

# Molecular architecture and functional dynamics of the pre-incision complex in nucleotide excision repair

Received: 27 March 2024

Accepted: 19 September 2024

Published online: 01 October 2024

 Check for updatesJina Yu <sup>1,2</sup>, Chunli Yan<sup>1,2</sup>, Tanmoy Paul<sup>1,2</sup>, Lucas Brewer <sup>1,2</sup>, Susan E. Tsutakawa <sup>3</sup>, Chi-Lin Tsai <sup>4</sup>, Samir M. Hamdan <sup>5</sup>, John A. Tainer <sup>3,4</sup>  & Ivaylo Ivanov <sup>1,2</sup> 


Nucleotide excision repair (NER) is vital for genome integrity. Yet, our understanding of the complex NER protein machinery remains incomplete. Combining cryo-EM and XL-MS data with AlphaFold2 predictions, we build an integrative model of the NER pre-incision complex (PInC). Here TFIIH serves as a molecular ruler, defining the DNA bubble size and precisely positioning the XPG and XPF nucleases for incision. Using simulations and graph theoretical analyses, we unveil PInC's assembly, global motions, and partitioning into dynamic communities. Remarkably, XPG caps XPD's DNA-binding groove and bridges both junctions of the DNA bubble, suggesting a novel coordination mechanism of PInC's dual incision. XPA rigging interlaces XPF/ERCC1 with RPA, XPD, XPB, and 5' ssDNA, exposing XPA's crucial role in licensing the XPF/ERCC1 incision. Mapping disease mutations onto our models reveals clustering into distinct mechanistic classes, elucidating xeroderma pigmentosum and Cockayne syndrome disease etiology.

Nucleotide excision repair (NER) is a genome maintenance pathway critical for human health. NER repairs a vast array of structurally unrelated DNA lesions caused by ultraviolet radiation, reactive oxygen species, environmental carcinogens, and chemotherapeutic agents such as cis-platinum<sup>1–4</sup>. The variety of lesions processed by NER has led to the evolution of a complex protein machinery<sup>5</sup> that performs damage recognition, DNA unwinding, damage verification, precise dual incision of lesioned DNA, and gap-filling DNA synthesis. Despite numerous biochemical and genetic studies, knowledge of the inner workings of this machinery remains fragmentary. Predictably, functional impairment of NER complexes by mutations causes severe human genetic diseases<sup>6–11</sup>—xeroderma pigmentosum (XP) and Cockayne syndrome (CS). The vastly divergent disease phenotypes range from normal development with an extreme cancer predisposition to developmental abnormalities, accelerated neurodegeneration, and

premature ageing without cancer predisposition. Yet, a molecular basis for this striking heterogeneity in clinical outcomes has not emerged.

There are two NER sub-pathways—global genome NER (GG-NER) and transcription-coupled NER (TC-NER). GG-NER is transcription-independent and broadly probes the genome for damage. By contrast, TC-NER is specialized in bulky lesion removal from actively transcribed DNA<sup>12–14</sup>. The two sub-pathways differ in the early damage recognition steps but converge prior to lesion scanning and incision. Therefore, here we focus exclusively on the GG-NER sub-pathway. In GG-NER, the lesion recognition factor XPC aided by human Rad23B and Centrin 2 (CETN2)<sup>15–17</sup> scans the genome for DNA damage and in a twist-to-open mechanism exposes the lesion, creating a nascent DNA bubble<sup>18–21</sup>. Afterward, transcription factor IIIH (TFIIH)<sup>11,22–29</sup> is recruited to the damaged site, forming a pre-unwound NER complex<sup>15,16</sup>. TFIIH contains

<sup>1</sup>Department of Chemistry, Georgia State University, Atlanta, GA, USA. <sup>2</sup>Center for Diagnostics and Therapeutics, Georgia State University, Atlanta, GA, USA.

<sup>3</sup>Molecular Biophysics and Integrated Bioimaging, Lawrence Berkeley National Laboratory, Berkeley, CA, USA. <sup>4</sup>Department of Molecular and Cellular Oncology, The University of Texas MD Anderson Cancer Center, Houston, TX, USA. <sup>5</sup>Bioscience Program, Biological and Environmental Sciences and Engineering Division, King Abdullah University of Science and Technology (KAUST), Thuwal, Saudi Arabia.  e-mail: [jtainer@mdanderson.org](mailto:jtainer@mdanderson.org); [iivanov@gsu.edu](mailto:iivanov@gsu.edu)

ten protein subunits – seven forming core TFIIH (XPB, XPD, p44, p34, p8, p62, and p52) and three comprising the CAK complex (MAT1, Cdk7 and Cyclin H)<sup>30</sup>. While CAK is key for TFIIH's function in transcription, its dissociation from core TFIIH is required for functional NER. Two subunits, XPB and XPD, are DNA translocases<sup>31</sup> with key functional roles in DNA remodeling during NER. Assisted by XPA<sup>32–34</sup>, XPB first unwinds the DNA duplex upstream of the lesion, expanding the NER bubble<sup>35</sup>. XPD then verifies the lesion<sup>31,34,36</sup> by scanning the damage-containing DNA strand. Since NER accuracy is paramount for genome integrity, all these dynamic reversible transformations must be completed before actual lesion removal<sup>37</sup>. DNA damage removal relies on the coordinated action of two structure-specific endonucleases, XPG and XPF/ERCC1<sup>38–44</sup>. XPG is a member of the FEN1 (flap endonuclease 1) family of nucleases<sup>38</sup>. The enzyme recognizes the 3' junction of the NER bubble and cleaves the damaged strand one nucleotide into the DNA duplex. Correspondingly, XPF belongs to the Mus81 nuclease family and functions as a stable heterodimer with the ERCC1 protein<sup>43</sup>. The XPF/ERCC1 complex acts on the 5' junction, making an incision 3–5 nucleotides into the DNA duplex. Strand incision constitutes a point of no return<sup>4</sup> and commits the NER machinery to complete the repair process. Failure to conclude repair post-incision has dire consequences for the cell as the resulting abortive intermediates are more mutagenic or lethal than the original DNA damage. To prevent unlicensed incisions, NER relies on an intricately orchestrated assembly of a pre-incision complex (PlnC)<sup>4,5,42</sup> comprised of TFIIH, XPG, XPF/ERCC1, XPA, and RPA. Within PlnC, the coordinated action of XPG and XPF/ERCC1 on both sides of the lesion results in a ~27-nucleotide ssDNA gap substrate<sup>45</sup>. Subsequently, Pol $\delta$ , RFC, and PCNA are loaded at the 5' junction of this substrate to initiate gap-filling DNA synthesis and restore the excised region. Lastly, DNA ligase seals the nicked DNA to complete the repair process. Correct assembly of the PlnC is critical for the accuracy of the dual incision process and, therefore, vital to avoid genomic injury.

Here we have used integrative molecular modeling to synthesize available structural data on NER constituent proteins and sub-assemblies and combined it with biochemistry and cross-linking mass spectrometry (XL-MS) restraints to create a practically complete structural model of the human NER pre-incision complex. The new model enables comprehensive side-by-side structural comparison to the previously determined lesion scanning complex (LSC), thus shedding light on the reorganization of the NER protein machinery from the middle through the late stages of the pathway. The model also serves as a starting point for microsecond-timescale molecular dynamics (MD) simulations, which reveal PlnC's global motions. We also employ graph-theoretical algorithms to partition the assembly into dynamic communities and network analysis to reveal the key differences between the functional dynamics of the pre-incision and lesion-scanning complexes. Importantly, mapping of disease mutations onto our network models allows us to cluster the mutants into distinct mechanistic classes impacting DNA binding, protein stability, and dynamics at PlnC protein interfaces. Thus, our integrative modeling provides a roadmap for future experiments to test the interplay between the structural disruption of NER complexes by mutations and the etiology of devastating genetic diseases. Collectively, our results unveil how the NER machinery dynamically reshapes itself and self-regulates to achieve precise lesion removal and preserve genome integrity.

## Results

### TFIIH serves as a molecular ruler defining the NER bubble size and PlnC's overall organization

We synthesize diverse structural data to create an integrative model of the most crucial intermediate in NER—the pre-incision complex. To construct the model, we systematically evaluated the cryo-EM structures and densities of human apo-TFIIH<sup>22</sup>, TFIIH/XPA/DNA<sup>46</sup>, XPF/ERCC1<sup>43</sup>, and XPA/ERCC1<sup>47</sup> along with the X-ray structures of the XPG catalytic core<sup>38</sup>

and RPA–ssDNA (RPA70, RPA32 and RPA14)<sup>48</sup>. Missing unmodelled regions that could be traced in the respective EM densities were built de novo. Predicted folded protein regions not resolved in the experimental structures were modeled with AlphaFold2<sup>49,50</sup>. Inclusion of these structured regions was necessary for the molecular dynamics simulations to unravel the functional dynamics of the pre-incision complex. Positioning of the newly modeled modules and protein domains within the PlnC assembly was guided by existing mutational and cross-linking mass spectrometry (XL-MS) data<sup>46</sup> on NER protein interfaces and by AlphaFold2-multimer predictions. Our combined analysis yields a defined structural model of the pre-incision complex (Fig. 1 and Supplementary Movie 1) and provides unanticipated insights into the functional interactions of its constituent proteins.

The pre-incision complex is organized around transcription factor IIH (TFIIH)<sup>11,22–27,30</sup>, which serves as a platform for the assembly and reorganization of the NER machinery. TFIIH's adaptable modular architecture underpins its multiple cellular functions in GG-NER, TC-NER, and transcription initiation. While the structure of the pre-wound NER complex (TFIIH/XPC/HRAD23/CETN2/DNA)<sup>15,16</sup> closely resembles the open subunit arrangement of apo-TFIIH<sup>22</sup>, the PlnC features a closed circular arrangement of the seven core TFIIH subunits (Fig. 1a). The two translocase subunits, XPB and XPD, are bound to a DNA bubble substrate and share an extended interface braced by contacts with the p44 subunit. This interface is essential for the sequential coordination of the XPB and XPD activities during the transitions of the NER protein machinery from DNA unwinding to lesion scanning and then to strand incision<sup>37</sup>. The closed circular TFIIH architecture imparts structural rigidity, making TFIIH into a stable platform for the assembly of XPF/ERCC1, XPG, XPA, and RPA (Fig. 1a).

TFIIH also serves as a molecular ruler defining the spatial extent of the NER bubble (Fig. 1a and c) and the relative positioning of the XPG and XPF nucleases (Fig. 1a). The DNA bubble region of PlnC has 23 nucleotides between the 3' and 5' junctions, matching the most probable length (27 nucleotides) of the excision products (Supplementary Note 1). Thus, our model explains the remarkable precision of the NER dual incision<sup>51</sup> as observed in NER gel assays.

In turn, the small size of the DNA bubble dictates the exceedingly compact spatial arrangement of XPF/ERCC1, XPG, XPA, and RPA in the PlnC. Notably, the XPF/ERCC1 complex is positioned on the anterior side of TFIIH's XPD subunit, capping the 5' DNA junction just above the XPB–XPA interface. Correspondingly, XPB binds the duplex leading into the 5' junction and accommodates dsDNA in a groove between its RecA-like ATPase domains. In this orientation, the XPF nuclease active site is poised for incision into the 5' duplex. Conversely, our PlnC model places the XPG catalytic core at the extreme opposite end of the NER bubble. The XPG core binds the 3' junction, facing the posterior side of XPD. RPA inserts itself between the XPG core and the XPA DNA-binding domain and protects the undamaged DNA strand. Remarkably, the spatial extent of XPG between its catalytic core and the XPG anchor domain (residues 157–296) is such that the nuclease spans the entire distance between the 3' and the 5' DNA junctions (Fig. 1a and b). Thus, XPG stretches along the XPD–RPA interface and caps XPD's DNA-binding groove. A critical component of the PlnC, XPA extends from the RPA trimer core to TFIIH's p8 subunit, making extensive contacts with DNA, XPD, XPB, and XPF/ERCC1. XPA also provides a  $\beta$ -hairpin, which separates the two DNA strands at the 5' junction<sup>46</sup>. The damaged strand passes through XPD while the non-damaged strand is directed toward the XPA zinc-finger domain and RPA. Notably, our integrative model shows that XPG and XPF/ERCC1 can both be accommodated in catalytically competent orientations, providing a structural basis for the coordinated dual incision activity of PlnC.

### XPG is positioned for incision at the 3' DNA junction of PlnC

We first modeled the XPG nuclease catalytic core<sup>41</sup> on the back side of XPD poised for incision at the 3' ss/dsDNA junction. XPG is a member

of the 5' nuclease family that includes FEN1, EXO1, and GEN1. A defining feature in these enzymes is a two-helix gateway or Arch domain, by which the nuclease selects for a single-stranded DNA feature in its substrate. In FEN1, it “threads” DNA flaps with free ends through a “needle” hole formed by this gateway and a helical cap<sup>52</sup>. Like FEN1, XPG harbors a helical arch (Fig. 2a and b) comprised of two gateway helices, GH1 (residues 33–41) and GH2 (residues 82–129). GH2 also forms an extended coiled coil with a capping helix CH (residues 734–763) as predicted by AlphaFold2 calculations. Unlike FEN1, XPG acts on a bubble substrate. This imposes a topological restriction on the gateway helices interacting with DNA, making it impossible to ‘thread’. Thus, it is key that we could model the single-stranded DNA slotting between the AlphaFold2-predicted XPG gateway helices (Fig. 2a and b). In this positioning, the coiled-coil region clamps down on top of the lesion-containing strand.

**XPG is poised to facilitate DNA strand separation during XPD unwinding**

XPG’s conserved structural elements (Fig. 2b and Supplementary Fig. 1) are ideally positioned not just for incision<sup>41</sup> but also for strand separation. A hallmark of 5' nucleases is that DNA binding is dominated by duplex DNA interaction. Thus, in our model, XPG engages the 3' duplex of the NER bubble substrate via a conserved electrostatically compatible surface suitable to recognize the helical features of dsDNA (Supplementary Fig. 2). The XPG active site is situated in a groove near the center of enzyme (Fig. 2b). On one side of the active site, ssDNA is perched on a K<sup>+</sup> ion coordinated by the helix-2-turn-helix (H2TH) motif (residues 848–880). DNA binding is reinforced through XPG-specific contacts to an  $\alpha$ 12b motif (residues 912–918) through a basic patch (K913, K916, and K917). On the other side of the active site, a hydrophobic wedge motif (residues 31–67) and a  $\beta$ -hairpin motif (residues 820–836) form a positively-charged groove, expanding the DNA-binding surface (Fig. 2b and Supplementary Fig. 2). Relevant to both XPG nuclease activity and XPD translocase activity, the

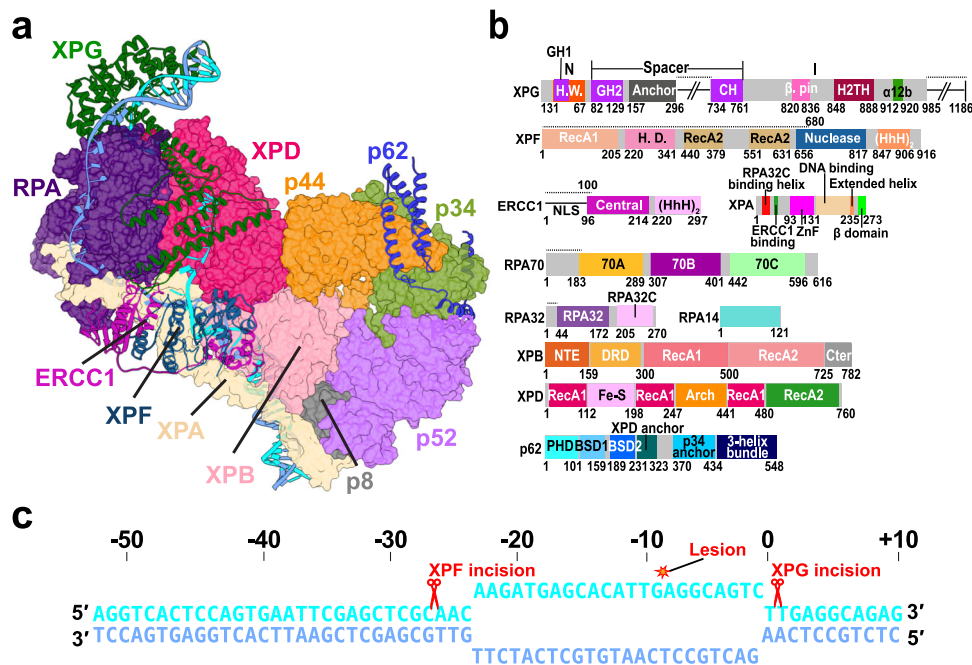
hydrophobic wedge and  $\beta$ -hairpin motifs ensure placement of the 3' junction with a  $\sim 100^\circ$  DNA bending angle. Based on work with FEN1, this substrate is validated by having single-stranded DNA between the gateway helices and junction DNA at the hydrophobic wedge prior to incision<sup>41,53</sup>. Importantly, in this position the XPG hydrophobic wedge and  $\beta$ -hairpin motifs could act as a helicase pin to facilitate strand separation during XPD’s DNA unwinding activity.

In our PinC model, we maintained the gateway arch angle with the XPG core. In this arrangement, the core makes minimal contacts with TFIIH. We also present an alternative model (Supplementary Fig. 3) where a modest adjustment to the gateway arch orientation allows the XPG core to tilt and make contacts with the XPD Arch domain. As the TFIIH/XPA/DNA structure was obtained without a true bubble substrate, no such contacts are present in the XL-MS data. Yet, the possibility cannot be excluded as XPD was shown to stimulate XPG incision after completion of lesion scanning<sup>54</sup>.

Bulky lesion blockage within XPD serves as the basis for lesion verification in NER. XPG cleaves DNA with modest variability, typically 5–8 nucleotides from a bulky lesion<sup>51,55</sup>. In our models, such blockage occurs 8 nucleotides from the 3' junction at a constriction in the XPD’s DNA-binding groove proximal to His135 (Fig. 2f) and is accompanied by closure of the Fe-S and Arch domain spacing around the lesion. Thus, we posit a mechanism where termination of XPD’s unwinding activity enables XPG repositioning for productive 3' incision.

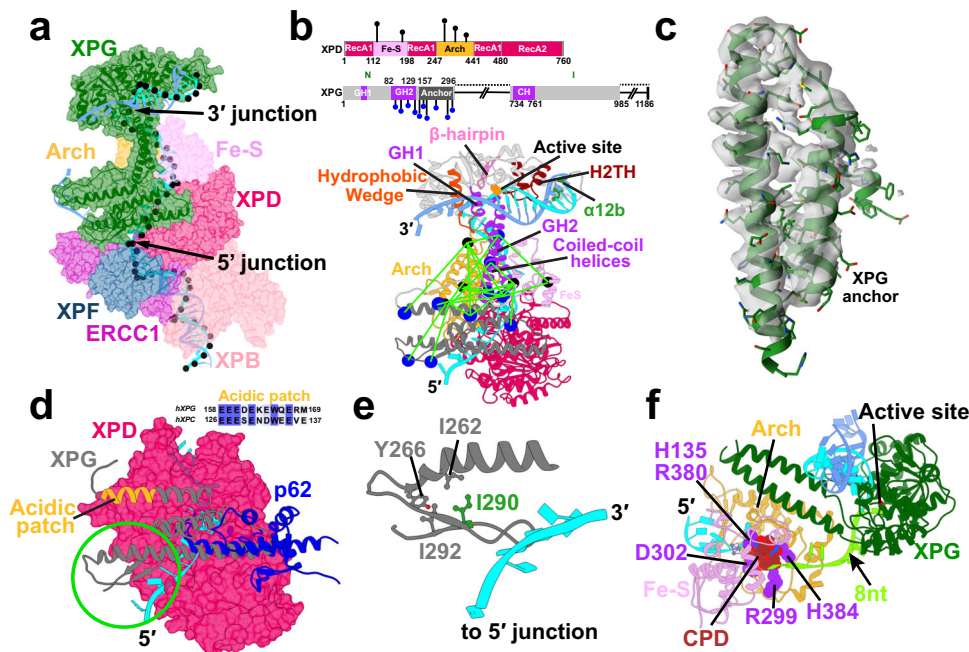
**XPG bridges the two DNA junctions offering an unexpected 5' incision sensing mechanism**

The XPG nuclease has a highly unusual architecture<sup>41</sup>—its catalytic core is formed by an N-terminal (N) and internal (I) nuclease domains, which are proximal in structure but distal in sequence, separated by  $\sim 600$ -residue insertion denoted as the “spacer region” or R-domain (Fig. 2b). The I- and N-domains exhibit high conservation to flap endonuclease 1<sup>52</sup> and other members of the FEN1 nuclease family (Supplementary Fig. 1). By contrast, the spacer region is not conserved with other FEN1



**Fig. 1 | Integrative model of the PinC unveils the overall structural organization of the assembly.** **a** View of the PinC assembly colored by subunits. XPG, XPF/ERCC1, p62, and DNA are shown in cartoon representation. TFIIH, XPA, and RPA are shown in surface representation. The lesion containing DNA strand is shown in cyan; the undamaged strand is shown in blue. **b** Domain organization of the PinC constituent proteins XPG, XPF, ERCC1, XPA, RPA, XPB, XPD, and p62 mapped onto

their respective sequences. Abbreviations denote H.W.—hydrophobic wedge; GH—gateway helix; CH—coiled-coil helix; H2TH—helix-2-turn-helix; H.D.—helicase domain; HhH—helix-hairpin-helix; DRD—damage recognition domain; NTE—N-terminal extension. **c** Schematic showing the DNA substrate of PinC, the length of the NER bubble, and the positions of the lesion site (red star) and the two incision sites (red scissor symbols).



**Fig. 2 | The XPG nuclease bridges the opposing ends of the NER bubble, providing an unexpected 5' incision sensing mechanism.** **a** View of XPG at the 3' DNA junction. XPG is depicted as cartoon and transparent surface and colored in dark green. The path of the DNA through TFIIH's XPD subunit is shown by black spheres placed at the phosphorus atoms. **b** Crosslinks between XPG and XPD mapped onto the PlnC structure. The XPD Arch and Fe-S domains are colored in orange and purple, respectively. XPG and XPD residues participating in the crosslinks are depicted as blue and black spheres, respectively. A schematic of the crosslinks mapped onto XPD's and XPG's sequences is shown above. Conserved structural motifs in the XPG catalytic core are shown; the two XPG coiled-coil helices are shown in purple; the XPG anchor domain is shown in gray. **c** The XPG anchor

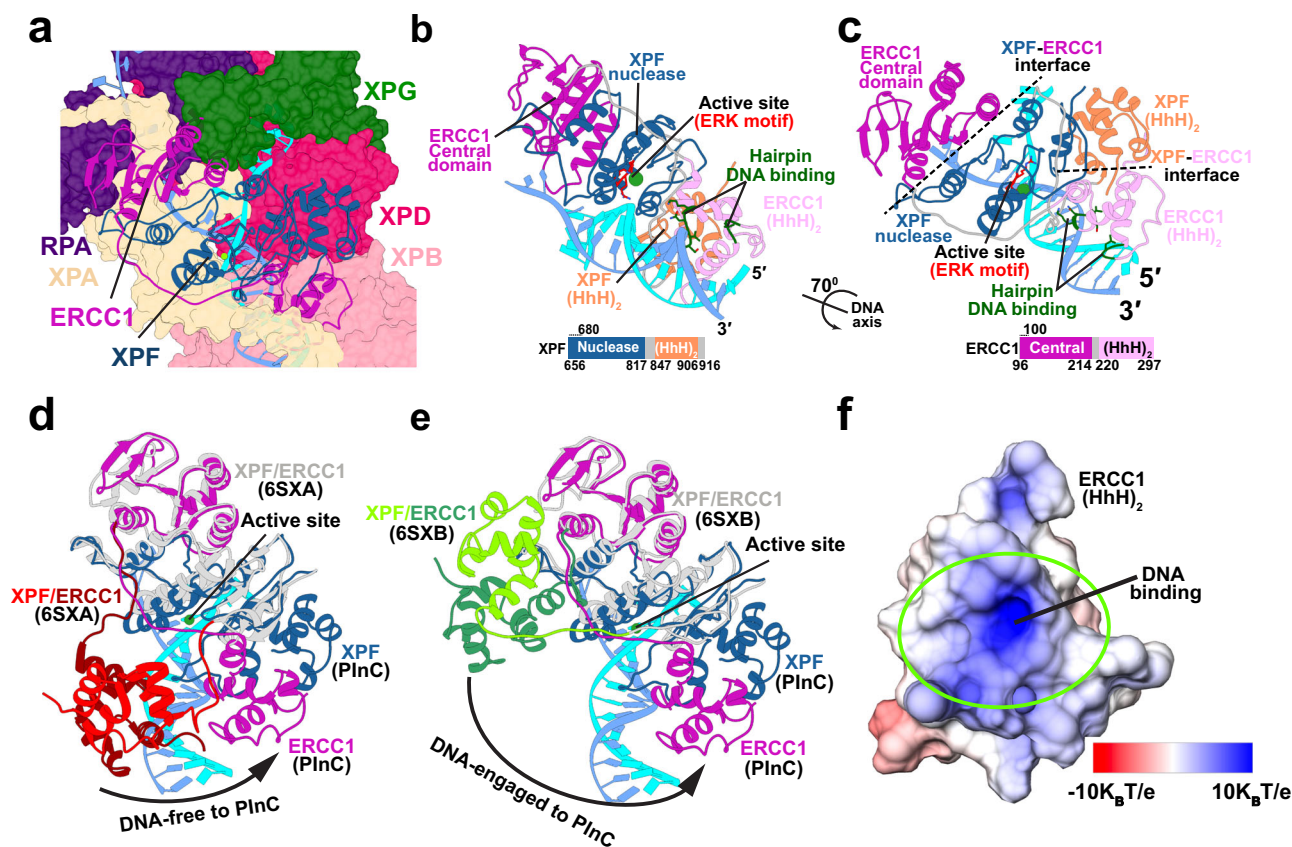
domain fitted into a segment from the TFIIH/XPA/DNA cryo-EM density (EMDB accession code: EMD-4970). **d** Overlay of AlphaFold2 docking models of the XPG-anchor and the p62 XPD-anchor with XPD. The  $\beta$ -hairpin, packed against the XPG-anchor domain is highlighted by a green circle. The acidic patch of XPG is highlighted in orange. The inset displays the sequence alignment of the acidic patches of human XPG and XPC. **e** Zoomed-in view of the  $\beta$ -hairpin interacting with the XPD anchor and ssDNA at the 5' junction. A hydrophobic cluster involving residue I290 colored in green. I290N substitution is a recognized XP/CS disease mutant. **f** A CPD lesion blocked inside XPD's DNA-binding groove near His135 and 8 nucleotides away from the 3' junction.

family enzymes, suggesting it evolved to fulfill functional requirements specific to XPG. The uniquely extended spacer region has been shown biochemically as important in establishing XPG's specificity for ssDNA-dsDNA junctions<sup>56</sup>. Additionally, structured elements of the spacer mediate protein interactions essential for NER progression: e.g., XPG contacts with TFIIH's XPD, XPB, p62, and p44 subunits as well as contacts with the RPA DNA-binding domains<sup>46,57-60</sup>. No experimental structures are available for any part of the spacer region. Therefore, we used AlphaFold2 to determine predicted folded segments within the XPG spacer region (Fig. 2). As there is no substrate-bound XPG structure, the XPG-bound DNA was based on the FEN1-bound substrate structure (PDB ID:SUM9)<sup>61</sup>.

The TFIIH/XPA/DNA cryo-EM structure<sup>46</sup> was pivotal for our hybrid modeling as it offered an unprecedented molecular view of the interactions of the core TFIIH subunits during lesion scanning. While XPG was not modeled in the electron density, the study provided valuable cross-linking data<sup>46</sup>. The XL-MS data suggests XPG remains flexibly tethered on the back side of XPD close to the ssDNA exit point between the XPD Arch and Fe-S domains. By carefully examining all XPG-TFIIH crosslinks and using AlphaFold2<sup>49,50</sup> to guide the modeling protocol, we were able to position all structured segments of XPG relative to TFIIH. Thus, placing the XPG coiled-coil helices between the Fe-S and Arch domains of XPD (Fig. 2b) is consistent with the cross-linking data<sup>46</sup>. AlphaFold2 also identified a previously unrecognized helical bundle domain of XPG (residues 157-296), which serves to anchor XPG to XPD and displace the p62 XPD-anchor and BSD2 domains. Remarkably, we were able to perfectly fit this domain into the cryo-EM density of the TFIIH/XPA/DNA complex (Fig. 2c), matching even the side chain densities. In Fig. 2d we show an overlay of

XPG-XPD and p62-XPD models generated with AlphaFold2-multimer. The overlay shows that the XPG-anchor domain and the p62 XPD-anchor compete for the same binding site on the XPD surface, consistent with a recent study which modeled p62-XPD interactions.<sup>62</sup> The cryo-EM density unambiguously shows that the XPG-anchor domain is bound to XPD. Yet, the XL-MS data, in addition to 23 above-threshold XPD-XPG crosslinks, also features 4 p62-XPD crosslinks. This may indicate that 1) XPG and p62 compete for the same XPD binding site, or 2) p62's anchor helices and BSD2 domain remain flexible near XPD's Fe-S domain after displacement. Since XL-MS is subject to sample conformational and compositional variability, presence of minor p62-bound species different from the dominant cryo-EM structure cannot be excluded—a point that remains to be addressed by future studies. Notably, P62 displacement from TFIIH is only partial—p62's p34-anchor and 3-helix bundle domains (Fig. 1a) are clearly resolved in the TFIIH/XPA/DNA EM density and their placement is fully supported by the XL-MS data.

With the XPG-core and XPG-anchor domains bridged by the two coiled-coil helices, XPG forms a bi-lobed structure whose spatial extent perfectly matches the distance between the 3' and the 5' DNA junctions. XPG has been shown to bind early in the PlnC assembly process and stabilize the NER bubble. Yet, the first incision occurs 5' to the lesion by the action of XPF/ERCC1. In this "XPG binds first, cleaves last" model, the two nucleases are expected to sequentially coordinate their activities. However, no molecular mechanism for such coordination had emerged. Remarkably, our PlnC model suggests that XPG serves as a bridge between the opposite ends of the NER bubble, reaching all the way to the XPF/ERCC1 interface (Fig. 2a and b). The XPG-anchor domain also inserts a  $\beta$ -hairpin (residues 262-296) near the 5' junction,



**Fig. 3 | XPF/ERCC1 is positioned for precise incision at the 5' DNA junction.** **a** View of XPF/ERCC1 at the 5' junction. ERCC1 and XPF are depicted in cartoon representation and colored in magenta and dark blue, respectively. The damaged and undamaged DNA strands are shown in cyan and light blue, respectively. **b** View of the XPF/ERCC1 heterodimer interacting with DNA at 5' junction. The ERK active site motif is shown explicitly and colored in red. XPF and ERCC1 are colored by domain. A schematic representation of the XPF and ERCC1 sequences, indicating the domains, is shown below. **c** View of the XPF/ERCC1 heterodimer rotated by 70° around the dsDNA axis. **d** Conformational shift of XPF/ERCC1 relative to the DNA-

free XPF/ERCC1 structure (6SXA). **e** Conformational shift of XPF/ERCC1 relative to the DNA-engaged XPF/ERCC1 structure (6SXB). The XPF nuclease and ERCC1 central domains are overlaid for comparison. The XPF/ERCC1 (HhH)<sub>2</sub> domains are colored in red and dark red in the DNA-free structure and green and dark green in the DNA-engaged structure. ERCC1 and XPF in the PlnC are colored in magenta and blue, respectively. **f** Electrostatics of the ERCC1 (HhH)<sub>2</sub> DNA-binding surface. The electrostatic potential is mapped onto the molecular surface and colored from red (negative) to blue (positive).

which is oriented in parallel to the ssDNA and directly contacts the DNA bases (Fig. 2d and e). This finding provides an unanticipated mechanism for XPG to sense the XPF/ERCC1 incision event at the 5' junction. Importantly, such nuclease coordination mechanism would be mediated entirely by XPG, instead of relying on the flexible ssDNA to transmit the signal post-incision.

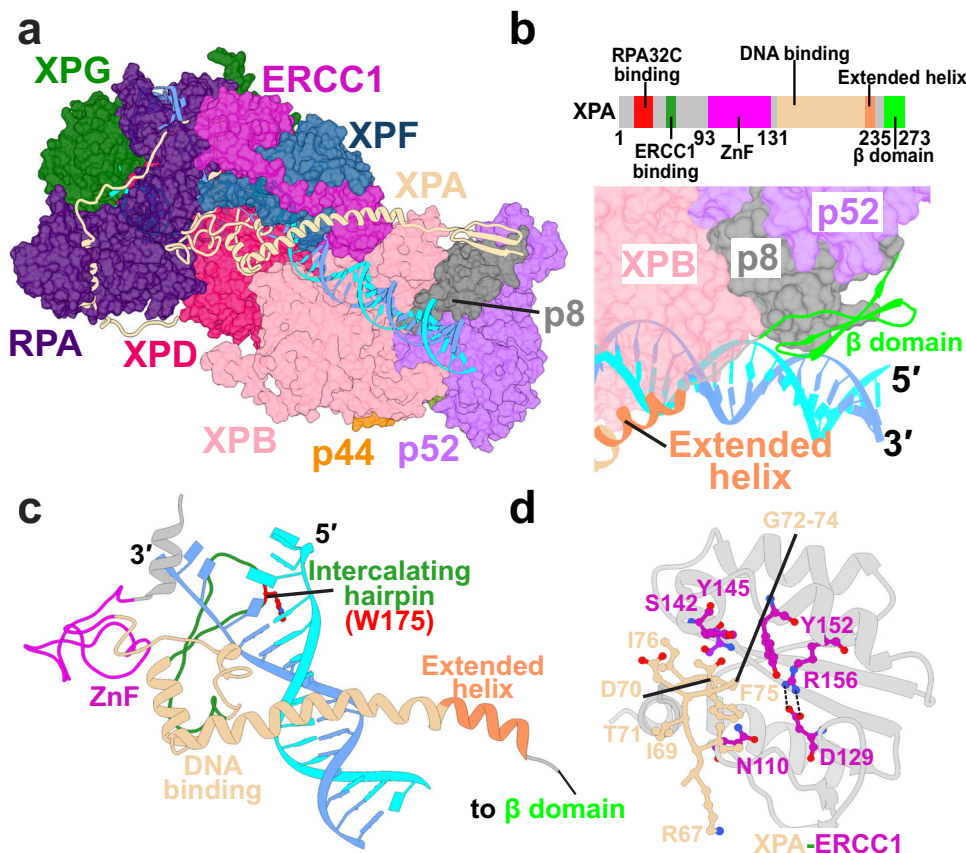
### XPG competes with XPC for binding to the 3' DNA duplex and the p62 PH domain

In the PlnC model, the XPG core occupies the position of XPC in the pre-unwound complex<sup>15,16</sup>. The fact that XPC and XPG both compete for binding to the 3' DNA duplex and do not coexist in NER complexes<sup>63,64</sup>, implies a mechanism must exist for XPG to displace XPC during the progression from lesion recognition to scanning and strand incision. Recruitment of TFIIH to the pre-unwound complex is essential for NER and requires XPC binding to p62's PH domain via a conserved acidic patch<sup>60</sup> (Supplementary Fig. 4). We identified a highly similar acidic patch in XPG (residues 151–164) on an outer loop of the XPG-helical bundle domain (Fig. 2d and Supplementary Fig. 4). We note this acidic patch aligns with residues 128–148 in the Rad2 sequence, which features multiple acidic segments competing for p62 binding<sup>60</sup>. If XPG and XPC use similar acidic patches to interact with the same interface on the flexible p62 PH domain, their binding would be competitive and mutually exclusive. The XL-MS data<sup>46</sup>

supports PH domain placement proximal to the XPG coiled-coil helices and the Fe-S domain of XPD (Supplementary Fig. 4). As the PH domain is key for XPC-TFIIH association<sup>15</sup>, this observation provides a mechanism for handoff of the DNA substrate from XPC to XPG mediated by acidic patch-p62 interactions.

### XPF/ERCC1 is strategically positioned for incision at the 5' DNA junction

A defining feature of the XPF nuclease is its modular domain architecture (Figs. 1b and 3b) that encompasses an SF2 helicase-like N-terminal domain, a central nuclease domain, and two C-terminal helix-hairpin-helix domains (HhH)<sub>2</sub>. This flexible domain arrangement allows the nuclease to adapt its structure to accommodate diverse protein partners and substrate DNA. XPF functions as an obligate heterodimer with the ERCC1 protein, mediated by their (HhH)<sub>2</sub> domains. The ERCC1 central domain has high structural homology to the XPF nuclease domain. In our PlnC model, the XPF/ERCC1 caps both sides of the 5' DNA junction and is positioned for incision (Figs. 3a–c). Binding to duplex DNA at the junction is mediated mainly by the ERCC1 (HhH)<sub>2</sub> domains, which present a conserved electrostatically complementary DNA-binding surface (Fig. 3f and Supplementary Fig. 5). By contrast, the ERCC1 central domain binds ssDNA, associating mostly with the undamaged strand. The XPF nuclease domain binds directly at the ssDNA-dsDNA junction, facing the DNA duplex and placing the catalytic metal



**Fig. 4 | XPA rigging interlaces XPF/ERCC1 with DNA, XPD, XPB, and RPA at the 5' junction and is key for licensing the XPF incision.** **a** View of XPA within the PlnC assembly. XPA is depicted in cartoon representations and colored in tan. **b** Zoomed-in view highlighting of the interface of XPA's  $\beta$ -domain and p8. TFIIH's p8, p52, and XPB subunits are shown in surface representation. A schematic of

XPA's sequence colored by domain is shown above. **c** XPA interacting with DNA at the 5' junction. XPA is colored by domains. The intercalating hairpin is shown in green and labeled. **d** Detailed residue interactions between the glycine-rich loop of XPA and the V-shaped groove of ERCC1. Residues are depicted in ball and stick representation and colored in purple and magenta.

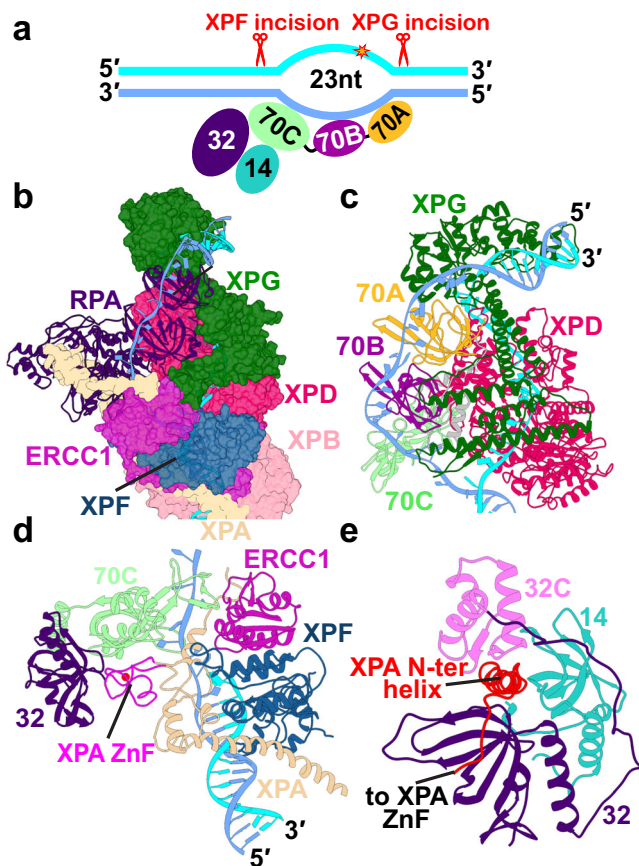
3 Å away from the scissile phosphodiester bond. The  $Mg^{2+}$  ion that affects catalysis is held by a  $GDX_nERKX_3D$  active site motif<sup>55</sup>, conserved among XPF family nucleases.

Achieving a catalytically competent state requires a major conformational shift of the XPF/ERCC1 complex relative to the known human XPF structures<sup>43</sup>, DNA-free, and DNA-engaged XPF/ERCC1. While the XPF nuclease and ERCC1 central domains overlay well among the two structures and our PlnC model, the tandem (HhH)<sub>2</sub> domains of both proteins are completely repositioned (Fig. 3d and e). The long flexible linkers connecting the (HhH)<sub>2</sub> domains to the nuclease core and the ERCC1 central domain make this feasible. We posit that completing this conformational switch is key for licensing the 5' incision and requires interactions with other PlnC constituent proteins (Fig. 3a). For example, the (HhH)<sub>2</sub> of XPF binds both XPB and XPD to ensure correct orientation of the ERCC1 (HhH)<sub>2</sub> for binding to the 5' duplex. Additionally, the XPF nuclease and ERCC1 central domains form numerous contacts with XPA, which ensures stable association of XPF/ERCC1 within the PlnC. Thus, XPA serves as an indispensable platform for the recruitment of XPF/ERCC1 to the pre-incision complex.

#### XPA orchestrates recruitment of core NER factors to the PlnC and licenses XPF for incision

XPA is a recognized master coordinator of nucleotide excision repair, controlling the assembly and coordinated exchange of core NER factors including XPC/Rad23B, RPA, TFIIH and XPF/ERCC1<sup>5</sup>. Despite its modest size, XPA spans the length of the PlnC assembly, traversing and interlacing the surfaces of RPA, XPF/ERCC1, XPD, XPB, p8 and p52

(Fig. 4)<sup>46</sup>. In our model, XPA is strategically positioned at the 5' edge of the NER bubble where it plays a key role in coordinating the repair process. XPA features a central DNA binding domain (DBD) flanked by largely disordered and flexible N- and C-terminal regions. The DBD inserts a  $\beta$ -hairpin and a tryptophan residue (W175) into the 5' junction of the NER bubble (Fig. 4c). Thus, XPA provides the helicase pin to facilitate XPB's DNA-unwinding activity<sup>46</sup>. The DBD also includes a zinc-binding motif that is proximal to the ssDNA of the undamaged strand and forms key interactions to RPA<sup>32</sup>, XPF/ERCC1<sup>47</sup> and XPD<sup>46</sup> (Figs. 1a and 4a). XPA's C-terminal end contains an extended helix that acts as a clamp on dsDNA and prevents the dissociation of the upstream duplex from XPB. This helical clamp concludes with an antiparallel  $\beta$ -sheet<sup>46</sup>, which we modeled with AlphaFold2 (Fig. 4a and b). The interaction has been previously identified to anchor the XPA C-terminus to p52 and p8 but without structural detail<sup>46</sup>. The newly modeled XPA N-terminus is predominantly unstructured, except for a previously identified helix (residues 21–40)<sup>32</sup> that inserts into the RPA trimer core and binds the RPA32C domain (Fig. 5). The RPA32C-binding motif on XPA is situated >50 residues away from the DBD. Thus, it is key that we could model XPA's N-terminus connecting the helix and DBD in this particular RPA binding mode. Besides RPA, XPA mediates interactions with XPF/ERCC1 that are essential for NER progression. Specifically, our model replicates a critical interaction between the glycine-rich loop of XPA (residues 67–80) and a V-shaped groove of ERCC1<sup>47</sup>, which serves as a validation point of our modeling (Fig. 4d). This 14-amino acid stretch of XPA has been shown biochemically as both necessary and sufficient for ERCC1 recruitment to the PlnC. The intricate entanglement of XPA with



**Fig. 5 | RPA binds and protects the undamaged strand, engaging ssDNA with its 70A, 70B and 70C domains.** **a** Schematic representation of RPA's DNA-binding core with the ssDNA of the undamaged strand. The ssDNA is engaged by three OB-fold domains (70A-70B-70C). **b** View of RPA within the PlnC assembly. RPA is shown in cartoon representation and colored in purple. **c** View of RPA70AB situated near the 3' junction and interacting with XPG and XPD. RPA70AB is colored by domains. **d** View of the RPA trimer core engaging ssDNA at the 5' junction. The XPA zinc finger is wedged between RPA70C and 32D. **e** A view of XPA's N-terminal XPA helix lodged between the RPA32C and RPA14 domains.

XPF/ERCC1, RPA and other key PlnC components underscores its critical role in licensing the XPF incision and, more generally, in orchestrating NER progression.

### RPA matches the ssDNA length of the undamaged strand

A key component of the NER machinery, the single-strand DNA binding protein RPA protects the undamaged strand of the NER bubble. RPA features four DNA-binding OB-fold domains (RPA70A, 70B, 70C, and 32D) that modularly engage ssDNA in three attainable binding modes<sup>48,66</sup>. The first binding mode involves only the 70A and 70B domains and has an 8–10 nucleotide footprint on ssDNA. The second mode engages the 70C domain, increasing RPA's footprint to 20–24 nucleotides. Full RPA engagement covers up to 30 nucleotides. In our PlnC model, RPA inserts itself between XPG, XPD, and XPA's DBD core. The length of the undamaged strand is surprisingly consistent with only three OB-fold domains (70A-70B-70C) specifically engaging ssDNA (Fig. 5). By contrast, previous NER models had assumed RPA binding with all four OB domains. There have also been conflicting models regarding the ordering of the RPA domains along the undamaged strand, some suggesting RPA70AB association near the 5' junction and others—to the 3' junction. Furthermore, interactions between the XPA DBD and RPA70AB have been reported as essential for the completion of the NER reaction<sup>32</sup>. In our model, RPA70AB

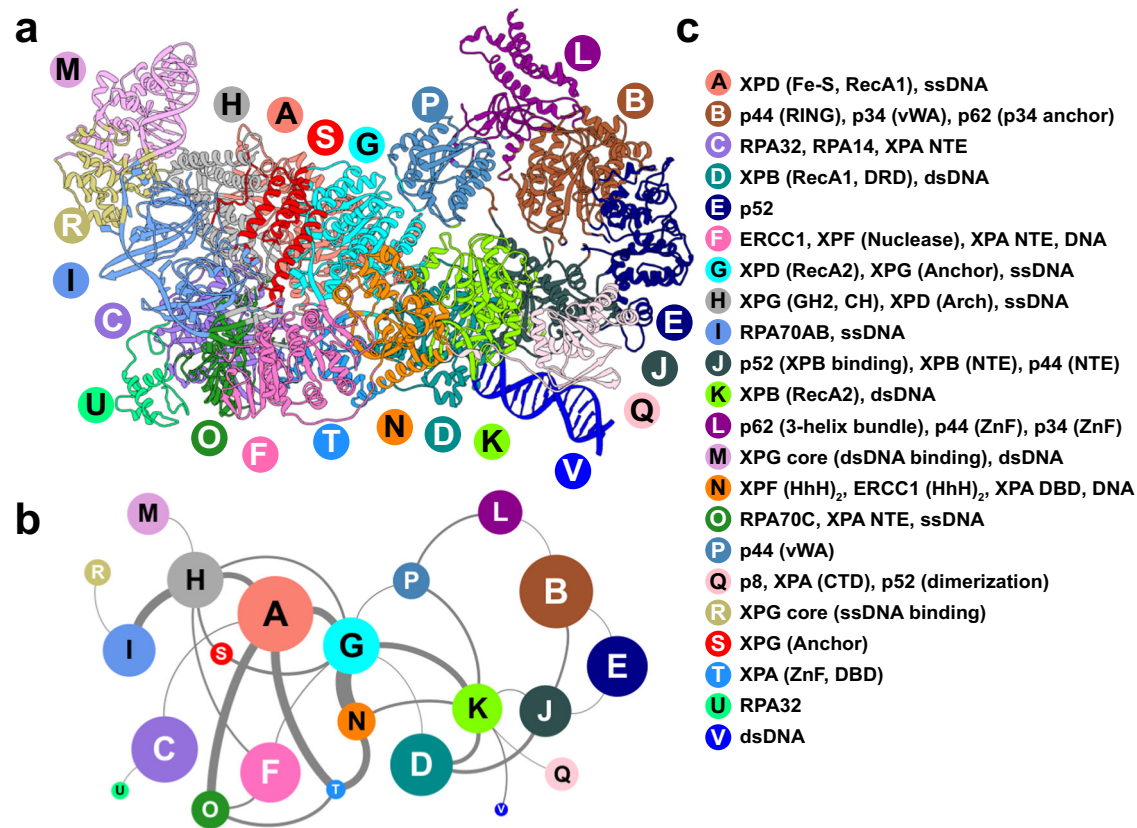
interacts with XPG near the 3' junction (Fig. 5b and c), while the XPA zinc finger domain is positioned between RPA70C and RPA32D (Fig. 5d). The N-terminal XPA helix lodges between the RPA32C and RPA14 domains (Fig. 5b and e). This arrangement is consistent with the expected polarity of RPA on ssDNA and precludes direct binding of RPA70AB at the 5' junction. To resolve this conundrum, we posit that RPA initially inserts only the 70A,B domains between the closely spaced junctions of the nascent NER bubble. The RPA70A,B module forms early contacts with the XPA zinc finger, explaining why these interactions are essential for NER. Subsequently, XPA moves with the 5' edge of the expanding NER bubble while RPA70A,B remain bound to ssDNA near the 3' junction. This process creates an opening on ssDNA for the insertion of RPA70C between RPA70A,B, and XPA and orients RPA32 and RPA14 near the 5' junction.

### PlnC's global motions and dynamic modules are key for dual incision coordination

The NER machinery undergoes a dramatic structural shift between the lesion scanning and strand incision phase. To reveal the impacts of this conformational switching on the functional dynamics of the NER complexes, we performed microsecond-timescale molecular dynamics simulations of PlnC, LSC and apo-TFIIH. We compared their relative flexibility by mapping computed B-factors from MD onto the structural models (Supplementary Fig. 6 and Supplementary Note 2). TFIIH mobility is markedly reduced in the PlnC compared to LSC, yielding exceptionally low B-factors (Supplementary Fig. 6c and f). We observe a ridge of stability that extends from the XPG core and encompasses TFIIH's XPD, XPB and p44 subunits, most of RPA, XPA's DBD, and XPF/ERCC1. Notably, in the LSC the opening/closing dynamics of XPD's Arch and Fe-S domains is preserved (Supplementary Fig. 6e), which is essential for lesion scanning. By contrast, this dynamics is abolished in the PlnC (Supplementary Fig. 6f). Thus, the NER machinery progressively loses mobility from DNA-unwinding to lesion scanning and strand incision. Loss of XPD's residual mobility converts PlnC into a rigid platform for the assembly of XPF/ERCC1 and XPG.

To analyze and visualize the concerted global motions of the PlnC, we relied on two methods—dynamic network analysis<sup>67,68</sup> and principal component analysis (PCA)<sup>69</sup>. Using covariance data from the MD simulations, network analysis partitions PlnC into communities representing the assembly's dynamically independent structural modules. Once we identify PlnC's moving parts, PCA helps us determine the directionality of their motions (Supplementary Fig. 7).

Network analysis identifies 22 dynamic communities, which are color-coded and mapped onto the PlnC structure (Figs. 6a and 7). The edge betweenness graph (Fig. 6b) encodes the magnitude of allosteric communication between pairs of communities. Notably, XPG separates into four communities (H, M, R, S). While XPG's I- and N-domains are structurally entangled in forming the nuclease core, the M and R communities encompassing XPG's core do not separate by domain. Instead, the catalytic core splits along the central  $\beta$ -hairpin leading into the XPG active site. Community M, carrying both the H2TH and  $\alpha$ 12b motifs, is sculpted to bind two turns of duplex DNA. Predictably, the motions of the 3' DNA duplex closely mirror community M's dynamics (Supplementary Fig. 7). By contrast, community R, containing the hydrophobic wedge and adjacent helices, associates with the ssDNA side of the 3' junction. Importantly, XPG's coiled-coil helices form a single dynamic module with XPD's Arch domain and the ssDNA passing through XPD (community H). Thus, helical arch dynamics is decoupled from the XPG core (Fig. 6b), allowing the coiled-coil to serve as a rigid cap on XPD's DNA-binding groove. The two long helices effectively block the opening/closing dynamics of the Arch and Fe-S domains required for ssDNA translocation. Consistently, the first four PCA modes (Supplementary Fig. 7) show the Arch, Fe-S, and coiled-coil helices predominantly moving concertedly with the same directionality. We also observe a strong edge between communities H and



**Fig. 6 | Network of dynamic communities underlie PlnC's functional dynamics.** **a** Communities identified from dynamic network analysis that transcend subunit divisions. **b** Graph of allosteric communication among communities. Nodes are sized by the number of residues in each community. Edge thickness represents the

magnitude of dynamic communication between communities (betweenness). **c** Labels identifying the domains or structural elements participating in each dynamic community.

A, suggesting tight coupling of the XPG helical arch with XPD's Arch, Fe-S, and RecA1 domains. We also note the plug region of XPD – a functionally important, but also a highly mobile and conformationally variable region of PlnC<sup>70</sup>. Modeling XPD with AlphaFold2 shows the plug to be consistent with the conformation found in the preinitiation complex (Supplementary Fig. 8a). However, the low pLDDT values suggest low fold prediction confidence for this region. Independent simulations of XPD in isolation also show that the plug is mobile and undergoes partial unfolding (Supplementary Fig. 8b and c). Intriguingly, the XPG-anchor domain forms a separate community (S), which is dynamically coupled to community G that encompasses mainly the RecA2 domain (Fig. 6b). Through communities G and A, the XPG-anchor serves as an essential link to the communities encircling the 5' DNA junction: F (ERCC1 central, XPF nuclease domains, XPA-NTE), T (XPA DBD and ssDNA) and N (XPF and ERCC1 (HhH)<sub>2</sub>, XPA DBD and dsDNA). Thus, our network analysis and PCA results corroborate our postulated model for the XPG-Arch domain acting as a sensor of the 5' incision and as a bridge between XPF/ERCC1 and the XPG core.

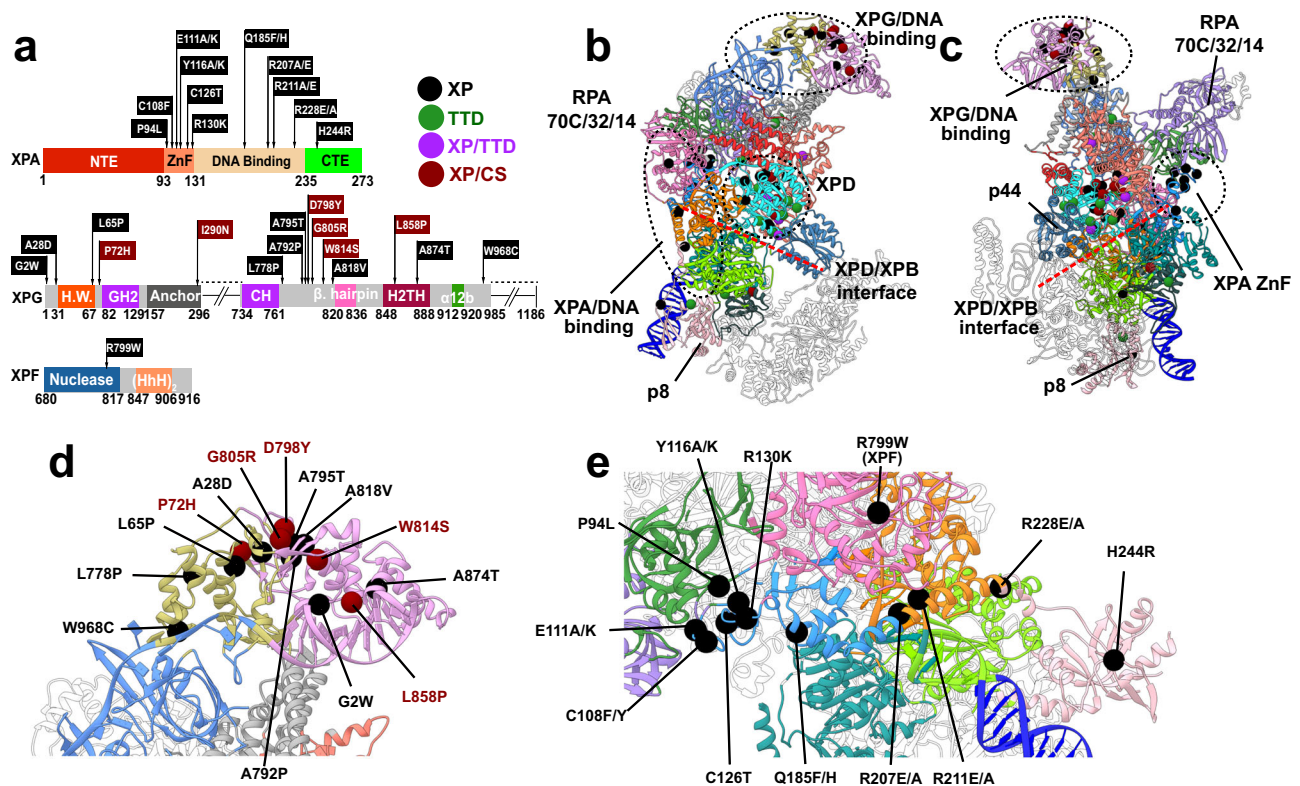
At the 5' DNA junction, two large communities dominate—1) community F encompassing ERCC1's central and XPF nuclease domains; and community N, which includes the XPF and ERCC1 (HhH)<sub>2</sub> and XPA's extended helix domains. A segment of XPA's DBD and intercalating  $\beta$ -hairpin form their own smaller community with ssDNA (community T). The extent to which XPA is entangled with other core NER proteins (RPA, ERCC1, XPF, and p8) is remarkable, with XPA contributing to six distinct dynamic modules (Fig. 6a, communities C, O, F, T, N, and Q). In essence, XPA serves as a structural adhesive to link RPA with ERCC1 and XPF and to organize them at the 5' junction. This finding supports and extends current understanding of XPA's multifaceted function in NER: 1) XPA acts as a clamp on

dsDNA to facilitate NER bubble expansion; 2) recruits and positions RPA at the NER bubble; and 3) serves as a platform to assemble XPF/ERCC1 within the PlnC. Unlike XPA, RPA subdivides mostly by domains, highlighting its modular architecture. The RPA 70A and 70B domains have motions that are well correlated. Predictably the domains form a single dynamic module (community I). By contrast, the 70C domain is less coupled to the 70A, 70B pair and forms its own community O. Intriguingly, community O is strongly coupled to community T, which incorporates XPA's zinc-finger domain (Fig. 6b). The zinc-finger is key for stabilizing RPA at the 5' DNA junction. RPA14 and RPA32 form a single dynamic module (community C), which also includes XPA's N-terminal helix.

### Disease mutations cluster at critical junctures of the PlnC dynamic network

The pre-incision complex is arguably the most critical NER intermediate. Correspondingly, disruption of this intricate molecular machine by mutations gives rise to devastating human genetic diseases. Specifically, there are three diseases linked to defects in PlnC constituent proteins, including TFIIH: xeroderma pigmentosum (XP), Cockayne-syndrome (CS), or trichothiodystrophy (TTD). XP, TTD, and XP/CS are autosomal recessive genetic disorders. Patients are often compound heterozygotes having two distinct mutations in each allele. Both alleles contribute to the expressed phenotype with combined phenotypes possible (e.g., XP/TTD, XP/CS)<sup>71</sup>. In general, XP mutants are GG-NER defective, TTD mutants cause partial transcription defects, XP/TTD mutations exhibit both defects and XP/CS mutations are defective in both GG-NER and TC-NER.<sup>6-8,72,73</sup> To link molecular features to disease phenotypes, we mapped missense disease mutations (Fig. 7, Supplementary Table 1, and Supplementary Movie 2) onto our integrative PlnC model. Mutations are





**Fig. 7 | Human disease mutations mapped onto the PlnC model are shown to localize at key community interfaces. a** Missense disease mutations mapped onto the sequences of XPG, XPA, and XPF. Disease phenotypes are labeled as XP, XP/CS, TTD, and XP/TTD and identified by symbols and color. **b** Map of human disease mutations (spheres) onto the PlnC structure (anterior view) shown in cartoon

representation and colored by community. The three primary clusters of mutations associated with XPG, TFIIH, and XPA are outlined with black dashed lines. **c** Map of human disease mutations onto the PlnC structure shown in posterior view. **d** Close-up view of mutations within XPG. **e** Close-up view of mutations within XPA and XPF.

distributed in an irregular spatial pattern with three major clusters emerging: 1) the XPG catalytic core harbors 13 mutations; 2) 34 are confined within TFIIH's XPD, XPB, and p8 subunits; and 3) 11 fall within XPA. The XPG anchor and XPF nuclease core each carry a single mutation. With all TFIIH-resident mutations previously annotated<sup>27,37</sup>, we focus our analysis on the XPG, XPF, and XPA clusters. The XPG nuclease core splits into two dynamic communities (M and R) with the community interface particularly rich in disease mutations (A28D, A795T, A792P, A818V are XP type; G805R, L858P, D798Y, P72H are XP/CS type)<sup>41</sup>. Many of these involve substitutions of hydrophobic for polar or charged residues. This not only disrupts the local dynamics close to the XPG active site but also affects the stability of the central  $\beta$ -sheet bridging the M and R communities of the nuclease core. To determine the effect of mutations on XPG protein stability, we used the Rosetta ddG protocol (Supplementary Table 1)<sup>74</sup>. There is a striking difference in the ddG scores between XPG and XPA resident mutations (Supplementary Fig. 9 and Supplementary Movie 3), suggesting XPG mutants act largely by protein destabilization. By contrast, XPA mutants act by disrupting key interactions and dynamics at community interfaces. Notably, mutations within the XPG catalytic core produce large computed ddG scores and are thus predicted to be highly destabilizing (Supplementary Fig. 9 and Supplementary Movie 3) consistent with mutant expression levels<sup>41</sup>. XPG also contains mutations that are internal to communities M and R: 1) G2W, W814S, A814T, and W968C directly affect DNA binding; 2) L778P and L65P alter the backbone of key helices, disrupting the nuclease fold. Intriguingly, the single disease mutation located outside the XPG core is I290N<sup>8</sup>, which disrupts a hydrophobic cluster next to the  $\beta$ -hairpin (residues 265–296; Fig. 2e) at the edge of the XPG anchor domain. Proximity to the 5' junction ssDNA and the XPF/ERCC1 interface suggests that the  $\beta$ -hairpin is functionally significant. We posit that this

structural element may be key for sensing incision at the 5' edge of the NER bubble.

All XPA disease mutations are XP phenotype and are distributed across the entire length of the protein. Notably, XPA disease mutations lie primarily at community interfaces (Fig. 7 and Supplementary Movie 2). Six mutants cluster within the zinc-finger domain (P94L, C108F, C126T, E111A/K, Y116A/K, R130K) (Figs. 7c and f). Structural disruption of the zinc finger impairs XPA–DNA binding, impacting NER activity<sup>75</sup>. The zinc-finger domain is also the linchpin of the RPA70C interface with XPA's DBD, bridging dynamic communities O, C, and T. Destabilization or dynamic disruption of this interface could abolish the interaction of XPA and RPA, which is critical for NER. The C-terminal domain of XPA harbors a single XP mutant (H244R), which affects XPA–TFIIH association and decreases NER activity. From our model, we note the proximity of H244 to dsDNA. Switching the residue to an arginine increases DNA binding at the expense of weakening the interface between XPA's C-terminal  $\beta$ -sheet and TFIIH's p8 subunit. The XPA R228A/E mutation lies in the extended helix of XPA at the interface of communities Q and N. Disruption in that region not only affects DNA binding but could also impact the dynamic module formed by XPA and ERCC1's (HhH)<sub>2</sub> or its interactions with XPB. The remaining XP mutants in XPA lie in the DBD. Strikingly, all three are situated at dynamic community interfaces (Q185F/H between communities T and G; R211E/A and R207E/A between communities N and T) and are poised to disrupt DNA binding or interfere with the complex entanglement of XPA with XPF and ERCC1 at the 5' DNA junction (Fig. 7e and f). By contrast, the lone mutation in XPF (R799W) is internal to community N but lies at the interface between XPF nuclease and ERCC1 central domains and is C-capping an ERCC1 helix. Mutation to tryptophan would disrupt the helix and the XPF–ERCC1 interface to affect recruitment of XPF/ERCC1 to TFIIH.

## Discussion

NER complexes are amazingly dynamic protein machines that actively reshape themselves and self-regulate to achieve precise lesion removal from genomic DNA. By synthesizing cryo-EM and XL-MS data with advanced computational modeling, we built a practically complete hybrid model of the NER pre-incision complex and analyzed its functional dynamics. The results support a structural basis for XPF and XPG nuclease licensing and coordination of dual incision in NER.

In early NER, the XPC/Rad23B/CETN2 complex first detects lesioned DNA, recruits TFIIH in an open apo-like conformation, and stabilizes a nascent NER bubble. At this stage, TFIIH's XPD subunit is distal from DNA while the XPB subunit acts as a translocase and unwinds DNA downstream of the lesion site. TFIIH's CAK module presents an obstacle to further NER progression and is removed upon subsequent XPA recruitment to the 5' end of the NER bubble. This conformational switch involves displacement of CAK's MAT1 subunit by the N-terminus of XPA. It has been previously proposed that MAT1 could serve as XPB-XPD spacer, and its removal could allow XPD to approach DNA<sup>15,37</sup>. Additionally, XPA stimulates XPB unwinding and orchestrates initial recruitment of RPA. In turn, RPA binds and protects the undamaged strand between its 70A and 70B domains. Throughout this process, XPC remains bound at the upstream DNA duplex.

Our computationally informed mechanism (Fig. 8), sheds light on the late-stage reorganization of the NER protein machinery from lesion scanning to dual incision. Once the NER bubble reaches a critical size, XPD threads the lesioned DNA strand through its DNA-binding groove, triggering closure of TFIIH and a collapse of the XPD-XPB spacing. The ensuing close association of XPD, XPB, and p44 blocks XPB's unwinding activity<sup>37</sup>, leading to the formation of the LSC. Thus, the action of XPB and XPD in NER is sequential and strictly coordinated. During the lesion-scanning phase, XPD reels in ssDNA toward the 5' junction, unwinding the upstream DNA duplex. When the lesion becomes blocked in the narrow space between XPD's Fe-S and Arch domains, DNA unwinding stops. This event triggers assembly of the PlnC. Intriguingly, the helicase pins assisting strand separation at the edges of the NER bubble are not internal to the two translocases. Instead, they are provided by other proteins (XPG and XPA), ensuring XPB and XPD unwinding activities are optimally efficient only in the context of a fully assembled NER machinery.

In the next stage of NER, XPD activity is disabled, allowing the XPG and XPF/ERCC1 nucleases to assemble on the circularized and rigidified TFIIH scaffold. Concomitantly, RPA expands its footprint on the undamaged strand to include the 70C domain. XPG replaces XPC at the 3' edge of the NER bubble, perhaps as early as the lesion scanning stage as there is evidence XPG could stimulate XPD activity<sup>46,54</sup>. The newly identified XPG anchor domain is critical for this handoff, which involves XPG binding to XPD, competing with p62's XPD anchor and BSD2 domains, and displacing XPC in its interaction with the PH domain of p62. Remarkably, XPG spans the entire length of the NER bubble and inserts a  $\beta$ -hairpin from the XPG-anchor into the 5' junction close to the XPF/ERCC1 interface. Thus, our findings support an unexpected mechanism for XPG to sense the action of XPF and coordinate the dual incision of the lesioned strand. XPG is recruited early to the PlnC, but XPF incision occurs first. Importantly, mere recruitment of XPF is insufficient for DNA cleavage. We posit that licensing of the XPF incision requires conformational switching to achieve the observed complex entanglement of XPA DBD, ssDNA, XPF and ERCC1 at the 5' junction. Similarly, XPG nuclease licensing is triggered by termination of XPD unwinding that may involve tilting of the XPG core (see alternative XPG model). Reorientation of the XPG core could allow interaction with XPD's Arch domain to stimulate the 3' incision. The final stage of NER involves PlnC disassembly and departure of TFIIH with the excision product. Gap-filling DNA synthesis ensues and restores DNA to its original condition.

Collectively, these findings elucidate the structure and dynamics of a critically important state of the NER machinery – the pre-incision complex. The practically complete model and computational analyses yield key mechanistic insights into PlnC's assembly and regulation, the structural basis of XPF and XPG nuclease coordination, and the licensing of the NER dual incision.

Linking molecular mechanisms to disease phenotypes is a grand challenge for structural biology. This challenge is often unmet as it requires knowledge of dynamic conformations and assemblies that resist purely experimental approaches. Our integrative methods and results provide a framework for meeting this challenge and for designing future experiments to uncover the intricate molecular choreography of global genome NER. Our dynamic network models powerfully elucidate the etiology of devastating human genetic syndromes. Notably, we find that XP and XP/CS disease mutations cluster at key interfaces of PlnC's dynamic communities, impacting NER protein stability, functional dynamics, DNA binding, nuclease licensing, and/or community integrity.

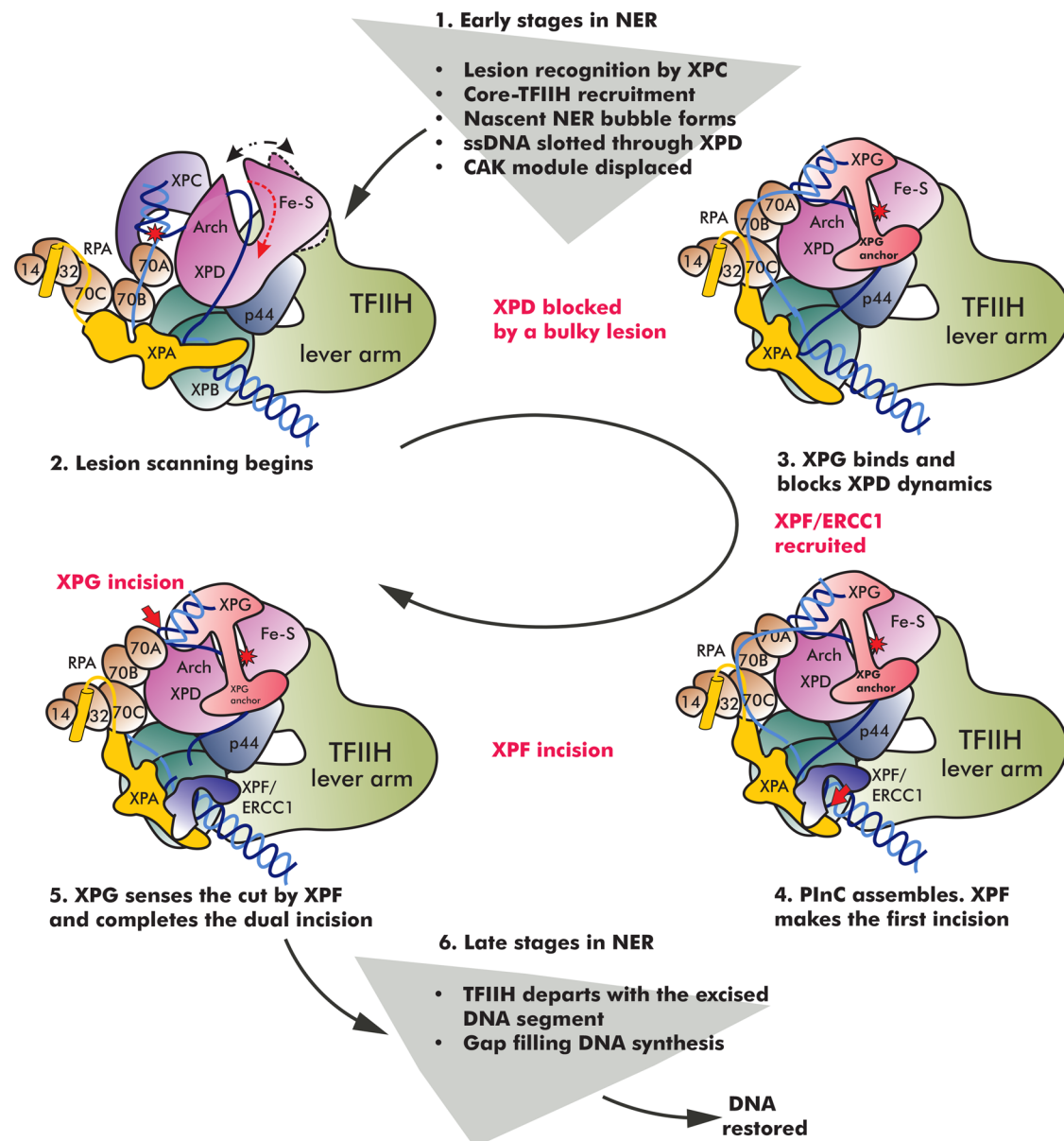
## Methods

### Model building

To construct a model of the pre-incision complex (PlnC), we systematically examined the cryo-EM structures and densities of human apo-TFIIH<sup>22</sup>, TFIIH/XPA/DNA<sup>46</sup>, and XPF/ERCC1<sup>43</sup>, the NMR structure of XPA-ERCC1<sup>47</sup>, and the X-ray structures of the XPG catalytic core<sup>38</sup> and RPA-ssDNA (RPA70, RPA32, and RPA14)<sup>48</sup>. The TFIIH/XPA/DNA structure<sup>46</sup> (PDB ID: 6RO4 and EMD accession code: EMD-4970) was the starting point for model building. The PlnC hybrid model has an NER bubble size of 23 nucleotides, matching the 27-nucleotide optimal length of the excision products and the XPF and XPG incision patterns. Sources of experimental structural information used in constructing the integrative model are summarized in Supplementary Fig. 10 and Supplementary Table 2. Regions modeled with AlphaFold2 are shown in Supplementary Fig. 11 with pLDDT scores mapped onto the structures. Additionally, maps showing geometric and electrostatic complementarity as well as conservation across the newly modeled PlnC interfaces are shown in Supplementary Fig. 12.

FEN1 shares 30% sequence identity with the XPG catalytic core<sup>76,38</sup> (PDB ID: 6TUR, 6TUW, and 6VBH). Thus, we modeled DNA-bound XPG based on the human FEN1/DNA X-ray structure<sup>61</sup> (PDB ID: SUM9). XPG positioning into the hybrid model was based on existing XL-MS data<sup>22</sup>. In addition, positioning of the XPG core required placement of the 3' DNA junction 8 nucleotides away from the expected position of the DNA lesion near XPD's His135 residue. The two XPG gateway helices (GH1 residues 33–41 and GH2 residues 82–129) and the capping helix (CH, residues 734–763) were predicted with AlphaFold2<sup>49,50</sup> and positioned in the gap between XPD's Arch and Fe-S domains in accordance with the crosslink data<sup>22</sup>. The XPD-anchor domain (residues 157–296) was predicted by AlphaFold2 and fitted into the TFIIH/XPA/DNA cryo-EM density. The loop connecting GH2 and the XPD-anchor was built with Modeler<sup>77</sup>.

To model XPF/ERCC1, we used the cryo-EM structures of XPF/ERCC1<sup>43</sup> (PDB ID: 6SXA and 6SXB). We first docked the XPF nuclease domain to the 5' junction. The catalytic metal was oriented 3 Å away from the scissile phosphodiester bond. Mg<sup>2+</sup> ion coordination was based on the *Aeropyrum pernix* SNF2 structure<sup>65</sup> (PDB ID: 2BGW). A water molecule was placed between Mg<sup>2+</sup> ion and the DNA backbone phosphate group. The ERCC1 (HhH)<sub>2</sub> domain was oriented to interact with the ssDNA through two DNA hairpins based on the 6SXB structure. The long linkers from the ERCC1 central domain to the ERCC1 (HhH)<sub>2</sub> (residues 214–230) and from the XPF nuclease domain to the XPF (HhH)<sub>2</sub> (residues 817–847) were built with Modeler. The SF2 helicase-like N-terminal domain of XPF was omitted from the hybrid PlnC model due to lack of sufficient structural or biochemical restraints.



**Fig. 8 | NER protein machinery undergoes a dramatic structural reorganization from lesion recognition to lesion scanning and strand incision.** The schematic represent key steps in the NER pathway—XPC lesion recognition, NER bubble extension, XPD-mediated lesion scanning, PlnC assembly and dual incision of the damaged DNA segment, gap-filling synthesis, and DNA restoration. Core NER factors are shown in cartoon representation and color-coded. The position of the

lesion is indicated by a red star. Red dashed arrows show the direction of ssDNA movement during the different stages of NER. White dotted arrow denotes the opening/closing dynamics of XPB during bubble expansion. Black dotted arrow denotes the opening/closing dynamics of XPD during lesion scanning. A red cross denotes the blocking of a lesion inside XPD and damage verification. Red arrows indicate the incision points on DNA by XPF/ERCC1 and XPG, respectively.

To model RPA, we used following X-ray structures: *Ustilago maydis* RPA/ssDNA<sup>48</sup> (PDB ID: 4GOP), yeast RPA/ssDNA<sup>78</sup> (PDB ID: 6I52), and human RPA<sup>79,80</sup> (PDB ID: 1JMC and 1L1O). The RPA70AB/ssDNA complex was modeled by superimposing the yeast RPA/ssDNA structure<sup>79</sup> (PDB ID: 6I52) onto the human apo-RPA 70AB<sup>78</sup> (PDB ID: 1JMC). Within PlnC, only RPA70A, 70B, and 70C can engage DNA due to the size of the NER bubble. RPA70AB was placed close to the 3' junction where it interacts with XPG. We reoriented RPA70C to bind ssDNA near the 5' junction. The RPA70C/ssDNA was modeled by aligning the *Ustilago maydis* RPA/ssDNA structure<sup>48</sup> (PDB ID: 4GOP) with the human trimer core structure<sup>1</sup> (PDB ID: 1L1O). The orientation of RPA32D and RPA14 follows from the placement of the RPA70C module as they are all connected, forming the trimer core (70C/32D/14).

To model XPA, we used the following structures: the cryo-EM TFIIH/XPA/DNA structure<sup>46</sup> (PDB ID: 6RO4), the NMR structure of

XPA/ERCC1<sup>47</sup> (PDB ID: 2JNW), and the human X-ray structure of RPA32C/Smarcal1 N-terminus<sup>81</sup> (PDB ID: 4MQV). The XPA N-terminal extension (residues 1–103), which includes the RPA32C binding helix (residues 22–40), and the C-terminal extension ( $\beta$ -domain) (residues 235–273) lacked known structural homologs and were modeled using AlphaFold2. The  $\beta$ -domain was fitted into the TFIIH/XPA/DNA density. To position XPA's N-terminal helix (residues 22–40) we used the X-ray structure of RPA32C/Smarcal1 N-terminus.

To assemble the complete PlnC model, we also modeled loop regions of TFIIH's core subunits (XPB, XPD, p44, p34, and p52) into the TFIIH/XPA/DNA density.

#### Molecular dynamics

Molecular dynamics simulations of PlnC complex in the presence and absence of a CPD lesion were performed on the Summit machine of the

Oak Ridge Leadership Computing Facility. The systems were set up with the TLeap module of AMBER<sup>82</sup> and solvated with TIP3P water molecules<sup>83</sup>. To balance the overall charge of each system, we used Na<sup>+</sup> counterions. Extra Na<sup>+</sup> and Cl<sup>-</sup> ions were introduced to produce 150 mM salt concentration as needed to mimic physiological conditions. We then used the NAMD code to carry out energy minimization for 5000 steps with positional constraints imposed on the protein and DNA backbone atoms. NVT simulations were used to gradually bring the temperature of the systems to 300 K over a period of 100 ps. During the NVT run positional restraints ( $k = 10 \text{ kcal mol}^{-1} \text{ \AA}^{-2}$ ) were imposed on all heavy atoms of PlnC. Equilibration was continued for another 5 ns in the NPT ensemble while the restraints were gradually released. Production simulations were carried out in the NPT ensemble (1 atm, 300 K) for 1 $\mu$ s for each complex. The particle mesh Ewald (PME) method was employed to compute the long-range electrostatic interactions. The r-RESPA multiple-time-step method<sup>84</sup> was used with a 2-fs timestep for bonded interactions, a 2-fs timestep for short-range non-bonded interactions, and 4-fs timestep for long-range electrostatic interactions. A short-range non-bonded interaction cutoff of 10 Å and a switching function at 8.5 Å were used for the simulations. All covalent bonds to hydrogen atoms were constrained using the SHAKE method. The simulations were performed with the NAMD 2.14 code<sup>85</sup> and the AMBER forcefields: Parm14SB<sup>86</sup> and OL15<sup>87</sup>. All figures were generated with UCSF Chimera<sup>88</sup>.

### Covariance-based community network analysis

Covariance-based community network analysis<sup>67,68</sup> was performed on the PlnC trajectories. Network analysis is a graph-theoretical method, which maps the PlnC assembly onto a protein network graph wherein residues are represented as nodes and edges connect contacting residues. Two non-adjacent residues are considered in contact if they are within 4.5 Å for 75% or more of the trajectory. The MDTraj package<sup>89</sup> was used to obtain contact maps from the MD ensembles. Edges are weighted by residue-residue covariances given by:  $w_{ij} = -\ln(|c_{ij}|)$ , where  $c_{ij}$  are pairwise correlation coefficients. Partitioning this graph into strongly connected components with the Girvan-Newman algorithm<sup>90</sup> defines dynamic communities, which are PlnC's dynamically independent functional modules. Edge betweenness among communities then recapitulates allosteric communication within the assembly.

### Principal component analysis

PCA is a dimensionality reduction technique, which diagonalizes the residue-residue covariance matrix from the simulation trajectories, yielding corresponding eigenvectors (principal modes) and eigenvalues (mean square fluctuations). The first few principal modes recapitulate the functionally significant large-scale motions of the PlnC assembly. PCA was performed using the CPPTRAJ module in AmberTools16<sup>91</sup>.

### Rosetta protein stability analysis

XP and XP/CS disease mutations in XPA, XPG, and XPF were assessed for their impact on protein stability using the Rosetta Cartesian ddG protocol<sup>74</sup>. The wild type (WT) structure of the PlnC complex was relaxed in Cartesian space using the Rosetta FastRelax protocol. Mutations were then introduced and the FastRelax protocol used to repack the side chains within 6 Å of the mutation site. The protein backbone within 3 residues of the mutation site is also allowed to readjust. The ddG value were determined by the Rosetta score differences between the relaxed mutant protein and the relaxed WT protein for each of the evaluated missense mutations.

### Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

## Data availability

The models of the PlnC complex have been deposited in the PDB-dev database with accession codes: PDBDEV\_00000373, [9A88](#), and PDBDEV\_00000374, [9A89](#). The final configuration of the PlnC molecular dynamics trajectory is provided as a plain text file pre-incision-complex-final-MD-configuration\_PDB.txt in PDB format as Supplementary Data 1 file. PDB accession codes of all the publicly available datasets used in the study: [6RO4](#), [6TUR](#), [6TUW](#), [6VBH](#), [5UM9](#), [6SXA](#), [6SXB](#), [2BGW](#), [4GOP](#), [6I52](#), [1JMC](#), [1L10](#), [2JNW](#), and [4MQV](#).

## References

- Scharer, O. D. Nucleotide excision repair in eukaryotes. *Cold Spring Harb. Perspect. Biol.* **5**, a012609 (2013).
- Wang, W., Xu, J., Chong, J. & Wang, D. Structural basis of DNA lesion recognition for eukaryotic transcription-coupled nucleotide excision repair. *DNA Repair* **71**, 43–55 (2018).
- Brueckner, F., Hennecke, U., Carell, T. & Cramer, P. CPD damage recognition by transcribing RNA polymerase II. *Science* **315**, 859–862 (2007).
- Marteijn, J. A., Lans, H., Vermeulen, W. & Hoeijmakers, J. H. Understanding nucleotide excision repair and its roles in cancer and ageing. *Nat. Rev. Mol. Cell Biol.* **15**, 465–481 (2014).
- D'Souza, A., Blee, A. M. & Chazin, W. J. Mechanism of action of nucleotide excision repair machinery. *Biochem. Soc. Trans.* **50**, 375–386 (2022).
- Berneburg, M. & Lehmann, A. R. Xeroderma pigmentosum and related disorders: defects in DNA repair and transcription. *Adv. Genet.* **43**, 71–102 (2001).
- Lehmann, A. R. The xeroderma pigmentosum group D (XPD) gene: one gene, two functions, three diseases. *Genes Dev.* **15**, 15–23 (2001).
- Fasshi, H. et al. Deep phenotyping of 89 xeroderma pigmentosum patients reveals unexpected heterogeneity dependent on the precise molecular defect. *Proc. Natl Acad. Sci. USA.* **113**, E1236–E1245 (2016).
- Pugh, J. et al. Use of Big Data to estimate prevalence of defective DNA repair variants in the US Population. *JAMA Dermatol.* **155**, 72–78 (2019).
- Bradford, P. T. et al. Cancer and neurologic degeneration in xeroderma pigmentosum: long term follow-up characterises the role of DNA repair. *J. Med. Genet.* **48**, 168–176 (2011).
- Compe, E. & Egly, J. M. TFIIH: when transcription met DNA repair. *Nat. Rev. Mol. Cell Biol.* **13**, 343–354 (2012).
- Yan, C. et al. Mechanism of Rad26-assisted rescue of stalled RNA polymerase II in transcription-coupled repair. *Nat. Commun.* **12**, 7001 (2021).
- Nieto Moreno, N., Olthof, A. M. & Svejstrup, J. Q. Transcription-coupled nucleotide excision repair and the transcriptional response to UV-induced DNA damage. *Annu Rev. Biochem.* **92**, 81–113 (2023).
- van den Heuvel, D., van der Weegen, Y., Boer, D. E. C., Ogi, T. & Luijsterburg, M. S. Transcription-coupled DNA repair: from mechanism to human disorder. *Trends Cell Biol.* **31**, 359–371 (2021).
- van Eeuwen, T. et al. Cryo-EM structure of TFIIH/Rad4–Rad23–Rad33 in damaged DNA opening in nucleotide excision repair. *Nat. Commun.* **12**, 3338 (2021).
- Kim, J. et al. Lesion recognition by XPC, TFIIH and XPA in DNA excision repair. *Nature* **617**, 170–175 (2023).
- Sugasawa, K. et al. Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. *Mol. Cell* **2**, 223–232 (1998).
- Sugasawa, K. et al. A multistep damage recognition mechanism for global genomic nucleotide excision repair. *Genes Dev.* **15**, 507–521 (2001).
- Shell, S. M. et al. Xeroderma pigmentosum complementation group C protein (XPC) serves as a general sensor of damaged DNA. *DNA Repair* **12**, 947–953 (2013).

20. Min, J. H. & Pavletich, N. P. Recognition of DNA damage by the Rad4 nucleotide excision repair protein. *Nature* **449**, 570–575 (2007).
21. Velmurugu, Y., Chen, X., Slogoff Sevilla, P., Min, J.-H. & Ansari, A. Twist-open mechanism of DNA damage recognition by the Rad4/XPC nucleotide excision repair complex. *Proc. Nat. Acad. Sci. USA* **113**, E2296–E2305 (2016).
22. Greber, B. J. et al. The cryo-electron microscopy structure of human transcription factor IIH. *Nature* **549**, 414–417 (2017).
23. Rimel, J. K. & Taatjes, D. J. The essential and multifunctional TFIIH complex. *Protein Sci.* **27**, 1018–1037 (2018).
24. Compe, E. & Egly, J. M. Nucleotide excision repair and transcriptional regulation: TFIIH and beyond. *Annu. Rev. Biochem.* **85**, 265–290 (2016).
25. Zurita, M. & Cruz-Becerra, G. TFIIH: New discoveries regarding its mechanisms and impact on cancer treatment. *J. Cancer* **7**, 2258–2265 (2016).
26. Schaeffer, L. et al. DNA repair helicase: a component of BTF2 (TFIIH) basic transcription factor. *Science* **260**, 58–63 (1993).
27. Yan, C. et al. Transcription preinitiation complex structure and dynamics provide insight into genetic diseases. *Nat. Struct. Mol. Biol.* **26**, 397–406 (2019).
28. Tsutakawa, S. E. et al. Envisioning how the prototypic molecular machine TFIIH functions in transcription initiation and DNA repair. *DNA Repair* **96**, 102972 (2020).
29. He, Y. et al. Near-atomic resolution visualization of human transcription promoter opening. *Nature* **533**, 359–365 (2016).
30. Luo, J. et al. Architecture of the human and yeast general transcription and DNA repair factor TFIIH. *Mol. Cell* **59**, 794–806 (2015).
31. Fuss, J. O. & Tainer, J. A. XPB and XPD helicases in TFIIH orchestrate DNA duplex opening and damage verification to coordinate repair with transcription and cell cycle via CAK kinase. *DNA Repair* **10**, 697–713 (2011).
32. Kim, M. et al. Two interaction surfaces between XPA and RPA organize the preincision complex in nucleotide excision repair. *Proc. Nat. Acad. Sci. USA*. **119**, e2207408119 (2022).
33. Topolska-Wos, A. M. et al. A key interaction with RPA orients XPA in NER complexes. *Nucleic Acids Res.* **48**, 2173–2188 (2020).
34. Li, C. L. et al. Tripartite DNA lesion recognition and verification by XPC, TFIIH, and XPA in nucleotide excision repair. *Mol. Cell* **59**, 1025–1034 (2015).
35. Fan, L. et al. Conserved XPB core structure and motifs for DNA unwinding: Implications for pathway selection of transcription or excision repair. *Mol. Cell* **22**, 27–37 (2006).
36. Fan, L. et al. XPD helicase structures and activities: Insights into the cancer and aging phenotypes from XPD mutations. *Cell* **133**, 789–800 (2008).
37. Yu, J. et al. Dynamic conformational switching underlies TFIIH function in transcription and DNA repair and impacts genetic diseases. *Nat. Commun.* **14**, 2758 (2023).
38. O'Donovan, A., Davies, A. A., Moggs, J. G., West, S. C. & Wood, R. D. XPG endonuclease makes the 3' incision in human DNA nucleotide excision repair. *Nature* **371**, 432–435 (1994).
39. Matsunaga, T., Mu, D., Park, C. H., Reardon, J. T. & Sancar, A. Human DNA repair excision nuclease. Analysis of the roles of the subunits involved in dual incisions by using anti-XPG and anti-ERCC1 antibodies. *J. Biol. Chem.* **270**, 20862–20869 (1995).
40. Mocquet, V. et al. Sequential recruitment of the repair factors during NER: the role of XPG in initiating the resynthesis step. *EMBO J.* **27**, 155–167 (2008).
41. Tsutakawa, S. E. et al. Human XPG nuclease structure, assembly, and activities with insights for neurodegeneration and cancer from pathogenic mutations. *Proc. Natl Acad. Sci. USA* **117**, 14127–14138 (2020).
42. Fagbemi, A. F., Orelli, B. & Scharer, O. D. Regulation of endonuclease activity in human nucleotide excision repair. *DNA Repair* **10**, 722–729 (2011).
43. Jones, M. et al. Cryo-EM structures of the XPF-ERCC1 endonuclease reveal how DNA-junction engagement disrupts an auto-inhibited conformation. *Nat. Commun.* **11**, 1120 (2020).
44. Staresincic, L. et al. Coordination of dual incision and repair synthesis in human nucleotide excision repair. *EMBO J.* **28**, 1111–1120 (2009).
45. Li, W., Adebali, O., Yang, Y., Selby, C. P. & Sancar, A. Single-nucleotide resolution dynamic repair maps of UV damage in *Saccharomyces cerevisiae* genome. *Proc. Nat. Acad. Sci. USA* **115**, E3408–E3415 (2018).
46. Kokic, G. et al. Structural basis of TFIIH activation for nucleotide excision repair. *Nat. Commun.* **10**, 2885 (2019).
47. Tsodikov, O. V. et al. Structural basis for the recruitment of ERCC1-XPF to nucleotide excision repair complexes by XPA. *EMBO J.* **26**, 4768–4776 (2007).
48. Fan, J. & Pavletich, N. P. Structure and conformational change of a replication protein A heterotrimer bound to ssDNA. *Genes Dev.* **26**, 2337–2347 (2012).
49. Jumper, J. et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–589 (2021).
50. Mirdita, M. et al. ColabFold: making protein folding accessible to all. *Nat. Methods* **19**, 679–682 (2022).
51. Evans, E., Fellows, J., Coffey, A. & Wood, R. D. Open complex formation around a lesion during nucleotide excision repair provides a structure for cleavage by human XPG protein. *EMBO J.* **16**, 625–638 (1997).
52. Tsutakawa, S. E. et al. Human flap endonuclease structures, DNA double-base flipping, and a unified understanding of the FEN1 superfamily. *Cell* **145**, 198–211 (2011).
53. Rashid, F. et al. Single-molecule FRET unveils induced-fit mechanism for substrate selectivity in flap endonuclease 1. *Elife* **6**, e21884 (2017).
54. Bralic, A. et al. A scanning-to-incision switch in TFIIH-XPG induced by DNA damage licenses nucleotide excision repair. *Nucleic Acids Res.* **51**, 1019–1033 (2023).
55. Bessho, T., Mu, D. & Sancar, A. Initiation of DNA interstrand cross-link repair in humans: the nucleotide excision repair system makes dual incisions 5' to the cross-linked base and removes a 22- to 28-nucleotide-long damage-free strand. *Mol. Cell Biol.* **17**, 6822–6830 (1997).
56. Scherly, D. et al. Complementation of the DNA-repair defect in xeroderma-pigmentosum group-G cells by a human cDNA related to yeast Rad2. *Nature* **363**, 182–185 (1993).
57. Iyer, N., Reagan, M. S., Wu, K. J., Canagarajah, B. & Friedberg, E. C. Interactions involving the human RNA polymerase II transcription/nucleotide excision repair complex TFIIH, the nucleotide excision repair protein XPG, and Cockayne syndrome group B (CSB) protein. *Biochemistry* **35**, 2157–2167 (1996).
58. Hohl, M., Thorel, F., Clarkson, S. G. & Schärer, O. D. Structural determinants for substrate binding and catalysis by the structure-specific endonuclease XPG. *J. Biol. Chem.* **278**, 19500–19508 (2003).
59. Dunand-Sauthier, I. et al. The spacer region of XPG mediates recruitment to nucleotide excision repair complexes and determines substrate specificity. *J. Biol. Chem.* **280**, 7030–7037 (2005).
60. Lafrance-Vanasse, J., Arseneault, G., Cappadocia, L., Legault, P. & Omichinski, J. G. Structural and functional evidence that Rad4 competes with Rad2 for binding to the Tfb1 subunit of TFIIH in NER. *Nucleic Acids Res.* **41**, 2736–2745 (2013).
61. Tsutakawa, S. E. et al. Phosphate steering by Flap Endonuclease 1 promotes 5'-flap specificity and incision to prevent genome instability. *Nat. Commun.* **8**, 15855 (2017).
62. Barnett, J. T., Kuper, J., Koelmel, W., Kisker, C. & Kad, N. M. The TFIIH subunits p44/p62 act as a damage sensor during nucleotide excision repair. *Nucleic Acids Res.* **48**, 12689–12696 (2020).

63. Wakasugi, M. & Sancar, A. Assembly, subunit composition, and footprint of human DNA repair excision nuclease. *Proc. Natl Acad. Sci. USA* **95**, 6669–6674 (1998).
64. Riedl, T., Hanaoka, F. & Egly, J. M. The comings and goings of nucleotide excision repair factors on damaged DNA. *EMBO J.* **22**, 5293–5303 (2003).
65. Newman, M. et al. Structure of an XPF endonuclease with and without DNA suggests a model for substrate recognition. *EMBO J.* **24**, 895–905 (2005).
66. Brosey, C. A. et al. A new structural framework for integrating replication protein A into DNA processing machinery. *Nucleic Acids Res.* **41**, 2313–2327 (2013).
67. Eargle, J. & Luthey-Schulten, Z. NetworkView: 3D display and analysis of protein center dot RNA interaction networks. *Bioinformatics* **28**, 3000–3001 (2012).
68. Hudson, W. H. et al. Distal substitutions drive divergent DNA specificity among paralogous transcription factors through subdivision of conformational space. *Proc. Natl Acad. Sci. USA*. **113**, 326–331 (2016).
69. David, C. C. & Jacobs, D. J. Principal component analysis: a method for determining the essential dynamics of proteins. *Protein Dyn. Methods Protoc.* **1084**, 193–226 (2014).
70. Kuper, J. et al. XPD stalled on cross-linked DNA provides insight into damage verification. *Nat. Struct. Mol. Biol.* <https://doi.org/10.1038/s41594-024-01323-5> (2024).
71. Ueda, T., Compe, E., Catez, P., Kraemer, K. H. & Egly, J. M. Both XPD alleles contribute to the phenotype of compound heterozygote xeroderma pigmentosum patients. *J. Exp. Med.* **206**, 3031–3046 (2009).
72. Boyle, J. et al. Persistence of repair proteins at unrepaired DNA damage distinguishes diseases with ERCC2 (XPD) mutations: cancer-prone xeroderma pigmentosum vs. non-cancer-prone trichothiodystrophy. *Hum. Mutat.* **29**, 1194–1208 (2008).
73. DiGiovanna, J. J. & Kraemer, K. H. Shining a light on xeroderma pigmentosum. *J. Investig. Dermatol.* **132**, 785–796 (2012).
74. Park, H. et al. Simultaneous optimization of biomolecular energy functions on features from small molecules and macromolecules. *J. Chem. Theory Comput.* **12**, 6201–6212 (2016).
75. van den Heuvel, D. et al. A disease-associated XPA allele interferes with TFIIH binding and primarily affects transcription-coupled nucleotide excision repair. *Proc. Natl Acad. Sci. USA*. **120**, e2208860120 (2023).
76. González-Corrochano, R. et al. The crystal structure of human XPG, the xeroderma pigmentosum group G endonuclease, provides insight into nucleotide excision DNA repair. *Nucleic Acids Res.* **48**, 9943–9958 (2020).
77. Martí-Renom, M. A. et al. Comparative protein structure modeling of genes and genomes. *Annu. Rev. Biophys. Biomol. Struct.* **29**, 291–325 (2000).
78. Yates, L. A. et al. A structural and dynamic model for the assembly of replication protein A on single-stranded DNA. *Nat. Commun.* **9**, 5447 (2018).
79. Bochkarev, A., Pfuetzner, R. A., Edwards, A. M. & Frappier, L. Structure of the single-stranded-DNA-binding domain of replication protein A bound to DNA. *Nature* **385**, 176–181 (1997).
80. Bochkareva, E., Korolev, S., Lees-Miller, S. P. & Bochkarev, A. Structure of the RPA trimerization core and its role in the multistep DNA-binding mechanism of RPA. *EMBO J.* **21**, 1855–1863 (2002).
81. Xie, S. et al. Structure of RPA32 bound to the N-terminus of SMARCA1 redefines the binding interface between RPA32 and its interacting proteins. *FEBS J.* **281**, 3382–3396 (2014).
82. Case, D. A. et al. The Amber biomolecular simulation programs. *J. Comput. Chem.* **26**, 1668–1688 (2005).
83. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
84. Tuckerman, M., Berne, B. J. & Martyna, G. J. Reversible multiple time scale molecular dynamics. *J. Chem. Phys.* **97**, 1990–2001 (1992).
85. Kale, L. et al. NAMD2: Greater scalability for parallel molecular dynamics. *J. Comput. Phys.* **151**, 283–312 (1999).
86. Maier, J. A. et al. ff14SB: improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).
87. Galindo-Murillo, R. et al. Assessing the current state of amber force field modifications for DNA. *J. Chem. Theory Comput.* **12**, 4114–4127 (2016).
88. Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
89. McGibbon, R. T. et al. MDTraj: a modern open library for the analysis of molecular dynamics trajectories. *Biophys. J.* **109**, 1528–1532 (2015).
90. Girvan, M. & Newman, M. E. Community structure in social and biological networks. *Proc. Natl Acad. Sci. USA*. **99**, 7821–7826 (2002).
91. Roe, D. R. & Cheatham, T. E. III. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**, 3084–3095 (2013).

## Acknowledgements

We thank Wei Yang (NIDDK) and Orlando D. Schärer (UNIST, South Korea) for insightful discussions. This work was supported by the National Institute of General Medical Sciences grant R35GM139382 (I.I.), the National Institute of Environmental Health Sciences grant R01 ES032786 (I.I., S.E.T and C-L.T.), the National Science Foundation grant MCB-2027902 (I.I.), and NCI grants P01 CA092584 (J.A.T., I.I., S.E.T. and C-L.T.), R35 CA220430 (J.A.T.) and core funding by BAS/1/1002-01-01 to S.M.H. An award of computer time to I.I. was provided by the INCITE program. This research also used resources of the Oak Ridge Leadership Computing Facility, which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725.

## Author contributions

I.I. directed the study. J.Y., C.Y., S.E.T, and I.I. contributed to the design of the study. J.Y., C.Y., T.P. and L.B. performed model building and molecular simulations of the models. J.Y., C.Y., T.P., L.B., S.E.T., C-L.T., S.M.H., J.A.T and I.I. analyzed the data. S.E.T., C-L.T., S.M.H. and J.A.T provided critical comments during the analysis of the data. All authors discussed the results and were involved in the editing of the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-024-52860-y>.

**Correspondence** and requests for materials should be addressed to John A. Tainer or Ivaylo Ivanov.

**Peer review information** *Nature Communications* thanks Karl-Peter Hopfner and the other, anonymous, reviewer(s) for their contribution to the peer review of this work. A peer review file is available.

**Reprints and permissions 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 licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence 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 licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2024