Plant PAXX has an XLF-like function and stimulates DNA end joining by the Ku-DNA ligase IV/XRCC4 complex

Hira Khan and Takashi Ochi[*] (iD)

Faculty of Biological Sciences, School of Molecular and Cellular Biology, The Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, UK

SUMMARY

Non-homologous end joining (NHEJ) plays a major role in repairing DNA double-strand breaks and is key to genome stability and editing. The minimal core NHEJ proteins, namely Ku70, Ku80, DNA ligase IV and XRCC4, are conserved, but other factors vary in different eukaryote groups. In plants, the only known NHEJ proteins are the core factors, while the molecular mechanism of plant NHEJ remains unclear. Here, we report a previously unidentified plant ortholog of PAXX, the crystal structure of which showed a similar fold to human 'PAXX'. However, plant PAXX has similar molecular functions to human XLF, by directly interacting with Ku70/80 and XRCC4. This suggests that plant PAXX combines the roles of mammalian PAXX and XLF and that these functions merged into a single protein during evolution. This is consistent with a redundant function of PAXX and XLF in mammals.

Keywords: DNA repair, non-homologous end joining, DNA double-strand break, PAXX.

INTRODUCTION

Non-homologous end joining (NHEJ) is a major pathway to repair DNA double-strand breaks (DSBs; Chang et al., 2017). NHEJ is evolutionarily conserved between bacteria and eukaryotes, and is comprised of two core proteins, Ku and DNA ligase (Shuman & Glickman, 2007). In eukaryotes, most organisms have minimal core proteins; Ku70/80 heterodimer and DNA ligase IV/XRCC4 (LIG4/XRCC4) heterotrimer. Other NHEJ-specific proteins, particularly in mammals, include DNA-dependent protein kinase catalytic subunit (DNA-PKcs), XRCC4-like factor (XLF), Paralog of XRCC4 and XLF (PAXX), Artemis, X-family DNA polymerases (POL), Aprataxin and PNK-like factor (APLF) and cell cycle regulator of non-homologous end joining (CYREN)/the modulator of retrovirus infection (MRI) (Chang et al., 2017). During evolution, NHEJ gained complexity by incorporating more proteins to maintain genome stability in order to cope with challenges during the development of vertebrates. NHEJ is also important for genome editing, for instance, by CRISPR–Cas9 because it introduces DSBs by enzymes. The current understanding of eukaryotic NHEJ mostly comes from studies in yeasts and mammals. However, our knowledge of NHEJ may not be directly applied to different organisms because they often possess different sets of NHEJ proteins (Daley et al., 2005).

Plants provide examples of such organisms, which have Ku70, Ku80, LIG4 and XRCC4 (Riha et al., 2002; Tamura et al., 2002; West et al., 2000) but lack apparent orthologs of other core proteins such as DNA-PKcs and XLF. Plant NHEJ proteins play important roles in DSB repair, telomere maintenance, genome editing via CRISPR–Cas9, as observed in other organisms, and also in the stable integration of *Agrobacterium* transfer DNA (T-DNA) into plant genomes (Gehrke et al., 2022; Gelvin, 2017; Puchta, 2005). A number of studies in mammalian NHEJ highlighted the importance of DNA-PKcs and XLF in DNA end synapsis (Andres et al., 2012; Brouwer et al., 2016; Cary et al., 1997; Chaplin et al., 2021; Chen et al., 2000, 2021; DeFazio et al., 2002; Graham et al., 2016, 2018; Spagnolo et al., 2006; Wang et al., 2018; Zhao et al., 2019). Although yeasts do not have DNA-PKcs, they have XLF orthologs Nej1 (Callebaut et al., 2006; Hentges et al., 2006), which play a similar role to XLF. This raises the question of how plant NHEJ works only with Ku70/80 and LIG4/XRCC4 complexes at the molecular level.

Here, our evolutionary analysis of NHEJ proteins identified the plant ortholog of PAXX. The crystal structure of *Arabidopsis thaliana* PAXX (AtPAXX) shows that it has a similar fold to human PAXX and interacts with Ku70/80 and, unlike human PAXX, with XRCC4. *In vitro*, these

interactions stimulated DNA end joining by the LIG4/XRCC4 complex. Thus, our results suggest that AtPAXX plays a hybrid role in human XLF and PAXX.

RESULTS

Overview of the conservation pattern of NHEJ proteins

We searched for amino acid sequences of known NHEJ proteins to understand their conservation in eukaryotes. A broad group of eukaryotes has Ku70, Ku80, DNA-PKcs, DNA ligase IV, XRCC4, XLF, PAXX, Artemis, POL including polμ/λ/TdT, APLF and CYREN/MRI (Figure 1; Table S1). CYREN/MRI is the least conserved protein. Sea urchin has polynucleotide kinase–phosphatase-like protein fused to the motif that is similar to the N-terminal Ku-binding motif of CYREN/MRI (Table S1). APLF orthologs are also limited to multicellular organisms but are present more broadly than those of CYREN/MRI. A choanoflagellate *Salpingoeca rosetta* has an APLF-like protein possessing a poly(ADP-ribose)-binding zinc-finger (PBZ) domain, but the protein also has a BRCT domain and core domains of DNA ligase (Table S1). In most opisthokonta and ciliates, the full set of NHEJ proteins except for Artemis, APLF and CYREN/MRI is present. *Caenorhabditis elegans* has a unique DSB-repair system among multicellular organisms because it has Ku70, Ku80 and LIG4 orthologs but lacks XRCC4 orthologs (Figure 1). A structural homologue search using the

DALI server against AlphaFold models (Holm, 2022) of *C. elegans* proteins yields two DNA ligases; DNA ligase I and IV (accession numbers: Q27474 and Q95YE6 respectively). The latter has one BRCT domain like DNA ligase III (Ellenberger & Tomkinson, 2008) and lacks the XRCC4-interacting motif (Sibanda et al., 2001) in line with the lack of XRCC4. Apart from Ku70/80 and DNA-PKcs, core NHEJ proteins are absent in most organisms belonging to the excavata group. A recent evolutionary analysis of DNA-PKcs identified its orthologs in a wide range of eukaryotes including *Naegleria gruberi* (Lees-Miller et al., 2021) and *Leishmania major* (Silva et al., 2021). In addition, we found that *Trypanosoma cruzi* has a putative DNA-PKcs ortholog that does not have the FAT, kinase and FATC domains but the heat-repeat domain, of which fold is predicted to be similar to that of human DNA-PKcs (Figure 1; Figure S1b). Note that we could not identify a similar DNA-PKcs ortholog in *T. brucei*. Uniquely, as reported previously (Nenarokova et al., 2019), core NHEJ proteins are absent in *Giardia lamblia* (Figure 1).

Some of the end-processing enzymes in NHEJ also work in different DNA-repair pathways apart from Artemis and DNA polymerases (μ, λ and TdT). Therefore, we here discuss the evolutionary analysis of these proteins only. Metallo-β-Lactamase/β-CASP domains-containing proteins can be found in all kingdoms and work in DNA and RNA processing (Callebaut et al., 2002). Among them, a unique
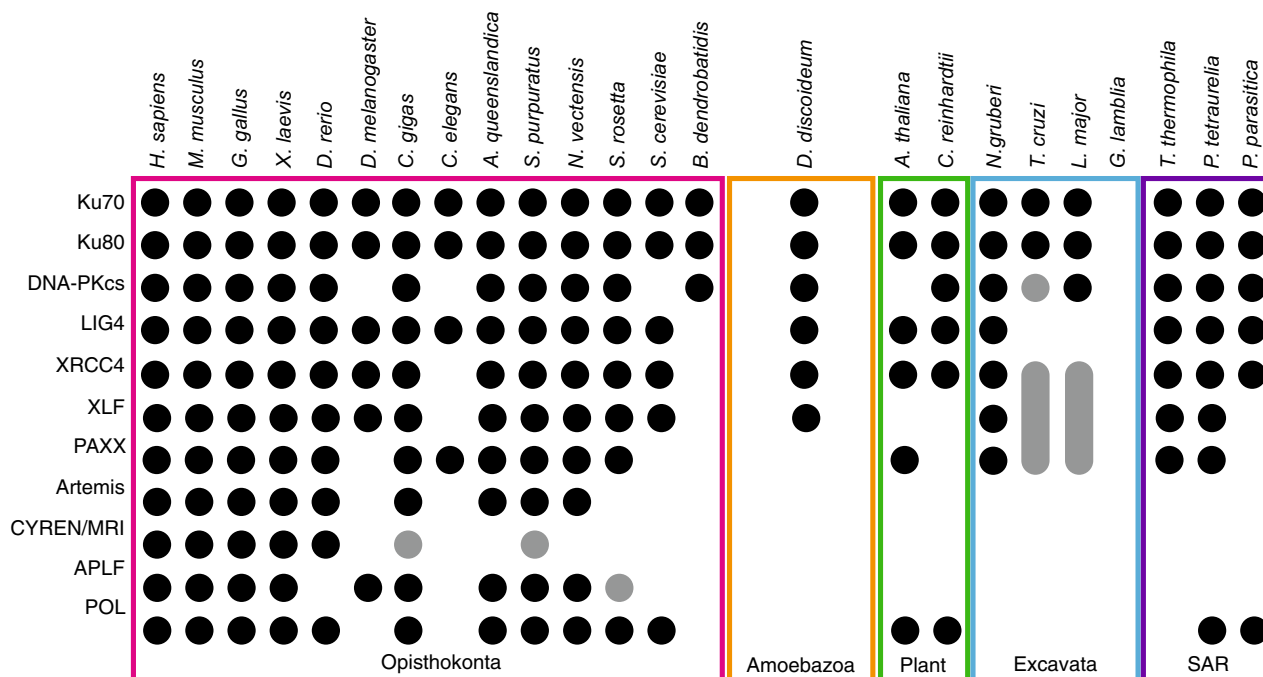

Figure 1. Evolutionarily conserved NHEJ proteins in eukaryotes.
A diagram summarises the presence of NHEJ proteins in various groups of eukaryotes. Black circles represent the presence of NHEJ proteins with confidence, whereas grey circles represent uncertainty. NHEJ, non-homologous end joining.

structural feature of Artemis is a zinc-finger motif (H228, H254, C256 and C272 in human) in the β-CASP domain (Yosaatmadja et al., 2021). Metallo-β-Lactamase/β-CASP domains-containing proteins having the zinc-finger motif were limited to jawed organisms (Figure 1). X-family DNA polymerases that are specific to NHEJ carry BRCT domains (Moon et al., 2007). Outside opisthokonta, such polymerases are found in plants (Bienstock et al., 2014) and ciliates (Figure 1).

Altogether, our evolutional analysis suggested that Ku70/80, DNA-PKcs, LIG4, XRCC4, XLF and POL are core NHEJ proteins present in both unicellular and multicellular eukaryotes.

PAXX is conserved among uni- and multicellular organisms

We had previously suggested that PAXX orthologs were present in Filozoa excluding insects and nematodes (Ochi et al., 2015). Further sequence mining identified the orthologs in *C. elegans*, amoeboflagellates, ciliates, excavates and plants (Figure 1). We found that the recently identified NHEJ factor in *C. elegans* Nhj-1 (Vujin et al., 2020) is likely to be a PAXX ortholog because of its similarity to the PAXX fold (Figure S1a). In amoebozoa, PAXX is not present in *Dictyostelium discoideum* but is present in other amoebozoa such as *Acanthamoeba castellanii* and *Planoprotostelium fungivorum* (accession numbers: L8H683 and A0A2P6NT14). Ciliates such as *Tetrahymena thermophila* and *Paramecium tetraurelia* have PAXX orthologs (Figure 1). Based on sequences of the ciliate PAXX, we found that the orthologs had been known as Die5p (defective in internal eliminated sequence [IES] excision 5) (Matsuda et al., 2010). Ciliates have complex genomic structures and have two nuclei, called the macronucleus and micronucleus. During macronucleus differentiation, DNA rearrangements occur by eliminating IESs from chromosomes (Chalker & Yao, 2011), and generated DSBs are joined by NHEJ. Die5p was proposed as a DSB-repair protein working downstream of Ku80 (Shieh & Chalker, 2013), which is consistent with the role of PAXX in humans. Interestingly, *T. cruzi* has two putative XRCC4-like proteins that are neither centrosomal protein SAS6 (Kitagawa et al., 2011; van Breugel et al., 2011) nor CCDC61 (Ochi et al., 2020) and have phosphodiesterases at the C-termini (Table S1; Figure S1c). A similar domain arrangement was also found in *L. major*, implying this phosphodiesterase fusion is conserved in those organisms. *N. gruberi*, another organism belonging to excavata, has most of the core NHEJ proteins including XLF and PAXX (Figure 1).

These results suggest that PAXX was an NHEJ factor in the early eukaryotes but probably limits its function to a certain type of DSBs at least in mammals due to redundancy with other proteins such as XLF (Abramowski et al., 2017; Balmus et al., 2016; Hung et al., 2016; Kumar et al., 2016; Lescale et al., 2016; Liu et al., 2017).

Plant NHEJ lacks XLF but has PAXX

The plant is the only group that does not have XLF but PAXX in the organisms with LIG4/XRCC4 shown in Figure 1. This implies that PAXX might play a major role in plant NHEJ. We therefore investigated the structure and function of plant PAXX.

Plant PAXX is well-conserved among streptophyta (Table S2) and has a putative Ku-binding site at the C-terminus, which is similar to PAXX orthologs (Figure 2a). Since the XLF C-terminus also interacts with Ku (Yano & Chen, 2008), plant PAXX is still possibly the ortholog of XLF. An amino acid sequence analysis of PAXX and XLF orthologs indicated that AtPAXX is more similar to the PAXX family than the XLF family (Figure 2b). To define whether plant PAXX is the human PAXX ortholog, we determined the crystal structure of full-length *A. thaliana* PAXX (AtPAXX) at 3.6-Å resolution (Table 1). AtPAXX has a short coiled-coil domain that mediates homodimerisation as observed in the structure of human PAXX (Figure 2c) and is more similar to human PAXX than XLF (Figure 2d): Unlike XLF, the α-helical region does not fold back to the coiled-coil domain, confirming that this AtPAXX is not an ortholog of XLF but that of PAXX. As observed in the crystal structure of full-length human PAXX (Ochi et al., 2015), we were unable to see the electron density of residues after ~150, indicating that they were disordered *in crystallo*. The lack of XLF orthologs in *A. thaliana* was further confirmed by searching *A. thaliana* proteins that have a similar fold to XRCC4 using DALI server against AlphaFold models (Holm, 2022), which identified *A. thaliana* XRCC4 (AtXRCC4) and AtPAXX only. These results suggest that plants have orthologs of PAXX but not those of XLF.

Plant PAXX interacts with Ku and XRCC4

Human PAXX interacts with Ku, which is essential for the PAXX function (Ochi et al., 2015). Since AtPAXX has a similar motif to the Ku-interacting motif of human PAXX (Figure 2a), we tested whether AtPAXX interacts with *A. thaliana* Ku70/80 (AtKu) by a GST-pulldown experiment using recombinant AtKu expressed in insect cells (Valuchova et al., 2017). The result showed that AtPAXX, particularly its C-terminus (residues 190–236), interacts with AtKu (Figure 3a). The interaction was further confirmed by an electrophoretic mobility shift assay, which showed that bands of the AtKu-DNA complex were further shifted by adding full-length AtPAXX but not its truncated construct (residues 1–154; AtPAXX[1–154]) (Figure 3b). An AlphaFold model of the AtPAXX-Ku complex indicated that L228 of AtPAXX plays a key role in the interaction (Figure 3c). We therefore mutated the leucine to glutamate and showed that the L228E mutation abolished the interaction (Figure 3b).
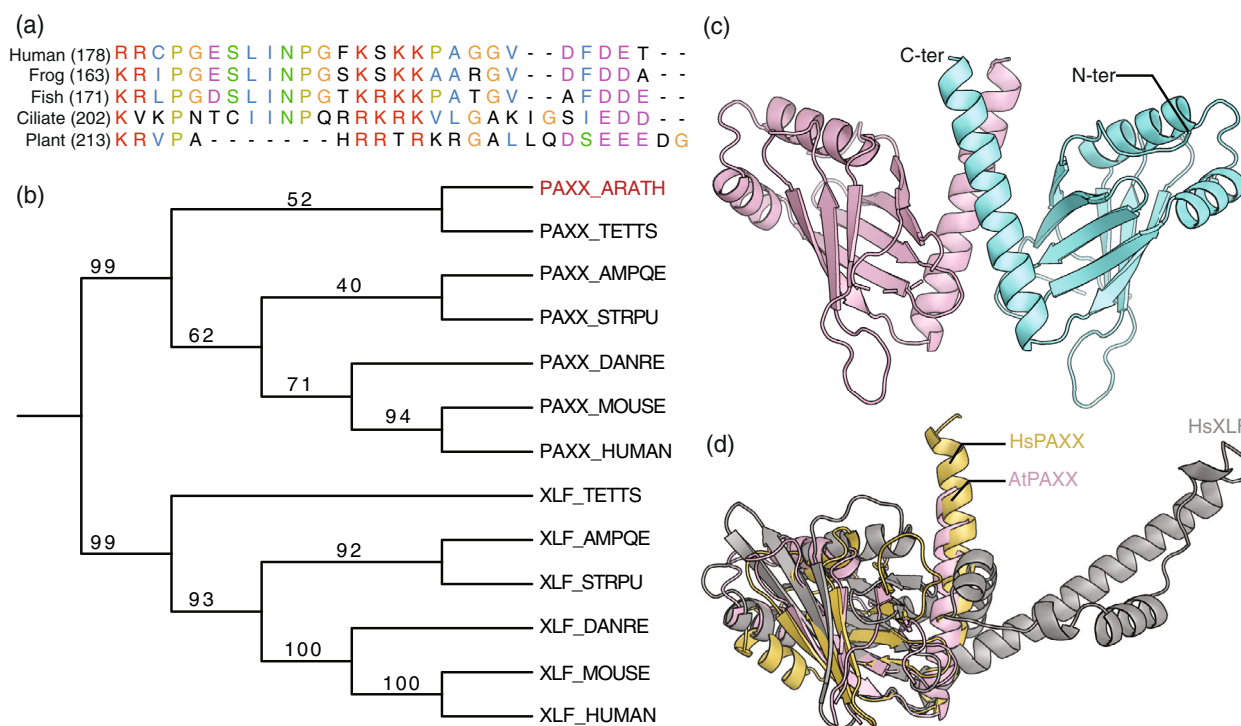
Figure 2. Plants have orthologs of PAXX.

(a) Sequence alignment of the C-termini of PAXX orthologs. Numbers in brackets are residue numbers.

(b) Phylogenetic tree of PAXX and XLF orthologs. The phylogenetic tree was generated by aligning sequences of PAXX and XLF orthologs from human (HUMAN), mouse (MOUSE), zebrafish (DANRE), sea urchin (STRPU), sponge (AMPQE), *Tetrahymena* (TETTS) and AtPAXX (ARATH in red). Numbers on the three are bootstrap values ($n = 100$).

(c) Crystal structure of AtPAXX dimer. Two chains are shown in a cartoon representation with two different colours. The electron density of residues ~150–236 was not observed due to the flexibility of the C-terminal region of AtPAXX.

(d) Structural comparison of AtPAXX with human PAXX and XLF. The head domain of the structure of AtPAXX was superimposed on those of human PAXX and XLF (PDB codes: 3WTD and 2QM4).

These suggested that AtKu interacts with AtPAXX *via* its C-terminal region, particularly L228.

We speculated that AtPAXX might interact with AtXRCC4 in a similar manner to the interaction of human XRCC4 and XLF. We therefore purified AtXRCC4 and tested for its interaction with AtPAXX by a GST-pulldown assay. The result showed that AtXRCC4 interacts with AtPAXX (Figure 3d). An AlphaFold model of the AtPAXX/XRCC4 complex predicted that both N- and C-termini of AtPAXX could interact with AtXRCC4 (Figure 3e). However, AtXRCC4 did not interact with the C-terminal low complexity region (residues 190–236) but with the folded region (residues 1–190) of AtPAXX (Figure 3d). We further confirmed this result by size exclusion chromatography using truncated constructs of AtPAXX$^{1-154}$ and AtXRCC4 (residues 17–163; AtXRCC4$^{17-163}$) (Figure 3f). Thus, these results support our model of AtPAXX interacting with AtXRCC4 *via* their head domains.

Altogether, we have biochemically shown that AtPAXX interacts with AtKu and AtXRCC4 and that AtPAXX has functions equivalent to the human XLF protein.

Plant PAXX stimulates end joining by DNA ligase IV

We next tested whether AtPAXX could stimulate the end joining by AtLIG4/XRCC4 *in vitro*. Blunt-ended linearised plasmids were incubated with different combinations of AtNHEJ proteins. The results showed that the end joining of AtLIG4/XRCC4 was inefficient but was slightly stimulated by AtKu and further by adding AtPAXX (Figure 4a). AtPAXX^{L228E}, which does not interact with AtKu (Figure 3b), also stimulated the end joining by AtLIG4/XRCC4 in the presence of AtKu but ~1.4-fold less efficient than the wild type AtPAXX (Figure 4a), indicating that the interaction between AtPAXX and AtXRCC4 is likely to play a key role in the end joining but that between AtPAXX and AtKu is also important. However, the end joining barely occurred without AtKu, which plays a central role in bridging two DNA ends.

Collectively, these results show that AtPAXX interacts with both AtXRCC4 and AtKu, and suggest that these interactions are important for DNA end joining by AtLIG4/XRCC4.

Table 1 Data collection, phasing and refinement statistics of the AtPAXX crystal structures

Crystal	AtPAXX
Beamline	DLS I04-1
Wavelength (Å)	0.91788
Resolution (Å)	
Overall	187.36–3.60
Outer shell	3.76–3.60
Space group	$P22_12_1$
Unit cell parameters (Å)	
a	107.03
b	152.419
c	187.087
Number of unique reflections	36 248
Completeness (%)	100 (100)[a]
Redundancy	13.6
R_{merge}[b]	0.242 (1.401)
$<I/\sigma>$	8.0 (2.0)
$CC^{1/2}$	0.997 (0.889)
Phasing method	Molecular replacement
Refinement	
Resolution (Å)	
Overall	152.42–3.60
Outer shell	3.70–3.60
PDB code	8P26
R_{work}[c] (highest shell)	0.2480 (0.3344)
R_{free}[d] (highest shell)	0.2859 (0.3807)
Number of atoms	
Protein atoms	10 120
Water molecules	22
Average B-factors (Å^2)	119.85
Ramachandran favoured (%)	95.37
Clashscore	16.76
R.M.S.D.	
Bond (Å)	0.006
Angle (°)	0.906

[a]The statistics in parenthesis are for the highest resolution shell.
[b]$R_{merge} = \Sigma_h|I_h - <I>|/\Sigma_h I_h$, where I_h is the intensity of reflection h, and $<I>$ is the mean intensity of all symmetry-related reflections.
[c]$R_{work} = \Sigma||F_{obs}| - |F_{calc}||/\Sigma|F_{obs}|$, F_{obs} and F_{calc} are observed and calculated structure factor amplitudes.
[d]R_{free} as for R_{work} using a randomly selected 5% of the data excluded from the refinement.

DISCUSSION

We showed evolutional conservation of core NHEJ proteins, Ku70, Ku80, DNA-PKcs, LIG4, XRCC4, XLF, PAXX and POL and that plants have PAXX orthologs that interact with both Ku70/80 and XRCC4, and stimulate DNA end joining by LIG4/XRCC4. Thus, through our evolutionary analysis of NHEJ proteins, we discovered previously unidentified plant PAXX orthologs that play a hybrid role equivalent to human PAXX and XLF in NHEJ (Figure 4b).

NHEJ is conserved between prokaryotes and eukaryotes. Bacterial NHEJ utilises Ku homodimer and DNA ligase D, which is a multi-functional single polypeptide comprised of DNA ligase, polymerase and phosphoesterase domains (Shuman & Glickman, 2007). Its exact molecular mechanism remains to be elucidated, but these proteins are likely to carry out three NHEJ steps similar to those in mammals. The minimum core NHEJ complexes in eukaryotes are Ku70/80 and LIG4/XRCC4. LIG4 is essential for NHEJ and works only in the pathway. However, kinetoplastids lack the ligase while having Ku70/80 and putative DNA-PKcs orthologs and XRCC4-like proteins (Figure 1). In those organisms, Ku is not important for NHEJ but for telomere maintenance (Genois et al., 2014), a known conserved function of DNA-PK (Sui et al., 2020). Interestingly, the putative DNA-PKcs orthologs of *T. cruzi* possess the HEAT-repeat domain only (Figure S1) and therefore might play roles in DNA end bridging and protection, which are non-kinase functions of DNA-PKcs (Chaplin et al., 2021; Chen et al., 2021; Jiang et al., 2015; Wang et al., 2018). Since these functions are regulated by the kinase activity of DNA-PKcs in mammals (Liu et al., 2022), other kinases might control DNA-PK in *T. cruzi*. Thus, this analysis highlights that NHEJ and its constituent proteins seem to have evolved differently to adapt to different eukaryotes.

A unique feature of eukaryotic NHEJ is the XRCC4 superfamily, including XRCC4, XLF and PAXX. These scaffold proteins play crucial roles in NHEJ by mediating protein–protein interactions. Since XRCC4 has an essential role in the stability of LIG4 (Bryans et al., 1999; Herrmann et al., 1998), these proteins are almost always present as a pair in eukaryotes except for *C. elegans* (Figure 1). XLF and PAXX have a redundant function (Abramowski et al., 2017; Balmus et al., 2016; Hung et al., 2016; Kumar et al., 2016; Lescale et al., 2016; Liu et al., 2017), and PAXX is often absent in eukaryotic organisms (Figure 1). We showed that plants do not possess XLF but PAXX orthologs. How can we conclude that the orthologs are not XLF but PAXX? We concluded this because their structures are more similar to human PAXX than XLF. One distinguishing feature in the human proteins is the relative angle of the head and coiled-coil domain, with XLF different from XRCC4 and PAXX (Ochi et al., 2015). The angle of AtPAXX is similar to human PAXX and unlike that of XLF. Also, XLF has a fold-back α-helical structure, which was not observed in AtPAXX. Based on the observation that AtPAXX interacts with both XRCC4 and Ku so does XLF, we consider that PAXX and XLF might diverge from a common origin during evolution.

Recent cryo-EM structures of the short-range NHEJ complex showed that XLF bridges two Ku70/80-LIG4/XRCC4 complexes bound on two different DNA ends (Chaplin et al., 2021; Chen et al., 2021), suggesting a key role of XLF in the end bridging/synapsis step. How does this step take place in plants? Our results suggest that AtPAXX is likely to play a similar role to XLF by interacting with both AtXRCC4 and AtKu70/80. Since mammalian PAXX and XLF have been shown to have a redundant function, it is unsurprising that one of them is enough for
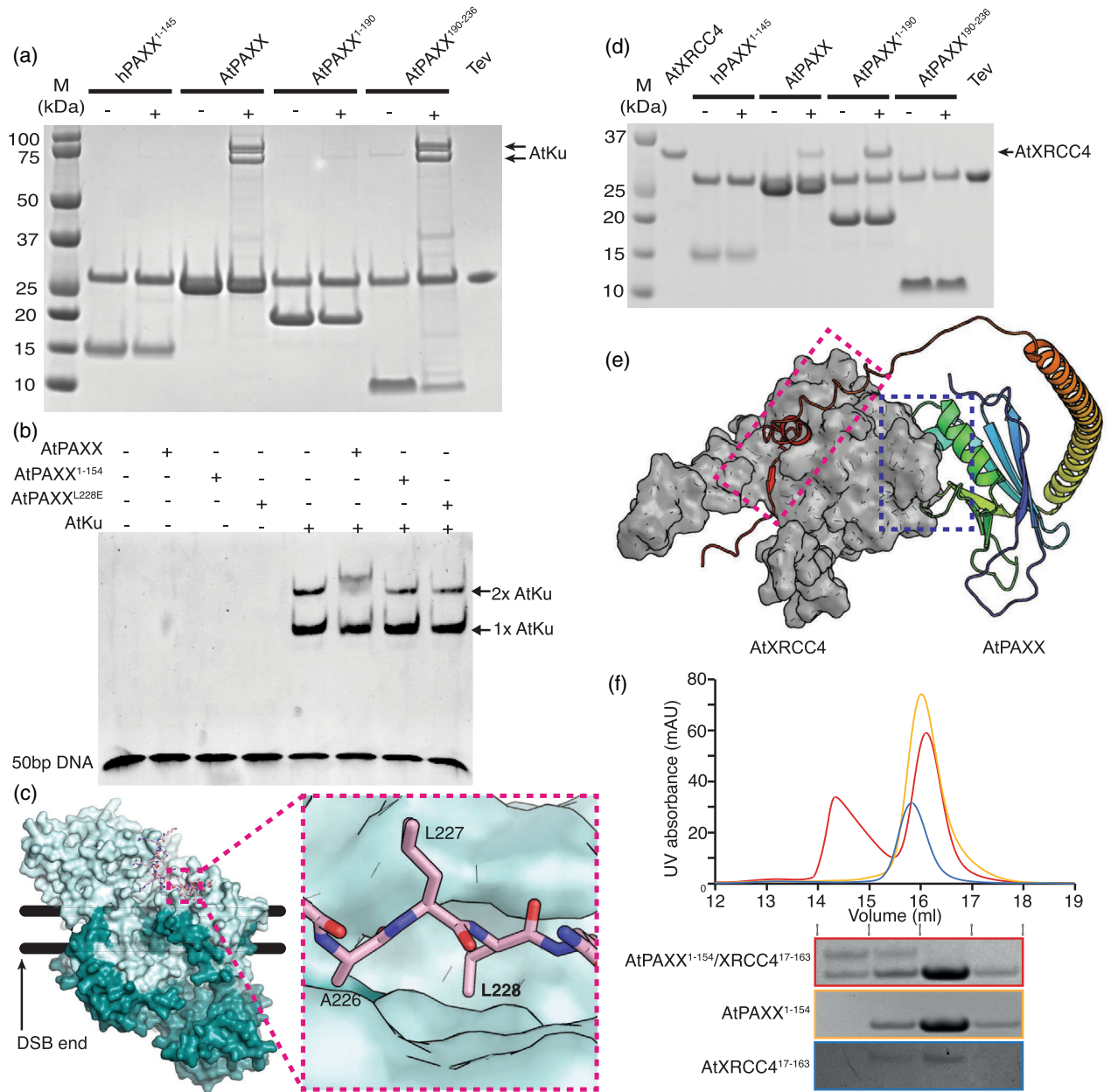
Figure 3. AtPAXX interacts with AtKu and AtXRCC4.

(a) The C-terminus of AtPAXX interacts with AtKu. An SDS-PAGE gel shows the result of a GST-pulldown experiment. '+' indicates GST-tagged PAXX constructs were incubated with insect cell lysates expressing AtKu. The PAXX constructs were released from glutathione beads by digesting the linker between GST and PAXX by Tev.

(b) AtPAXX interacts with AtKu bound on DNA. An electrophoretic mobility shift assay result shows that AtPAXX but not AtPAXX$^{1-154}$ and AtPAXX^{L228E}, interacts with mainly two AtKu molecules (2x AtKu) bound on 50 bp DNA.

(c) AlphaFold model of AtPAXX-Ku complex. The model shows that the C-terminus of AtPAXX interacts (pink) with the surface predominantly created by AtKu70 (cyan; AtKu80 is shown in deep teal). Two black lines represent double-stranded DNA.

(d) AtPAXX interacts with AtXRCC4. An SDS-PAGE gel shows the result of a GST-pulldown experiment. '+' indicates GST-tagged PAXX constructs were incubated with AtXRCC4. The experiment was done in a similar manner to (b).

(e) AlphaFold model of AtPAXX/XRCC4 complex. Potential-interacting regions of AtPAXX and AtXRCC4 (grey) were indicated with blue (N-terminal interaction) and red (C-terminal interaction) dotted rectangles. AtPAXX is presented in a gradient of blue (N-terminus) and red (C-terminus).

(f) The N-terminal domain of AtPAXX interacts with that of AtXRCC4. A graph shows size exclusion chromatography analyses of the complex and individual proteins (red) of AtPAXX (yellow) and AtXRCC4 (blue). SDS-PAGE gels show four fractions between elution volumes between 14 and 18 ml of the complex and individual samples.
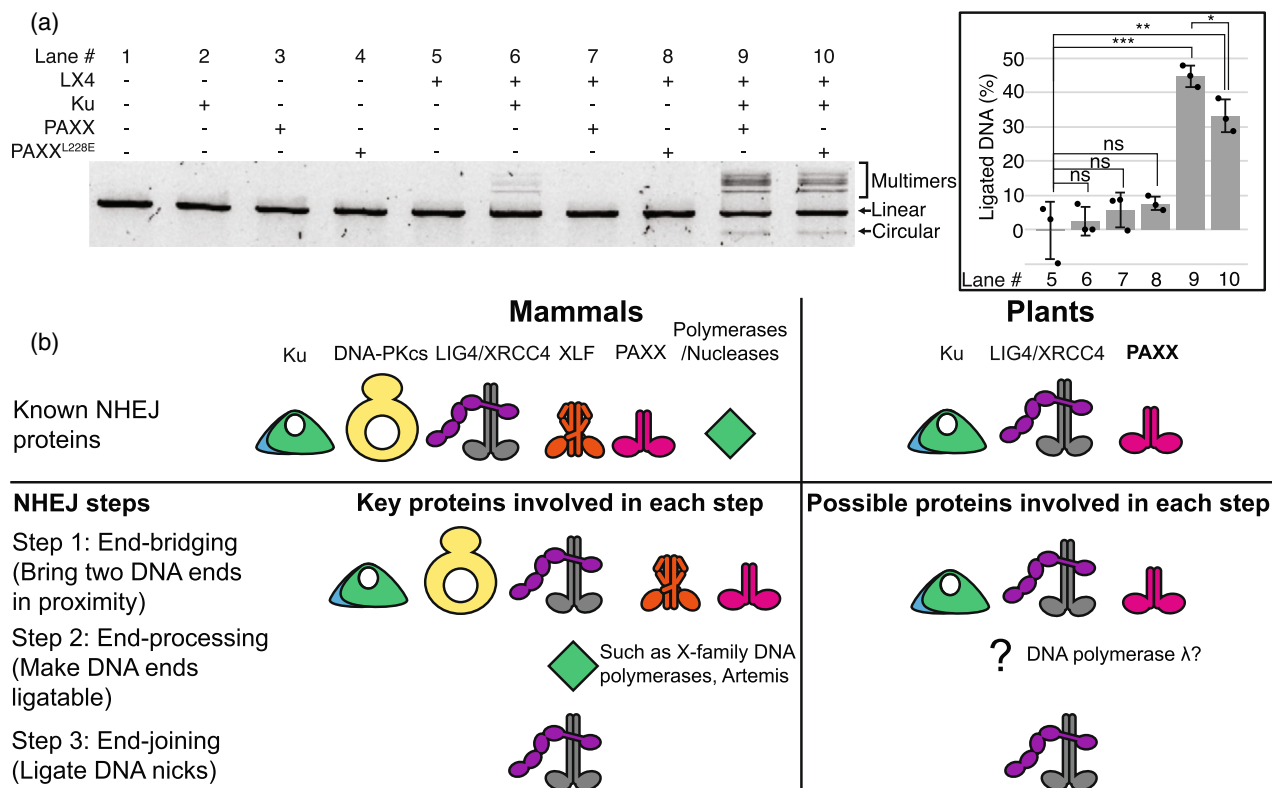
Figure 4. AtPAXX is a plant NHEJ factor.

(a) AtPAXX stimulates blunt-ended DNA-joining by AtKu and AtLIG4/XRCC4. An agarose gel shows the result of a DNA ligation assay. Linearised blunt-ended plasmids (linear indicated by the top arrow) were incubated with different protein combinations indicated by '+'. The bottom arrow and a bracket indicate ligated DNA substrates (circular and multimers). The bar graphs on the right show gel densitometry analysis results ($n = 3$). Bars and error bars are averages and standard deviations, and data points are shown in black dots. P-values of t-test results calculated against the result of AtLIG4/XRCC4 only (lane 5) are indicated by ns, *, ** and *** ($P > 0.05$, ≤ 0.05, ≤ 0.01 and ≤ 0.001 respective).

(b) Comparison of mammalian and plant NHEJ. Protein icons indicate known NHEJ proteins involved in three different steps. NHEJ, non-homologous end joining.

NHEJ. We cannot rule out that plants have other redundant proteins to PAXX. It is intriguing to further investigate how the NHEJ synapse complex is formed in plants and how this impacts the efficiency of NHEJ.

In summary, our results suggest that NHEJ and its constituent proteins evolved to meet requirements in genome maintenance in different eukaryotes while retaining key conserved functions. This suggests that there is no universal model to explain NHEJ in eukaryotes. Understanding how exactly NHEJ proteins work in different organisms needs detailed experimental studies in these organisms.

METHODS

Sequence analysis

Homologues of NHEJ proteins were searched using a combination of Psi-Blast (Altschul et al., 1997) and Jackhmmer (Johnson et al., 2010). Sequence alignments were generated using ClustalOmega (Sievers et al., 2011) and Muscle (Edgar, 2004) on Seaview (Gouy et al., 2010). Phylogenetic trees are created using PhyML (Guindon et al., 2010) on Seaview and visualised using FigTree.

Single chains of AlphaFold models were downloaded from Uniprot, whereas models of homo- and heteromultimers were generated using AlphaFold Colab (Jumper et al., 2021; Mirdita et al., 2022).

Constructs

The open-reading frame of *A. thaliana paxx* was synthesised and codon-optimised for the *Escherichia coli* expression (Thermo-Fisher Scientific, Waltham, MA, USA), whereas *A. thaliana lig4* and *xrcc4* genes were gifts from Dr Chris West (University of Leeds). *Atpaxx* and *Atxrcc4* constructs were cloned into pHAT4 and pGAT3 vectors respectively (Peränen et al., 1996). *A. thaliana ku70/80* co-expression vector pFastDual-*ku70/80* was a gift from Dr Karel Říha (CEIT, MU) (Valuchova et al., 2017). *A. thaliana lig4* and *xrcc4* genes were cloned into the pFastBac Dual vector.

Protein purification

AtPAXX fused to a hexahistidine tag (pHAT4-*atpaxx*) was expressed in BL21(DE3) cells (NEB). The proteins were purified using Ni-NTA (Qiagen), and the His-tag was cleaved by the Tev protease. Untagged AtPAXX was further purified using a HiTrap Q column (Cytiva, Little Chalfont, UK). The single peak containing the protein was collected and concentrated using Vivaspin 20

(Cytiva). The protein in a storage buffer (20 mM Tris pH 7.4, 150 mM NaCl, 2 mM DTT) was snap-frozen in liquid nitrogen and stored at −80°C. AtPAXX$^{1-154}$ was also purified in a similar way. AtPAXX^{L228E} fused to hexahistidine followed by a GST tag (pGAT3-*atpaxx*^{L228E}) was expressed in BL21(DE3) (NEB) and purified using a 5 ml GSTrap HP column (Cytiva). The His and GST tag were cleaved by the Tev protease and separated from AtPAXX^{L228E} using Ni-NTA (Qiagen). AtPAXX^{L228E} was loaded onto a HiTrap Q HP column equilibrated with 20 mM Tris pH 7.4, 50 mM NaCl, 2 mM DTT eluted using a linear gradient of the same buffer but 500 mM NaCl. Peak fractions containing AtPAXX^{L228E} were collected and added to 20 mM Tris–HCl pH 7.4, 2 mM DTT to adjust the NaCl concentration to 150 mM based on the chromatography profile of the peak fractions. The protein was concentrated using Vivaspin 20, snap-frozen in liquid nitrogen and stored at −80°C (Figure S2).

AtXRCC4 fused to a hexahistidine followed by a GST tag (pGAT3-*atxrcc4*) was expressed in BL21(DE3) cells (NEB). The proteins were purified using GST sepharose (Cytiva), and the his-GST tag was cleaved by the Tev protease. The tag was separated from untagged AtXRCC4 using Ni-NTA. Purified AtXRCC4 was stored in the same final buffer used for AtPAXX and stored at −80°C after snap-frozen in liquid nitrogen (Figure S2). A shorter version of AtXRCC4$^{17-163}$ was also purified in a similar way (Figure S2).

AtKu70/80 was expressed in Sf9 cells and purified as described previously (Ochi et al., 2015). The frozen cells were suspended in 50 mM sodium phosphate buffer pH 7.4, 150 mM NaCl, 5% (v/v) glycerol, 2 mM 2-mercaptoethanol, 10 mM imidazole, 20 units of benzonase (Millipore, Burlington, MA, USA) supplied with cOmplete, EDTA-free protease inhibitor cocktail (Roche, Burgess Hill, UK) and lysed by sonication. The lysate was incubated for 30 min on ice and diluted fivefold with 50 mM sodium phosphate buffer pH 8.0, 1 M NaCl, 5% (v/v) glycerol, 400 mM ammonium acetate, 2 mM 2-mercaptoethanol, 10 mM imidazole. The cell debris was removed by centrifugation for 15 min at 35 000 *g*, 4°C. His-tagged AtKu70/80 was isolated from the supernatant using HisPur Ni-NTA resins (Thermo Fisher) and dialysed in 20 mM HEPES pH 7.5, 200 mM NaCl, 2 mM 2-mercaptoethanol, 30 mM imidazole overnight after the Tev protease was added to cleave the His-tag. Untagged AtKu70/80 was isolated from the His-tag using the Ni-NTA resin and loaded onto a HiTrap Heparin column (Cytiva) equilibrated with the dialysis buffer. AtKu70/80 bound on the column was eluted with 20 mM HEPES pH 7.5, 500 mM NaCl, 2 mM DTT, and concentrated and buffer-exchanged to 20 mM HEPES pH 7.5, 150 mM NaCl, 2 mM DTT before being snap-frozen in liquid nitrogen to be stored at −80°C (Figure S2).

AtLIG4/XRCC4 (pFastBacDual-*atlig4/xrcc4*) were co-expressed in Sf9 cells, which were suspended in 50 mM Tris–HCl pH 7.4, 300 mM NaCl, 2 mM 2-mercaptoethanol, 10 mM imidazole, 5% (v/v) glycerol supplied with cOmplete, EDTA-free protease inhibitor cocktail (Roche). The cells were lysed by sonication, and the cell debris was removed by centrifugation for 30 min at 35 000 *g*, 4°C. The supernatant was collected and was incubated with HisPur Ni-NTA resins (Thermo Fisher) for 90 min at 4°C. The resins were washed with 50 mM HEPES pH 7.0, 300 mM NaCl, 2 mM 2-mercaptoethanol, 30 mM imidazole, 5% (v/v) glycerol, followed by 50 mM HEPES pH 7.0, 200 mM NaCl, 2 mM 2-mercaptoethanol, 30 mM imidazole, 5% (v/v) glycerol. Bound proteins on the resins were eluted using 50 mM HEPES pH 7.0, 200 mM NaCl, 2 mM 2-mercaptoethanol, 300 mM imidazole, 5% (v/v) glycerol. The eluted proteins were diluted twofold using water and loaded onto a 5 ml HiTrap Heparin HP column (Cytiva) equilibrated with 20 mM HEPES pH 7.5, 100 mM NaCl, 2 mM DTT. Bound proteins on the column were eluted using a linear gradient of 20 mM HEPES

pH 7.5, 1 M NaCl, 2 mM DTT. Fractions containing AtLIG4/XRCC4 were collected, concentrated and loaded onto a Superdex 200 increase 10/300 GL (Cytiva) equilibrated with 20 mM HEPES pH 7.5, 200 mM NaCl, 2 mM DTT. Fractions containing AtLIG4/XRCC4 were collected and concentrated before being snap-frozen in liquid nitrogen and stored at −80°C (Figure S2).

X-ray crystallography

Crystals of AtPAXX were obtained using the sitting drop vapour diffusion method where 17 mg ml^{-1} of AtPAXX was mixed with 20% (w/v) PEG550MME, 100 mM NaCl, 100 mM Bicine pH 9.0 in the 1:1 ration. The crystals were cryo-protected in 33% (w/v) PEG550MME, 100 mM NaCl, 100 mM Bicine pH 9.0 and snap-frozen in liquid nitrogen. Their diffraction data were collected at I04-1 in Diamond Light Source and were processed using their automatic pipeline Dial (Waterman et al., 2016) and reduced using Aimless (Evans, 2011) (Table 1). Initial phases of structure factors were obtained by the molecular replacement using the Phaser module in PHENIX suite (Adams et al., 2010), using a search model of the AtPAXX dimer generated using AlphaFold Colab (Jumper et al., 2021; Mirdita et al., 2022). The model of AtPAXX was manually and computationally refined using Coot and the phenix.refine module until no further improvement of the electron density map of AtPAXX was observed. TLS groups were selected as each chain, and non-crystallographic symmetry restraints were applied for the model refinement.

GST-pulldown assay

Three hundred micrograms of GST-AtPAXX constructs or GST-hPAXX$^{1-145}$ were mixed with 20 µl of glutathione sepharose 4B beads (Cytiva) and incubated for 30 min at room temperature in a GST-pulldown buffer (20 mM Tris–HCl pH 8.0, 200 mM NaCl, 5 mM DTT, 10% [v/v] glycerol, 0.5% [v/v] NP-40). After washing unbound proteins, the beads were suspended in 500 µl of the pulldown buffer and incubated with 100 µg of purified AtXRCC4 for 30 min at room temperature or with lysates of 50 ml cell culture of Sf9 cells expressing AtKu70/80 described above for 60 min at 4°C. After washing beads, Tev protease was added to them to release PAXX constructs and their bound proteins from the beads. Supernatants were then analysed by SDS-PAGE.

Electrophoretic mobility shift assay

The protocol has been described previously (Ochi et al., 2015). 20 nM 6FAM-labelled 50 bp dsDNA was mixed with 20 nM AtKu70/80 or 200 nM AtPAXX, AtPAXX$^{1-154}$, AtPAXX^{L228E} or different combinations of them as described in the main text in 20 mM Tris–HCl pH 7.4, 100 mM KCl, 5% glycerol, 1 mM DTT, 10 µg ml^{-1} BSA for 30 min at room temperature. The samples were applied onto a 5% polyacrylamide gel in 0.5× TBE buffer for electrophoresis at 4°C. DNA was visualised by iBright imager (ThermoFisher Scientific).

Size exclusion chromatography

One hundred micrometres of AtPAXX$^{1-154}$, AtXRCC4$^{17-163}$ or their mixture were loaded onto a Superdex 200 increase 10/300 GL (Cytiva) column equilibrated with 20 mM Tris–HCl pH 7.4, 150 mM NaCl, 2 mM DTT. Fractions were collected every 1 ml and analysed by SDS-PAGE.

DNA ligation assays

Blunt-ended DNA substrates were prepared by digesting the pGEX6-p1 vector by EcoRV (New England Biolabs, Ipswich, MA,

USA) followed by heat inactivation and purification using a gel extraction kit (Qiagen, Germantown, MD, USA). Fifty nanograms of the DNA were incubated with different combinations of NHEJ proteins (25 nM AtKu70/80, 25 nM AtLIG4/XRCC4, 250 nM AtPAXX, 250 nM AtPAXX^{L228E}) in 20 µl of a ligation buffer (25 mM Tris–HCl pH 7.4, 100 mM KCl, 1 mM MgCl$_2$, 1 mM DTT, 5% [v/v] glycerol, 10 µM ATP, 10 µg ml^{-1} BSA). First, the NHEJ proteins except for AtLIG4/XRCC4 were incubated with DNA for 5 min and another 30 min at 37°C after the ligase was added. The reaction was stopped by adding 2 µL of 0.1% (w/v) SDS, 100 mM EDTA and 1 µl of 4 mg ml^{-1} Proteinase K (ApexBio Technology, Houston, TX, USA) and incubating for 30 min at 50°C. Reaction mixtures were analysed by being applied onto 0.8% agarose gel supplied with SYBR Safe (ThermoFisher Scientific) in TBE buffer for electrophoresis.

AUTHOR CONTRIBUTIONS

TO designed the project and experiments, and carried out computational analyses. HK and TO carried out the experiments. TO wrote the manuscript with the help of HK.

ACKNOWLEDGEMENTS

CONFLICT OF INTEREST STATEMENT

The authors do not have any conflict of interest.

DATA AVAILABILITY STATEMENT

The structure factors and coordinates of the crystal structure of AtPAXX are available in the Protein Data Bank (PDB code: 8P26).

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Table S1. Accession numbers of NHEJ protein orthologs used for Figure 1.
Table S2. PAXX orthologs in streptophyta.
Figure S1. AlphaFold models of NHEJ orthologs.
Figure S2. Purified recombinant Arabidopsis NHEJ proteins.

REFERENCES

Abramowski, V., Etienne, O., Elsaid, R., Yang, J., Berland, A., Kermasson, L. et al. (2017) PAXX and Xlf interplay revealed by impaired CNS development and immunodeficiency of double KO mice. *Cell Death and Differentiation*, 25, 444–452. Available from: https://doi.org/10.1038/cdd.2017.184

Adams, P.D., Afonine, P.V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N. et al. (2010) PHENIX: a comprehensive python-based system for macromolecular structure solution. *Acta Crystallographica Section D*, 66, 213–221. Available from: https://doi.org/10.1107/S0907444909052925

Altschul, S., Madden, T., Schäffer, A., Zhang, J., Zhang, Z., Miller, W. et al. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Research*, 25, 3389–3402. Available from: https://doi.org/10.1093/nar/25.17.3389

Andres, S.N., Vergnes, A., Ristic, D., Wyman, C., Modesti, M. & Junop, M. (2012) A human XRCC4–XLF complex bridges DNA. *Nucleic Acids Research*, 40, 1868–1878. Available from: https://doi.org/10.1093/nar/gks022

Balmus, G., Barros, A., Wijnhoven, P., Lescale, C., Hasse, H., Boroviak, K. et al. (2016) Synthetic lethality between PAXX and XLF in mammalian development. *Genes & Development*, 30, 2152–2157. Available from: https://doi.org/10.1101/gad.290510.116

Bienstock, R., Beard, W. & Wilson, S. (2014) Phylogenetic analysis and evolutionary origins of DNA polymerase X-family members. *DNA Repair (Amst)*, 22, 77–88.

Brouwer, I., Sitters, G., Candelli, A., Heerema, S., Heller, I., Melo de, A. et al. (2016) Sliding sleeves of XRCC4–XLF bridge DNA and connect fragments of broken DNA. *Nature*, 535, 566–569. Available from: https://doi.org/10.1038/nature18643

Bryans, M., Valenzano, M.C. & Stamato, T.D. (1999) Absence of DNA ligase IV protein in XR-1 cells: evidence for stabilization by XRCC4. *Mutation Research/DNA Repair*, 433, 53–58. Available from: https://doi.org/10.1016/S0921-8777(98)00063-9

Callebaut, I., Malivert, L., Fischer, A., Mornon, J.-P., Revy, P. & de Villartay, J.-P. (2006) Cernunnos interacts with the XRCC4 DNA-ligase IV complex and is homologous to the yeast nonhomologous end-joining factor Nej1. *The Journal of Biological Chemistry*, 281, 13857–13860. Available from: https://doi.org/10.1074/jbc.C500473200

Callebaut, I., Moshous, D., Mornon, J. & de Villartay, J. (2002) Metallo-β-lactamase fold within nucleic acids processing enzymes: the β-CASP family. *Nucleic Acids Research*, 30, 3592–3601. Available from: https://doi.org/10.1093/nar/gkf470

Cary, R.B., Peterson, S.R., Wang, J., Bear, D.G., Bradbury, E.M. & Chen, D.J. (1997) DNA looping by Ku and the DNA-dependent protein kinase. *Proceedings of the National Academy of Sciences of the United States of Amreica*, 94, 4267–4272.

Chalker, D. & Yao, M. (2011) DNA elimination in ciliates: transposon domestication and genome surveillance. *Annual Review of Genetics*, 45, 227–246. Available from: https://doi.org/10.1146/annurev-genet-110410-132432

Chang, H.H.Y., Pannunzio, N.R., Adachi, N. & Lieber, M.R. (2017) Non-homologous DNA end joining and alternative pathways to double-strand break repair. *Nature Reviews. Molecular Cell Biology*, 18, 495–506. Available from: https://doi.org/10.1038/nrm.2017.48

Chaplin, A.K., Hardwick, S.W., Stavridi, A.K., Buehl, C.J., Goff, N.J., Ropars, V. et al. (2021) Cryo-EM of NHEJ supercomplexes provides insights into DNA repair. *Molecular Cell*, 81, 3400–3409.e3. Available from: https://doi.org/10.1016/J.MOLCEL.2021.07.005

Chen, L., Trujillo, K., Sung, P. & Tomkinson, A.E. (2000) Interactions of the DNA ligase IV-XRCC4 complex with DNA ends and the DNA-dependent protein kinase. *The Journal of Biological Chemistry*, 275, 26196–26205. Available from: https://doi.org/10.1074/jbc.M000491200

Chen, S., Lee, L., Naila, T., Fishbain, S., Wang, A., Tomkinson, A.E. et al. (2021) Structural basis of long-range to short-range synaptic transition in NHEJ. *Nature*, 593, 294–298. Available from: https://doi.org/10.1038/s41586-021-03458-7

Daley, J.M., Palmbos, P.L., Wu, D. & Wilson, T.E. (2005) Nonhomologous end joining in yeast. *Annual Review of Genetics*, 39, 431–451. Available from: https://doi.org/10.1146/annurev.genet.39.073003.113340

DeFazio, L.G., Stansel, R.M., Griffith, J.D. & Chu, G. (2002) Synapsis of DNA ends by DNA-dependent protein kinase. *The EMBO Journal*, 21, 3192–3200. Available from: https://doi.org/10.1093/emboj/cdf299

Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, 32, 1792–1797. Available from: https://doi.org/10.1093/nar/gkh340

Ellenberger, T. & Tomkinson, A. (2008) Eukaryotic DNA ligases: structural and functional insights. *Annual Review of Biochemistry*, 77, 313–338. Available from: https://doi.org/10.1146/annurev.biochem.77.061306.123941

Evans, P.R. (2011) An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallographica Section D*, 67, 282–292. Available from: https://doi.org/10.1107/S090744491003982X

Gehrke, F., Schindele, A. & Puchta, H. (2022) Nonhomologous end joining as key to CRISPR/Cas-mediated plant chromosome engineering. *Plant Physiology*, 188, 1769–1779. Available from: https://doi.org/10.1093/PLPHYS/KIAB572

Gelvin, S.B. (2017) Integration of agrobacterium T-DNA into the plant genome. *Annual Review of Genetics*, **51**, 195–217. Available from: https://doi.org/10.1146/ANNUREV-GENET-120215-035320

Genois, M.-M., **Paquet, E.**, **Laffitte, M.-C.**, **Maity, R.**, **Rodrigue, A.**, **Ouellette, M.** *et al.* (2014) DNA repair pathways in trypanosomatids: from DNA repair to drug resistance. *Microbiology and Molecular Biology Reviews*, **78**, 40–73. Available from: https://doi.org/10.1128/mmbr.00045-13

Gouy, M., **Guindon, S.** & **Gascuel, O.** (2010) SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. *Molecular Biology and Evolution*, **27**, 221–224. Available from: https://doi.org/10.1093/molbev/msp259

Graham, T., **Walter, J.** & **Loparo, J.** (2016) Two-stage synapsis of DNA ends during non-homologous end joining. *Molecular Cell*, **61**, 850–858. Available from: https://doi.org/10.1016/j.molcel.2016.02.010

Graham, T.G.W., **Carney, S.M.**, **Walter, J.C.** & **Loparo, J.J.** (2018) A single XLF dimer bridges DNA ends during nonhomologous end joining. *Nature Structural & Molecular Biology*, **25**, 877–884. Available from: https://doi.org/10.1038/s41594-018-0120-y

Guindon, S., **Dufayard, J.-F.**, **Lefort, V.**, **Anisimova, M.**, **Hordijk, W.** & **Gascuel, O.** (2010) New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Systematic Biology*, **59**, 307–321. Available from: https://doi.org/10.1093/sysbio/syq010

Hentges, P., **Ahnesorg, P.**, **Pitcher, R.S.**, **Bruce, C.K.**, **Kysela, B.**, **Green, A.J.** *et al.* (2006) Evolutionary and functional conservation of the DNA non-homologous end-joining protein, XLF/Cernunnos. *The Journal of Biological Chemistry*, **281**, 37517–37526. Available from: https://doi.org/10.1074/jbc.M608727200

Herrmann, G., **Lindahl, T.** & **Schar, P.** (1998) *Saccharomyces cerevisiae* LIF1: a function involved in DNA double-strand break repair related to mammalian XRCC4. *The EMBO Journal*, **17**, 4188–4198. Available from: https://doi.org/10.1093/emboj/17.14.4188

Holm, L. (2022) Dali server: structural unification of protein families. *Nucleic Acids Research*, **1**, W210–W215. Available from: https://doi.org/10.1093/NAR/GKAC387

Hung, P., **Chen, B.-R.**, **George, R.**, **Liberman, C.**, **Morales, A.**, **Colon-Ortiz, P.** *et al.* (2016) Deficiency of XLF and PAXX prevents DNA double-strand break repair by non-homologous end joining in lymphocytes. *Cell Cycle*, **16**, 286–295. Available from: https://doi.org/10.1080/15384101.2016.1253640

Johnson, L. S., **Eddy, S. R.**, & **Portugaly, E.** (2010). Hidden Markov model speed heuristic and iterative HMM search procedure. *BMC Bioinformatics*, **11**. https://doi.org/10.1186/1471-2105-11-431

Jiang, W., **Crowe, J.**, **Liu, X.**, **Nakajima, S.**, **Wang, Y.**, **Li, C.** *et al.* (2015) Differential phosphorylation of DNA-PKcs regulates the interplay between end-processing and end-ligation during nonhomologous end-joining. *Molecular Cell*, **58**, 172–185. Available from: https://doi.org/10.1016/j.molcel.2015.02.024

Jumper, J., **Evans, R.**, **Pritzel, A.**, **Green, T.**, **Figurnov, M.**, **Ronneberger, O.** *et al.* (2021) Highly accurate protein structure prediction with alphafold. *Nature*, **596**, 583–589. Available from: https://doi.org/10.1038/s41586-021-03819-2

Kitagawa, D., **Vakonakis, I.**, **Olieric, N.**, **Hilbert, M.**, **Keller, D.**, **Olieric, V.** *et al.* (2011) Structural basis of the 9-fold symmetry of centrioles. *Cell*, **144**, 364–375. Available from: https://doi.org/10.1016/j.cell.2011.01.008

Kumar, V., **Alt, F.** & **Frock, R.** (2016) PAXX and XLF DNA repair factors are functionally redundant in joining DNA breaks in a G1-arrested progenitor B-cell line. *Proceedings of the National Academy of Sciences of the United States of America*, **113**, 10619–10624. Available from: https://doi.org/10.1073/pnas.1611882113

Lees-Miller, J.P., **Cobban, A.**, **Katsonis, P.**, **Bacolla, A.**, **Tsutakawa, S.E.**, **Hammel, M.** *et al.* (2021) Uncovering DNA-PKcs ancient phylogeny, unique sequence motifs and insights for human disease. *Progress in Biophysics and Molecular Biology*, **163**, 87–108. Available from: https://doi.org/10.1016/J.PBIOMOLBIO.2020.09.010

Lescale, C., **Lenden Hasse, H.**, **Blackford, A.**, **Balmus, G.**, **Bianchi, J.**, **Yu, W.** *et al.* (2016) Specific roles of XRCC4 paralogs PAXX and XLF during V(D)J recombination. *Cell Reports*, **16**, 2967–2979.

Liu, L., **Chen, X.**, **Li, J.**, **Wang, H.**, **Buehl, C.J.**, **Goff, N.J.** *et al.* (2022) Autophosphorylation transforms DNA-PK from protecting to processing DNA ends. *Molecular Cell*, **82**, 177–189.e4. Available from: https://doi.org/10.1016/J.MOLCEL.2021.11.025

Liu, X., **Shao, Z.**, **Jiang, W.**, **Lee, B.J.** & **Zha, S.** (2017) PAXX promotes KU accumulation at DNA breaks and is essential for end-joining in XLF-deficient mice. *Nature Communications*, **8**, 1–13. Available from: https://doi.org/10.1038/ncomms13816

Matsuda, A., **Shieh, A.**, **Chalker, D.** & **Forney, J.** (2010) The conjugation-specific Die5 protein is required for development of the somatic nucleus in both *Paramecium* and *Tetrahymena*. *Eukaryotic Cell*, **9**, 1087–1099. Available from: https://doi.org/10.1128/ec.00379-09

Mirdita, M., **Schütze, K.**, **Moriwaki, Y.**, **Heo, L.**, **Ovchinnikov, S.** & **Steinegger, M.** (2022) ColabFold: making protein folding accessible to all. *Nature Methods*, **19**, 679–682. Available from: https://doi.org/10.1038/s41592-022-01488-1

Moon, A.F., **Garcia-Diaz, M.**, **Batra, V.K.**, **Beard, W.A.**, **Bebenek, K.**, **Kunkel, T.A.** *et al.* (2007) The X family portrait: structural insights into biological functions of X family polymerases. *DNA Repair (Amst)*, **6**, 1709–1725. Available from: https://doi.org/10.1016/j.dnarep.2007.05.009

Nenarokova, A., **Záhonová, K.**, **Krasilnikov, M.**, **Gahura, O.**, **McCulloch, R.**, **Zíková, A.** *et al.* (2019) Causes and effects of loss of classical nonhomologous end joining pathway in parasitic eukaryotes. *mBio*, **10**, e01541–19. Available from: https://doi.org/10.1128/MBIO.01541-19/ASSET/C043F9D2-F1C4-4B62-9618-66100EB73E6B/ASSETS/GRAPHIC/MBIO.01541-19-F0003.JPEG

Ochi, T., **Blackford, A.N.A.N.**, **Coates, J.**, **Jhujh, S.**, **Mehmood, S.**, **Tamura, N.** *et al.* (2015) PAXX, a paralog of XRCC4 and XLF, interacts with Ku to promote DNA double-strand break repair. *Science (1979)*, **347**, 185–188. Available from: https://doi.org/10.1126/science.1261971

Ochi, T., **Quarantotti, V.**, **Lin, H.**, **Jullien, J.**, **Rosa E Silva, I.**, **Boselli, F.** *et al.* (2020) CCDC61/VFL3 is a paralog of SAS6 and promotes ciliary functions. *Structure*, **28**, 674–689. Available from: https://doi.org/10.1016/j.str.2020.04.010

Peränen, J., **Rikkonen, M.**, **Hyvönen, M.** & **Kääriäinen, L.** (1996) T7 vectors with a modified T7lac promoter for expression of proteins in *Escherichia coli*. *Analytical Biochemistry*, **236**, 371–373. Available from: https://doi.org/10.1006/abio.1996.0187

Puchta, H. (2005) The repair of double-strand breaks in plants: mechanisms and consequences for genome evolution. *Journal of Experimental Botany*, **56**, 1–14. Available from: https://doi.org/10.1093/JXB/ERI025

Riha, K., **Watson, J.M.**, **Parkey, J.** & **Shippen, D.E.** (2002) Telomere length deregulation and enhanced sensitivity to genotoxic stress in Arabidopsis mutants deficient in Ku70. *The EMBO Journal*, **21**, 2819–2826. Available from: https://doi.org/10.1093/EMBOJ/21.11.2819

Shieh, A. & **Chalker, D.** (2013) LIA5 is required for nuclear reorganization and programmed DNA rearrangements occurring during *Tetrahymena* macronuclear differentiation. *PLoS One*, **8**, e75337. Available from: https://doi.org/10.1371/journal.pone.0075337

Shuman, S. & **Glickman, M.S.** (2007) Bacterial DNA repair by non-homologous end joining. *Nature Reviews. Microbiology*, **5**, 852–861. Available from: https://doi.org/10.1038/nrmicro1768

Sibanda, B.L., **Critchlow, S.E.**, **Begun, J.**, **Pei, X.Y.**, **Jackson, S.P.**, **Blundell, T.L.** *et al.* (2001) Crystal structure of an Xrcc4-DNA ligase IV complex. *Nature Structural & Molecular Biology*, **8**, 1015–1019. Available from: https://doi.org/10.1038/nsb725

Sievers, F., **Wilm, A.**, **Dineen, D.**, **Gibson, T.J.**, **Karplus, K.**, **Li, W.** *et al.* (2011) Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal omega. *Molecular Systems Biology*, **7**, 539. Available from: https://doi.org/10.1038/msb.2011.75

Silva, G.L.A., **Tosi, L.R.O.**, **McCulloch, R.** & **Black, J.A.** (2021) Unpicking the roles of DNA damage protein kinases in Trypanosomatids. *Frontiers in Cell and Development Biology*, **9**, 2110. Available from: https://doi.org/10.3389/fcell.2021.636615

Spagnolo, L., **Rivera-Calzada, A.**, **Pearl, L.H.** & **Llorca, O.** (2006) Three-dimensional structure of the human DNA-PKcs/Ku70/Ku80 complex assembled on DNA and its implications for DNA DSB repair. *Molecular Cell*, **22**, 511–519. Available from: https://doi.org/10.1016/j.molcel.2006.04.013

Sui, J., **Zhang, S.** & **Chen, B.P.C.** (2020) DNA–dependent protein kinase in telomere maintenance and protection. *Cellular & Molecular Biology Letters*, **25**, 1–14. Available from: https://doi.org/10.1186/S11658-020-0199-0

Tamura, K., **Adachi, Y.**, **Chiba, K.**, **Oguchi, K.** & **Takahashi, H.** (2002) Identification of Ku70 and Ku80 homologues in *Arabidopsis thaliana*: evidence for a role in the repair of DNA double-strand breaks. *The Plant Journal*,

29, 771–781. Available from: https://doi.org/10.1046/j.1365-313x.2002.01258.x

Valuchova, S., **Fulnecek, J.**, **Prokop, Z.**, **Stolt-Bergner, P.**, **Janouskova, E.**, **Hofr, C.** *et al.* (2017) Protection of Arabidopsis blunt-ended telomeres is mediated by a physical association with the Ku heterodimer. *Plant Cell*, **29**, 1533–1545. Available from: https://doi.org/10.1105/tpc.17.00064

van Breugel, M., **Hirono, M.**, **Andreeva, A.**, **Yanagisawa, H.**, **Yamaguchi, S.**, **Nakazawa, Y.** *et al.* (2011) Structures of SAS-6 suggest its organization in centrioles. *Science (1979)*, **331**, 1196–1199. Available from: https://doi.org/10.1126/science.1199325

Vujin, A., **Jones, S.J.** & **Zetka, M.** (2020) NHJ-1 is required for canonical non-homologous end joining in *Caenorhabditis elegans*. *Genetics*, **215**, 635–651. Available from: https://doi.org/10.1534/genetics.120.303328

Wang, J.L., **Duboc, C.**, **Wu, Q.**, **Ochi, T.**, **Liang, S.**, **Tsutakawa, S.E.** *et al.* (2018) Dissection of DNA double-strand-break repair using novel single-molecule forceps. *Nature Structural & Molecular Biology*, **25**, 482–487. Available from: https://doi.org/10.1038/s41594-018-0065-1

Waterman, D.G., **Winter, G.**, **Gildea, R.J.**, **Parkhurst, J.M.**, **Brewster, A.S.**, **Sauter, N.K.** *et al.* (2016) Diffraction-geometry refinement in the DIALS framework. *Acta Crystallographica Section D*, **72**, 558–575. Available from: https://doi.org/10.1107/s2059798316002187

West, C.E., **Waterworth, W.M.**, **Jiang, Q.** & **Bray, C.M.** (2000) Arabidopsis DNA ligase IV is induced by γ-irradiation and interacts with an Arabidopsis homologue of the double strand break repair protein XRCC4. *The Plant Journal*, **24**, 67–78. Available from: https://doi.org/10.1046/J.1365-313X.2000.00856.X

Yano, K. & **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**, 1321–1325. Available from: https://doi.org/10.4161/cc.7.10.5898

Yosaatmadja, Y., **Baddock, H.T.**, **Newman, J.A.**, **Bielinski, M.**, **Gavard, A.E.**, **Mukhopadhyay, S.M.M.** *et al.* (2021) Structural and mechanistic insights into the Artemis endonuclease and strategies for its inhibition. *Nucleic Acids Research*, **49**, 9310–9326. Available from: https://doi.org/10.1093/NAR/GKAB693

Zhao, B., **Watanabe, G.**, **Morten, M.J.**, **Reid, D.A.**, **Rothenberg, E.** & **Lieber, M.R.** (2019) The essential elements for the noncovalent association of two DNA ends during NHEJ synapsis. *Nature Communications*, **10**, 1–12. Available from: https://doi.org/10.1038/s41467-019-11507-z