Cryo-EM structures of the XPF-ERCC1 endonuclease reveal how DNA-junction engagement disrupts an auto-inhibited conformation

Morgan Jones1, Fabienne Beuron2, Aaron Borg3, Andrea Nans4, Christopher P. Earl1, David C. Briggs1, Ambrosius P. Snijders3, Maureen Bowles1, Edward P. Morris2, Mark Linch5 & Neil Q. McDonald1,6

The structure-specific endonuclease XPF-ERCC1 participates in multiple DNA damage repair pathways including nucleotide excision repair (NER) and inter-strand crosslink repair (ICLR). How XPF-ERCC1 is catalytically activated by DNA junction substrates is not currently understood. Here we report cryo-electron microscopy structures of both DNA-free and DNA-bound human XPF-ERCC1. DNA-free XPF-ERCC1 adopts an auto-inhibited conformation in which the XPF helical domain masks the ERCC1 (HhH)2 domain and restricts access to the XPF catalytic site. DNA junction engagement releases the ERCC1 (HhH)2 domain to couple with the XPF-ERCC1 nuclease/nuclease-like domains. Structure-function data indicate xeroderma pigmentosum patient mutations frequently compromise the structural integrity of XPF-ERCC1. Fanconi anaemia patient mutations in XPF often display substantial in-vitro activity but are resistant to activation by ICLR recruitment factor SLX4. Our data provide insights into XPF-ERCC1 architecture and catalytic activation.
**S**tructure-specific endonucleases (SSEs) are found in all branches of life and play crucial roles in genome repair, replication and recombination. These endonucleases act on similar DNA structures with defined polarity but use different catalytic mechanisms. The structurally related XPF/MUS81 family are an important group of human 3′-nucleases that associate to form two active endonuclease heterodimers (XPF–ERCC1 and MUS81–EME1) and a DNA translocase (FANC–FAAP24) with a pseudo-nuclease architecture. XPF–ERCC1 recognises double-stranded/single-stranded (ds/ss) DNA junctions which have a 3′-ssDNA overhang, nicking the dsDNA backbone to produce a substrate for subsequent steps in DNA repair pathways. XPF–ERCC1 activity is essential for removing helical DNA distortions arising from ultraviolet-induced damage and bulky adducts as part of the nucleotide excision repair (NER) pathway. In this context XPF–ERCC1 nicks the damaged DNA strand 3′ of the lesion at the ds/ss junction of an NER repair bubble. It is also required for interstrand cross-link repair (ICL), some double-stranded break repair processes, base excision repair, Holliday junction resolution, gene-conversion and telomere maintenance. Mutations in XPF and ERCC1 genes are associated with genetic disorders exhibiting diverse phenotypes. These pathologies are caused by defects in the genome maintenance pathways that involve XPF–ERCC1, including xeroderma pigmentosum (XP), Cockayne’s syndrome, Fanconi anaemia (FA), XPFE progeria and cerebro-oculo-facio-skeletal syndrome. The genotype–phenotype correlations of XPF–ERCC1 driven diseases are still poorly understood.

XPF is the enzymatically active subunit of the heterodimeric XPF–ERCC1 endonuclease and is comprised of a helicase-like module (HLM) and a catalytic module (CM) (Fig. 1a). The XPF HLM is related to the superfamily 2 helicases, with two divergent RecA-like domains that flank an all α-helical domain (Fig. 1a). Both XPF RecA-like domains, termed RecA-like domain 1 (RecA1) and RecA-like domain 2 (RecA2) lack the residues necessary to bind and hydrolyse ATP. Despite this, the HLM is required for full XPF activity and binds both the ICLR recruitment factor SLX4 and ds/ssDNA structures. The XPF CM consists of a nuclease domain containing a metal-dependent GDX,ERKxD active site motif and a tandem helix–hairpin–helix, termed an (HhH)2 domain. The smaller ERCC1 subunit has no catalytic activity but is structurally related to the XPF CM, consisting of a nuclease-like domain (NLD) and a dsDNA-binding (HhH)2 domain. Both ERCC1 domains heterodimerise with their equivalent domains in the XPF CM, forming discrete nucleases–NLD and 2×(HhH)2 functional units. As well as contributing to XPF stability, ERCC1 can recognise ds/ssDNA substrates and engages the XPA repair protein that is required for XPF–ERCC1 recruitment to sites of NER. Currently, there are no available structures of the XPF HLM or of any full-length XPF–Mus81 family members. By solving the structure of a near full-length human XPF–ERCC1 we have defined its overall architecture and uncovered a previously unreported auto-regulatory mechanism. We show XPF–ERCC1 adopts an auto-inhibited conformer in the absence of DNA in order to prevent promiscuous cleavage and provide structural evidence for the initial steps of XPF–ERCC1 activation upon binding a DNA junction.

**Results**

**Structure determination of human XPF–ERCC1 endonuclease.** A single particle cryo-electron microscopy (cryo-EM) density map of purified recombinant XPF–ERCC1 complex (128 kDa) (Fig. 1b) was determined at a global resolution of 4.0 Å (Supplementary Fig. 3a, c, f and Supplementary Movie 1) enabling the assignment of XPF–ERCC1 domain organisation (Fig. 1c, d). The overall structure of human XPF–ERCC1 endonuclease. The cryo-EM structure of near full-length XPF–ERCC1 reveals a compact conformation with extensive interactions between the XPF HLM and CM modules (Fig. 1d). Overall, the HLM adopts a “C”-shape that has dimensions of approximately 70 × 40 × 60 Å. The two RecA-like domains form a rigid platform and lack a nucleotide cleft characteristic of many ATP-driven helicases. Instead the two XPF RecA-like domains are linked through the intimate intertwining of secondary structural elements that extend beyond their globular portion (Supplementary Fig. 4d). While RecA1 caps one edge of the HLM and engages the XPF nuclease domain in the CM, the helical domain caps the other HLM extremity and engages the CM and the dsDNA-binding ERCC1 (HhH)2 domain (Figs. 1d and 2a). This arrangement serves to separate and uncouple both functional domains of ERCC1 through its connecting linker. These interactions confirm the key regulatory role for the HLM by engaging crucial elements within the XPF CM and ERCC1. Interfaces observed in the XPF–ERCC1 structure were largely validated using cross-linking mass spectrometry (XL-MS) (Fig. 2f, g) (Supplementary Table 2). Cross-links are found predominately between both the XPF (HhH)2 domain and the ERCC1 NLD, and between the XPF RecA2 and ERCC1 NLD. In addition, several cross-links exceeding the distance cut-off are consistent with two principal vectors of dynamic movement in solution.

**Structure of the XPF HLM.** The XPF HLM is typical of other helicase superfamilies 2 (SF2) members with a RecA1–helical domain–RecA2 organisation, but with substantial inserts within RecA2 (Fig. 1a). In the absence of ATP binding and hydrolysis motifs or a nucleotide binding cleft, RecA1–RecA2 are linked together through a predominantly polar interface (2007 Å2). Major interface contributions are made by secondary structural elements β8 and α20 that form a C-terminal extension to RecA1 and RecA2, respectively, as well as the XPF amino-terminus (Supplementary Fig. 4d). β8 extends the smaller RecA2 four parallel β-stranded sheet while α20 packs against the larger RecA1 seven-stranded parallel beta sheet (β1–β7). Additional RecA1–RecA2 contacts centre on a π-ring stacking interaction between RecA1 domain Y71XPF and RecA2 domain Y564XPF at one interface edge (Supplementary Fig. 4c) and L39XPF and I592XPF on the other edge. Polar residues make up the remaining contacts with a small cavity.
No protein expression was observed for a Y71A XPF mutant (Table 1). The observed structural rigidity of the RecA1–RecA2 unit is structurally homologous to equivalent domains in nucleosome-bound chromatin remodellers ISW1 and INO8023,24.

XPF RecA2 has two large inserts with unknown functions. Insert one (residues 345–377) separates the helical and RecA2 domains and insert two (residues 441–550) interrupts the RecA2 fold. There is sufficient density in our map to trace the backbone of residues 345–362 and 366–377 from insert one projecting away from the body of the structure. However, no density was recovered for insert two, in agreement with predictions that this region is intrinsically disordered in the absence of DNA. Futhermore,
XL-MS data identified a large number of intra-insert cross-links within inserts one and two, consistent with these highly basic regions being flexible (Supplementary Table 2).

The XPF helical domain is an integral part of the HLM and folds as a five anti-parallel helical bundle. This domain packs tightly against RecA2 and is anchored through an interface centred close to residues Q300XPF/D302XPF and S412XPF/Q419XPF (Supplementary Fig. 4b). The Q300AXPF mutant significantly reduces XPF–ERCC1 expression and increases aggregation (Table 1). Helix α17 (residues 426–440) also contributes to tethering the helical domain to RecA2. The observed position of the helical domain determines the orientation and angle of the extended RecA2 C-terminal α20 helix.
Fig. 2 Architecture of the XPF helicase-like module and coupling with the catalytic module. a View of the XPF–ERCC1 structure showing the helicase-like module (HLM as surface rendering) contacts the XPF nuclease domain (gold ribbon cartoon) at two interfaces (dashed red boxes). Domains are coloured according to the scheme used in Fig. 1. b, c Close-up view of interaction interfaces overlaid with the composite cryo-EM map. Selected residues are displayed as sticks and coloured by heteroatom: blue—N, Red—O. b The hydrophobic interaction interface between XPF RecA2 (blue) and XPF nuclease domain (gold). c Interaction of XPF helical domain residues 273–275 (green) with the XPF nuclease domain (gold). d The XPF–ERCC1 HLM (surface rendered) contacts with the ERCC1 (HhH2) domain at a single interface (dashed red box). e Interaction of XPF helical domain helix α13 (green) and the ERCC1 (HhH2) domain close to its dsDNA-binding residues (pink). f Two orthogonal views of the XPF–ERCC1 structure with XL-MS distance constraints overlaid. Distances within the allowed Ca–Ca cut-off distance of 30 Å are displayed in blue, distances greater than this cut-off displayed in red. Blue dotted line indicates a cluster of allowed distances between the XPF helical and ERCC1 (HhH2) domains. g Cartoon schematic representing inter-domain cross-links detected by mass spectrometry. Each black line indicates a single unique cross-link between residues in different domains. Domains within the pink ellipsoid form the XPF HLM, whereas domains within the XPF CM and ERCC1 are within the pale blue ellipsoid.

| Protein | Mutant | Mutation rationale | Effect on protein expression |
|---------|--------|--------------------|----------------------------|
| XPF     | L608P  | XP                 | Soluble aggregates         |
| XPF     | T567A  | XP                 | Soluble aggregates         |
| XPF     | R799W  | XP                 | No expression              |
| XPF     | Y71A   | Structural integrity | No expression             |
| ERCC1   | T248A  | Autoinhibition     | No expression              |
| ERCC1   | T252A  | Autoinhibition     | No expression              |

(Supplementary Fig. 4d), stabilising the HLM conformation through interaction between Q226XPF and T614XPF.

The XPF helical domain regulates XPF–ERCC1 activity. The XPF HLM is coupled to the CM through contacts from RecA1 and the helical domain (Fig. 2a). RecA1 forms a substantial interface (1684 Å2) with the XPF nuclease domain involving aromatic and hydrophobic residues from RecA1 α5 and α6 helices and XPF nuclease domain α4 and α21 helices and β14 strand (Fig. 2b). The hydrophobic nature of the contact suggests that anchoring of the HLM to the XPF nuclease domain through RecA1 forms a permanent part of the XPF–ERCC1 architecture. The XPF helical domain forms a contact with the XPF nuclease domain that sterically prevents the ds/ssDNA substrate from reaching the XPF active site (Fig. 2c and Supplementary Movie 3). A key contact within this auto-inhibited conformation is between sidechains of H275XPF and S730XPF. A H275AXPF, W274AXPF double mutant, likely to disrupt this contact, displays a 1.5-fold increase in catalytic efficiency relative to the wild type (Table 2).

A second autoinhibitory interface exists between the XPF helical domain and the ERCC1 (HhH2) domain (Fig. 2d, e and Supplementary Movie 3). This interface is formed through predominantly polar contacts involving the highly conserved T248ERCC1, T252ERCC1 residues and both S312XPF and T316XPF. Previous structural and biochemical data suggest that the ERCC1 (HhH2) domain binds dsDNA through hairpin residues S244ERCC1–N246ERCC1 and G276ERCC1–G278ERCC1 mainchain atoms. These motifs are proximal to T248ERCC1 and T252ERCC1, and are not accessible in the DNA-free conformation of XPF. The S312A mutant displays a 1.5-fold higher catalytic efficiency than the wild type likely due to the disruption of this autoinhibitory interaction (Table 2). Equally, shortening the connecting linker between the XPF nuclease and (HhH2) domain would be predicted to shift the 2α-(HhH2) unit towards the nuclease domain releasing the DNA-binding residues. Indeed, a 829–833ΔXPF mutant displayed a modest 1.2-fold increase in catalytic efficiency and a 7.5-fold tighter Kₘ relative to wild type (Table 2).

Heterodimerisation of XPF and ERCC1 through two interfaces. ERCC1 is intimately coupled to the XPF CM through two obligate dimerisation surfaces at the equivalent domains of each molecule. The XPF nuclease domain uses a helix–strand–helix motif (α25–β19–α26) to heterodimerise with the equivalent surface of the ERCC1 NLD (α3–β8–α4) forming a kidney-shaped dimer with an extensive interaction interface (1684 Å²) (Supplementary Fig. 4a). The contact is predominantly hydrophobic and is flanked by three salt bridges (Supplementary Fig. 4a). This interface uses equivalent elements to those mediating heterodimerisation of homologous domains from Mus81–Eme1 and FANCM–FAAP24 complexes. We note that the XPF (HhH2) domain hetero-dimers with the ERCC1 (HhH2) domain through predominantly hydrophobic contacts close to F851XPF and F900XPF as previously observed. The (HhH2) domain from XPF and ERCC1 are connected to their XPF nuclease domain/ERCC1 NLD domain through ordered linker sequences. There is sufficient density in our cryo-EM map to trace the mainchain atoms for both linkers (Fig. 1d). The ERCC1 linker makes unexpected interactions with the XPF nuclease domain via Y215ERCC1 and D221ERCC1 (Fig. 3b). We note that Y215ERCC1 lies adjacent to S786XPF suggesting the FA mutation S786FXPF would disrupt this contact with ERCC1. Despite the close association of XPF CM and ERCC1 through heterodimerisation, their respective functional domains remain uncoupled and held apart through the extended conformation of their connecting linkers. This is important to consider when comparing with the DNA-bound conformations (see later).

Structural context of XP and FA patient mutations in XPF. Recruitment of XPF–ERCC1 into either NER or ICLR pathway complexes is dependent on interaction with partner proteins XPA or SLX4 at their respective damaged DNA structures (Fig. 3a). A previous study mapped the XPA-binding site to a cleft within the ERCC1 NLD (Fig. 3a)30. This interaction is spatially distinct from the proposed SLX4 site centred within the helical domain at L230XPF19. Insights from disease mutations have shown that repair pathway recruitment can be disrupted by separation-of-function (FA) or partial loss-of-function (XP) mutations, however the structural basis for this is unclear31.

With the availability of a three-dimensional XPF–ERCC1 structure, it was possible to explore the location and structural environment of disease-causing mutations and correlate this with their impact on enzyme stability and catalytic activity. Patient-derived XP or FA-associated mutations were characterised in vitro using a previously reported fluorescence incision assay20. Mutations associated with XP mapped primarily to the XPF RecA2 domain and its inserts15,32,33. L608XPF, R589XPF and T567XPF are located in the folded region of the RecA2 domain, with the latter two forming structurally important intra-domain contacts32 (Fig. 3c). Indeed, L608XPF and T567ΔXPF mutant proteins formed soluble aggregates when expressed recombinantly, as measured by analytical size...
exclusion chromatography (SEC) and an R589W<sub>XPF</sub> mutant exhibited 35-fold reduction in catalytic efficiency (Table 2). The R799W<sub>XPF</sub> XP mutation failed to express recombinantly and lies on the periphery of the heterodimeric nuclease–NLD interface with ERCC1 (Fig. 3b). These data, taken in the context of our structure, suggest the L608P<sub>XPF</sub>, T567A<sub>XPF</sub>, R589W<sub>XPF</sub> and R799W<sub>XPF</sub> XP disease mutants compromise XPF–ERCC1 structural stability (Table 1). I225XPF is also associated with XP<sup>32</sup> and maps onto the hydrophobic core of the helical domain (Fig. 3d) suggesting it is also likely to contribute to XPF–ERCC1 structural integrity.

FA patients are proficient in NER but deficient in ICLR, indicating a likely separation of function<sup>19,34</sup>. Our structure indicates the FA point mutations within XP such as L230R<sub>XPF</sub>, C236RF<sub>XPF</sub> and G325EX<sub>XPF</sub> cluster within the XPF helical domain (Fig. 3d)<sup>11</sup>. These mutations, when expressed recombinantly, were found to have a similar level of endonuclease activity to wild-type XPF–ERCC1 against a stem–loop substrate (Table 2). Previous studies indicated these FA mutations are unable to engage SLX4<sup>19</sup>. We found that XPF<sup>–</sup>ERCC1 mutants compromised XPF–ERCC1 activity (ERCC1<sup>ΔS244ERCC1</sup> and G276 ERCC1<sup>ΔS244ERCC1</sup>) indeed co-expressed with a truncated form of human SLX4 (XPF<sup>–</sup>ERCC1–SLX4<sup>NTD</sup>) indeed showed a six-fold increase in catalytic efficiency (Table 3 and Supplementary Fig. 5b, c). ERCC1, the most dramatic being the FA/CS mutants (Table 3 and Supplementary Fig. 5b, c).

### Table 2 Kinetic data for purified XPF–ERCC1 mutants.

| Protein | Mutant | Mutation rationale | \( V_{\text{max}} \) (fmol min<sup>−1</sup>) | \( K_m \) (nM) | \( k_{\text{cat}} \) (s<sup>−1</sup>) | \( k_{\text{cat}}/K_m \) (nM<sup>−1</sup> min<sup>−1</sup>) |
|---------|--------|-------------------|-----------------|-------------|-------------|-----------------|
| XPF     | WT     |                    | 104.7 ± 2.8     | 12.1 ± 1.4  | 20.9 ± 0.66  | 1.73            |
| XPF     | L230R  | FA                | 65 ± 2.1        | 8.1 ± 1.1   | 13.0 ± 3.3   | 1.60            |
| XPF     | L236R  | FA/CS             | 91.2 ± 2.5      | 10.8 ± 1.2  | 18.2 ± 0.5   | 1.69            |
| XPF     | E239K  | FA                | 88.9 ± 1.7      | 12.4 ± 2.1  | 16.9 ± 0.7   | 1.36            |
| XPF     | S766F  | ICLR deficient    | 80.9 ± 3.8      | 32.5 ± 1.5  | 7.9 ± 0.8    | 0.24            |
| XPF     | 323–326 Δ | ICLR deficient | 120.2 ± 1.4    | 20.5 ± 0.8  | 24.2 ± 0.3   | 1.18            |
| XPF     | S312A  | Autoinhibition    | 126.1 ± 28.2    | 5.1         | 135 ± 1.8    | 2.65            |
| XPF     | W274A, H275A | Autoinhibition | 193.4 ± 8.6    | 15.2 ± 2.4  | 38.7 ± 3.1   | 2.55            |
| XPF     | R112A  | DNA binding       | 24.5 ± 1.1      | 4.32 ± 0.63 | 5.6 ± 0.2    | 1.30            |
| XPF     | R829–R833 Δ | Autoinhibition linker disruption | 20.2 ± 0.5 | 1.9 ± 1.2 | 3.8 ± 0.2 | 2.00 |
| XPF     | R589W  | XP                | 8.22 ± 2.6      | 25.9 ± 6.1  | 1.4 ± 0.9    | 0.05            |
| ERCC1   | L235A  | Autoinhibition    | 33.1 ± 2.7      | 28.7 ± 22.3 | 3.4 ± 0.3    | 0.12            |

Each kinetic value was obtained from 3 technical replicates (n = 3) ± standard deviation (SD).

XPF–ERCC1 conformational activation on DNA-junction binding. We hypothesised that the autoinhibitory interactions formed by the XPF helical domain need to be released following XPF–ERCC1 DNA-junction engagement, prior to the incision reaction. To probe the nature of such potential conformational changes, we assembled a complex of XPF–ERCC1 bound to a DNA stem–loop model substrate (10-duplex 20-T single-strand stem–loop) that we previously showed presents a single incision site to XPF–ERCC1<sup>30</sup>. Using an electrophoretic mobility shift assay (EMSA) we observed 1:1 stoichiometric binding of the stem–loop DNA to XPF–ERCC1 (Supplementary Fig. 5b, c).

This sample was used for cryo-EM data collection leading to a single-particle cryo-EM density map at a global resolution of 7.7 Å (Supplementary Fig. 6b). Signal subtraction of the dimeric 2×(HhH)<sub>2</sub> domain and DNA density, followed by local refinement, improved the resolution of the resulting sub-volume to 5.9 Å (Supplementary Fig. 6c). The locally refined map shows evidence of helical features, with the local resolution highest in the core of the RecA domains (Supplementary Fig. 6a). 3DFSC (Supplementary Fig. 6e, f and Supplementary Movie 4) analysis indicates that the map does not suffer heavily from anisotropy and the lower resolution of the DNA-bound map relative to the DNA-free is as a result of increased flexibility. Indeed, XPF–ERCC1 does not engage DNA in vivo unless recruited by XPA in complex with TFIHH<sup>37</sup>. It is likely that the DNA-bound XPF–ERCC1 complex only becomes fully stabilised in the presence of these additional factors.

The DNA-bound reconstruction enabled the placement of all XPF–ERCC1 domains using the DNA-free structure as an initial template (Fig. 4a and Supplementary Movie 5). Aligning the DNA-bound and DNA-free maps identified key changes in the architecture of XPF–ERCC1, the most dramatic being the disengagement of the 2×(HhH)<sub>2</sub> domain from the XPF helical domain and it’s repositioning adjacent to the XPF nuclease–ERCC1 NLD dimer, as seen for other XPF/Mus81 family endonucleases<sup>27,28</sup> (Fig. 4e). An additional region of density was identified adjacent to the 2×(HhH)<sub>2</sub> domain but segmented into a distinct volume (Fig. 4b). This density was assigned as the duplex portion of the stem–loop substrate due to the unambiguous presence of a 19 Å concave major groove and its length measuring the distance of 10 base pairs (Fig. 4c). In order to correctly position the 2×(HhH)<sub>2</sub> domain with respect to the dsDNA, the structure of the Aeropyrum pernix XPX homodimer in complex with dsDNA was fit into the map and used to align the human 2×(HhH)<sub>2</sub>—dsDNA functional unit (Fig. 4b, c). The fit to density was then optimised for the human structure using Flex-EM<sup>38</sup>. This positions the 2×(HhH)<sub>2</sub> domain–dsDNA-binding residues S244<sub>ERCC1</sub>–N246<sub>ERCC1</sub> and G276<sub>ERCC1</sub>–G278<sub>ERCC1</sub> in close proximity to the dsDNA minor groove in a homologous fashion to other family members (Fig. 4e). Furthermore, comparison of the DNA-free and DNA-bound 2D class averages clearly indicates a repositioning of the 2×(HhH)<sub>2</sub> domain upon substrate engagement (Fig. 4a).

The remaining domains of XPF–ERCC1 can be fit unambiguously into the density. The RecA1–RecA2 unit remains structurally rigid, with high-resolution features present in 2D class...
**Fig. 3 Mapping XPF–ERCC1 disease mutations and DNA repair pathway recruitment sites.**

**a** Top, a ribbon model of XPF–ERCC1 highlighting the spatially distinct binding sites of XPA and SLX4. XPA binds to the ERCC1 NLD (orange) and SLX4 binds to the XPF helical domain (green). The XPA peptide (residues 66–77) atoms are displayed as red spheres (PDB: 2JNW). The key SLX4 binding residue L230 sidechain atoms are also displayed as red spheres. Bottom, representative DNA structures targeted by XPF–ERCC1 through SLX4 (interstrand cross-link) or XPA (intrastrand cross-link) recruitment.

**b–d** The molecular environment of patient-derived disease mutations are indicated on the structure, superposed with the cryo-EM map displayed close to the mutation site. Selected residues are displayed using stick rendering coloured by heteroatom. Residues associated with Fanconi anaemia (FA) patient mutations are coloured red whilst those associated with XP are coloured black. Black or orange dashed ellipses indicate the environment close to XP or FA mutations respectively.

**b** Mutations in the XPF nuclease domain and ERCC1 NLD lie close to their interface and give rise to both FA and XP. **c** XP-associated mutations disrupt key structural contacts in the XPF RecA2 domain are shown overlaid with the composite cryo-EM map. **d** FA-associated mutations cluster within the XPF helical domain. The helical domain also contains the XP-associated mutant, I225 (black).

---

**Table 3 SLX41–758–XPF–ERCC1 mutation data summary.**

| Protein                  | Mutant | V<sub>max</sub> (fmol min<sup>−1</sup>) | K<sub>m</sub> (nM) | k<sub>cat</sub> (min<sup>−1</sup>) | k<sub>cat</sub>/K<sub>m</sub> (nM<sup>−1</sup> min<sup>−1</sup>) |
|--------------------------|--------|--------------------------------------|-------------------|-------------------------------|-----------------------------------------------|
| XPF(–ERCC1–SLX4<sup>N</sup>TD) | WT     | 219.2 ± 9.3                         | 4.2 ± 0.7         | 43.8 ± 1.1                    | 10.4                                          |
| XPF(–ERCC1–SLX4<sup>N</sup>TD) | 323–326 Δ | 49 ± 6.5                              | 32.2 ± 9.6         | 9.8 ± 1.4                      | 0.30                                          |
| XPF(–ERCC1)               | WT     | 104.7 ± 2.8                          | 12.1 ± 1.4        | 20.9 ± 0.66                  | 1.73                                          |

Each kinetic value was obtained from 3 technical replicates (n = 3) ± the standard deviation (SD).
averages (Fig. 4f, g), reaffirming its role as an inactive helicase. Whilst the remainder of the complex increases in flexibility upon substrate engagement (Fig. 4g), the interface between the XPF RecA1 and nuclease domains remains intact (Fig. 4d). Comparison with the DNA-free structure reveals that the XPF helical domain pivots by approximately 15°, rotating ~11 Å away from the nuclease domain (Supplementary Fig. 7a). The increased flexibility of the XPF helical domain following its disengagement with the XPF nuclease domain can be visualised by the loss of high-resolution features in 2D class averages following substrate engagement (Fig. 4g). This conformational change breaks the autoinhibitory contact formed between H275XP and S730XP as predicted from the DNA-free structure. The remaining unmodelled density likely corresponds to the flexible first RecA2 domain insert (Fig. 4b).

A model for DNA junction-based activation. Tight regulation of endonuclease catalytic activity is needed to prevent inappropriate DNA cleavage. Indeed XPF–ERCC1 displays no activity towards DNA duplexes, ssDNA or an equimolar mixture of ds and ssDNA substrate (Fig. 5b). This implies that it is the proximity of the ssDNA and dsDNA elements in a junction context that is uniquely required to stimulate XPF–ERCC1 activation and overcome complex autoinhibition. Analysis of our DNA-bound structure reveals that the presence of a junction shifts the dimeric 2x(HhH)2 domain by 47 Å to contact the XPF nuclease–ERCC1 NLD dimer, disrupting contacts with the XPF helical domain (Fig. 5a, b, Supplementary Movies 9 and 10). In this configuration the dimeric 2x(HhH)2 domain lies proximal to the ERCC1 NLD domain, coupling both known ssDNA-binding elements of the endonuclease25,27,28,39 within the ERCC1 NLD and XPF (HhH)2 domain, coupling both known ssDNA-binding elements of the nuclease domain40, however, the precise arrangement of multiple ssDNA and dsDNA domains required for DNA-junction recognition remains to be determined. The final DNA-bound model lacks the single-stranded portion of the stem–loop and places the scissile phosphodiester bond approximately 15 Å from the XPF active site motif (residues 725–727) (Fig. 5b). We interpret the DNA-bound structure as showing important features of an initial step towards full DNA-junction recognition prior to the incision reaction. The low resolution of the DNA component within the cryo-EM map (approximately 9 Å) suggests that the dimeric 2x(HhH)2–DNA complex can adopt multiple conformers. Equally, the accessibility...
of the dsDNA major groove opposite to the 2×(HhH)₂ minor groove interaction could be re-oriented towards the positively charged concave surface within the XPF HLM (Fig. 5c, d).

The closest structural homologue of both DNA-bound and DNA-free structures, as identified by the DALI protein structural comparison server⁴¹, is the helicase/translocase MDA5 that binds dsRNA ⁴⁰–⁴² (rmsd of 4.1 Å over 283 C-alphas) (Fig. 5d and Supplementary Fig. 7b, c). MDA5 binds to the major groove of A-form dsRNA (PDB: 4GL2) through an equivalent positively charged concave surface within the XPF HLM (Fig. 5c, d). The superposition of an XPF (HhH)₂ domain bound to ssDNA (PDB: 2KN7) with our DNA-free structure reveals that the XPF Rec2 β-sheet and could represent part of the ssDNA-binding sites on the XPF (HhH)₂ domain and the ERCC1 NLD is too far (>50 Å) to be engaged simultaneously by the 20-thymine residue stem–loop (Fig. 5a). Movement of the 2×(HhH)₂ domain in the presence of a stem–loop shortens this distance to approximately 30 Å (Fig. 5b). This is consistent with changes in (HhH)₂ domain position and linkers observed in published structures for A. pernix XPF and Mus81–Eme1 in the presence and absence DNA. It is also supported by both our 3D variability analysis (Supplementary Movies 6–8) and by XLMS data. We therefore speculate that longer junction substrates may reveal even further dynamic rearrangements sufficient to place a junction at the nucleosome active site (Fig. 5e).

Discussion
The structural and functional studies described in this report provide insights into XPF–ERCC1 architecture, regulation and activation. The XPF–ERCC1 endonuclease catalyses the first...
irreversible step in NER repair by nicking the 5’-edge of the repair bubble structure on the damaged strand. The structure of DNA-free XPF–ERCC1 reveals how the heterodimer is auto-inhibited by blocking both DNA binding and active site access through contacts with the XPF helical domain. This structure reveals inter-domain interfaces not previously described and rationalises our previous report that the HLM impacts on endonuclease activity and substrate interaction. Whilst the ssDNA-binding surfaces of XPF (HhH)2 and ERCC1–NLD are fully solvent accessible in the auto-inhibited structure, they are uncoupled from their respective dsDNA-binding surfaces (ERCC1 (HhH)2 and XPF–HLM), which are sterically blocked. The structure also confirms the presence of a heterodimeric interface between the XPF nuclease and ERCC1 NLD as described for other family members.

This study provides evidence linking conformational activation of XPF–ERCC1 through DNA-junction recognition, with a likely contribution from recruitment partner proteins at DNA-junction sites prepared for either NER or ICLR pathways. Mapping the XPA interaction site within ERCC1 and the Slx4 site within XPF HhH2 domain reveals spatial separation of each recruitment partner site in the auto-inhibited state. It suggests the critical binding determinants are non-overlapping, but full structures of XPF–ERCC1 with SLX4 or XPA combined with competition binding studies are required to prove this. XPF–ERCC1 activation by SLX4 is disrupted by some FA mutations that map to the helical domain, in agreement with previous in vivo work. Given its proposed regulatory role, the helical domain may be repositioned on binding SLX4 to stimulate activity. In contrast, XP-associated mutations were found to generally reduce endonuclease activity in vitro towards an NER substrate by destabilising the complex whereas FA mutants exhibited activity similar to wild type. Interestingly, our XPF–ERCC1 preparations were found to contain a significant amount of active XPF–ERCC1 heterotetramer (Supplementary Fig. 1a, c, d). Cryo-EM data was collected for this sample, although it was not possible to obtain a reconstruction below 14 Å resolution due to intrinsic flexibility (Supplementary Fig. 1e, f). Despite this, future work will seek to address whether the XPF–ERCC1 heterodimer and heterotetramer play distinct roles in DNA repair pathways.

XPF–ERCC1 cryo-EM structures described here reveal how binding a DNA-junction substrate is able to disengage the XPF helical domain from the XPF CM and release the heterodimeric 2×(HhH)2 domain. A role for the linker regions in enabling this release is likely. The released 2×(HhH)2 domain is then able to engage a minor groove in a dsDNA duplex adjacent to the DNA ds/ss junction and packs against the XPF nuclease–ERCC1 NLD dimer, as observed for structures of Mus81–Em1 and A. pernix XPF. The repositioning of the dimeric 2×(HhH)2 domain has three consequences. First, it destabilises the autoinhibition interface with the XPF helical domain. Second, it exposes the DNA-binding surface of ERCC1 (HhH)2. Third, it enables the proper coupling of the ERCC1 ssDNA and dsDNA-binding functions by shortening the linker regions and forming a compact conformation with ERCC1–NLD–(HhH)2 domain contacts. The structures described here do not reveal the full basis for DNA-junction recognition or the extent of conformational flexing required to place the scissile bond proximal to the XPF catalytic centre. We speculate that the similarities between XPF HLM and the MDA5 helicase point to a concave surface that could engage the major groove of a DNA duplex within a DNA junction to promote movement of the ds/ssDNA discontinuity into the XPF catalytic site. Evidently further high-resolution structures are required with longer DNA substrates and recruitment partner complexes in order to fully understand how the scissile phosphodiester bond is presented to the XPF catalytic site and the extent of the conformational alterations required.

Whilst this paper was in preparation, the structure of a ds/ssDNA-bound TFIIH–XPA (PDB code: 6RO4) was published representing a 5’-NER pre-incision complex that can recruit XPF–ERCC1. Superposition of the ERCC1 (HhH)2 domain–dsDNA complex onto the exposed DNA minor groove at the TFIIH–XPA–ds-ssDNA junction (Supplementary Fig. 8b) revealed a non-overlapping complementarity in DNA binding with XPA. ERCC1 engaged precisely the available DNA elements that were not engaged by XPA (Supplementary Fig. 8a). The resulting model predicts extensive interfaces between the XPF–ERCC1 and TFIIH–XPA–DNA with few steric clashes, many of which were within the flexible XPA loop region (residues 104–131). In this model, the dimeric 2×(HhH)2 domain lies adjacent to the TFIIH subunit XPB and DNA whilst the XPF nuclease–ERCC1 NLD dimer is positioned close to XPD, XPA and DNA. The highly basic and flexible Rea2 insert one (residues 349–377) is oriented to interact with either the extended XPA helix or dsDNA. Further structural studies are required to validate such a model.

Finally, there is a pressing need to explore chemical inhibition of XPF–ERCC1 to sensitise cancer cells to platinum-based therapeutics and reduce drug resistance mediated by XPF–ERCC1. Equally, XPF-ERCC1 inhibitors could target cancer cell vulnerabilities including XPF-FANCm synthetic lethality relevant to FANCm-deficient tumours and potentially other platinum-sensitive contexts. The availability of an atomic structure for human XPF–ERCC1 described here will encourage efforts to develop new precision medicines as well as to overcome cancer chemoresistance.

Methods

XPF–ERCC1 expression, purification and complex assembly. All reagents purchased from Sigma-Aldrich unless otherwise stated. A pFastBac Dual vector containing full length, wild type, human XPF (NCBI reference sequence: NM_005236.2) and ERCC1 (NCBI reference sequence: NM_001166049.2) cDNA was modified to include a C-terminal ERCC1 5′-Strep-tag using restriction enzyme cloning. All primer sequences used in this study are shown in Supplementary Table 3. This plasmid was transformed into competent DH10 β competent Escherichia coli cells (Thermo-Fisher) and recombinant bacmid DNA purified. Recombinant baculoviruses expressing XPF and ERCC1 were generated using standard protocols (Oxford Expression Technologies). In short, 1 × 106 SF21 cells (Thermo-Fisher) grown in SF91 media (Thermo-Fisher) and 10 μg/ml gentamycin (Life Technologies) were infected at a multiplicity of infection (MOI) of 2 and harvested after 72 h. Cell pellets were resuspended in extract buffer (20 mM HEPES pH 7.8, 150 mM NaCl, 1 mM tris(2-carboxyethyl)phosphine (TCEP), 10% glycerol, 0.25 tablet of EDTA-free protease-inhibitor cocktail per litre of culture, and 1 μl per 250 ml lysate BaseMuncher (Expedeon) and lysed by sonication. The lysate was cleared of insoluble cell debris by centrifugation at 35,000g for 45 min and incubated with Strep-tactin resin (GE Healthcare) for 1 h at 4 °C. The resin was extensively washed with extract buffer minus protease inhibitors and BaseMuncher and incubated for 12 hours with Tobacco Etch Virus protease (supplier NEB). The eluate, containing XPF–ERCC1 was concentrated and loaded onto an anion-exchange column (HiTrap-Q, GE Healthcare) and XPF–ERCC1 containing fractions eluted using a gradient across 20 ml of extract buffer + 1 M NaCl before a final SEC step using a Superdex 200 column (GE Healthcare) in cryo buffer (20 mM HEPES pH 7.8, 150 mM NaCl, 1 mM TCEP, 0.01% CHAPS, 0.1% CHAPS). Mutants were cloned using the Q5 site-directed mutagenesis kit (New-England Biotech) and were then expressed using the same protocol as described above for wild-type XPF–ERCC1.

XPF–ERCC1 DNA complex assembly. DNA with a modified phosphorothioate backbone (SL+, DNA) was resuspended in DNA resuspension buffer (10 mM Tris, pH 7.4, 1 mM EDTA and 75 μM dNTPs) and annealed at 85 °C for 5 min and allowed to cool to 37 °C. The structure of the DNA duplex (SL+ DNA) was examined by gel electrophoresis, and purified DNA was amplified by PCR using 10 μM Tris, pH 7.8, 150 mM NaCl, 1 mM TCEP, 5 μM MgCl2, 0.01% 3-(d-3-chloroanilinopropyl) dimethylammonio)-1-propanesulfonate (CHAPS), 0.25 tablet of EDTA-free protease-inhibitor cocktail per litre of culture, and 1 μl per 250 ml lysate BaseMuncher (Expedeon) and lysed by sonication. The lysate was cleared of insoluble cell debris by centrifugation at 35,000g for 45 min and incubated with Strep-tactin resin (GE Healthcare) for 1 h at 4 °C. The resin was extensively washed with extract buffer minus protease inhibitors and BaseMuncher and incubated for 12 hours with Tobacco Etch Virus protease (supplier NEB). The eluate, containing XPF–ERCC1 was concentrated and loaded onto an anion-exchange column (HiTrap-Q, GE Healthcare) and XPF–ERCC1 complex eluted. After the eluate was thoroughly dialysed against cryo buffer, 20 mM HEPES pH 7.8, 150 mM NaCl, 1 mM TCEP, 0.01% CHAPS, 5 mM CaCl2, 0.5 mM TCEP and then incubated with SL+ DNA at a 1:2 protein:DNA molar ratio for 10 min at 4 °C followed by cross-linking with 0.05% (v/v) glutaraldehyde for 10 min at 4 °C. The cross-linking reaction was quenched by the addition of 1 mM Tris-HCl, pH 7.8 and the complex further purified via SEC using a Superdex 2000 column.
XPF–ERCC1 cryo-EM grid preparation and data collection. For cryo-EM analysis, 4 μl of the purified XPF–ERCC1 heterodimer at 1.5 mg/ml was applied to both R1.2/1.3 400 mesh Ultrafoil® and QuantiFoil® grids that had been previously glow discharged for 45 s at 42 mA. The grids were blotted for 4 s at 100% humidity and 4°C and plunged into liquid ethane cooled by liquid nitrogen using a FEI Vitrobot Mk IV. The grids were then loaded onto a Titan Krios transmission electron microscope at 300 kV (Thermo-Fisher). Images were collected in counting mode using a Gatan K2 Summit direct electron detector camera mounted behind a GIF Quantum energy filter operating in zero-loss mode. Exposures were 15 s, with a total dose of 63 e⁻/Å² dose-fractionated into 40 frames with a calibrated pixel size of 1.38 Å. Images were recorded with a defocus of 1.5 μm to 4 μm. A total of 15,315 micrographs were collected from three separate data collection sessions.

XPF–ERCC1 cryo-EM image processing. Movie frames were corrected for motion using MotionCor248 and contrast transfer function was estimated using CTfSift49 within Scipion250. The total number of movies used for processing was 14,453. Two-hundred micrographs were selected from the first collection from which 82,412 particles were picked using Xmipp351 semi-automated picking and extracting using RELION-352. The particles were sorted and aligned using RELION-352, sorted using Xmipp351 screen particles followed by three rounds of reference-free 2D classification in CryoSPARC-253. A subset of six 2D classes were selected that represented different views of the molecule and used as templates for reference-based particle picking using GautoMatch72 on the full dataset. This approach yielded 396,106, 1,201,881 and 2,391,908 data particles for data collection runs one, two and three, respectively. The particles were extracted and binned twofold using RELION-352, sorted using Xmipp351 to screen particles and then submitted for three rounds of reference-free 2D classification in CryoSPARC-253. This reduced the particle numbers to 151,412, 390,007 and 1,074,111 particles for data collection runs one, two and three, respectively. All 361,516 particles from the final data collection were re-extracted in an un-binned 200 × 200 pixel box using RELION-352 and e3map2star and then merged. The data then underwent 3D classification without alignment in RELION-352 to identify the highest-resolution features, comprising 405,339 particles, were refined to 4.1 Å resolution in CryoSPARC-253 using non-uniform refinement. Per-particle motion correction was carried out using Bayesian polishing in RELION-352. A mask was generated using UCSF Chimera55 that excluded the dimeric XPF–ERCC1 2×(HhHh) domain as the lowest resolution region of the map, suggesting some degree of mobility. A mask which excluded the low-resolution XPF–ERCC1 2×(HhHh) hairpins was generated in Chimera55 and using the particle subtraction tool in CryoSPARC-253 the portion of the particle images aligning to the hairpin density in the map was removed. Non-uniform local refinement in CryoSPARC-253 was performed on the subtracted particles, re-aligning them to the masked reference volume, leading to a reconstruction at 3.6 Å resolution which excluded the hairpin portion of the 4.0 Å map.

All resolutions reported here were determined by Fourier shell correlation (at FSC = 0.143) based on the "gold-standard" protocol using a soft mask around the complex density56. To avoid over-masking, the masked maps were visually inspected to exclude the possibility of clipping. In addition, the occurrence of over-masking was monitored by inspecting the shapes of FSC curves. The two-half maps had their phases randomised beyond the resolution at which the no-mask FSC drops below the FSC = 0.143 criterion. The tight mask is applied to both half maps, and an FSC is calculated. This FSC is used along with the original FSC before phase randomisation to compute the corrected FSC. Local resolution was calculated using Blocres within CryoSPARC-253. For visualisation, maps were sharpened by randomisation to compute the corrected FSC. Local resolution was calculated using an FSC is calculated. This FSC is used along with the original FSC before phase randomisation to compute the corrected FSC. Local resolution was calculated using邦res within CryoSPARC-253. The shiny, polished particles were then moved onto the Titan Krios transmission electron microscope and detector as described above. A total of 8965 movies were collected from a single data collection using the same freezing and imaging conditions used to collect the reference-free 2D classification. Movie frames were corrected for motion using MotionCor248, and contrast transfer function was estimated using Contrast Transfer Function was estimated using CTfSift49 within Scipion250. The total number of movies used for processing was 14,453. Two-hundred micrographs were selected from the first collection from which 82,412 particles were picked using Xmipp351 semi-automated picking and extracting using RELION-352. The particles were sorted and aligned using RELION-352, sorted using Xmipp351 screen particles followed by three rounds of reference-free 2D classification in CryoSPARC-253. This reduced the particle numbers to 151,412, 390,007 and 1,074,111 particles for data collection runs one, two and three, respectively. All 361,516 particles from the final data collection were re-extracted in an un-binned 200 × 200 pixel box using RELION-352 and e3map2star and then merged. The data then underwent 3D classification without alignment in RELION-352 to identify the most stable, high-resolution class. The two classes that displayed the highest-resolution features, comprising 405,339 particles, were refined to 4.1 Å resolution in CryoSPARC-253 using non-uniform refinement. Per-particle motion correction was carried out using Bayesian polishing in RELION-352. The shiny, polished particles were then refined to 4.0 Å resolution in CryoSPARC-253 using non-uniform refinement.

Inspection of the 4.0 Å resolution map rendered by local resolution in Chimera55 identified the dimeric XPF–ERCC1 2×(HhHh) domain as the lowest resolution region of the map containing some degree of mobility. A mask which excluded the low-resolution XPF–ERCC1 2×(HhHh) hairpins was generated in Chimera55 and using the particle subtraction tool in CryoSPARC-253 the portion of the particle images aligning to the hairpin density in the map was removed. Non-uniform local refinement in CryoSPARC-253 was performed on the subtracted particles, re-aligning them to the masked reference volume, leading to a reconstruction at 3.6 Å resolution which excluded the hairpin portion of the 4.0 Å map.

All resolutions reported here were determined by Fourier shell correlation (at FSC = 0.143) based on the "gold-standard" protocol using a soft mask around the complex density56. To avoid over-masking, the masked maps were visually inspected to exclude the possibility of clipping. In addition, the occurrence of over-masking was monitored by inspecting the shapes of FSC curves. The two-half maps had their phases randomised beyond the resolution at which the no-mask FSC drops below the FSC = 0.143 criterion. The tight mask is applied to both half maps, and an FSC is calculated. This FSC is used along with the original FSC before phase randomisation to compute the corrected FSC. Local resolution was calculated using邦res within CryoSPARC-253. For visualisation, maps were sharpened by applying an automated local resolution weighted negative factor using the local filtering function of CryoSPARC-253.

XPF–ERCC1 cryo-EM model building. Initially the crystal structures of the ERCC1 NLD (PDB code: 2A1I) and the tandem helix–hairpin–helix domains comprising XPF and ERCC1 chains (PDB code: 2A1I) were rigid body fitted into the locally filtered and sharpened map obtained at 4.0 Å resolution. Homology models were generated for the XPF RecA1 domain and rigid body fit into the map using the same procedure. Subsequently, the fitted domains were rebuilt manually using COOT59 optimising the fit where sidechain densities were evident prior to using FlexEM58 and real-space refinement as implemented in PHENIX58 whilst imposing secondary structural and geometric restraints to prevent overfitting (Table 4). The RecA2 and helical domains were built de novo and subjected to PHENIX58 real-space refinement. A further 6 cycles of rebuilding and refinement in COOT57 and PHENIX58 lead to a model containing 743 residues from XPF and 195 from ERCC1. Linkers regions connecting the XPF nuclease and ERCC1 NLD domains to their respective (HhH), domains were built manually into the map and the N-terminal portion of the XPF nuclease domain homology model was rebuilt in COOT57 to fit the map. The final atomic model was evaluated using MolProbity59 (Table 4). The location of patient mutations and sidechains referred to in the text are mapped onto the primary sequence, together with sequence conservation within XPF and ERCC1 homologues respectively (Supplementary Figs. 10 and 11).

XPF–ERCC1–DNA complex cryo-EM grids and data collection. XPF–ERCC1–DNA complex concentration to 1.3 mg/ml and applied to Quantifoil R1.2/1.3 300 mesh copper grids. The freezing and imaging conditions used were the same as for the DNA-free XPF–ERCC1 complex described above. A total of 9896 movies were collected from a single data collection using the same electron microscope and detector as described above.
excluded both the DNA and hairpin domain density which was used to carry out masked refinement improving the resolution of the sub-volume to 5.9 Å (Table 4).

**XPF-ERCC1-DNA complex model building.** Individual domains of XPF-ERCC1 were taken from the DNA-free structure and fitted into the DNA-bound cryo-EM map density as rigid bodies using the UCSF Chimera tensor fit-in-map tool. The homodimeric *A. pernix* XPF (PDB:2BWG) bound to dsDNA through its (HhH)2 hairpins was fitted into the DNA-bound map density and the subsequent position of the DNA-bound *A. pernix* hairpins used as a reference to align the human hairpins using MatchMaker in UCSF Chimera39. The DNA from the *A. pernix* structure was reduced to a 10 base-pair duplex and modelled into the map whilst preserving the hairpin domain–DNA contacts. The sequence conservation of the functional human ERCC1 and *A. pernix* (HhH) domains is high: 25.5% identical and 69.1% similar residues. The ds-DNA bound structure of MDA5 (PDB: 4CL2) was included into the DNA-bound map density as a guide to place the helical domain of XPF by inspecting the position of the homologous domain in MDA5.

**XPF-ERCC1-DNA-TFIIH-XPA complex modelling.** The XPF-ERCC1–DNA structure was aligned to the TFIIH–DNA–XPA structure (PDB code: 6RO4) through structural super-imposition in UCSF Chimera38 and alignment with the two DNA strands of a single duplex from each structure. The ds/ss DNA junction was defined by the high-resolution DNA structure in the TFIIH–XPA complex and demarcated by the position of the XPA β-hairpin.

**XPF-ERCC1 cross-linking mass spectrometry.** All chemicals were purchased from Sigma–Aldrich unless otherwise stated. A total of 100 µg XPF–ERCC1 heterodimer at a concentration of 1 mg/ml in 20 mM HEPES, pH 7.8, 10% Glycerol, 0.01% CHAPS, 150 mM NaCl, 1 mM TCEP, 0.5 mM EDTA was cross-linked using 1 mM DSSO in 50% minimum sulfoxide (DSSO) (Thermo-Fisher) with mild shaking for 30 min at 37 °C. The cross-linking was quenched using a final concentration of 5 mM ammonium bicarbonate for a further 20 min at 37 °C. To remove potential aggregates, gradient ultracentrifugation was employed using a 5–30% glycerol gradient in 20 mM HEPES, 150 mM NaCl, mixed using a Gradient Master (Bio–Comp), and centrifuged for 16 h at 4 °C at 200,000 g using a SW 55 Ti Rotor (Beckman Coulter)39. Totally, 100 µl fractions were collected and silver stained to identify fractions containing cross-linked non-aggregated XPF–ERCC1. Fractions containing cross-linked proteins were then pooled and buffer exchanged into 8 M urea using a Vivaspin 500, 300,000 molecular weight cut off (MWCO) PES filter (Sartorius, V50212). Cysteine reduction was carried out using 2.5 mM TCEP for 30 min at 37 °C and alkylated in the dark using 5 mM iodocacetamide at room temperature. The urea was then buffer exchanged for 50 mM ammonium bicarbonate and proteins were proteolyzed using trypsin (Promega) at 150 µg w/w trypsin: protein overnight at 37 °C. The solution was acidified using 2% formic acid and peptides were then purified by using the MWCO filter and desalted using in-house built STAGE tips made using Empore SPE C18 discs (3 M, 6883-U). The eluent was then dried to completion. Peptides were reconstituted in 0.1% trifluoroacetic acid (TFA) and chromatographically resolved using an Ultimate 3000 RSLCnano (Dionex) HPLC. Peptides were first loaded onto an Acclaim PepMap 100 C18, 3 µm particle size, 100 Å pore size, 20 mm × 75 µm ID (Thermo Scientific, 164353) trap column using a loading buffer (2% acetonitrile (MeCN) and 0.05% TFA in 97.05% H2O) with a flow rate of 7 µl/min. Chromatographic separation was achieved using an EASY-Spray column, PepMap C18, 2 µm particles, 100 Å pore size, 500 mm × 75 µm ID (Thermo Scientific, ES803). The gradient utilised a flow of 0.3/30µl/min from mobile A (0.1% formic acid, 5% dimethyl sulfoxide (DMSO) in H2O) and 2% mobile B (0.1% formic acid, 75% MeCN, 5% DMSO and 19.9% H2O). After 6 min, mobile B was increased to 30% over 69 min, to 45% over 30 min, further increased to 90% in 16 min and held for 4 min. Finally, Mobile B was reduced back to 5% over 1 min for the rest of the acquisition. Data were acquired in real time over 140 min using an Orbitrap Fusion Lumos Tribrid mass spectrometer in positive, top speed mode with a cycle time of 5 s. The chromatogram (MS1) was captured using 60,000 resolution, a scan range of 375–1500 with a 50 ms maximum injection time, and 4e5 AGC target. Dynamic exclusion with repeat count 2, exclusion duration of 30 s, 20 ppm tolerance window was used, along with isotope exclusion, a minimum intensity exclusion of 2e4, charge state inclusion of 3–8 ions and peptide mono isotopic precursor selection. Precursors within a 1.6 m/z isolation window were then fragmented using 25% normalised CID, 100 ms maximum injection time and 5e4 AGC target. Scans were recorded using 30,000 resolution in centroid mode, with a scan range of 120–2000 m/z. Spectra with a mass difference Δm,m/z of 0.091721 Da were further fragmented with a 30% normalised higher collision dissociation, using a 2 m/z isolation window, 150 ms maximum injection time and 2e4 AGC target. Four scans were recorded using an ion trap detection in rapid mode starting at 120 m/z.

**XL-MS data analysis.** Data processing were carried out using Proteome Discoverer Version 2.4 (Thermo Scientific) with the XlinkX40 node where the minimum XlinkX score was set to 63. The acquisition strategy was set to MS2_MS3 mode. The database comprised solely of the specific XPF and ERCC1 sequences. Trypsin was selected as the proteolytic enzyme allowing up to two missed cleavages with a minimal peptide length of five residues. Masses considered were in the range of 300–10000 Da. The precursor mass tolerance, FTMS fragment mass tolerance, and FTMS Fragment Mass Tolerance were set to 10 ppm, 10 ppm and 0.6 Da, respectively. A static carbamylation (+57.021 Da) modification was utilized for cysteine residues, with additional dynamic modifications considered including: amidated and hydroxyl DSSO (+142.050 and +176.014 Da, respectively) on lysine serine and threonine residues, oxidation (+15.995 Da) on methionine residues, and protein N-terminal acetylation (+42.011 Da). The FDR threshold was set to one with the strategy set to simple. The list of reported cross-linked spectral matches were manually examined and cross-links with spectra that did not contain acceptable b and y ion coverage were excluded. We note that this method requires accessible lysine sidechains therefore predominately hydrophobic interfaces, such as the Rcc4–nucleic, did not return any cross-links50. A number of cross-links were observed that exceed the permitted the Ca–Cα cut-off distance of 30 Å.

**XPF-ERCC1–SLX4 homodimer complex assembly.** cDNA encoding the SLX4PTD (residues 1–758) (NCBI reference sequence: NM_032444) was shuttled into a pGEX-1 vector (Sigma). Recombinant baculoviruses expressing the SLX4PTD were generated as previously described and used to infect 1 × 10⁶ SF21 cells (Thermo-Fisher) grown in SFIII media (Thermo-Fisher) and 10 µg/ml gentamycin (Life Technologies) at an MOI of 0.5. These cells were co-infected with XPF–ERCC1 expressing baculovirus at an MOI of 2. Cells were pelleted after 72 h and protein extracted as previously described for XPF–ERCC1. Following Strep-tactin affinity purification, the complex was purified using anion-exchange (HiTrap Q, GE Healthcare) using a gradient of 150 mM NaCl to 500 mM NaCl over 20 ml of extract buffer minus protease inhibitors and BaseMuncher. This separated the SLX4NTD–XPF–ERCC1 complex from unbound XPF–ERCC1. Following the purification step, the SLX4NTD–XPF–ERCC1 complex was pooled and concentrated prior to a final SEC step using a Superose-6 increase column equilibrated in extract buffer minus protease inhibitors and BaseMuncher (GE Healthcare). Fractions containing both SLX and XPF were identified via Western blot.

**Real-time fluorescence incision assay.** Fluorescently labelled stem-loop (SL) DNA substrates, containing a 5′ 6-FAM fluorophore and 3′ BHQ1 quench, were purified by SEC (Superdex 200i, GE Healthcare) in assay buffer (5 mM HEPES, 10% glycerol, 0.5 mM DTT, 1 mM MnCl2 and 40 mM NaCl. The purified substrates were then annealed by heating to 95 °C for 1 min followed by cooling to 4 °C and dispensed into the assay plate. Reactions were carried out in 384-well black, flat-bottomed microtiter plates (Corning 3854). Purified XPF–ERCC1 was buffer exchanged into assay buffer and 5 nM added to each in a total volume of 20 µl to initiate the endonuclease reaction. Fluorescence measurements were carried out using the CLARIOstar plate reader (BMG Labtech) using an excitation wavelength of 483 nm and an emission wavelength of 525 nm. Sixty readings were collected at 30-s intervals and the linear response range for each substrate was used to determine the change in fluorescence per unit time. Kinetic parameters were calculated using the Michaelis–Menten equation. Experimental product release was quantified by plotting the relative fluorescence units produced by known amounts of the cleavage products against their concentration to generate a standard curve. 

**References**

1. Dehe, P. M. & Gaillard, P. H. Control of structure-specific endonucleases to maintain genome stability. *Nat. Rev. Mol. Cell Biol.* 18, 315–330 (2017).
2. Ciccia, A., McDonald, N. & West, S. C. Structural and functional relationships of the XPF/MUS81 family of proteins. *Annu Rev. Biochem.* 77, 259–287 (2008).
3. Faridounnia, M., Folkins, G. E. & Boelens, R. Function and interactions of ERCC1–XPF in DNA damage response. *Molcules* 23, E3205 (2018).
6. Wyatt, H. D. M., Laister, R. C., Martin, S. R., Arrowsmith, C. H. & West, S. C. Nature Commun. 36, 1–9 (2020).
7. Ahmad, A. et al. ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Mol. Cell 28, 5082–5092 (2008).
8. Jaspers, N. G. J. et al. First reported patient with human ERCC1 deficiency. J. Hum. Genet. 92, 800–806 (2019).
9. Jaspers, N. G. J. et al. First reported patient with human ERCC1 deficiency. J. Hum. Genet. 92, 800–806 (2019).
10. Ahmad, A. et al. ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Mol. Cell 28, 5082–5092 (2008).
11. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 NUCLEIC ACIDS RES. 36, 1–9 (2007).
12. Ahmad, A. et al. ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Mol. Cell 28, 5082–5092 (2008).
13. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 NUCLEIC ACIDS RES. 36, 1–9 (2007).
14. Ahmad, A. et al. ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Mol. Cell 28, 5082–5092 (2008).
15. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 NUCLEIC ACIDS RES. 36, 1–9 (2007).
16. Ahmad, A. et al. ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Mol. Cell 28, 5082–5092 (2008).
17. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 NUCLEIC ACIDS RES. 36, 1–9 (2007).
18. Ahmad, A. et al. ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Mol. Cell 28, 5082–5092 (2008).
19. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 NUCLEIC ACIDS RES. 36, 1–9 (2007).
20. Ahmad, A. et al. ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Mol. Cell 28, 5082–5092 (2008).
21. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 NUCLEIC ACIDS RES. 36, 1–9 (2007).
22. Ahmad, A. et al. ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Mol. Cell 28, 5082–5092 (2008).
23. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 NUCLEIC ACIDS RES. 36, 1–9 (2007).
24. Ahmad, A. et al. ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Mol. Cell 28, 5082–5092 (2008).
25. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 NUCLEIC ACIDS RES. 36, 1–9 (2007).
26. Ahmad, A. et al. ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Mol. Cell 28, 5082–5092 (2008).
27. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 NUCLEIC ACIDS RES. 36, 1–9 (2007).
28. Ahmad, A. et al. ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Mol. Cell 28, 5082–5092 (2008).
29. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 NUCLEIC ACIDS RES. 36, 1–9 (2007).
30. Ahmad, A. et al. ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Mol. Cell 28, 5082–5092 (2008).
31. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 NUCLEIC ACIDS RES. 36, 1–9 (2007).
32. Ahmad, A. et al. ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Mol. Cell 28, 5082–5092 (2008).
33. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 NUCLEIC ACIDS RES. 36, 1–9 (2007).
34. Ahmad, A. et al. ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Mol. Cell 28, 5082–5092 (2008).
35. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 NUCLEIC ACIDS RES. 36, 1–9 (2007).
36. Ahmad, A. et al. ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Mol. Cell 28, 5082–5092 (2008).
37. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 NUCLEIC ACIDS RES. 36, 1–9 (2007).
38. Ahmad, A. et al. ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Mol. Cell 28, 5082–5092 (2008).
39. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 NUCLEIC ACIDS RES. 36, 1–9 (2007).
40. Ahmad, A. et al. ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Mol. Cell 28, 5082–5092 (2008).
41. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 NUCLEIC ACIDS RES. 36, 1–9 (2007).
42. Ahmad, A. et al. ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Mol. Cell 28, 5082–5092 (2008).
43. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 NUCLEIC ACIDS RES. 36, 1–9 (2007).
44. Ahmad, A. et al. ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Mol. Cell 28, 5082–5092 (2008).
45. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 NUCLEIC ACIDS RES. 36, 1–9 (2007).
46. Ahmad, A. et al. ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Mol. Cell 28, 5082–5092 (2008).
47. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 NUCLEIC ACIDS RES. 36, 1–9 (2007).
48. Ahmad, A. et al. ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Mol. Cell 28, 5082–5092 (2008).
49. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 NUCLEIC ACIDS RES. 36, 1–9 (2007).
50. Ahmad, A. et al. ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Mol. Cell 28, 5082–5092 (2008).
51. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 NUCLEIC ACIDS RES. 36, 1–9 (2007).
52. Ahmad, A. et al. ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Mol. Cell 28, 5082–5092 (2008).
53. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 NUCLEIC ACIDS RES. 36, 1–9 (2007).
54. Ahmad, A. et al. ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Mol. Cell 28, 5082–5092 (2008).
55. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 NUCLEIC ACIDS RES. 36, 1–9 (2007).
56. Ahmad, A. et al. ERCC1-XPF endonuclease facilitates DNA double-strand break repair. Mol. Cell 28, 5082–5092 (2008).
57. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 NUCLEIC ACIDS RES. 36, 1–9 (2007).
58. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 NUCLEIC ACIDS RES. 36, 1–9 (2007).
59. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 NUCLEIC ACIDS RES. 36, 1–9 (2007).
60. Fairman-Williams, M. E., Guenther, U. P. & Jankowsky, E. SF1 and SF2 NUCLEIC ACIDS RES. 36, 1–9 (2007).
Texas) for stimulating and encouraging the project in its early phases. We acknowledge the helpful advice and support of Peter Rosenthal on all aspects of cryo-electron microscopy. M.J. was funded by a Crick/UCL joint PhD studentship. N.Q.M. acknowledges that this work was supported by the Francis Crick Institute, which receives its core funding from Cancer Research UK (FC001115), the UK Medical Research Council (FC001115) and the Wellcome Trust (FC001115). M.L. is supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre. E.P.M. and F.B. acknowledge support from Cancer Research UK (C12209/A16749).

Author contributions
M.J. cloned and purified the XPF–ERCC1 complexes, processed all cryo-EM data and built the atomic models. M.J. and F.B. carried out the EM grid preparations. F.B., M.J. and C.E. screened the EM grids, A.N. ran the data collections. F.B. and E.P.M. assisted the XPF–ERCC1 single particle reconstruction and interpretation. M.B. assisted with XPF–ERCC1 biochemistry, D.B. assisted in model refinement and validation. A.B. and A.S. carried out the XLMS and data analysis. M.L. contributed to the study design. M.J. and N.Q.M. designed the study, interpreted the results and wrote the paper.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information is available for this paper at https://doi.org/10.1038/s41467-020-14856-2.

Correspondence and requests for materials should be addressed to N.Q.M.

Peer review information Nature Communications thanks Gang Cai and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020