Non-homologous end-joining partners in a helical dance: structural studies of XLF–XRCC4 interactions

Qian Wu1*, Takashi Ochi*, Dijana Matak-Vinkovic†, Carol V. Robinson‡, Dimitri Y. Chirgadze* and Tom L. Blundell*

1Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA, U.K.; †University Chemical Laboratory, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, U.K., and ‡Chemistry Research Laboratory, University of Oxford, 12 Mansfield Road, Oxford OX1 3TA, U.K.

Abstract

XRCC4 (X-ray cross-complementation group 4) and XLF (XRCC4-like factor) are two essential interacting proteins in the human NHEJ (non-homologous end-joining) pathway that repairs DNA DSBs (double-strand breaks). The individual crystal structures show that the dimeric proteins are homologues with protomers containing head domains and helical coiled-coil tails related by approximate two-fold symmetry. Biochemical, mutagenesis, biophysical and structural studies have identified the regions of interaction between the two proteins and suggested models for the XLF–XRCC4 complex. An 8.5 Å (1 Å = 0.1 nm) resolution crystal structure of XLF–XRCC4 solved by molecular replacement, together with gel filtration and nano-ESI (nano-electrospray ionization)–MS results, demonstrates that XLF and XRCC4 dimers interact through their head domains and form an alternating left-handed helical structure with polypeptide coiled coils and pseudo-dyads of individual XLF and XRCC4 dimers at right angles to the helical axis.

Key words: double-strand break (DSB), non-homologous end-joining (NHEJ), X-ray cross-complementation group 4 (XRCC4), XRCC4-like factor (XLF).

Abbreviations used: ATM, ataxia telangiectasia mutated; BRCT, BRCA1 C-terminal; DNA–complementation group 4 (XRCC4), XRCC4-like factor (XLF).

XLF and XRCC4 play roles in recruiting and stabilizing DNA ligase IV at DSBs in NHEJ

DNA DSBs (double-strand breaks) can be caused by ionizing radiation or toxic chemical exposure, but are also present as intermediates in V(D)J recombination and class switch recombination for antigen receptor diversity formation. Unrepaired DSBs lead to chromosome fragmentation and rearrangement and are lethal to cells, changing cell gene regulation and expression, and often leading to cancer cell formation. The two major DSB repair pathways are HR (homologous recombination) and NHEJ (non-homologous end-joining).

Our current understanding of the NHEJ repair pathway (Figure 1) is that it comprises three major steps: first, the Ku heterodimer and DNA-PKcs (DNA-dependent protein kinase catalytic subunit) recognize DSBs and generate a protein-binding platform for XRCC4 (X-ray cross-complementation group 4), XLF (XRCC4-like factor) and other proteins [1,2]; secondly, Artemis containing endonuclease activity and other end-processing proteins, such as PNKP (polynucleotide kinase/phosphatase) and PolX family DNA polymerases, process the DSBs ends before ligation [3,4]; and thirdly, the XRCC4–LigIV (DNA ligase IV) complex ligates the two ends of the DNA promoted by XLF [5]. Understanding how these transient NHEJ complexes assemble structurally in both space and time is a challenging, but timely, research focus. In the past, we have defined the crystal structures of XRCC4 with LigIV peptide [6], XLF [7] and more recently DNA-PKcs [8]. The next step in exploring the NHEJ protein assembly is to study the complexes of these key protein components.

Although XLF itself cannot directly ligate DSBs, it performs an essential NHEJ function by interacting with XRCC4 and stabilizing XRCC4–LigIV broken DNA ends, thereby enhancing the LigIV end-joining process [9]. The mechanism of XLF mediating ligation enhancement is through enhancement of LigIV recharging following ligation in the presence of ATP [10]. How XLF is structurally involved in the NHEJ pathway is not clear. In the present paper, we focus on current biophysical studies of XLF and XRCC4 and XLF–XRCC4 interactions and recent results from the crystal structure of the XLF–XRCC4 complex, which shed light on this question.

XLF and XRCC4 are dimeric coiled-coil proteins with a common ancestor

Despite the low sequence identity (13.6%), the crystal structures of XLF and XRCC4 demonstrate that the two proteins are homologous homodimers comprising globular head domains and C-terminal helices that form coiled-coil tail structures [6,7,12,13]. However, the structural differences between the two are large. The head domains form seven-stranded antiparallel β-sheets sandwiching a helix–turn–helix motif between β4 and β5, but XLF contains an extra helix in the N-terminal region. Whereas the tail structure of XRCC4 comprises an elongated coiled-coil, the equivalent extended helix α4 of XLF is followed by further helices, α5 and α6,
which fold back around the coiled-coil formed by α4 so that the C-termini come close to the α1 helices of the head domains. The sequence and structural differences between XLF and XRCC4 tails explain why LigIV does not bind to XLF in the same way as XRCC4. A further significant structural difference is the angle formed between head domain and helical tail structures for XLF and XRCC4. There is an approximately 45° difference between XLF and XRCC4 coiled-coil tail structures when the head domains from both proteins were aligned. This is presumably because the helix α6 of XLF folds back and contacts the head domain, pushing it further away from the coiled-coil helices.

The highly flexible and disordered C-termini of both XLF (residues 234–299) and XRCC4 (residues 214–336) were removed for the crystal structure analyses [6,7,13]. The C-terminal sequence of XLF is important for DNA binding, and DNA-PKcs targets both protein C-terminal structures for phosphorylation [14,15]. DNA-PKcs phosphorylates XRCC4 to regulate its binding with DNA [14]. The phosphorylation of XLF residues in the unstructured C-terminal affects neither XLF DNA-binding ability nor DNA-repair efficiency [16]. The approximate location of the XLF C-terminal region is predicted to be near the N-terminal head domain according to the direction of helix α6. EM (electron microscopy) studies have revealed that the mouse XRCC4 C-terminal region is a dimeric globular domain [17]. SAXS (small-angle X-ray scattering) studies indicated that the disordered C-terminal of XRCC4 folds back as observed in XLF [18]. Characterization of the structures of these regions is needed in future in order to complete our understanding of the function of XLF and XRCC4 in NHEJ.

**XLF and XRCC4 interact through their head domains**

Interactions between XLF and XRCC4 identified through a yeast two-hybrid study led to the discovery of XLF [5], even though interactions are dynamic, salt-sensitive and not dependent on DNA [5,10,18,19]. Another yeast two-hybrid study demonstrated that XLF (residues 1–128) and XRCC4 (residues 1–119) are the minimal regions required for their interaction, implying that XLF and XRCC4 contacts are through their head domains [19]. Indeed, when XLF is immobilized to glutathione-conjugated Sepharose beads through its C-terminal GST (glutathione transferase) tag, it is still able to pull down XRCC4–LigIV, implying that the C-terminal of XLF is not important for interaction with XRCC4–LigIV [15]. Although both proteins are present in solution as stable homodimers, a heterodimeric interaction model between XLF and XRCC4 is unlikely [19]. Furthermore, domain swapping between XLF and XRCC4 indicated that the head domains and coiled-coil regions of XLF and XRCC4 are not interchangeable, but rather each has a specific role [20].
The first mutagenesis studies of XLF and XRCC4 revealed that the structurally exposed Leu\textsuperscript{115} located in the XLF β6–β7 loop is important for XLF–XRCC4 interaction. Lys\textsuperscript{63}, Lys\textsuperscript{65} and Lys\textsuperscript{99} of XRCC4 essential for XLF–XRCC4 interaction are located in the beginning of α2 (just after the loop in the helix–turn–helix structure) and the end of β6 (near the β6–β7 loop). These studies led to the first proposal of a linear side-by-side XLF–XRCC4 interaction model, in which XLF head domains slide into the space created by XRCC4 head domains and the N-terminal part of the coiled-coil tail structure [13].

Further extensive mutagenesis studies indicated that two more XLF head domain residues, Arg\textsuperscript{64} and Leu\textsuperscript{65}, both located in the loop between helix–turn–helix α2–α3, are important for interaction with XRCC4 [21]. Leu\textsuperscript{115}, Arg\textsuperscript{64} and Leu\textsuperscript{65} are located in XLF conserved regions (residues 57–65 and 108–123) [13]. Isothermal titration calorimetry of the interaction between XLF and XRCC4 in solution indicated weak enthalpic but significant entropic contributions, implying a hydrophobic interface [21]. Together with protein–protein docking analysis, a new XLF–XRCC4 interaction model was proposed in which XLF does not slide into the space created by the XRCC4 head domain and the N-terminal part of the coiled coil for interaction. Instead, interaction between XLF and XRCC4 is mediated through relatively small regions located at the sides of the head domains and contain the helix–turn–helix structures and the β6–β7 loop [21].

SAXS structural studies of XLF–(1–248)–XRCC4-(1–140), XLF–(1–248)–XRCC4 and XLF–(1–248)–XRCC4–LigIV BRCT (BRCA1 C-terminal) domains suggested a similar XLF–XRCC4 linear, rather than sliding, binding model. In addition, SAXS also revealed there is an approximately 45° rotation between XRCC4 and XLF coiled-coil tails [18].

**XLF–XRCC4 partners form an alternating helical fibre**

In order to study the XLF–XRCC4 complex formation, XLF–(1–233) and XRCC4–(1–164) have been expressed, purified individually and then run together on an analytical gel filtration column (Q. Wu, T. Ochi, D. Chirgadze and T.L. Blundell, unpublished work). An elution peak, indicated by SDS/PAGE (Figure 2A, left-hand panel) to be an XLF–XRCC4 complex, runs further to the left and separately from individual proteins. Increasing the concentration of the complex shifts the elution peak further to the left, indicating formation of larger complexes at higher concentrations (Figure 2A, right-hand panel). This XLF–XRCC4 concentration-dependent higher-order complex formation is confirmed by nano-ESI (nano-electrospray ionization)–MS (Figure 2B) (see the Supplementary Online Data at http://www.biochemsoctrans.org/bst/039/bst0391387add.htm for experimental details) [22]. The Authors

The heterogeneous XLF–(1–233)–XRCC4–(1–164) complex samples can be crystallized using the hanging drop method in 0.1 mM Tris/HCl (pH 7.5) and 2 M sodium formate (Figure 2C, left-hand panel) and SDS/PAGE of the washed protein crystals confirms the presence of both proteins (Figure 2C, right-hand panel). The crystal of the complex diffracts to a resolution of 8.5 Å (1 Å = 0.1 nm) at the Diamond beamline I04, and the structure of XLF–(1–233)–XRCC4–(1–164) complex structure was solved at this resolution by molecular replacement (see the Supplementary Online Data).

The interaction between XLF and XRCC4 is mediated through the helix–turn–helix and β6–β7 loop structures from the head domains of each protein. The binding of the two proteins generates a tilt angle between the pseudo-dyads relating head domains and coiled-coil tail structures (Figure 3A). The XLF–(1–233)–XRCC4–(1–164) proteins form a left-handed helical filament structure (Figure 3B).

In the crystals, such helical filaments together create a tubular structure with a 120 Å diameter central cylindrical cavity (Figure 3B). The crystal lattice is stabilized through contacts between the coiled-coil domains of XLF and XRCC4 (Figure 3A). The interactions appear to be mediated by hydrophobic contacts between XLF and XRCC4. The packing arrangement of the tubes, viewed along the c-axis, appears to be a series of engaged gear cogs (Figure 3A).

### The biological role of helical XLF–(1–233)–XRCC4–(1–164) assemblies

The helical alternating XLF–XRCC4 complex structure does not contain the region of XRCC4 that binds LigIV. The crystal structure of XRCC4–LigIV BRCT domains shows that the BRCT2 domain of LigIV interacts with the coiled-coil region of XRCC4 and is positioned close to one XRCC4 protomer head domain [23] where it can be accommodated without interfering with the observed helical structure of the XLF–XRCC4 complex. An EM study concluded that the catalytic domains of LigIV are located near the XRCC4 head domain and is connected to BRCT1 through a flexible linker [17]. As XLF interacts with the XRCC4 head domain, the location and flexibility of catalytic domains of LigIV when bound to XRCC4 requires further analysis in order to establish whether the presence of LigIV catalytic domains affects the interaction of XLF and XRCC4 in the helical fibre.

In classical chromosomal NHEJ, the function of XLF overlaps with that of ATM (ataxia telangiectasia mutated), which detects DSBs and activates DSB responses by phosphorylating histone H2AX and other substrates [24]. XLF, which is also targeted for phosphorylation by ATM in
Figure 2  XLF-(1–233)–XRCC4-(1–164) complex formation, identification and crystallization
(A) Gel filtration results show that XLF-(1–233) and XRCC4-(1–164) form a complex under 150 mM NaCl conditions. mAu, milli-absorbance units. Complex formation is concentration-dependent. Insets show SDS/PAGE of the samples, with molecular masses indicated in kDa. (B) Nano-ESI–MS shows that the major complex is a XLF-(1–233)–XRCC4-(1–164) heterotetramer with a measured molecular mass of 96661 ± 58 Da, composed of one XLF and one XRCC4 homodimer with peaks between 4000 and 5000 m/z. Both homodimers are present individually in solution (2500–4000 m/z region) with measured molecular masses of 42 190 ± 23 Da for XRCC4 and 53 455 ± 49 Da for XLF homodimers. Higher oligomers, hexamers and octamers appear above 5000 m/z and their intensities are concentration-dependent. The mass spectrum on the left-hand side was obtained from the sample of higher concentration and contains both hexamers and octamers of XLF–XRCC4 complexes, whereas the spectrum from the sample at half the concentration shows only hexamers. (C) The XLF-(1–233)–XRCC4-(1–164) protein sample was crystallized (left-hand panel), and the components of the complex was confirmed using SDS/PAGE (right-hand panel). Molecular masses are indicated in kDa.

its C-terminal region, may have a role in this process [24]. The crystal structure of the core nucleosome (PDB code 1KX5) with a diameter of approximately 100 Å can fit within the XLF–XRCC4 helical filament (diameter of approximately 120 Å), opening up the possibility that the XLF–XRCC4 fibre might wrap around chromatin interacting with DNA and histones. This would explain the earlier observation that the C-terminus of XLF, which would be located at the inner side of XLF–XRCC4 helical structure, is responsible for DNA interaction [10]. It is also possible to accommodate the Ku70/80 heterodimer (PDB code 1JEY) and DNA-PKcs (PDB code 3KGV) within the helical fibre. The C-terminal structures of XLF and XRCC4 are both targeted for phosphorylation by DNA-PKcs [15,16] and XLF can bind to the Ku heterodimeric core structure through its C-terminal structure [25]. Having both DNA-PKcs and Ku70/80 located
within the helical fibre would assist these functions. Indeed, the XLF–XRCC4 helical filament may act as a ‘reaction shell’, which stabilizes chromatin near-IR foci, and gathers Ku70/80 and DNA-PKcs together for efficient NHEJ function.

The recently defined crystal structures of the N-terminal regions of the centriole protein SAS-6 in Caenorhabditis elegans, Chlamydomonas reinhardtii and Danio rerio have revealed similar protein folds to those of XLF and XRCC4 [26,27]. The homodimeric SAS-6 proteins form nine-fold symmetrical ring structures with head domains interacting together. The coiled-coil tails of the SAS-6 dimers extend outwards towards the assemblies of microtubules. A mutagenesis study has shown that the head-to-head interaction of SAS-6 proteins during oligomerization is mediated by the β6–β7 loop inserting into the hydrophobic pocket created by the helix-turn–helix structure and β7 from the neighbouring homodimer head domain. This is very similar to the binding model described here between XLF and XRCC4. The interaction region between XLF-(1–233) and XRCC4-(1–164) is relatively small, which makes the helical complex structure rather flexible and could also allow the formation of a closed ring structure as in SAS-6.

In addition to the proteins bound within the XLF–XRCC4 helical structure, there may be other NHEJ proteins assembled around it interacting with the coiled-coil C-terminal regions as in SAS-6. One of these proteins is likely to be LigIV, which binds to the XRCC4 coiled-coil tail. This would be required at the DSBs and therefore might not be bound at every XRCC4, but rather could destabilize or rearrange the helical structure. Further proteins may interact with the extension at the C-terminus of XRCC4; for example, PNKP interacts with XRCC4, both through a site phosphorylated by protein kinase CK2, as well as with the unphosphorylated protein [28,29]. XLF does not bind to LigIV, but the folded-back loop sequence between XLFα4 and α5 is evolutionarily conserved [13]. Site-directed mutagenesis studies of XLF at Leu174, Arg178 and Leu179, which are all located in this evolutionarily conserved hinge region, reduces the stimulation of the DNA end ligation activity without affecting the association with XRCC4 or DNA [13]. This XLF conserved region of unknown function may bind to other as yet unidentified NHEJ proteins. XLF is also required for alignment-based gap filling by DNA polymerases λ and μ [30]. Thus XLF–XRCC4 may provide a similar safety-belt function elsewhere by securing key proteins close to the DSB site, therefore assisting DNA repair, which is crucial for cells to survive.

Indeed, the XRCC4–XLF assembly now described may be the first dance steps of the protein partners XLF and XRCC4. The next steps may involve further NHEJ proteins in a synchronized formal dance that will reveal more of the complex process of DNA DSB repair.

Acknowledgements
We thank Dr Victor Bolanos Garcia and Dr Lynn Sibanda for helpful discussions.
Funding

T.L.B. and D.C. thank the Wellcome Trust for funding through a programme grant.

References

15. Yu, Y. (2003) DNA-PK phosphorylation sites in Xrcc4 are not required for survival after radiation or for V(D)J recombination. DNA Repair 2, 1239–1252

Received 22 May 2011
doi:10.1042/BST0391387
SUPPLEMENTARY ONLINE DATA

Non-homologous end-joining partners in a helical dance: structural studies of XLF–XRCC4 interactions

Qian Wu¹*, Takashi Ochi*, Dijana Matak-Vinkovic†, Carol V. Robinson‡, Dimitri Y. Chirgadze* and Tom L. Blundell*¹

¹Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1GA, U.K., †University Chemical Laboratory, Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, U.K., ‡Chemistry Research Laboratory, University of Oxford, 12 Mansfield Road, Oxford OX1 3TA, U.K.

Nano-ESI MS

For MS of the intact complexes, 20 ml of the XLF–XRCC4 complex was buffer-exchanged into 200–250 mM ammonium acetate Biospin columns (Bio-Rad Laboratories). Nano-ESI mass spectra were acquired on a modified QSTAR XL (MDS Sciex, Applied Biosystems) mass spectrometer using a protocol described previously [1].

Typical instrument parameters used, in positive-ion mode, on the QSTAR XL were: capillary voltage 1.4 kV; declustering potential 100 V; focusing potential 150 V; quadrupole voltage (Q0) 20–90 V; collision gas (CAD) 3–8. Data were acquired using Analyst QS software (Applied Biosystems) and MassLynx v4.1 (Waters).

Molecular replacement

Phaser 2.3 of Phenix software suite [2] first identified two possible solutions for the position of the XLF using the XLF-(1–233)–XRCC4-(1–164) heterotetramer in the asymmetric unit, resulting in a high solvent content of 87 %, which probably explains the low resolution of the diffraction from the crystals.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Statistic</th>
</tr>
</thead>
<tbody>
<tr>
<td>X-ray diffraction data</td>
<td>P65</td>
</tr>
<tr>
<td>Space group</td>
<td>P65</td>
</tr>
<tr>
<td>Unit cell</td>
<td>a = b = 236.76 Å, c = 103.23 Å, α = 90°, β = 90°, γ = 120°</td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>50.0–8.50 (8.80–8.50)</td>
</tr>
<tr>
<td>Rsym (%)</td>
<td>0.05 (44.5)</td>
</tr>
<tr>
<td>Completeness (%)</td>
<td>99.9 (100)</td>
</tr>
<tr>
<td>Number of unique reflections</td>
<td>3019</td>
</tr>
<tr>
<td>Average redundancy</td>
<td>6.4 (6.2)</td>
</tr>
<tr>
<td>Average intensity, &lt;I/σ(I)&gt;</td>
<td>12.0</td>
</tr>
<tr>
<td>Refinement</td>
<td></td>
</tr>
<tr>
<td>Resolution range (Å)</td>
<td>49.8–8.49</td>
</tr>
<tr>
<td>Number of reflections: work/test</td>
<td>2717/278</td>
</tr>
<tr>
<td>Rcryst (%)</td>
<td>40.3</td>
</tr>
<tr>
<td>Rfree (%)</td>
<td>41.5</td>
</tr>
</tbody>
</table>

The presence of the two non-crystallographic two-fold axes perpendicular with the c-axis (Figure S1B, Calculated) was confirmed by calculating the self-rotation function using the X-ray diffraction data (Figure S1B, Observed). The tubular XLF–XRCC4 structure has solvent gaps between the
Figure S1 | Crystallographic packing of the XLF–XRCC4 complex

(A) Crystallographic packing of the XLF–XRCC4 complex viewed along the c-axis. Black arrows indicate pseudo-two-fold axes. The blue line is a unit cell. XLF and XRCC4 are represented in red and blue respectively. (B) Observed and calculated self-rotation maps of the XLF–XRCC4 complex. The images show $\chi = 180^\circ$ of the self-rotation maps that were calculated using Molrep [4]. The structural factors of the model were calculated using SFALL.

filaments (Figure S1A). The gaps can be filled with another set of the six filaments related by pseudo-translation along the c-axis to the original ones. However, since strong peaks were absent from the native Patterson map except for the origin (results not shown), 12-filament tubes are unlikely to exist in our crystal.

References


Received 22 May 2011
doi:10.1042/BST0391387