Dissection of DNA double-strand-break repair using novel single-molecule forces

Jing L. Wang1,9, Camille Duboc1,9, Qian Wu1,9, Takashi Ochi2,8, Shikang Liang2, Susan E. Tsutakawa3, Susan P. Lees-Miller4, Marc Nadal1, John A. Tainer3,5, Tom L. Blundell2 and Terence R. Strick1,6,7*

Repairing DNA double-strand breaks (DSBs) by nonhomologous end joining (NHEJ) requires multiple proteins to recognize and bind DNA ends, process them for compatibility, and ligate them together. We constructed novel DNA substrates for single-molecule nanomanipulation, allowing us to mechanically detect, probe, and rupture in real-time DSB synthesis by specific human NHEJ components. DNA-PKcs and Ku allow DNA end synthesis on the 100 ms timescale, and the addition of PAXX extends this lifetime to ~2 s. Further addition of XRCC4, XLF and ligase IV results in minute-scale synthesis and leads to robust repair of both strands of the nanomanipulated DNA. The energetic contribution of the different components to synaptic stability is typically on the scale of a few kilocalories per mole. Our results define assembly rules for NHEJ machinery and unveil the importance of weak interactions, rapidly ruptured even at sub-picoNewton forces, in regulating this multicomponent chemomechanical system for genome integrity.

Results

DNA molecular forces sustain a functional NHEJ reaction. To objectively and quantitatively measure the properties of complex molecular synthesis and interactions, we developed a DNA-scaffold-based single-molecule assay that permits the observation of repeated cycles of synthesis and rupture. Microscopic properties such as the kinetics of formation and lifetime of synthesis can be derived from such assays (Fig. 1a). Briefly, the DNA system consists of two linear double-strand DNA (dsDNA) segments ~1,510 bp in length that are connected to each other by a third double-strand DNA segment, termed a ‘bridge’, of ~690 bp (Methods and Supplementary Fig. 1). The bridge is anchored 38 bp away from the ends of each DNA segment it tethers, allowing the ends to fluctuate freely as they face each other and providing ample space for loading of NHEJ components. We attached one end of this construct to a digoxigenin-modified glass surface and the other end to a streptavidin-coated magnetic bead via 1-kb digoxigenin- or biotin-labeled DNA fragments attached to either end of the ~3.6-kbp construct. We applied a vertical extending force, \( F \), to the bead and, thus, to the construct, while imaging the position of the magnetic bead above the surface. This system allows real-time monitoring of DNA end synthesis by measuring the molecule’s overall extension, \( l \). If there is no synapsis between the DNA ends, then the extension of the ~3.6-kbp construct under a 1.4 picoNewton (pN) force (1 pN = 1 × 10⁻¹² N) is predicted to be 1,085 nm, on the basis of the worm-like chain (WLC) model that describes the mechanical properties of DNA. If there is synapsis between the DNA ends, then the bridge no longer contributes to extension of the system, which is predicted to be only 913 nm, or 172 nm shorter than in the absence of synapsis.

We first show that this system can faithfully recapitulate complete NHEJ of blunt-end DSBs (Fig. 1b). We subjected the DNA to digestion with the SmaI restriction enzyme, assembled constructs in the magnetic trap (~25–50 constructs per field of view), and introduced the full complement of human NHEJ components into the reaction chamber: 10 nM Ku, 100 pM DNA-PKcs, 20 nM PAXX, 20 nM XRCC4–ligase IV (Methods and Supplementary Fig. 1c,d). We repeatedly cycled the force applied to the DNA between a low and high value with a 450-s period (\( F_{\text{low}} = 0.04 \text{pN for} \sim 225 \text{s,}\ F_{\text{high}} = 1.4 \text{pN for} \sim 225 \text{s} ;\}

© 2018 Nature America Inc., part of Springer Nature. All rights reserved.

1Institut Jacques Monod, Université Paris Diderot, CNRS, UMR 7592, Paris, France. 2Department of Biochemistry, Cambridge University, Cambridge, UK. 3Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. 4Department of Biochemistry and Molecular Biology, Arnie Charbonneau Cancer Institute, University of Calgary, Calgary, Alberta, Canada. 5Department of Molecular and Cellular Oncology, The University of Texas M.D. Anderson Cancer Center, Houston, TX, USA. 6Ecole Normale Supérieure, PSL Research University, CNRS, INSERM, Institute of Biology (IBENS), Paris, France. 7Programme Equipes Labellisées, Ligue Contre le Cancer, Paris, France. 8Present address: MRC Laboratory of Molecular Biology, Cambridge, UK. 9These authors contributed equally: Jing L. Wang, Camille Duboc, Qian Wu. *e-mail: strick@ens.fr
in the low-force state, the DNA ends are free to encounter each other, whereas in the high-force state, we can reliably interrogate the status of the DNA ends. After one or two traction cycles, we frequently observed a DNA state with reduced extension (Fig. 1b) that never recovered its initial extension length, indicative of successful DNA ligation. Successful ligation of the blunt DNA ends of the construct generates a SmaI site. Thus, after disrupting residual NHEJ components with 0.2% SDS and washing out the detergent (Methods), we added SmaI into the reaction, and the initial DNA state was restored via an abrupt increase in DNA extension, $\Delta l = 161 \pm 2\text{ nm}$ (s.e.m., $n = 129$ events; Fig. 1c). This repair-specific cleavage reaction provides a length-change calibration that we can then use as a signature of bona fide, specific interactions between the ends of the DNA construct. Most DNA molecules (36 out of 50, or 72%) were repaired within the first few traction cycles (Fig. 1d), underscoring the robust nature of DBS repair in this system when all components are present. Ligation was still observed, albeit at lower efficiency, when PAXX, XLF, or the XRCC4 N- and C-terminal domains were individually omitted from the reaction (Table 1). No ligation was observed when DNA-PKcs or ligase IV was omitted from the assay (Table 1) or if the DNA ends were dephosphorylated (discussed below).

The ligated DNA’s ability to supercoil demonstrated that a double ligation event occurred, in which both strands of the DNA ligated to their counterparts at the other DNA end (Supplementary Fig. 2a). We also show (Supplementary Fig. 2b–d) that T4 DNA ligase, although capable of efficiently ligating the construct prepared with overhanging ends (via XmaI digestion), was unable to efficiently ligate the construct with blunt ends, underscoring the requirement in vivo for robust molecular pathways to repair non-complementary DSBs. We conclude that this DNA system functionally recapitulates human NHEJ at single-molecule resolution and thus provides a means to evaluate the contribution of each reaction component to synaptic stability.

**Stepwise assembly of the human NHEJ machinery.** We progressively assembled the NHEJ machinery and assayed for DNA end synthesis using the force-modulation method described above. We found that the minimal combination of DNA-PK holoenzyme (DNA-PKcs+Ku) and PAXX is necessary to observe frequent, second-scale interactions between the two DNA ends (Fig. 2a,b and Supplementary Table 1). In this assay, an interaction was detected as a transient plateau in the DNA extension signal obtained upon increasing the force from $F_{\text{low}}$ to $F_{\text{high}}$ where the duration of the plateau reflects the lifetime of the end interaction. The histogram depicting the distribution of observed $\Delta l$ values (Fig. 2c) displays a Gaussian peak located close to the expected amplitude (166 ± 1 nm s.e.m., $n = 129$ events). Comparing the change in DNA extension at the end of the plateau to our calibrated length-change standard obtained by specific cleavage of the repaired DNA ends confirmed that we are detecting bona fide interactions between the ends of the DNA construct. We designated those interaction events within three s.d. of the likeliest value as ‘synaptic.’ The lifetime distribution for synaptic events follows single-exponential behavior with an average, $t_{\text{synaps}}$ of 2.2 ± 0.3 s (s.e.m., $n = 98$ events) (Fig. 2d).

Synapsis frequency, defined as the fraction of force-modulation cycles in which we observed a synaptic event, is ~6% (102 synaptic events out of 1,611 force-modulation cycles collected on ~90 DNA substrates; Supplementary Table 1).

These synaptic events did not substantially depend on the phosphorylation state of the DNA ends (Fig. 3a). Notably, synaptic events were abolished when we replaced wild-type PAXX with a mutant (V199A and F201A) incapable of interacting with the Ku subunit of the DNA-PK holoenzyme (Fig. 3b). XLF was unable to substitute for PAXX (Fig. 3c) despite its ability to interact with Ku$^{26,29}$. Neither DNA-PK holoenzyme nor PAXX on their own generated synapsis on the 2-s scale (Fig. 3d,e). We conclude that the ‘upstream’ components Ku, DNA-PKcs and PAXX sustain formation of DNA end synapsis that can resist up to ~1 pN traction forces (0.15 kCal/mol×nm) for second time scales and that PAXX plays a central role in stable bridging of the gap between the two DNA ends.

To assess the role of the ‘downstream’ NHEJ components XRCC4, XLF and ligase IV, we added them in combination with the previously examined ‘upstream’ components to generate a ‘complete’ reaction (Fig. 4). However, in contrast to the prior assays containing all components, in this case, we dephosphorylated the DNA ends to prevent ligation and to enable us to repeatedly interrogate the construct. We observed long-lived synaptic interactions:
Supplementary Table 2. Efficiency of observed NHEJ ligation reactions as a function of NHEJ components present

| Assay                                      | # molecules | Number of molecules ligated at # pulling cycle | Total | Efficiency % | Normalized efficiency % |
|-------------------------------------------|-------------|-----------------------------------------------|-------|--------------|------------------------|
| Complete reaction                         | 50          | 26 7 2 1 - - - - -                          | 36    | 72 ± 12      | 100                    |
| Complete but for DNA-PKcs                 | 40          | 0 0 0 - - - - -                             | 0     | 0            | 0                      |
| Complete but for Ku                       | 69          | 1 0 1 - - - - -                             | 2     | 3 ± 2        | 4                      |
| Complete but for PAXX                    | 57          | 6 5 1 1 1 0 1 0 1                          | 16    | 28 ± 7       | 39                     |
| Complete but for XLF                      | 184         | 0 1 0 1 1 3 2 0 9                          | 9     | 5 ± 2        | 7                      |
| Complete but for XRCC4 NTD/CTD            | 193         | 1 0 1 1 0 0 0 0 3                          | 3     | 2 ± 1        | 3                      |
| Complete but for ligase IV               | 102         | 0 0 0 0 0 0 0 0 0                          | 0     | 0            | 0                      |

Ligation events were identified as an irreversible shortening of the DNA scaffold’s extension by Δl = 165 nm. For the complete reaction, 28 of the 36 molecules scored for repair were further confirmed by DNA cleavage via SmaI digest (Fig. 1c); the remaining eight molecules were lost from tracking during SDS washes, which can destabilize streptavidin and antidigoxigenin links between DNA and surfaces.

We report that synaptic junctions of varied composition can resist piconewton-scale forces for times ranging from seconds to minutes. Presumably, this is a reflection of actual commitment to repair. On the basis of these observations, we propose a simple model for NHEJ synapsis (Fig. 5). Taking the 66-s lifetime of the complete synaptic junction as a baseline, we find that removing PAXX is responsible for an approximately seven-fold destabilization of the complete synaptic junction (9-s synapsis in the absence of PAXX). We likewise determined that disabling the XRCC4–XLF–ligase IV system is responsible for an ~30-fold destabilization of the complete synaptic junction (2-s synapsis when DNA-PK and PAXX were not supplemented with XRCC4–XLF–ligase IV). Because the lifetime of a complex is related to its free energy of activation by Boltzmann’s law, we estimate that PAXX contributes $k_B T \ln(66 s/9 s) \approx 2 k_B T$ or 1.2 kcal/mol to the stability of the complete synaptic junction, whereas XRCC4–XLF–ligase IV contributes $k_B T \ln(66 s/2 s) \approx 3.5 k_B T$ or 2.1 kcal/mol to the stability of the complete synaptic junction. Together, these two sets of components would stabilize the synaptic state by roughly 5.5 $k_B T$, increasing its lifetime nearly 200-fold. These results suggest that the primary synapsis formed by the DNA-PK holoenzyme alone is short-lived, in the range of hundreds of milliseconds. Lastly, in an attempt to observe this short-lived interaction, we increased the rate at which the force is switched between low and high values by a factor of ten. The large number of traction cycles generated allowed us to detect rare and short-lived synaptic events of expected synaptic amplitude (Supplementary Fig. 6). The lifetime of this synapsis is, as predicted, in the 100 ms range. Control experiments showed that this short-lived and specific synapsis is only observed in the presence of DNA-PKcs and Ku (Supplementary Fig. 6 and Supplementary Table 2).

The additive nature of these short-lived and weak interactions, revealed by the present single-molecule experiments, suggests a multivalent system in which multiple protein interfaces stabilize DNA end synapsis. Use of multiple, weak interactions to obtain high affinity balanced by regulation is an emerging common theme in DNA repair systems; one example is RPA, whose subnuclease affinity is built on multiple weak interactions\(^{45}\). Longer synapsis effectively increases the probability of ligation.

Our work supports and broadens the observation that PAXX stabilizes the core NHEJ proteins at damaged chromatin\(^{11,13,12}\) and suggests PAXX as a potential target for cancer drug development. Indeed, removing PAXX from the complete reaction reduces the probability of repair from 72% to 28% but does not completely ablate repair—as is the case when XLF or XRCC4 are omitted from the NHEJ reaction (Table 1). It remains to be determined whether PAXX can incorporate into NHEJ complexes only at the
Fig. 2 | Ku, DNA-PKcs and PAXX are necessary to stabilize DNA-end synapsis. a. Experimental design. DNA is prepared with blunt ends using SmaI digest, stars represent phosphate groups. b. Representative time trace obtained upon application of the force-modulation pattern (red). Inset shows an expanded view of an end-interaction rupture event, which can be characterized by both the change in DNA extension upon rupture, \( \Delta l \), and the duration of the synaptic event before rupture, \( t_{\text{synapsis}} \). c. Histogram of DNA extension change, \( \Delta l \), upon rupture event. Red line is a fit to a Gaussian distribution, with a maximum (red arrow) at 166 ± 10 nm (s.d.). The entire histogram contains \( n = 129 \) events, of which 98 are within three s.d. from the peak. d. Lifetime distribution of the synaptic state is fit to a single-exponential distribution (red line), giving a lifetime of 2.2 ± 0.3 s (s.e.m., \( n = 98 \) events).

Fig. 3 | DNA synapsis on the 2-s timescale in the presence of Ku, DNA-PKcs, and PAXX (control experiments). a, Synapsis by Ku, DNA-PKcs and PAXX does not require phosphorylated DNA ends. Left, time trace showing rupture events. Middle, amplitude distribution of rupture events (\( n = 69 \)). Distribution peak is fit to a Gaussian distribution (red line) with a maximum at 163 ± 2 nm (s.e.m., \( n = 40 \)). The distribution displays additional density for \( \Delta l \) values of 75 nm and 400 nm. Values of 400 nm are consistent with the extension of a 1,500-bp DNA segment at the force employed and thus likely correspond to loop-back interactions between the tip of a 1,500-bp DNA segment and the surface to which that segment is anchored. The smaller peak is consistent with local bending or wrapping deformations of DNA with ~50 nm persistence length. Right, lifetime distribution of synaptic events follows a single-exponential distribution (red line), giving a lifetime of 0.5 ± 0.5 s (s.e.m., \( n = 40 \)). b–e Representative time traces show that the combinations of Ku, DNA-PKcs, and the PAXX mutant (b), Ku, DNA-PKcs, and XLF (c) and PAXX alone (e) do not lead to 2-s synapsis. The previously reported cross-contamination of DNA-PKcs preparations by Ku\(^{35,36}\) precludes the possibility of testing DNA-PKcs alone. P+/P– denotes the presence or absence, respectively, of 5' phosphates at the DNA ends.

initial stage of assembly or whether it may also join at later stages of assembly, for instance, once XRCC4–XLF–ligase IV has assembled. With this validated single-molecule system, it will now be feasible to examine the impacts of other components that act in NHEJ, such as Artemis and APLF\(^{34}\) proteins, as well as noncoding RNA\(^{35}\), and relate their energetic and mechanistic roles to cancer predispositions arising from mutational defects or aberrant regulation.
molecules. Specifically, we propose a strategy of chemical inhibitors to compete with the functionally important yet individually weak interactions in this major system for DNA double-strand break repair in humans.

Methods

Methods, including statements of data availability and any associated accession codes and references, are available at https://doi.org/10.1038/s41594-018-0065-1.

Received: 5 March 2018; Accepted: 13 April 2018; Published online: 21 May 2018

Fig. 4 | Stabilization of DNA end synapsis by Ku, DNA-PKcs, PAXX, XLF, XRCC4, and ligase IV. a, Experimental design. DNA is prepared with blunt ends generated by SmaI digestion followed by dephosphorylation (Methods). b, Representative time trace obtained upon application of the force-modulation pattern (red) in the presence of Ku, DNA-PKcs, PAXX, XLF, XRCC4, and ligase IV. Transient synapsis is detected as an intermediate DNA extension state that spontaneously reverts to the maximum extension state, allowing determination of the duration of the intermediate state and the change in extension upon reversion to the maximum extension. c, Histogram of DNA extension change (Δl) upon a rupture event. The red line is a fit to a Gaussian distribution, with a maximum (red arrow) at 165 ± 9 nm (s.d., n = 324 events in the histogram, of which 183 are within three s.d. from the peak). d, Lifetime distribution of the synaptic state is fit to a single-exponential distribution (red line), giving a lifetime of 66.4 ± 7.6 s (s.e.m., n = 183 events).

Fig. 5 | Model for multivalent stabilization of DNA-end synapsis by the NHEJ machinery. Schematic representation of a model for synapsis and repair involving formation of an initial synaptic complex that is stabilized either by incorporation of PAXX (left path) or by XRCC4-XLF-ligase IV (right path). Bottom, a complete complex that is stabilized by both PAXX and XRCC4-XLF-ligase IV has the longest lifetime and leads to efficient DSB ligation.

Our present results show that novel and functional DNA scaffolds provide unique and unexpected kinetic insights into the protein machinery that is responsible for carrying out NHEJ in human cells. Our studies thus provide a solid foundation upon which to further explore both the order of assembly and the stoichiometry of functional repair complexes using fluorescently labeled protein variants and correlative single-molecule nanomanipulation and fluorescence (NanoCOSM®). Importantly, the measurements of synapsis lifetime and $k_t^T$-scale energetics uncover the nature of multicomponent DNA-break repair complexes. This knowledge has broad scientific and biomedical implications, including strategies for targeting the DNA damage response by small molecules.
14. Tadi, S. K. et al. PAXX is an accessory c-NHEJ factor that associates with Ku70 and has overlapping functions with XLF. Cell Reports 17, 541–553 (2016).
15. Ropars, V. et al. Structural characterization of filaments formed by human XRCC4-Cernunnos/XLF complex involved in nonhomologous DNA end-joining. Proc. Natl Acad. Sci. USA 108, 12663–12668 (2011).
16. Hammel, M. et al. 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, 32638–32650 (2011).
17. Wu, Q. et al. Non-homologous end-joining partners in a helical dance: structural studies of XLF-XRCC4 interactions. Biochem. Soc. Trans. 39, 1387–1392 (2011).
18. Andres, S. N. et al. A human XRCC4-XLF complex bridges DNA. Nucleic Acids Res. 40, 1868–1878 (2012).
19. Bouchiat, C. & Mezard, M. Elasticity model of a supercoiled DNA molecule. Phys. Rev. Lett. 80, 1556 (1998).
20. Liang, S. et al. Achieving selectivity in space and time with DNA double-strand break repair. Nat. Struct. Mol. Biol. 23, 509–510 (2016).
21. Mahaney, B. L., Hammel, M., Meek, K., Tainer, J. A. & Lees-Miller, S. P. 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 (2013).
22. Lieber, M. R. The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway. Annu. Rev. Biochem. 79, 181–211 (2010).
23. Liang, S. et al. Achieving selectivity in space and time with DNA double-strand-break response and repair: molecular stages and scaffolds come with strings attached. Struct. Chem. 28, 161–171 (2017).
24. Yaneva, M., Kowalewski, T. & Lieber, M. R. Interaction of DNA-dependent protein kinase with DNA and with Ku: biochemical and atomic-force microscopy studies. EMBO J. 16, 5098–5112 (1997).
25. West, R. B., Yaneva, M. & Lieber, M. R. Productive and nonproductive complexes of Ku and DNA-dependent protein kinase at DNA termini. Mol. Cell. Biol. 18, 5908–5920 (1998).
26. Strick, T. R., Allemand, J. F., Bensimon, D., Bensimon, A. & Croquette, V. The elasticity of a single supercoiled DNA molecule. Science 271, 1835–1837 (1996).
27. Bouchiat, C. et al. Estimating the persistence length of a worm-like chain molecule from force-extension measurements. Biophys. J. 76, 409–413 (1999).
28. Yano, K. et al. Ku recruits XLF to DNA double-strand breaks. EMBO Rep. 9, 91–96 (2008).
29. Grundy, G. J. et al. The Ku-binding motif is a conserved module for recruitment and stimulation of non-homologous end-joining proteins. Nat. Commun. 7, 11242 (2016).
30. Hammel, M., Yu, Y., Fang, S., Lees-Miller, S. P. & Tainer, J. A. XLF regulates filament architecture of the XRCC4-Ligase IV complex. Structure 18, 1431–1442 (2010).
31. Balmus, G. et al. Synthetic lethality between PAXX and XLF in mammalian development. Genes Dev. 30, 2152–2157 (2016).
Methods

**DNA construct.** The construct is assembled from two precurser dsDNA molecules. For the first molecule, we assemble two ssDNA oligos, one short and one long, into a covalently linked, asymmetric branched structure. In a PCR reaction using a pair of oligo branches, the longer oligo branch is PCR while the shorter one remains unused. The branch points end up 67 bp from the end of dsDNA. This 'break' PCR product is 3kb in length, and by hybridization, one can convert the remaining ssDNA branches into dsDNA branches with 4-bp overhangs. This fragment can then be circularized by hybridization and ligation to a ~600 bp bridge DNA fragment, which is simply a standard dsDNA fragment digested so as to have overhangs compatible with those of the short dsDNA branches. The construct, assembled to a micro-size magnetic bead at the biotin-labeled end and a treated glass coverslip at the dig-labeled end can then be handled and observed in a magnetic trap (~30–50 constructs per field of view). The ends can be processed into blunt or overhang ends by restriction digest or after ligation to the biotin- and dig-labeled fragments or directly under the magnetic trap.

**Oligonucleotides.** DBCO-Hyb1 5′ YCCATGGCACATGATGCTTGGG and DBCO-Hyb2 5′ YGAGCCAAAGGCCTCTAGCATTGC, where Y indicates DBCO-dT (courtesy of J. Randolph, Glen Research, USA), and products corresponding to covalently coupled oligos were separated on an 8 M urea, 8% acrylamide/bisacrylamide (29:1) gel running at 10 V/cm in 1X Tris-borate-EDTA buffer. Gel slices corresponding to product were excised from the gel, using a surgical imaging lane to carry out UV shadowing to identify the relevant bands. DNA was purified by first eluting DNA from the crushed gel fragment overnight in 10 mM phosphate buffer pH 8.0 with agitation at 4°C. A SepPak C18 cartridge (Waters) was conditioned with first acetonitrile and then water; the eluate was loaded onto the cartridge, which was then washed with water, and the oligo released from the cartridge by eluting with acetonitrile. Acetonitrile was evaporated under vacuum centrifugation, and oligo was resuspended to -10 μM. The oligo coupled oligos were then used as primers in a PCR reaction using Charomid 9.5-ΔSbf1 as template (i.e., a Charomid 9.5 plasmid derivative from which the native Sbf1 site has been removed by Sbf1 digestion, fill-in and religation), and -1 kb product DNA was purified using agarose gel electrophoresis and extraction (Macherey-Nagel). The construct is assembled from two precursor dsDNA molecules.

**Protein expression constructs.** Ku70 and Ku80 were both cloned into pFastBac Dual vector (Thermo Fisher) and coexpressed in insect S9 cells. Ku70 contains an N-terminal hexahistidine tag followed by a TEV cleavage site. XLF and XRCC4 are individually cloned into a HIS-Tag vector* and transiently expressed in COS7 and Rosetta (DE3) cells (Invitrogen). Both XLF and XRCC4 contain a C-terminal hexahistidine tag. Cell cultures were grown in LB medium until OD₆₅₀ was approximately 0.6. IPTG was added to a final concentration of 1 mM. Proteins were expressed at 16°C overnight for XLF and at 37°C for 4 h for XRCC4. The PAXX and full-length ligase IV–XRCC4 coexpression constructs were as described previously42. XLF-ΔC (306-529), which cannot tranchead into XLF but can bind to PETG-41A (Gateway Destination vector, EMBL) and expressed as described previously. The catalytically dead ligase IV (K273A)–XRCC4 protein complex was generated from the ligase IV–XRCC4 coexpression plasmid (a gift from M.-D. Tsai, The Ohio State University, Columbus, OH) by the QuikChange method (Agilent Technologies). Full-length ligase IV fused with a hexahistidine tag at the N terminus and cxXRCC4 (resides 138–213) were amplified from the ligase IV–XRCC4 coexpression plasmid and cxXRCC4 plasmid, of which cysteines were mutated to alanines, respectively, and cloned into pRSFDuet1 vector (Novagen).

**Protein purifications.** Ku70/80. Insect S9 cells containing Ku70/80 were lysed by sonication in lysis buffer (50 mM Tris-HCl, pH 8.0, 5% glycerol, 150 mM NaCl, 2 mM β-mercaptoethanol, 20 mM imidazole, protease inhibitor (Roche) and 40 μg/ml DNAse I (Sigma)). After 30 min incubation in 4°C after sonication, the salt concentration of the lysate was adjusted to 500 mM NaCl. After centrifugation, Ku70/80 supernatant was mixed with Ni-NTA affinity resin (Qiagen) pre-equilibrated with binding buffer (50 mM Tris-HCl pH 8.0, 5% glycerol, 500 mM NaCl, 2 mM β-mercaptoethanol, and 20 mM imidazole). After beads were washed with binding buffer for 10 x column volumes, protein was eluted using elution buffer (50 mM Tris-HCl pH 8.0, 5% glycerol, 500 mM NaCl, 2 mM β-mercaptoethanol, and 100 mM imidazole). For assembly onto the microscope, the ligation reaction was first diluted 60-fold to 50 pM nominal concentration of DNA in Tris buffer (10 mM Tris-HCl, pH 8). Then, 0.5 μl of this dilution was mixed with 10 μl of Dynal MyOne C1 magnetic beads. The beads were prepared by taking 10 μl of stock solution, washing them with 10 μl of RB (20 mM KHEPES, pH 7.8 and TE buffer, MgCl₂, 1 mM ATP, 1 mM DTT, 0.05% Tween-20, and 0.5 mg/ml BSA) and concentrating the beads, discarding the supernatant, and resuspending the beads in 10 μl of RB. The bead-dNA mixture was diluted after 5–10s with an additional 15 μl of RB and injected onto functionalized surfaces.

**Protein expression constructs.** Ku70 and Ku80 were both cloned into pFastBac Dual vector (Thermo Fisher) and coexpressed in insect S9 cells. Ku70 contains an N-terminal hexahistidine tag followed by a TEV cleavage site. XLF and XRCC4 are individually cloned into a HIS-Tag vector and transiently expressed in COS7 and Rosetta (DE3) cells (Invitrogen). Both XLF and XRCC4 contain a C-terminal hexahistidine tag. Cell cultures were grown in LB medium until OD₆₅₀ was approximately 0.6. IPTG was added to a final concentration of 1 mM. Proteins were expressed at 16°C overnight for XLF and at 37°C for 4 h for XRCC4. The PAXX and full-length ligase IV–XRCC4 coexpression constructs were as described previously. XLF-ΔC (306-529), which cannot tranchead into XLF but can bind to PETG-41A (Gateway Destination vector, EMBL) and expressed as described previously. The catalytically dead ligase IV (K273A)–XRCC4 protein complex was generated from the ligase IV–XRCC4 coexpression plasmid (a gift from M.-D. Tsai, The Ohio State University, Columbus, OH) by the QuikChange method (Agilent Technologies). Full-length ligase IV fused with a hexahistidine tag at the N terminus and cxXRCC4 (resides 138–213) were amplified from the ligase IV–XRCC4 coexpression plasmid and cxXRCC4 plasmid, of which cysteines were mutated to alanines, respectively, and cloned into pRSFDuet1 vector (Novagen).
5 mM DTT). Eluted sample was further purified by running through Superdex 200 10/300 (GE Healthcare) equilibrated in Buffer GF (20 mM Tris-HCl, pH 8, 150 mM NaCl, 5% glycerol, and 5 mM DTT). Protein samples were analyzed on 4–12% NuPAGE Bis-Tris gels, concentrated and stored in ~−80°C.

XLF and XRCC4. XLF cells were purified by a similar process as for Ku70/80. The lysis buffer for XLF is buffer A (20 mM HEPES, pH 8.0, 10 mM NaCl, and 5 mM DTT) and buffer B (20 mM HEPES, pH 8.0, 1 M NaCl, and 5 mM DTT). The final gel filtration buffer is 20 mM HEPES, pH 8.0, 150 mM NaCl, and 5 mM DTT. Purification of XRCC4 was performed as described previously. Purification of XLF was identical to that of wild-type XLF, except that the mutant contained an N-terminal hexahistidine-MBP tag followed by a TEV cleavage site. After Ni affinity purification, the tag was cleaved before proceeding to the Q column.

Ligase IV–XRCC4, PAXX and DNA-PKcs. Purifications of all ligase IV–XRCC4 construıcts, PAXX and DNA-PKcs were performed as described previously.

Bulk ligation of DNA with overhang ends by XRCC4–ligase IV system. pHA T4 was double-digested with XhoI and NcoI to generate 4-bp overhangs. 200 ng of the digested plasmid was incubated with 25 nM of the indicated proteins in 20 ml of buffer containing 25 mM Tris-HCl, pH 7.5, 150 mM KCl, 1 mM MgCl₂, 1 mM DTT, 10 μM ATP, 10% (wt/vol) PEG10,000, and 10 μg/ml BSA. Mixtures were incubated at 37°C for 5 min before initiating ligation by XRCC4–ligase IV at 37°C for 30 min. The mixtures were incubated at 30°C for another 30 min after adding 2 μl of a reaction-stop solution (100 mM EDTA, 0.1% (wt/vol) SDS) and 0.2 μl of 20 mg/ml Protease K. Reaction mixtures were separated by electrophoresis on a 0.8% agarose gel in TBE buffer. The gel was stained with SYBR Gold, visualized using a UV imager and quantified using GeneTools (Synegene).

We note that the presence of 4-bp overhangs makes this assay more permissive than the blunt ends used in the scaffold assay.

Experimental conditions. All assays were conducted at 34°C in reaction buffer RB (20 mM K·HEPES, pH 7.8, 100 mM KCl, 5 mM MgCl₂, 1 mM ATP, 1 mM DTT, 0.05% Tween-20, and 0.5 mg/ml BSA), unless noted otherwise. When present, NHEJ components were used at a concentration of 10 nM Ku70/80 (dimer), 100 pM DNA-PKcs, 20 nM PAXX, 20 nM XRCC4, 20 nM XLF, and 20 nM ligase IV.

When mutated versions of proteins were used, they were at the same concentration as wild-type proteins.

To digest DNA molecules that have undergone NHEJ and repair under the magnetic trap, we first disrupted any remaining components of the repair machinery by rinsing the capillary with wash buffer supplemented with 0.2% SDS (wash buffer: 20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 5 mM MgCl₂, 1 mM DTT, 0.05% Tween-20, and 0.5 mg/ml BSA), unless noted otherwise. When present, NHEJ components were used at a concentration of 10 nM Ku70/80 (dimer), 100 pM DNA-PKcs, 20 nM PAXX, 20 nM XRCC4, 20 nM XLF, and 20 nM ligase IV.

When mutated versions of proteins were used, they were at the same concentration as wild-type proteins.

The datasets generated and analyzed during the current study are available from the corresponding author upon reasonable request.

References
39. Duboc, C., Fan, J., Graves, E. T. & Strick, T. R. Preparation of DNA substrates and functionalized glass surfaces for correlative nanomanipulation and colocalization (NanoGOSM) of single molecules. Methods Enzymol. 582, 275–296 (2017).
40. Peränen, J., Rikkonen, M., Hyvönen, M. & Kääriäinen, L. T7 vectors with modified T7lac promoter for expression of proteins in Escherichia coli. Anal. Biochem. 236, 371–373 (1996).
41. Ochi, T., Wu, Q., Chirgadze, D. Y., Grossmann, J. G., Bolanos-Garcia, V.-M. & Blundell, T. L. Structural insights into the role of domain flexibility in human DNA ligase IV. Structure 20, 1212–1222 (2012).
42. Li, Y. et al. Crystal structure of human XLF/Cernunnos reveals unexpected differences from XRCC4 with implications for NHEJ. EMBO J. 27, 290–300 (2008).
43. Murray, J. E. et al. Mutations in the NHEJ component XRCC4 cause primordial dwarfism. Am. J. Hum. Genet. 96, 412–424 (2015).
Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. figure legend, table legend, main text, or Methods section).

- **n/a** Confirmed

  - The exact sample size (**n**) for each experimental group/condition, given as a discrete number and unit of measurement
  - An indication of whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
  - The statistical test(s) used AND whether they are one- or two-sided
  - Only common tests should be described solely by name; describe more complex techniques in the Methods section.
  - A description of all covariates tested
  - A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
  - A full description of the statistics including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
  - For null hypothesis testing, the test statistic (e.g. F, t, r) with confidence intervals, effect sizes, degrees of freedom and **P** value noted. Give **P** values as exact values whenever suitable.
  - For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
  - For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
  - Estimates of effect sizes (e.g. Cohen's d, Pearson's r), indicating how they were calculated
  - Clearly defined error bars
  - State explicitly what error bars represent (e.g. SD, SE, CI)

*Our web collection on [statistics for biologists](https://nature.com/collections/statisticsforbiologists) may be useful.*

Software and code

Policy information about [availability of computer code](#)

| Data collection | N/A |
|-----------------|-----|
| Data analysis   | N/A |

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers upon request. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

The datasets generated and analysed during the current study are available from the corresponding author on reasonable request.
Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☑ Life sciences ☐ Behavioural & social sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences

Study design

All studies must disclose on these points even when the disclosure is negative.

| Sample size          | Data collected until 50-100 events obtained to achieve 10-15% standard error on mean. |
|----------------------|----------------------------------------------------------------------------------|
| Data exclusions      | No data were excluded from analysis                                              |
| Replication          | All attempts at replication were successful                                       |
| Randomization        | N/A                                                                               |
| Blinding             | N/A                                                                               |

Materials & experimental systems

Policy information about availability of materials

n/a Involved in the study
☐ ☑ Unique materials
☐ ☑ Antibodies
☐ ☑ Eukaryotic cell lines
☐ ☑ Research animals
☐ ☑ Human research participants

Unique materials

Obtaining unique materials All unique materials used are readily available from the authors or from standard commercial sources.

Method-specific reporting

n/a Involved in the study
☐ ☑ ChIP-seq
☐ ☑ Flow cytometry
☐ ☑ Magnetic resonance imaging