ross, **S. Chen**, B. Hoffman, A. Rosenzweig, “MbnH is a diheme MauG-like protein associated with microbial copper homeostasis”. *J. Biol. Chem.* (2019).

PRESENTATION & POSTER

Poster presenter, “Structural basis of long-range to short-range synaptic transition in NHEJ”, GRC-3DEM, Waterville Valley, US, 2021

Plenary speaker, “Structural basis of long-range to short-range synaptic transition in NHEJ”, 12th MBTP Annual Biophysics Symposium, Evanston, 2021

Invited plenary speaker, “Cryo-EM visualization of DNA-PKcs auto-phosphorylation dependent state transition in NHEJ”, 3rd International Webinar on Nucleic Acids and CRISPR, Virtual, 2021

Poster presenter, “Structural basis of long-range to short-range synaptic transition in NHEJ”, The Protein Society 35th Anniversary Symposium, Virtual, 2021 (Featured with flash talk preview in main session)

AWARD & HONOR

The Graduate School Conference Travel Grant for GRC Three-Dimensional Electron Microscopy (3DEM), 2021

IBiS Travel Award for GRC-3DEM, 2021

Rappaport Award for Research Excellence, 2021

Finn Wold & Protein Science Anniversary Awards, The Protein Society, 2021

IBiS Travel Award for The Protein Society 35th Anniversary Symposium, 2021

National Institute of General Medical Sciences (NIGMS) funded (5T32GM008382) Molecular Biophysics Training Program, 2017-2019