Evolutionary and functional conservation of the DNA non homologous end joining protein, XLF/Cernunnos


This version is available from Sussex Research Online: http://sro.sussex.ac.uk/id/eprint/458/

This document is made available in accordance with publisher policies and may differ from the published version or from the version of record. If you wish to cite this item you are advised to consult the publisher's version. Please see the URL above for details on accessing the published version.

Copyright and reuse:
Sussex Research Online is a digital repository of the research output of the University.

Copyright and all moral rights to the version of the paper presented here belong to the individual author(s) and/or other copyright owners. To the extent reasonable and practicable, the material made available in SRO has been checked for eligibility before being made available.

Copies of full text items generally can be reproduced, displayed or performed and given to third parties in any format or medium for personal research or study, educational, or not-for-profit purposes without prior permission or charge, provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.
EVOLUTIONARY AND FUNCTIONAL CONSERVATION OF THE DNA NON HOMOLOGOUS END-JOINING PROTEIN, XLF/CERNUNNOS

Pierre Hentges1, Peter Ahnesorg2, Robert S. Pitcher1, Chris K. Bruce3, Boris Kysela3, Andrew J. Green1, Julie Bianchi1, Thomas E. Wilson1, Stephen P. Jackson1 and Aidan J. Doherty1*

From 1Genome Damage and Stability Centre, University of Sussex, Brighton BN1 9RQ, UK. 2The Gurdon Institute and Department of Zoology, Cambridge University, Tennis Court Road, Cambridge CB2 1QN, UK. 3Division of Reproductive and Child Health, Section of Medical and Molecular Genetics, The Medical School, University of Birmingham, Birmingham, B15 2TT, UK. 4Department of Pathology, University of Michigan Medical School, Ann Arbor, Michigan 48109-0602, USA.

Non-homologous end-joining is a major pathway of DNA double-strand break repair in mammalian cells, deficiency in which confers radiosensitivity and immune-deficiency at the whole organism level. A core protein complex comprising the Ku70/80 hetero-dimer together with a complex between DNA ligase IV and XRCC4 is conserved throughout eukaryotes and assembles at double-strand breaks to mediate ligation of broken DNA ends. In Saccharomyces cerevisiae an additional NHEJ protein, Nej1p, physically interacts with the ligase IV complex and is required in vivo for ligation of DNA double-strand breaks. Recent studies with cells derived from radiosensitive and immune-deficient patients have identified the human protein, XLF (also named Cernunnos), as a crucial NHEJ protein. Here we show that XLF and Nej1p are members of the same protein super-family and that this family has members in diverse eukaryotes. Indeed, we show that a member of this family encoded by a previously uncharacterized open-reading frame in the Schizosaccharomyces pombe genome is required for NHEJ in this organism. Furthermore, our data reveal that XLF family proteins can bind to DNA and directly interact with the ligase IV-XRCC4 complex to promote DSB ligation. We therefore conclude that XLF family proteins interact with the ligase IV-XRCC4 complex to constitute the evolutionarily conserved enzymatic core of the NHEJ machinery.

The repair of DNA double-strand breaks (DSBs) is crucial for the maintenance of genomic integrity (1,2). In mammalian cells, non homologous end-joining (NHEJ) is the predominant pathway for DSB repair – particularly in the G0 and G1 phases of the cell cycle – and is mediated by the Ku70/Ku80 hetero-dimer, the DNA-dependent protein kinase catalytic subunit (DNA-PKcs), XRCC4 and DNA ligase IV (3,4). It is believed that during NHEJ, newly created DSB ends are bound by the Ku hetero-dimer (5,6) and subsequently, DNA-PKcs is recruited upon inward translocation of Ku70/Ku80 (7-9). The Ku-DNA-PKcs complex is thought to stabilize the DNA ends and prevent extensive end resection (10), and most likely also plays an important role in aligning the DNA termini (11-14). The DNA-PKcs-Ku complex then promotes the recruitment of a pre-formed complex between XRCC4 and DNA ligase IV (ligase IV), which mediates the final ligation steps of the NHEJ process (8,15).

XRCC4 is a ~36 kDa protein that consists of a globular N-terminal domain followed by a protruding coiled-coil arm (16,17). It has been well established that XRCC4 forms homo-dimers, as well as higher order multimers (16-19), and that an active ligase IV complex is most likely composed of one XRCC4 homo-dimer bound to one ligase IV molecule (17,18). Significantly, XRCC4 not only promotes ligase IV activity but is also required for ligase protein stability (20,21). In addition to the above core set of NHEJ proteins, a number of other factors have been identified that are required for the efficient repair of DSBs by NHEJ. These factors – most prominently Artemis (22,23), poly-nucleotide
kinase (24,25) and polymerases of the Pol X family (23,26,27) – are mainly required to process the DSB ends in order to create a ligatable substrate for the ligase IV-XRCC4 (LX) complex (3,4). NHEJ proteins are also involved in V(D)J recombination, the site-specific recombination process that generates antibody and T-cell receptor diversity in mammals (28-30).

Recently, a previously uncharacterized mammalian protein, XRCC4-like factor (XLF; also called Cernunnos), was identified as an XRCC4 interactor and was shown to bind to the LX complex (31-33). Strikingly, structural predictions indicated that XLF likely adopts a structure reminiscent of XRCC4 (31,33), and depletion of XLF in human cell lines has indicated that it indeed functions in NHEJ and V(D)J recombination (31,32). Furthermore, a number of radiosensitive, microcephalic and immune-deficient patients (32,34) have been shown to carry mutations in the XLF/Cernunnos gene, and these mutations appear to be responsible for various clinical features of these patients (31,32). Although the above reports have identified XLF as a core NHEJ factor required for efficient DSB repair in mammalian cells and have established that XLF can interact with the LX complex, the mechanism(s) by which XLF promotes NHEJ has thus far not been addressed. Here, we show that XLF homologues exist in a wide range of mammals and have established that XLF/Cernunnos plays a crucial role in NHEJ. Furthermore, we reveal that, like XRCC4 itself, S. pombe XLF and human XLF are able to multimerise in solution and can bind to DNA. Finally, we establish that XLF functions by directly stimulating the DNA ligation function of the LX complex.

**EXPERIMENTAL PROCEDURES**

**Construction of S. pombe mutants**- Media and standard genetic techniques were as described previously (35). To construct an S. pombe strain xlf1Δ, the ORF SPCCC24B10.14c was replaced by the nourseothricin-resistance cassette through genomic integration of a PCR fragment as previously described (36). The strain xlf1Δ lig4Δ was constructed by crossing xlf1::natMX ura4.D18 leu1.32 ade6.704 with lig4::kanMX ura4.D18 leu1.32 ade6.704 (37).

**Cloning, expression and purification of S. pombe Xlf1**- S. pombe xlf1 (SPCC24B10.14c) was PCR amplified from cDNA, and cloned into pET28a using BamHI and Xhol. Sequencing confirmed that the single intron predicted for xlf1 in the S. pombe genome annotation (www.genedb.org) was absent in the clone. His6-Xlf1 was expressed and purified from Escherichia coli B834 (DE3) pLyS5 cells. Briefly, B834(DE3) cells transformed with xlf1-pET28a were grown at 37°C to A600 of ~0.6 OD followed by induction for 4h at 30°C using 0.5mM isopropyl-1-thio-β-d-galactopyranoside (IPTG). Cells were harvested and stored at –80°C. Frozen cell pellets were resuspended in buffer A (50 mM Tris–HCl (pH 7.5), 500 mM NaCl, 10 mM imidazole, 10% (v/v) glycerol, 17 µg/ml of PMSF, 34 µg/ml of benzamidine) and cells broken by lysozyme digestion and sonication. Following centrifugation, the cell supernatant was loaded onto an NTA (Qiagen) column pre-equilibrated with buffer A. The sample was washed with 20% buffer B (50 mM Tris–HCl (pH 7.5), 500 mM NaCl, 300 mM imidazole, 10% (v/v) glycerol) to remove contaminants and then eluted with 300 mM imidazole (100% buffer B). Chromatographic fractions were analysed by SDS-PAGE and those containing Xlf1 were concentrated and loaded onto a Superdex S-200 analytical gel-filtration column equilibrated in buffer C (50 mM Tris–HCl (pH 7.5), 500 mM NaCl, 10% (v/v) glycerol, 1 mM DTT). After passage through the size-exclusion column Xlf1 eluted as a highly purified fraction.

**Expression and purification of human XLF-BL21(DE3) cells** were transformed with a pGEX plasmid containing GST-tagged human XLF (31) and grown overnight at 37°C in LB broth containing 100µg/ml ampicillin. After 16 hours 5ml of this culture was used to inoculate 2 litres of LB and grown at 27°C. When the culture reached an optical density of 1.2, it was rapidly cooled. To induce protein expression, IPTG was added to a final concentration of 0.5mM and grown at 19°C overnight for 18 hours. The cells were harvested and resuspended in Buffer A (50mM Tris pH8, 1mM DTT, 5mM EDTA, 400mM NaCl, 0.1% Triton X 100, 17 µg/ml of phenylmethylsulfonyl fluoride, 34 µg/ml of benzamidine). Cells were broken by sonication, and cleared by centrifugation. The supernatant was loaded onto a GSTrap™ HP Column (GE Healthcare). After washing, the GST tagged protein eluted with Buffer A plus 40mM reduced L-Glutathione (Sigma).
Chromatographic fractions were analysed by SDS-PAGE and Coomassie staining. Fractions containing XLF were concentrated and loaded onto a Superdex S-200 analytical gel-filtration column equilibrated in buffer C (50 mM Tris–HCl (pH 7.5), 500 mM NaCl, 10% (v/v) glycerol, 1 mM DTT). After passage through the size-exclusion column XIIf eluted as a highly purified fraction. The GST tag was removed by thrombin cleavage.

In vivo DSB plasmid repair assay - The plasmid DSB assay was carried out as described previously (37). Briefly plasmids pAL19, PS and PI were digested with PvuII, PstI or EcoRI, to make blunt-ended, 3’ overhang or 5’ overhang fragments, respectively. The linearised plasmid DNA was separated from uncut DNA by purification from agarose gels. pAL19 was used as the undigested control. 1µg of either linearised or uncut control plasmid DNA were transformed into log-phase cells using the lithium acetate method. NHEJ frequency was calculated as the percentage of leu+ colonies arising from cells transformed with linear over those transformed with undigested DNA. Three experiments were performed for each strain.

In vitro PCR-Based End-Joining Assays- pUC18 was cut with PvuI (which cuts twice in the plasmid) to give two-base 5’ overhangs, producing a linearised DNA duplex approximately 900 bp smaller than the uncut plasmid. Reaction mixtures contained 66 mM Tris-HCl (pH 7.5), 5 mM DTT, 1 mM ATP, 120 ng PvuU-cut pUC18 and various amounts of DNA ligation IV – XRCC4 complex (LX) and XLF as indicated in the figure legend. Initially, LX was titrated into this reaction without XLF to find a concentration at which it does not efficiently religate PvuI-cut pUC18. Reactions were incubated for 1 hr at 37°C, then were analyzed by PCR, using forward and reverse primers upstream and downstream respectively of the PvuI-restriction sites such that a 1.4 kbp product would be generated by uncut plasmid and one of approximately 500 bp if the cut plasmid was successfully religated. The PCR products were then separated and examined on a 1% agarose gel and visualized by ethidium bromide staining.

Spore survival assay- To prepare spores, strains of opposite mating type of the same mutants were mixed on ELN plates and incubated at 30°C. The mixture of vegetative and post-meiotic cells that grew after 2 days were scraped off the plates, resuspended in sterile water, and treated with Helix pomatia snail gut enzyme for 2 days to kill remaining vegetative cells. The spore preparation was then washed once with sterile water and resuspended in water. The spore concentration was determined using a haemocytometer. Spores were either gamma-irradiated or mock-irradiated, plated on rich media and incubated at 30°C. After 4 days the number of colonies that had formed were counted. Survival was determined as the percentage of colonies derived from irradiated spores in comparison to colonies from the mock-irradiated spores.

Agarose gel mobility shift assays- Double-stranded DNA (pUC18 or 1kb PCR-amplified DNA) was incubated with XLF at room temperature in EMSA buffer (20mM Tris pH8, 50mM KCl, 0.1mM DTT, 10µg/ml BSA, 5% glycerol), followed by electrophoresis of the samples in a 1% agarose-TBE gel at 100V for 1 hr. Gels were then stained with ethidium bromide and visualized with UV light.

DNA ligation assays- Indicated amounts of proteins were incubated for 10 min at 37°C in 20 µl reaction mixture (66 mM Tris-HCl, 5 mM MgCl₂, 1 mM DTT, 1mM ATP, pH 7.5) with 10 ng of the AflIII-PstI fragment of pBLUESRIPT as a substrate (γ-32P-ATP labelled on the 5’-end). Reactions with XLF were pre-incubated for 10 min on ice with indicated amounts of XLF, and ligation was started by adding the LX complex or T4 DNA ligase and transfer to 37°C. After incubation, the reactions were deproteinized by treating with 0.6% SDS and 0.6 mg/ml proteinase K (Sigma) for 10 min at 37°C, phenol/chloroform extracted and precipitated with Pellet-Paint co-precipitant (Novagen). Aliquots were then run on 0.8% agarose gels, dried, analyzed using a STORM phosphorimager (Molecular Dynamics). Reaction efficiencies were quantified using ImageQuant software (Molecular Dynamics) and expressed as the percentage of the radioactively labelled reaction substrate converted into higher order concatemers.

RESULTS

Identification of XLF homologues in a wide range of eukaryotes. The recent discovery of human XLF/Cernunnos prompted us to search for homologous proteins in other organisms for which extensive genome sequence information exists. By using Psi-Blast searches (38), we readily identified XLF homologues in...
vertebrates, insects as well as in filamentous fungi such as *Neurospora crassa* and *Aspergillus fumigatus*. Lower scoring, but significant, ‘hits’ were found in the genomes of the fission yeast *S. pombe* (E=4e-31, searched with human XLF amino acids (aa) 1-200) and the budding yeasts *Yarrowia lipolytica* (4e-32) and *Debaryomyces hansenii* (6e-6). However, these proteins differed considerably in size, ranging from *S. pombe* XLF (202 aa), metazoan XLFs (~300 aa) and *D. hansenii* (335 aa), up to 540-630 aa residues for the predicted XLF homologues in filamentous fungi.

To further explore the nature of these XLF-related proteins, we submitted the sequence of human XLF and the sequences of XLF homologues from *S. pombe* and *N. crassa* to the online sequence-structure threading programme PHYRE (39). For each of these proteins, the N-terminal portion was predicted to be most compatible with the XRCC4 fold among a library of solved structures (supplemental Fig. 1). Only the N-terminal 150-220 residues were modeled into the XRCC4 fold, encompassing the globular domain and the beginning of the C-terminal α-helical stalk, which is also the portion where most sequence similarity between XLF homologues resides. In *N. crassa*, the region further C-terminal is predicted to be largely disordered by several methods used in PHYRE; the C-terminal of human XRCC4 (204-265aa) has been found to be disordered in the crystal structure (16). These data suggest that, despite the considerable differences in size, the N-terminal portions of all these proteins are likely to adopt an XRCC4-like globular-stalk fold, as has previously been predicted for human XLF (31). We have therefore termed this group of predicted eukaryotic proteins the “XLF family”.

Although Psi-Blast searches with human XLF did not identify an XLF homologue in the budding yeast *Saccharomyces cerevisiae*, when the search was initiated with the *D. hansenii* sequence, a possible XLF homologue was identified in the genome of the ascomycete *Candida albicans*, and – strikingly – subsequent analyses revealed that this corresponds to the counterpart of the *S. cerevisiae* NHEJ gene, NEJ1. NEJ1p has been shown to serve as a transcriptionally controlled mating-type regulator of NHEJ in *S. cerevisiae*, and to function epistatically with Lif1p (S. *cerevisiae* XRCC4 homologue) and Dnl4p (S. *cerevisiae* ligase IV homologue) (40-43). Searches with the putative *C. albicans* XLF homologue also identified proteins in various other budding yeast species; and although these proteins displayed considerable sequence divergence from one another, they were more closely related to *S. cerevisiae* Nej1p than to any other protein in this organism. Taken together, the above data reveal that XLF-like proteins exist in a wide range of eukaryotic species, and furthermore suggest that *S. cerevisiae* Nej1p might correspond to the XLF counterpart in this organism, as has recently been suggested by Callebaut *et al.* (33).

To further characterize the sequences identified in the above Psi-Blast searches, we divided them into three groups: metazoan homologues of human XLF; yeast sequence homologues of human XLF; and yeast proteins related to Nej1p. We then carried out multiple sequence alignments within each group by using the N-terminal portions (~200 residues) of the proteins, and then performed secondary structure predictions for each group. Significantly, the three secondary structure predictions were found to be largely compatible between the three alignments (supplemental Fig. 2). These alignments were then all combined into a single alignment (supplemental Fig. 3), and a phylogenetic tree was calculated (Fig. 1A). In the final alignment, the XLF/Nej1p family is shown to be highly divergent, with only a small number of conserved sequence motifs between Nej1p and XLF. Interestingly, the structure of the phylogenetic tree relates to the order in which Psi-Blast searches identified homologues of human XLF; the phylogenetic tree illustrates that the *D. hansenii* and *C. albicans* sequences provide a link between human XLF and *S. cerevisiae* Nej1p, previously not suspected of sharing an evolutionary relationship. Notably, however, we have so far been unable to find XLF/Nej1p homologues in plants or in prokaryotes.

Given the high degree of sequence divergence within the XLF/Nej1p family and the predicted similarity to the XRCC4 family, we considered it essential to verify the relationships between members by using methods that did not rely solely on primary structure. Thus, we made four alignments of sequences belonging to either Nej1p, XLF,
Lif1p or XRCC4 using the programme Muscle (44) and submitted these alignments separately to the structure-structure threading programme Fugue (45). In each case, the best “hits” predicted a fold similar to human XRCC4. Since the sequences submitted were aligned to the structure of human XRCC4, this allowed us to make an alignment of the XLF/Nej1p and XRCC4/Lif1p families (supplemental Fig. 4).

To further explore the possibility that S. cerevisiae Nej1p is homologous to human XLF, we examined the syntenic relationship between the S. cerevisiae chromosomal interval bearing NEJ1 and related yeast genomes (Fig. 1B). Examination of the conservation of gene order and orientation in chromosomal regions of different ascomycete genomes allows inferences to be drawn as to the evolutionary origin of genes within such syntenic regions, even when the genes themselves are poorly conserved at the sequence level. Comparative genomic studies of syntenic regions have revealed that a whole genome duplication (WGD) occurred in a common ancestor of S. cerevisiae and C. glabrata (these are therefore named ‘post-duplication’ yeasts) (Fig. 1C; 46,47). However, the WGD occurred after divergence of the lineages leading to two other ascomycetes closely related to S. cerevisiae, A. gossypii and K. lactis (hence named ‘pre-duplication’ yeasts). Thus, a single chromosomal region in pre-duplication A. gossypii and K. lactis is generally represented by two syntenic regions in post-duplication S. cerevisiae and C. glabrata, containing either two diverging homologues or a single homologue if the duplicated copy has disappeared through gene loss. In addition, chromosomal regions syntenic to the S. cerevisiae genome can sometimes be discerned in the genomes of D. hansenii and Y. lipolytica, which are much more distantly related to S. cerevisiae (Fig. 1C). Syntenic relationships are generally much less well conserved in C. albicans due to extensive shuffling of chromosomal regions that has taken place since the lineages diverged (48).

The ORFs identified as NEJ1 homologues in Psi-Blast searches in the pre-duplication ascomycetes A. gossypii and K. lactis are located in a chromosomal region whose synteny is well conserved in S. cerevisiae and C. glabrata (Fig. 1B). In these four organisms, all of the previously identified NEJ1 homologues are clearly also syntenic homologues, with only a single copy present in the post-duplication yeasts. In pre-duplication A. gossypii and K. lactis, NEJ1 is located adjacent, and in a head-to-head orientation, to the coding sequence whose S. cerevisiae counterpart is YOR166c (uncharacterized gene product, well conserved within ascomycetes, containing a PINc nucleotide-binding domain). Although the region surrounding the NEJ1 locus in A. gossypii and K. lactis is not conserved as a syntenic block in D. hansenii and C. albicans, the coding sequence for Nej1p/XLF sequence homologues in these organisms are found next to YOR166c, with conserved transcriptional orientation as compared to the pre-duplication budding (Fig. 1B). Importantly, in Y. lipolytica and S. pombe, the clear sequence homologues of metazoan XLF are located adjacent to counterparts of S. cerevisiae YOR166c, although the relative orientation has changed in S. pombe. To our knowledge, this is the first time a syntenic relationship has been observed over such a large evolutionary distance within ascomycetes. Thus, despite the considerable divergence between yeast proteins in the XLF/Nej1 family, examination of the synteny of NEJ1 with other loci suggests that XLF and Nej1p are indeed derived from a common ancestor.

Deletion of fission yeast xlf1 abolishes NHEJ-Inactivation of human XLF or S. cerevisiae Nej1p leads to defects in NHEJ (31,32, 40-43). To explore whether the weak sequence homology described above between members of the XLF/Nej1p group of proteins reflects their functional conservation, we decided to characterize the putative S. pombe XLF homologue encoded by ORF SPC24B10.14c. This is the smallest member of the XLF/Nej1p family identified and lies between mammalian XLF and S. cerevisiae Nej1 in the phylogenetic tree (Fig. 1B). We constructed a S. pombe strain (xlf1Δ) in which the putative XLF homologue was deleted from the genome. Vegetative xlf1Δ cells displayed no growth defects (data not shown), as has been reported for the lig4Δ strain, in which NHEJ is inactivated by deletion of the gene encoding S. pombe DNA ligase IV (37).

To assess the NHEJ status of xlf1Δ cells, we carried out plasmid-based DSB repair assays (37,49). Thus, wild-type, xlf1Δ, and lig4Δ leucine auxotrophic strains were
transformed with plasmid PI that had been linearized by excision of an EcoRI fragment, thereby generating DSBs with 5'-overhanging termini. Since plasmid PI carries the LEU2 marker, cells that religate the linearized plasmid are able to grow in the absence of leucine, whereas cells that do not mediate plasmid ligation lose the plasmid DNA and do not form colonies. The above strains were transformed in parallel with undigested circular plasmid PI, allowing the efficiency of NHEJ to be calculated as the percentage of leu' colonies arising from a transformation with EcoRI-cut linearized plasmid as compared to those arising after transformation with the uncut plasmid. The values thus obtained were then normalized to the plasmid repair frequency determined for wild-type cells (Fig. 2A). Strikingly, compared to wild-type cells, the frequency of plasmid religation decreased to 2% in xlf1Δ cells, and to 1.5% in the lig4Δ cells, showing that the xlf1Δ mutant displays a DSB repair defect of similar magnitude to that of the lig4Δ strain.

To further investigate the NHEJ defects of Xlf1-deficient S. pombe strains and to ascertain whether Xlf1 and Lig4 function in the same pathway, we constructed a strain mutated for both xlf1Δ and lig4Δ (xlf1Δ lig4Δ), then carried out plasmid religation assays on this strain in parallel with analyses of the two single mutants and a wild-type strain. In these experiments, we transformed the strains with uncut plasmid pAL19 DNA or with pAL19 that had been linearized by treatment with PvuII or PstI, generating DSBS with blunt-ends or 3'-overhanging ends, respectively. As shown in figure 2B, the xlf1Δ and lig4Δ single mutants displayed a significant reduction in the repair of blunt-ended and 3'-overhanging plasmid ends that were similar in magnitude to those observed in the xlf1Δ lig4Δ double mutant. These data therefore imply that deletion of Xlf1 impairs the same DSB repair pathway as deletion of Lig4, thereby demonstrating that S. pombe Xlf1 functions in ligase IV-dependent NHEJ. Furthermore, the plasmid repair phenotype of xlf1Δ cells implies that NHEJ is non-functional in these cells. These data therefore indicate that S. pombe Xlf1 plays a key role in promoting the religation of DNA DSB ends, strikingly analogous to what has been established for human XLF and for S. cerevisiae Nej1p (31,32,40-43). Furthermore, the data provide strong evidence for extensive functional conservation of the XLF/Nej1p family of proteins between man and the two highly divergent yeasts, S. cerevisiae and S. pombe. In light of this, we conclude that all members of the XLF/Nej1p protein family are likely to carry out conserved functions in DNA ligase IV mediated DSB repair.

As has previously been reported for the S.pombe lig4Δ mutant (37), we found that vegetative xlf1Δ cells were not detectably hypersensitive to DNA damaging agents (supplemental Fig. 5). The failure to identify DNA damage sensitivity in S.pombe xlf1Δ cells is likely due to the very low proportion of G1 cells in asynchronous S.pombe cultures. This was suggested by studies on S.pombe cells arrested in G1 by nitrogen starvation, showing that NHEJ is preferred over homologous recombination as the DSB repair mechanism in the G1 phase of the cell-cycle (50). We reasoned that DNA damage sensitivity of S.pombe NHEJ mutants might appear in another state where only one haploid set of chromosomes is present, namely in spores. We found that spores derived from xlf1 deleted strains are hypersensitive to gamma irradiation as compared to the wild-type (Fig.2C). This DNA damage hypersensitivity of the xlf1 deletion mutant is similarly pronounced in the lig4Δ single mutant and the xlf1Δ lig4Δ double-deletion mutant, suggesting that xlf1 and lig4 function in the same genetic pathway. These data are also consistent with our findings in the plasmid DSB repair assay.

**Human and S. pombe XLFs form multimeric complexes and bind non-specifically to double-stranded DNA** - As discussed above, analysis of the amino acid sequence of human XLF with sequence-structure threading methods predicts that XLF adopts a protein fold resembling the experimentally determined tertiary structure of XRCC4 (16,17,31). To find out if this predicted structural similarity is paralleled by shared biochemical properties between XLF and XRCC4, we investigated the ability of XLF (human and S. pombe) to form multimers and to bind to double-stranded DNA (ds DNA). Thus, S. pombe XLF (spXlf1) was over-expressed as a hexa-histidine tagged fusion protein in E. coli and purified to near-homogeneity by sequential nickel affinity and size-exclusion chromatography. Human XLF was expressed as a GST fusion protein and purified to near-homogeneity by sequential
glutathione affinity and size-exclusion chromatography. Analysis by SDS-PAGE and Coomassie staining (Fig. 3A) showed that the resulting proteins migrated close to their predicted sizes (36.9 kDa for hXLF and 27.0 kDa for spXlf1). Western blotting with an anti-histidine tag antibody confirmed that the bands observed in the protein preparations corresponded to XLF (data not shown).

Notably, when we carried out size-exclusion chromatography on the recombinant XLF preparations by fractionating them on an analytical S-200 gel-filtration column that had been previously calibrated with molecular weight standards, this yielded theoretical molecular weights of 87.5 kDa and 53 kDa for hXLF and spXlf1, respectively (Fig. 3B). Size-exclusion analysis suggests that both human XLF and S. pombe Xlf1 are likely to exist as stable homodimeric complexes, as has been reported for XRCC4 (16-18). This conclusion is in line with our previous work on human XLF, which has indicated that it is able to multimerize when expressed in human cells (31).

It has been shown previously that XRCC4, in addition to complexing with ligase IV, can bind directly to dsDNA (51, 52). To determine whether XLF also has an affinity for dsDNA, we carried out electrophoretic mobility shift assays (EMSAs) with various DNA substrates. First, we incubated a 1 kbp ds DNA PCR fragment with increasing amounts of human XLF and then separated reaction mixtures on agarose gels, monitoring the formation of DNA-protein complexes by ethidium bromide staining (Fig. 3C). hXLF was able to band-shift this DNA substrate. To determine if XLFs bound specifically to linear DNA, we repeated the assay with closed-circular plasmid DNA and observed that both hXLF and spXLF also bound to this substrate (Fig. 3D). A similar electromobility shift was also observed with linearized plasmid in the presence of XLF (data not shown). These EMSA data suggest that XLF proteins bind to ds DNA in a non-specific way rather than specifically to the ends of the break, in contrast to the NHEJ Ku complex. Therefore, in common with what has been reported previously for XRCC4, XLF can also bind to both linearized and circular DNA with similar efficiencies. Together, these results demonstrate that XLF functionally parallels XRCC4 in terms of its ability to form homodimers and bind to ds DNA.

**Human XLF specifically stimulates DNA end-joining by ligase IV-XRCC4**- The homo-dimeric form of XRCC4 has been shown to bind tightly to ligase IV via an interaction between a conserved motif in its coiled-coil and the inter-BRCT linker region of ligase IV, through the wrapping of both BRCT domains around the XRCC4 molecule dimer (53,54). This interaction is believed to stabilize the DNA ligase IV protein as well as to promote the targeting of ligase to sites of DNA DSBs *in vivo*. Recently, it has been shown that XLF is required for efficient NHEJ and V(D)J recombination in mammalian cells (31,32), and appears to directly interact with XRCC4 and the LX complex (31,33). These data therefore raised the possibility that XLF serves to directly promote the end-joining activity of the LX complex. To address this possibility, we examined the effect of XLF on the DNA end-joining activity of the purified LX complex by using an *in vitro* PCR-based plasmid repair assay. In this assay, Pvu I-linearized pUC18 plasmid DNA was incubated with the LX complex, with XLF alone, or with LX complex together with increasing amounts of recombinant human XLF. The product of the reaction, circularized plasmid DNA, was then detected by PCR amplification with primers flanking the cleavage sites (see schematic in Fig. 4A). Strikingly, the addition of increasing amounts of XLF to reactions containing a fixed concentration of the LX complex (LX; 9.2 nM) markedly stimulated DNA end-joining (Fig. 4A). In contrast, no ligation was observed with samples containing XLF alone, demonstrating that XLF has no DNA ligase activity by itself and that the observed activity in the presence of LX was not due to a possible co-purifying ligase. These results thereby establish that XLF is able to promote the end-joining activity of the LX complex, either by a direct mechanism, such as by stimulation of ligase catalytic activity, or indirectly such as through helping to tether the LX complex to DNA. Notably, end-joining activity was inhibited by the addition of XLF at concentrations above 270 nM (Fig. 4A, lanes 6-9). Again, this may be the result of XLF directly modulating ligase IV enzymatic activity, or could reflect an indirect inhibitory effect possibly related to XLF DNA binding. A similar inhibitory effect on end-joining has previously been observed with other NHEJ proteins, such as Ku, at supra-optimal concentrations and this has been suggested to result from preventing
access of a functional ligase IV complex to the DSB ends (8,51).

While the PCR-based assay used above established that XLF stimulates end-joining by the LX complex, it could not be used to accurately measure the level of stimulation. To do this, and to address the specificity of the effects on LX, we employed a direct ligation assay. In this, we analyzed the effect of XLF on DNA end-joining of a 445 bp radio-labeled dsDNA fragment by the LX complex or by bacteriophage T4 ligase (here, reaction products were separated by acrylamide electrophoresis followed by autoradiography and quantification). Strikingly, addition of increasing amounts of XLF to constant amounts of LX complex promoted the conversion of the 445 bp DNA into dimeric and higher multimeric ligated products (Fig. 4B, left panel); and from these data, we calculated that XLF stimulates ligase IV-catalyzed end-joining by over 20-fold. Significantly, and in marked contrast to the reactions containing ligase IV and XRCC4, no stimulation of end-joining was observed when XLF was added to samples containing T4 ligase (Fig. 4B, right panel). These results therefore suggest that XLF does not promote DNA end-joining by a non-specific end-synapsis mechanism that simply brings neighbouring termini together via its DNA binding activity, but instead suggests that XLF specifically stimulates the ligase activity of the LX complex through a mechanism that involves specific, direct actions with this complex.

**DISCUSSION**

The core NHEJ machinery encompassing the Ku70-Ku80 heterodimer, DNA-PKcs, XRCC4 and ligase IV has been very well characterized over the past decade. These studies have not only addressed how these factors interact but have indicated the order in which the various NHEJ proteins assemble at the site of a DSB (4). Five years ago, a number of groups reported an additional NHEJ factor, Nej1p, in *S. cerevisiae* and this discovery implied that the basic NHEJ machinery, at least in yeast, is more complex than previously anticipated (40-43). However, failure to discover genes in higher eukaryotes with clear sequence homology to Nej1p, as well as Nej1p’s role as a diploid-haploid regulator of yeast NHEJ suggested that Nej1p and its function might be restricted to certain yeasts. The recent discovery of XLF/Cernunnos (31,32) as a NHEJ factor in mammalian cells, has raised speculation that human XLF might be related to Nej1p (33), particularly since both XLF and Nej1p bind to the DNA ligase IV co-factor XRCC4/Lif1p (31,33, 40-43). Here we show that both Nej1p and XLF belong to a conserved NHEJ protein family that we term the XLF family. Although these proteins share only very limited sequence homology, there is a clear phylogenetic link between the *S. cerevisiae* Nej1p and higher eukaryotic XLF family members. Moreover, we show that a thus far uncharacterized *S. pombe* protein (SPCC24B10.14c; named XII1) that is predicted to be a member of the XLF family by our analyses, is indeed required for NHEJ in *S. pombe*. These data provide a clear further functional link between the XLF family members and suggest that these proteins serve as XRCC4/Lif1p binding NHEJ factors in eukaryotes spanning yeast to man. Intriguingly, no XRCC4/Lif1p homologue has so far been identified in *S. pombe*. Although extensive sequence searches of the *S. pombe* genome did not reveal an XRCC4 homologue, the sequence divergence within the XRCC4/LIF1 family suggests it is possible that such a homologue remains to be uncovered in *S. pombe*. In particular, as the XRCC4/Lif1p protein seems to be crucial for DNA ligase IV protein stability in both *S. cerevisiae* and mammalian cells, while XLF/Nej1p are not required for DNA ligase stability (31,53,55). However, analysis of the synteny of Lif1 suggests that the pre-duplication yeasts *A. gossypii* and *K. lactis* may also lack a Lif1 homologue (supplemental Fig. 6). This raises the question if the NHEJ machinery may indeed not require Lif1/XRCC4 in some organisms, and NEJ1/XLF could substitute for the function of this factor. Searches for interaction partners of spXlf1 may help to address this point and clarify if indeed ligase IV-XRCC4-XLF are all conserved components central to the NHEJ machinery.

Our study and previous reports on human XLF (31,32) and Nej1p (40-43) suggest that proteins of the XLF family are required for efficient NHEJ and interact directly with the LX complex. However, the precise function of Nej1p and XLF still remains unclear. Here we have shown that, like XRCC4 itself, XLF family members are able to bind to DNA.
These shared properties are consistent with the predicted structural similarities between XRCC4- and XLF-family members (31), and suggest that these proteins share functions and/or may act cooperatively in the NHEJ process. Moreover, our data clearly indicate that XLF (and most likely all XLF family members) is directly involved in modulating the enzymatic process of DNA DSB ligation by DNA ligase IV, as addition of XLF greatly stimulates the ligation activity of recombinant LX in vitro. Intriguingly, previous work has shown that while XRCC4 is required for ligase IV protein stability in vivo, it only modestly stimulates ligase IV activity in vitro (21). We therefore conclude that XLF and XRCC4 are required together for optimal ligation of DSB ends by DNA ligase IV.

Precisely how XRCC4 and XLF family proteins act together to promote DNA end-ligation still remains to be addressed. Given that both have, or are predicted to have, a related globular N-terminal domain that may contain a DNA binding site (16) and both XRCC4 and XLF bind to DNA directly, we consider it likely that both proteins co-operate to promote the efficient recruitment of ligase IV to DNA ends. Such a model would explain not only the direct stimulatory effect of XLF on LX activity but also the striking predicted topological resemblance between XLF and XRCC4. Alternatively, or in addition, it could be that the LX complex normally exists in a low-activity conformation and that XLF family proteins bind to this and trigger its alteration into a more active state. Indirect support for such a model comes from the fact that XLF in human cells is far less abundant than XRCC4 and ligase IV (31) and from data found in previous reports on the stoichiometry of the LX complex, suggesting the existence of different XRCC4 and ligase IV containing complexes of varying composition (16-19). XLF family proteins could thus serve to regulate the equilibrium between these different complexes, activate ligase IV and serve to recruit such activated complexes to DSBs. Such a function in regulation of ligase IV is also suggested by findings that Nej1p is not only required for NHEJ but also suppresses aberrant DNA ligase activity at telomeres (56).

While XRCC4 and XLF family members are predicted to display significant topological similarity, sequence conservation among XLF family members is poor, as is sequence conservation between XRCC4 and Lif1p. Intriguingly, however, we have identified a motif that is well conserved in XLF family members, and that resembles a motif in the XRCC4 family (supplemental Fig. 7). The motif corresponds to a region that has previously been described to be involved in LX association and XRCC4 tetramerization (18). It is therefore possible that this region in both XRCC4 and XLF family members is crucial for ligase IV-XRCC4-XLF complex formation and/or its function in NHEJ. Consistent with this idea, previous studies have shown that mutation of both conserved lysine residues to aspartic acid impaired the LX interaction and that this mutated XRCC4 failed to fully complement the radiosensitivity of XRCC4 deficient XR-1 hamster cells (18). Similar studies with XLF and other XLF family members could help to uncover the roles of these and potentially other less conserved conserved motifs in the proteins and address how both XRCC4 and XLF modulate DNA ligase IV activity.

Ever since the observation that XRCC4 is a phospho-protein and a target for DNA-PKcs and related kinases in vitro (57-59), it has been speculated that such modification of XRCC4 could serve to regulate NHEJ. However, no functional relevance of such modifications has so far been uncovered (60). Interestingly, Nej1p is phosphorylated in response to DNA damage and this post-translational modification seems to regulate NHEJ activity in S. cerevisiae (61). It remains to be seen if other XLF family members are also subject to such posttranslational modifications but it is plausible that modification of both XRCC4 and XLF will serve to control the activity of NHEJ and thereby modulate the ability of cells to respond to DSBs. Finally, we note that Nej1p is down-regulated in diploid yeast cells, where NHEJ is strongly disfavored and HR is the preferred pathway of DSB repair. It will therefore be very interesting to study whether XLF activity or levels respond to environmental stimuli in metazoa, or whether they change during development or vary from one tissue to another.
REFERENCES


FOOTNOTES
AJD is a Royal Society University Research Fellow and AJD's laboratory is supported by grants from
BBSRC, AICR and CR-UK. The SPJ laboratory is supported by grants from Cancer Research UK and the European Union. PA was supported by a Cancer Research UK studentship. We thank Prof. Tony Carr and Dr Edgar Hartsuiker for advice and for providing the *lig4d* strain and plasmids pAL19, PS and PI and Dr Hiroshi Nojima for the *S.pombe* cDNA library, pLZ3. We are also grateful to Prof. Ken Wolfe for his advice on the syntenic analysis.
Figure 1. Identification of yeast XLF homologues. A, phylogenetic tree derived from alignment of the XLF/NEJ1 family. Yeast sequences identified as sequence homologues of human XLF are designated as ‘yeast XLF’, yeast sequences identified as sequence homologues of S. cerevisiae NEJ1 are designated as ‘yeast NEJ1’. The two groups of sequences overlap for D. hansenii and C. albicans. B, the synteny of the NEJ1 locus was examined in the pre-duplication yeasts A. gossypii and K. lactis, and the postduplication yeasts S. cerevisiae and C. glabrata, using Yeast Gene Order Browser (YGOb; http://wolfe.gen.tcd.ie/ygob). Information about the NEJ1/XLF locus was obtained from the Génolevures website (www.natchaug.labri.u-bordeaux.fr/Génolevures) for D. hansenii and Y. lipolytica; from the NCBI Map Viewer (http://www.ncbi.nlm.nih.gov/mapview) for D. hansenii, Y. lipolytica and C. albicans; and from the Sanger GeneDB website (www.genedb.org) for S. pombe. Due to the Whole Genome Duplication that occurred in the budding yeast lineage, a chromosomal region in A. gossypii and K. lactis (pre-duplication) is syntenic to two regions in S. cerevisiae and C. glabrata (post-duplication). These four budding yeasts contain one NEJ1 member each, in contrast to the ribosomal protein S28 (rps28) where the single copy of the pre-duplication organisms diverges into RPS28A and RPS28B in the post-duplication yeasts. YOR166C, an uncharacterised but conserved PINc domain-containing ORF is adjacent to NEJ1 in the pre-duplication budding yeasts, oriented in a head-to-head manner. In post-duplication S. cerevisiae and C. glabrata, YOR166C is no longer associated with NEJ1 but found on the syntenic region on a different chromosome. However, YOR166C remains adjacent to NEJ1/XLF in D. hansenii and C. albicans with identical transcriptional orientation. Similarly in the distantly related budding yeast Y. lipolytica, the clear homologue of human XLF is located next to YOR166C. Surprisingly, the clear XLF homologue in the fission yeast S. pombe is adjacent to YOR166C, although the two ORFs are now ordered in a tail-to-tail fasion, and are separated by two small RNA genes (not shown). ORFs were included only if they are homologous to the ORFs surrounding the NEJ1 locus in pre-duplication yeasts, but are designated by the name of the S. cerevisiae homologue, with the systematic name of each gene written below. The direction of transcription is indicated by an arrow; ORFs are not drawn to scale; the syntenic chromosomal portion is indicated by a grey bar. C, phylogenetic tree of ascomycetes. Adapted from http://natchaug.labri.u-bordeaux.fr/Génolevures.

Figure 2. Deletion of fission yeast XLF abolishes NHEJ. A, cultures of leucine auxotrophic cells from wild-type (Wt), xflΔ, or lig4Δ strains were transformed with uncut or linearised plasmid DNA containing the LEU2 marker and plated on agar without leucine. Colonies formed under leucine selection (stable maintenance of LEU2 plasmid, a result of re-circularisation of linearised plasmid) were counted. Religation efficiency is expressed as the ratio of colonies resulting from equal amounts of cut vs. uncut plasmid DNA. B, As described in A (above) using PvuII or PstI-cut plasmid DNA with wild-type cells (Wt), single-mutant strains xflΔ and lig4Δ, or the double deletion strain xflΔ lig4Δ. C, Spores from homozygous crosses of wild-type (Wt), xflΔ, lig4Δ, and xflΔ lig4Δ strains were either mock-irradiated or irradiated with gamma radiation, plated on rich agar plates and incubated at 30°C for 4 days. Survival was determined as the percentage of irradiated vs. mock-irradiated spores able to form colonies.

Figure 3. Human and S. pombe XLF form homodimers and bind non-specifically to ds DNA. A, the purity of the recombinant human and S. pombe XLF proteins on SDS-PAGE gels, stained with Coomassie Blue. MW, molecular weight markers. The lower band of the XLF sample also corresponds to XLF, as determined by mass spectrometry and Western-blotting analysis. B, XLF exists as a homodimer. A Superdex-200 HR 10/30 size-exclusion column was calibrated using molecular weight standards. Human and S. pombe XLFs, eluted from the sizing column at volumes that correspond to calculated molecular weights of 87.5 kDa and 53kDa, respectively. These values are more than double the predicted molecular weights of these proteins. Consequently, human and S. pombe XLFs exist predominantly as homodimeric species in solution. C, agarose electrophoretic mobility shift assays were employed to characterize the ability of recombinant human XLF to bind to linear ds DNA. Binding of increasing amounts of hXLF (lanes 2-6; 0, 0.1, 1, 5, 7 μg of protein) to 1 kbp ds DNA (100ng). D, agarose electrophoretic mobility shift assays showing the binding of
increasing amounts of either hXLF (lanes 1-8; 0, 1, 2, 3, 4, 6, 8 & 10 µg) or spXfl1 (lanes 9-15; 1, 2, 3, 4, 6, 8 & 10 µg) to uncut pUC18 plasmid DNA. Similar binding results were also observed for both XLFs using linearize plasmid DNA (data not shown).

**Figure 4. Human XLF specifically stimulates the end-joining activity of the human ligase IV-XRCC4 complex.** A, Human XLF stimulates LX mediated DSB repair at concentrations below 0.3 µM and appears to inhibit it at higher concentrations. Lane 1; molecular weight markers, lane 2-12; XLF concentrations were 0, 0.026 µM, 0.066 µM, 0.13 µM, 0.27 µM, 0.66 µM, 1.3 µM, 2.0 µM, 0.13 µM, 1.3 µM, 0 µM, respectively. When present, the LX was used at a concentration of 9.2 nM except in lane 13 (0.37 µM). Lane 14 shows the PCR product when the primers were used on uncut plasmid. B, The impact of pre-incubation of 445 bp DNA substrate with increasing amount of hXLF protein on ds ligation mediated by human ligase IV-XRCC4 (LX) complex (left panel) or T4 DNA ligase (right panel) was assessed on a 445 bp DNA substrate. Up to 20 fold stimulation of ligation was observed for LX complex, but no stimulation had been detected for T4 DNA ligase. Lane 5 in panel A (38.5 % of concatemers) represents the reaction mixture with approximately equimolar amounts of LX complex versus XLF.
Figure 1
Figure 2

A

% Cut/Uncut

Wt  xlf1Δ  lig4Δ

B

% Cut/Uncut

Wt  xlf1Δ  lig4Δ  xlf1Δ lig4Δ

C

% Survival vs IR Dose (Gy)

Wt  xlf1Δ  lig4Δ  xlf1Δ lig4Δ
Figure 3
Figure 4