

Peer Review File

Manuscript Title: Lesion recognition by XPC, TFIIH and XPA in DNA Excision Repair

Reviewer Comments & Author Rebuttals

Reviewer Reports on the Initial Version:

Referee #1:

Nucleotide excision repair (NER) is an important DNA repair pathway that removes bulky, helix destabilising lesions from DNA, either after their detection by specialised DNA damage sensor complexes or by stalled RNA polymerase II. In their manuscript, Kim et al. present structural and mechanistic analysis of lesion recognition and recruitment of TFIIH by the XPC complex and the structural transitions that occur in the complex following the recruitment of XPA.

The cryo-EM structures in this work are far superior to previous TFIIH-XPC structures, and based on these results, the manuscript under consideration extends the mechanistic understanding of the early steps of NER clearly beyond prior work. The work is important and interesting, and a candidate for publication in Nature. Specific comments listed below should be addressed before final evaluation of suitability for publication, although there are no major concerns.

Specific comments:

Page 3: "XPC protein comprises three subunits" - the "XPC protein" is XPC. What is referred to here is the XPC-RAD23-CETN2 complex or "trimeric XPC complex". The authors later introduce a disambiguation (XPC vs. trimeric XPC), which could be introduced here, as the statement as written on p. 3 is confusing.

Page 4: Double-Cy5 lesion mimic: Referencing Table S2 may be helpful (tables are referred to as Extended Data Tables in the manuscript but labeled Table SX in the PDF document).

Page 4: "The latest cryoEM structure of yeast Core7 and Rad4 complexed with a lesion DNA at 9 Å resolution suggests that the DNA lesion is placed between XPB and XPD, and thereby 5' to XPD 16. If so, the 5' to 3' translocation activity of XPD would result in XPD moving away from the lesion and being unable to carry out lesion verification." - This statement is potentially important, but not sufficiently and explicitly explained in the main text. More information (possibly including an Extended Data figure panel) should be added. This mechanistic analysis also has important implications on the location where lesions are verified in XPD, with a "lesion sensor pocket" suggested to reside near the FeS cluster of XPD. The authors should comment on this.

Page 7/Fig. 2a: It is not apparent from the figure that "only XPC directly contacts the DNA". The authors may need to provide an additional view (in the panel as shown, CETN2 appears to be sitting on the DNA as well).

Page 12: "heterodimer formed by p52 and p8 (Fig. 4f)" - Is that Fig. 3f?

Page 12: The authors suggest that two domains in p52 and the XPB NTD are DRDs (in addition to the "canonical" XPB DRD that has been described before). Can they provide superpositions (Extended Data figure panels) to show structural homology between all four domains?

Page 21: Nomenclature in "XPG-MBP, XPF and His-MBP-ERCC1" suggests a C-terminal MBP tag in XPG. Is that correct?

Extended Data Figs. 1 and 2: Model vs. map FSC curves need to be added.

Fig. 2b: R192 is referred to in the text and should be shown in the figure.

Fig. 4c: Please indicate how the nucleic acid bands were detected to the figure legend.

Referee #2:

In the submitted manuscript, Kim et al. describe the structures of NER intermediates comprised of XPC-RAD23-CENT2 trimer, core TFIIH (Core7) and XPA. Authors present three structures assembled on a lesion-mimicking DNA. The first structure shows the recruitment of Core7 to lesion-bound XPC trimer and reveals novel interactions between Core7 and XPC, as well as how XPC initially positions the Core7 relative to the lesion. The second structure additionally includes XPA that displaces the Core7 further away from XPC and kinks the DNA. The third structure results from XPC dissociation, reveals DNA entering into XPD and provides the basis for the model on DNA translocation during NER.

The structures provide novel insights into the first step of NER and are going to be valuable to the NER community. However, there are some technical and conceptual problems with the work presented. The most concerning is the C7AD structure that most likely does not represent displaced XPC during NER reaction but rather an artefact of Core7-XPA assembling on the DNA ends, and recapitulates previously solved structure of Core7-XPA on Y-shaped DNA (Kokic et al., Nat. Commun., 2019; further elaborated below). Since this structure is the basis for how authors describe the manipulation of DNA during NER, thereby significantly elevating the impact of their work, it is important that this is clarified.

Major issues

1. One very concerning issue is the C7AD structure. Authors claim that the structure is the result of an on-path XPC displacement by Core7-XPA. However, the structure is probably an artefact of Core7-XPA binding to the ends of DNA in their sample, and it is essentially a recapitulation of the Core7-XPA-DNA structure reported before (Kokic et al., Nat. Commun., 2019). There are several reasons why this might be the case:

(A) Authors make the dsDNA substrate from individually synthesized DNA strands. After annealing, the dsDNA has very heterogeneous ends (synthesis of such long DNA strands produces DNA of

various lengths, especially when internal modification is included, and the purity at a single nucleotide resolution is not achieved by PAGE purification provided by IDT). This means the DNA contains ds-ss junctions at the end.

(B) These junctions are very efficiently bound by Core7-XPA because XPB binds dsDNA and XPD ssDNA with high affinities, and the complex is stabilized by XPA.

(C) Authors do not see any density for the DNA past the junction (even when filtering the resolution of the EM density), further suggesting that DNA ends are bound. Moreover, the structure is identical to the previous Core7-XPA-DNA junction structure.

(D) The hairpin of XPA binds the junction in exactly the same way as observed when the complex was assembled on a lesion-free junction.

(E) Since the system was not supplemented with ATP, it does not make sense that DNA would open past the opening observed in the C7CAD complex, since DNA opening requires the activity of TFIIH ATPases (Coin et al., Mol. Cell, 2006).

(F) XPC displacement was reported to occur upon XPG recruitment (Riedl et al., EMBO J., 2003), which would make sense considering that XPG binds the same side on the DNA as XPC, yet XPG is not present in the reaction.

(G) The resolution is insufficient to sequence the DNA or demonstrate the presence of the lesion, so the identity of the intermediate containing displaced XPC over assembly at the DNA ends cannot be very well argued.

2. Map and model quality:

a) C7CD complex: The map densities for parts of XPC, centrin2, and the DNA are not of high quality compared to the rest of the complex, and that is not well presented in the paper. In some regions of interest even the secondary structures cannot be confidently placed in the density. This becomes particularly relevant when authors describe amino acid side chains interacting with the DNA around the lesion site. For example, authors claim T28 is sandwiched between Y656 and H685, yet the density in this region is fragmented and of suboptimal quality, and mentioned amino acids are far from being clearly resolved. Similarly, claiming that R192 that is mutated in cancer forms a slat bridge with a non-lesion strand phosphate is not possible based on the EM density – the density is not good enough to track the LHN register. Again, the claim that K839 and R843 contact DNA cannot be made not only due to limited resolution of the density for LHC, which does not allow for confident register tracking, but also because these side chains are positioned too far from the DNA. How was the XPC fragment (933-940) assigned to the density – the density does not appear good enough for de novo modelling? The same question extends to the identity and the register of XPC helices M and L.

b) C7CAD complex: A similar problem with bad density for the DNA and XPC 3' of the DNA lesion. Authors claim that DNA is kinked and unwound for 7 bp upstream of the lesion, yet the density in

this region is of insufficient quality to see the lesion, and DNA strands cannot be tracked with confidence, so how did the authors determine the extent of unpairing? The models also do not show unpaired 7 bp. Why is the density in the region so bad; it seems almost like oversampled noise? Such density interpretation is not a good practice; authors should show all those densities in the supplementary material with fitted models. Another important point for this structure is that W175 on the XPA hairpin intercalates into the DNA and causes the DNA to kink. How was the tip of the hairpin with W175 modelled: it is not fitting well into the density and W175 cannot even be seen? This is poorly modelled. Also, does this intercalation or the repositioning of XPC kink the DNA? Authors claim that XPA in C7CAD binds DNA duplex unlike ds-ss DNA junction in the previous structure. Yet, due to poor density in the region it is not clear if W175 inserts into locally unwound DNA that would mimic a minimal DNA junction.

3. One conceptual issue that was not really discussed is the TFIIH kinase module. Core7 is recruited to the lesion site as a holo-complex with the kinase module (Coin et al., Mol. Cell, 2008). The presence of the kinase module affects the structure of the Core7 (Greber et al., eLife, 2019) and clashes with bound factors (Kokic et al., Nat. Commun., 2019), and it is unclear how that relates to the structures presented. If authors claim to have solved the issue of TFIIH recruitment and the lesion hand-off during NER, this important point should be addressed. For example, why was the kinase module not included in the first structure?

4. It is also a bit problematic that authors extend their findings to TCR on several occasions to generalize their findings on lesion handoff yet there is currently no conclusive experimental data on how or when TFIIH is recruited during TCR. Using TFIIH positioning within PIC as a reference is misleading as the factors that position TFIIH during initiation are not present during TCR. This should be made very clear if authors want to pursue this comparison.

Other issues

1. (Lines 51-53) "How two different lesion-recognition pathways converge and how the XPB and XPD helicases of Core7 move lesion DNA for verification are unclear.": This was not exactly addressed in this manuscript; the role for TFIIH in TCR is not known or analysed in this work.

2. (Lines 56-57) "The DNA lesion is thus positioned outside of Core7, as also occurs with RNA polymerase": Do authors want to say that after XPA binding during GG-NER the positioning of TFIIH is the same as for TCR relative to the lesion? It is a strange comparison because the mode of lesion recognition is completely different; during TCR the lesion is trapped within RNA polymerase, and it is unknown how or when TFIIH is recruited.

3. (Lines 57-59) "push and pull the lesion-containing strand": Unclear what that means. Who is pushing, who is pulling, how does this push-and-pull work?

4. (Lines 76-78) Authors say that XPB enzymatic activity is required for the formation of PIC. This is not correct; PIC can form without any enzymatic activity by TFIIH.

5. (Lines 103-104) Authors state the affinity of XPC towards Cy5 DNA without showing any data. How

was the affinity determined? The same goes for the statement that core7 and XPA bind with lower affinity.

6. (Lines 113-114) "Previously unknown human trimeric XPC structure was determined at 3.9 Å in the C7CD complex.": Looking at the density it seems the resolution is exaggerated and varies strongly within the trimeric complex.

7. (Lines 124-125) If DNA is unwound for 7 bp it is not 'primed' to form ds-ss junction, but the junctions are already formed.

8. (Lines 130-131) "Dissociation of XPC from DNA may release RAD23 as cellular studies indicate.": I am not sure why this statement is placed here; this was not addressed by the study, and the structure does not reveal anything about it.

9. (Lines 220-221) "Due to severe unwinding...the major groove" "...": This sentence is unclear; how are these contacts explained by severe unwinding? How does severe unwinding differ from 'canonical' DNA unwinding?

10. (Lines 258-272) In the two paragraphs authors describe TFIIH structure that was already described in detail before (Greber, 2019; Kokic, 2019), without properly referencing the previous work. This includes the XPA-p8/p52 interaction that was also noted previously.

11. (Lines 303-311) "To test this hypothesis...": Firstly, the assay does not directly test their hypothesis about how XPB and XPD motors are working; it is basically a dual incision assay. Secondly, they did not develop a dual incision assay but just used a fluorescent substrate. This assay is being performed for decades now with same results.

I hope the comments are constructive and are going to help to further improve the manuscript.

Author Rebuttals to Initial Comments:

We deeply appreciate reviewers questions, comments and suggestions, which enable us to improve the manuscript extensively.

Referee #1:

Nucleotide excision repair (NER) is an important DNA repair pathway that removes bulky, helix destabilising lesions from DNA, either after their detection by specialised DNA damage sensor complexes or by stalled RNA polymerase II. In their manuscript, Kim et al. present structural and mechanistic analysis of lesion recognition and recruitment of TFIIH by the XPC complex and the structural transitions that occur in the complex following the recruitment of XPA.

The cryo-EM structures in this work are far superior to previous TFIIH-XPC structures, and based on these results, the manuscript under consideration extends the mechanistic understanding of the early steps of NER clearly beyond prior work. The work is important and interesting, and a candidate for publication in Nature. Specific comments listed below should be addressed before final evaluation of suitability for publication, although there are no major concerns.

We thank the reviewer for insightful comments and suggestions.

Specific comments:

Page 3: "XPC protein comprises three subunits" - the "XPC protein" is XPC. What is referred to here is the XPC-RAD23-CETN2 complex or "trimeric XPC complex". The authors later introduce a disambiguation (XPC vs. trimeric XPC), which could be introduced here, as the statement as written on p. 3 is confusing.

Yes, in the revised version we have modified the sentence as "The XPC complex is trimeric and composed of XPC, RAD23 and CETN2 subunits" on page 3.

Page 4: Double-Cy5 lesion mimic: Referencing Table S2 may be helpful (tables are referred to as Extended Data Tables in the manuscript but labeled Table SX in the PDF document).

Thanks for the suggestion. The diagram of Cy5 has been added to the revised Extended Data Fig. 2 and referred to on page 4. The typo in "Table SX" has been corrected.

Page 4: "The latest cryoEM structure of yeast Core7 and Rad4 complexed with a lesion DNA at 9 Å resolution suggests that the DNA lesion is placed between XPB and XPD, and thereby 5' to XPD 16. If so, the 5' to 3' translocation activity of XPD would result in XPD moving away from the lesion and being unable to carry out lesion verification." - This statement is potentially important, but not sufficiently and explicitly explained in the main text. More information (possibly including an Extended Data figure panel) should be added. This mechanistic analysis also has important implications on the location where lesions are verified in XPD, with a "lesion sensor pocket" suggested to reside near the FeS cluster of XPD. The authors should comment on this.

In the revised version, we have added ED Fig. 1a-b to show the previously published structures, and ED Fig. 1c-d, showing the new C7CD and C7CAD structures for comparison. The compiled image shows the different lesion locations relative to Fe₄S₄ cluster of XPD. In the text, we have added references to the

"lesion sensor pocket" near the FeS cluster and suggested an additional lesion-verification site at the exit point of XPD, which is shown in Fig. 4b.

Page 7/ Fig. 2a: It is not apparent from the figure that "only XPC directly contacts the DNA". The authors may need to provide an additional view (in the panel as shown, CETN2 appears to be sitting on the DNA as well).

An orthogonal view of figure 2a has been inserted into Fig. 2a, which shows that CETN2 and RAD23 are far away from DNA and stabilize LHN and LHC of XPC.

Page 12: "heterodimer formed by p52 and p8 (Fig. 4f)" - Is that Fig. 3f?

Thank you. We have corrected the typo. It is shown in new Fig. 3e (original 3f) and also shown in ED Fig. 8f.

Page 12: The authors suggest that two domains in p52 and the XPB NTD are DRDs (in addition to the "canonical" XPB DRD that has been described before). Can they provide superpositions (Extended Data figure panels) to show structural homology between all four domains?

In the revised manuscript a new ED Fig. 9 has been added to show these homologous DRD domains side-by-side after superposition.

Page 21: Nomenclature in "XPG-MBP, XPF and His-MBP-ERCC1" suggests a C-terminal MBP tag in XPG. Is that correct?

In the revised manuscript we have expanded the phrased as following "The C-terminal MBP-tagged XPG (XPG-MBP), XPF and the N-terminal His₆ and MBP tandemly tagged ERCC1 (His-MBP-ERCC1) were cloned".

Extended Data Figs. 1 and 2: Model vs. map FSC curves need to be added.

The agreements between model and map FSC have been added in ED Fig. 2 (original ED Fig. 1) and 3 (original ED Fig. 2).

Fig. 2b: R192 is referred to in the text and should be shown in the figure.

A stick model and label of R192 have been added to Fig. 2b. We have also included a zoom-in view of the region in revised ED Fig. 5b.

Fig. 4c: Please indicate how the nucleic acid bands were detected to the figure legend.

Dual incision of Cy5-DNA was detected by fluorescence of Cy5. The AP-DNA was internally ³²P labeled. Each lesion containing top strand was made by ligation of two to three oligos. These fragment top strand oligos and bottom ligation template oligo are listed in the revised ED Table 2. The procedure is briefly described in the revised Methods.

Referee #2:

In the submitted manuscript, Kim et al. describe the structures of NER intermediates comprised of XPC-RAD23-CENT2 trimer, core TFIIH (Core7) and XPA. Authors present three structures assembled on a lesion-mimicking DNA. The first structure shows the recruitment of Core7 to lesion-bound XPC trimer and reveals novel interactions between Core7 and XPC, as well as how XPC initially positions the Core7 relative to the lesion. The second structure additionally includes XPA that displaces the Core7 further away from XPC and kinks the DNA. The third structure results from XPC dissociation, reveals DNA entering into XPD and provides the basis for the model on DNA translocation during NER.

The structures provide novel insights into the first step of NER and are going to be valuable to the NER community. However, there are some technical and conceptual problems with the work presented. The most concerning is the C7AD structure that most likely does not represent displaced XPC during NER reaction but rather an artefact of Core7-XPA assembling on the DNA ends, and recapitulates previously solved structure of Core7-XPA on Y-shaped DNA (Kokic et al., Nat. Commun., 2019; further elaborated below). Since this structure is the basis for how authors describe the manipulation of DNA during NER, thereby significantly elevating the impact of their work, it is important that this is clarified.

We thank the reviewer for carefully examining the structures and suggestions for clarification. We see where our descriptions were unclear and have revised the manuscript accordingly. In this manuscript, the most important findings are in the first two structures, the initial recruitment of Core7 by XPC (C7CD) and reposition of the lesion relative to Core7 by XPA (C7CAD). The third structure is very similar to and a step prior to the Y-shaped non-lesion DNA complexed with Core7 and XPA (Kokic et al., 2019) before XPC is fully dissociated and ssDNA is bound by XPD. The main differences of C7AD from Kokic's are the presence of a lesion mimic in our DNA duplex and the absence of a long ssDNA bound by XPD. We agree with the reviewer that XPC is not fully disassociated in the third structure (C7AD) because (a) C7AD is a subspecies of the second structure (C7CAD), (b) the C-terminal region of the XPC subunit is still anchored in Core7 and the rest of XPC present as a mobile blob in the cryoEM map, and (c) only 2 nt of the lesion strand have been detected at the ssDNA exit of XPD (adjacent to XPA) in C7AD. We are rather sure that the C7AD structure is not an artefact of Core7-XPA assembling on the DNA ends. In the revised version, we clarify that XPC is not fully disassociated from Core7 and the lesion DNA strand is not bound by XPD yet. We also have clarified the similarity and difference between Kokic's and C7AD structures.

Major issues

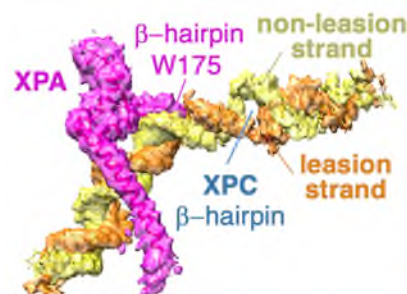
1. One very concerning issue is the C7AD structure. Authors claim that the structure is the result of an on-path XPC displacement by Core7-XPA. However, the structure is probably an artefact of Core7-XPA binding to the ends of DNA in their sample, and it is essentially a recapitulation of the Core7-XPA-DNA structure reported before (Kokic et al., Nat. Commun., 2019). There are several reasons why this might be the case:

Indeed, our lesion-containing C7AD structure resembles the lesion-free Y-shaped DNA-Core7-XPA complex structure reported by Kokic et al. But C7AD is not the same as Kokic's structure because XPC is not dissociated yet and ssDNA is not bound by XPD. Furthermore, the following experimental results argue against the hypothesis that our structure is an artefact due to XPA's binding to the DNA end.

Firstly, Kokic et al. used a Y-shaped DNA with unpaired 20 nt on each strand appended to a 29 bp duplex to obtain the DNA-Core7-XPA complex structure. They couldn't obtain a complex structure with a blunt-ended DNA. In our study a lesion is in the middle of a DNA duplex and flanked by 25 and 29 bp on each

side. As shown in the EMSA assay (Fig. 1c) without XPC, XPA and Core7 bind such duplexes very poorly and cannot form a stable DNA complex. Core7 and XPA are recruited onto a lesion-containing DNA only in the presence of XPC.

Secondly, in our C7CD and C7CAD structures, DNA duplexes are unambiguously detected surrounding the lesion and interacting with XPB and XPC (Fig. 1d-e). Moreover, in C7CAD, XPA binds the middle of a DNA duplex, not the DNA end. The map covering XPA and the DNA duplex is shown on the right. The C7AD (third) structure is isolated as a subspecies of C7CAD on cryoEM grids (ED Figures 2 and 3). C7AD and C7CAD are nearly identical except for the displaced (not dissociated) XPC in C7AD. We still can see the incompletely dissociated XPC in the C7AD map. Therefore, we believe that Kokic's structure is a step after C7AD, with XPC dissociated and ssDNA formed and bound to XPD.



(A) Authors make the dsDNA substrate from individually synthesized DNA strands. After annealing, the dsDNA has very heterogeneous ends (synthesis of such long DNA strands produces DNA of various lengths, especially when internal modification is included, and the purity at a single nucleotide resolution is not achieved by PAGE purification provided by IDT). This means the DNA contains ds-ss junctions at the end.

We had the same concern as the reviewer. Therefore, we purified each oligo synthesized by IDT ourselves using denaturing PAGE gels. We further purified annealed DNA duplex on native gels to eliminate excess ssDNA. Failed oligos of 1-5 nt shorter than the intended at each 5' end may still exist after purification, but the resulting 1-5 nt 3' overhang at each DNA end is unlike Kokic's Y-shaped DNA and cannot be bound by XPC for Core7 and XPA recruitment.

Further supporting the purity of our DNA substrate, 50 out of the total 56 bp DNA duplex (including the Cy5 in the middle) are observed in the cryoEM map of C7CD (Fig. 1d). Even though the map resolution at the DNA ends is ~ 7 Å, the rod-shaped DNA of 50 bp in length is apparent.

(B) These junctions are very efficiently bound by Core7-XPA because XPB binds dsDNA and XPD ssDNA with high affinities, and the complex is stabilized by XPA.

ssDNA is absent in all our structures. dsDNA ends may be bound by XPA alone as shown with yeast homolog Rad14 by Kisker and Carell et al., but to form Core7 and XPA complex on DNA, Kokic et al. needed to use a Y-shaped DNA containing 29 bp duplex followed by 20 nt ssDNA strands. As described above, (1) in our EMSA assay (Fig. 1c) the Cy5-DNA substrate we used in this study do not form a discrete stable complex with Core7 and XPA, and (2) in the C7CD and C7CAD maps, DNA duplexes are observed to surround the DNA lesion and XPA binding site.

(C) Authors do not see any density for the DNA past the junction (even when filtering the resolution of the EM density), further suggesting that DNA ends are bound. Moreover, the structure is identical to the previous Core7-XPA-DNA junction structure.

In C7CAD (the second structure), we observe DNA duplexes flanking both sides of the lesion, Cy5 or AP sites (please see the figure on this page), which dispel the notion that our Cy5-DNA contains a ds-ss junction. The fact that we don't observe ssDNA bound by XPD in C7AD, unlike Kokic's structure, supports that the DNA is not a ds-ss junction as in Kokic's. In the C7AD (the third structure), the absence of ssDNA

bound by XPD is because the DNA handing off from XPC to XPD is incomplete. We still observed the C-terminal 50 residues of XPC bound to Core7 (Fig. 1f). The “disappearance of” the rest of XPC along with the lesion and downstream DNA duplex is due to their mobility and incomplete strand separation and DNA handoff.

To clarify this point, when first reporting C7AD on p. 5, we have revised the text to indicate that XPC is “displaced but not fully dissociated”. On p. 6, when describing the C7AD structure, we added the explanation for disordered Cy5 and surrounding DNA in C7AD as “due to their mobility and incomplete transfer from XPC to XPD”.

(D) The hairpin of XPA binds the junction in exactly the same way as observed when the complex was assembled on a lesion-free junction.

Indeed, XPA binds the DNA in C7AD (the third structure) as in Kokic’s Y-shaped DNA. However, the hairpin of XPA also binds the kinked and locally unwound DNA duplex in C7CAD as shown in the figure on p. 4 here. Therefore, one cannot tell if the DNA is a ds-ss junction or a DNA duplex based on XPA binding alone. The requirement of XPC (here) and DNA structures (DNA duplex containing a lesion mimic vs. a Y-shaped DNA) are critical for consideration.

(E) Since the system was not supplemented with ATP, it does not make sense that DNA would open past the opening observed in the C7CAD complex, since DNA opening requires the activity of TFIIH ATPases (Coin et al., Mol. Cell, 2006).

We agree with the reviewer. Our C7AD is not the same as Kokic’s, where the single stranded DNAs pre-existed in the Y-shaped substrate and one of the ssDNAs is bound by XPD. Because of the absence of ATP and other NER cofactors, we do not observe ssDNA except for 2 nt bound by XPD in our C7AD structure because XPC held on to the dsDNA downstream of the lesion. However, the similarity of C7AD to Kokic’s structure suggests that in the presence ATP and other NER cofactors, a large DNA bubble would form and our lesion-containing duplex will melt eventually as in the Kokic’s structure.

(F) XPC displacement was reported to occur upon XPG recruitment (Riedl et al., EMBO J., 2003), which would make sense considering that XPG binds the same side on the DNA as XPC, yet XPG is not present in the reaction.

We agree with the reviewer. Because XPG, RPA, XPF and ATP are not included in this study, the handoff of DNA from XPC to Core7 is incomplete. The C-terminal helices of XPC are still anchored in Core7. The N-terminal four domains of XPC and its associated DNA (the lesion and downstream duplex) are mobile in the vicinity of Core7 in the C7AD complex as observed in the cryoEM map volume. To clarify this point, we have revised the text to state that XPC is “not fully dissociated” and DNA is “incompletely handed off to Core7” on p. 5 and 6. Interestingly, XPG was included in Kokic’s TFIIH, XPA and Y-shaped DNA complex, but XPG was “invisible” still.

(G) The resolution is insufficient to sequence the DNA or demonstrate the presence of the lesion, so the identity of the intermediate containing displaced XPC over assembly at the DNA ends cannot be very well argued.

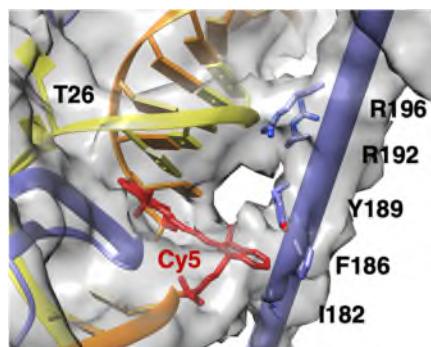
The above statement should be qualified. The moderate resolution of 3.5 to 6 Å is insufficient to sequence DNA but sufficient to determine (a) the location and length of DNA duplex, (b) whether and where DNA is

kinked, (c) where the two DNA strands are separated, (d) where the flipped-out DNA bases are, and (e) where the lesion is based on the unpaired DNA strands and flipped-out bases opposite it. The map also unequivocally shows that the DNA kink generated by the XPA hairpin is surrounded by base paired DNA duplex on both sides (please see the figure on p. 4). There is no doubt where the DNA lesion is because the XPC-DNA volume in the C7CAD map (at lower resolution than that of C7CD) has the same shape as and can be superimposed with the map volume of C7CD. As XPC only binds DNA duplex downstream of a DNA lesion and bases opposite the lesion in the non-lesion strand are always flipped out, we know where the lesion is even in the absence of high-resolution maps. In all crystal structures of yeast XPC (Rad4), there is no electron density corresponding to the DNA lesions, but we have no trouble identifying where the lesion is despite it being disordered.

2. Map and model quality:

a) C7CD complex: The map densities for parts of XPC, centrin2, and the DNA are not of high quality compared to the rest of the complex, and that is not well presented in the paper. In some regions of interest even the secondary structures cannot be confidently placed in the density. This becomes particularly relevant when authors describe amino acid side chains interacting with the DNA around the lesion site. For example, authors claim T28 is sandwiched between Y656 and H685, yet the density in this region is fragmented and of suboptimal quality, and mentioned amino acids are far from being clearly resolved. Similarly, claiming that R192 that is mutated in cancer forms a slat bridge with a non-lesion strand phosphate is not possible based on the EM density – the density is not good enough to track the LHN register. Again, the claim that K839 and R843 contact DNA cannot be made not only due to limited resolution of the density for LHC, which does not allow for confident register tracking, but also because these side chains are positioned too far from the DNA. How was the XPC fragment (933-940) assigned to the density – the density does not appear good enough for de novo modelling? The same question extends to the identity and the register of XPC helices M and L.

We have improved the map of troubled regions around the N- and C-terminal part of XPC-DNA, which are shown below as well as in Extended Data figure 5. After additional masked classifications to sort out conformation variations and further refinement of our structure models, the structural interpretation remains the same for C7CD and the other two complexes. As the reviewer suggested, we have added the following sentence when introducing the XPC structure, “Due to the mobility, peripheral domains and DNA ends are at lower resolutions” on p. 5 and pointed out that “the flexibility of XPC and XPC-DNA interactions resulted in the reduced resolution” on p. 8. We have also removed K839 and R843 in the description and discussion because of the uncertainty in the revised manuscript.

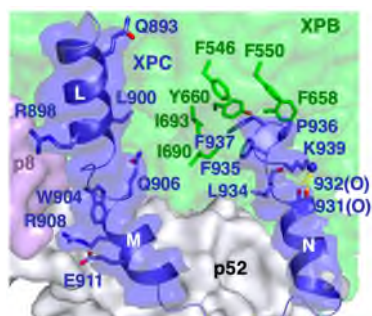
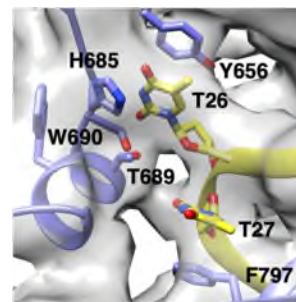


It is a common practice that structural biologists combine other information and knowledge to make reliable interpretations. Taking a page from X-ray crystallography in structure refinement, when experimental data are insufficient (which are >99% cases), the known protein geometry is used to supplement and improve model quality. Furthermore, the correctness of a structural model is often tested and validated by whether the model agrees with mutagenic, biochemical or genetic data. In our case we are greatly helped by the AlphaFold prediction of the four structural domains of XPC as well as helical nature of LHN, LHC and the C-terminal L,

M and N. Even at moderate resolutions, these helices are clearly defined, sometimes with large sidechains sticking out, which allow us to assign protein sequence.

R192 is assigned to contact the non-lesion strand based on regularity of the alpha helical structure and the hydrophobic sidechains on LHN (I182, F186 and Y189) surrounded the hydrophobic moiety of Cy5 (the lesion strand). The close contacts between LHN and both DNA strands across the minor groove are obvious even in the moderate-resolution map, which are shown on the previous page. [In the revised manuscript](#), the above figure is included as ED. Fig. 5b. The known defects of R192G mutation is consistent with the placement of R192 adjacent to the DNA backbone. We use “appears” [in the revised text](#) to qualify our structural model.

Given that the structure of human XPC is previously unknown, we believe that one ought to interpret the moderate resolution map (4-6 Å) as best one can but take the result with a grain of salt. The example of T26 with the improved experimental map is shown on the right. Taking good stereo geometry and hydrophobicity of bases and protein sidechains into account, we don't find a credible interpretation other than that the T26 base is sandwiched between Y656 and H685. To acknowledge the uncertainty, we have [revised the text](#) as “T26 appears sandwiched between”. This figure is included as ED Fig. 5a.



The C-terminal 50 residues of XPC are assigned as helices L, M and H based on the improved cryoEM map, which is shown on the left superimposed onto the structural model of XPC (blue). A few sidechains are well defined. The penultimate C-terminal residue K939, which caps the C terminus of N helix, is anchored by the proceeding hydrophobic residues in XPB at the interface of XPB, p52 and p8. [In the revised manuscript](#) a simplified version of this figure appears as Fig. 3e, and the same figure as shown here is added as Extended Data Figure 5d.

b) C7CAD complex: A similar problem with bad density for the DNA and XPC 3' of the DNA lesion. Authors claim that DNA is kinked and unwound for 7 bp upstream of the lesion, yet the density in this region is of insufficient quality to see the lesion, and DNA strands cannot be tracked with confidence, so how did the authors determine the extent of unpairing? The models also do not show unpaired 7 bp.

We are sorry if the sentence of “kinked and unwound (at) 7 bp upstream of the Cy5” is taken as “unpaired for 7 bp”. The kink occurs 7 bp upstream and is thus 5' and not 3' to the lesion in C7CAD. The corresponding map volume is shown in Fig. 1e and also here on p. 4. We [have revised the text](#) as “kinked and locally unwound (at) 7 bp upstream of the Cy5”.

Why is the density in the region so bad; it seems almost like oversampled noise? Such density interpretation is not a good practice; authors should show all those densities in the supplementary material with fitted models. Another important point for this structure is that W175 on the XPA hairpin intercalates into the DNA and causes the DNA to kink. How was the tip of the hairpin with W175 modelled: it is not fitting well into the density and W175 cannot even be seen? This is poorly modelled. Also, does this intercalation or the repositioning of XPC kink the DNA? Authors claim that XPA in C7CAD binds DNA duplex unlike ds-ss DNA junction in the previous structure. Yet, due to poor density in the region it is not clear if W175 inserts into locally unwound DNA that would mimic a minimal DNA junction.

It is well known that macromolecular mobility decreases structure resolution. It is often the most interesting part of a complex has the lowest resolution. The XPC-DNA portion and the β -hairpin of XPA

are no exception. We have rebuilt and improved the fitting of W175 and the β -hairpin of XPA. Although the resolution is 4 Å and map connectivity has a break at Gly176 mainchain of XPA, but the β -strands before (K163-K169) and after (D177-L182) and the rest of the $\alpha\beta$ domain fit the cryoEM map very well and is superimposable with Kokic's structure. In the revised manuscript, we have included the figure shown on p. 4 as ED Fig. 5c. In the figure legend, we have included the following explanation, "The unambiguous docking of the $\alpha\beta$ domain including β -strands before (K163-K169) and after (D177-L182) constrains the position of the 7-residue β -hairpin (P170 – G176) of XPA to the DNA kink site."

We are not sure what "oversampled noise" is. However, we calculated various low-pass maps, in which noises would disappear, and found that connectivity is invariably improved in our maps rather than degraded, which suggests that our map volumes are not due to noise. In the revised manuscript, we point out the correlation between the mobility and reduced resolution.

3. One conceptual issue that was not really discussed is the TFIIH kinase module. Core7 is recruited to the lesion site as a holo-complex with the kinase module (Coin et al., Mol. Cell, 2008). The presence of the kinase module affects the structure of the Core7 (Greber et al., eLife, 2019) and clashes with bound factors (Kokic et al., Nat. Commun., 2019), and it is unclear how that relates to the structures presented. If authors claim to have solved the issue of TFIIH recruitment and the lesion hand-off during NER, this important point should be addressed. For example, why was the kinase module not included in the first structure?

We thank the reviewer for the suggestion to discuss the CAK module in our comparison between TCR and GGR. The CAK kinase module inhibits NER as shown by many groups including us. Coin's Mol Cell paper showed that upon XPA binding CAK is dissociated from TFIIH (reference #20 in the original submission). In the revised Introduction, we add "XPD's DNA binding is blocked by the CAK module" and "CAK dissociation upon XPA binding" on p. 3. On p. 5, when introducing C7CD complex, we have added the following sentence. "If present, the CAK module would not interfere the C7CD formation". In the following paragraph describing the C7CAD complex, we have added that C7CAD forms "in the presence of XPA and absence of CAK", and as a result, "XPD contacting both DNA and XPB" with reference to Coin et al.

4. It is also a bit problematic that authors extend their findings to TCR on several occasions to generalize their findings on lesion handoff yet there is currently no conclusive experimental data on how or when TFIIH is recruited during TCR. Using TFIIH positioning within PIC as a reference is misleading as the factors that position TFIIH during initiation are not present during TCR. This should be made very clear if authors want to pursue this comparison.

It is unfortunate that there is no experimental data on how and what recruits TFIIH in TCR. However, the interactions between TFIIH and RNA polymerase in PIC foretells how TFIIH may be recruited when RNA polymerase is stalled by a transcription-blocking lesion (see revised Figure 5). We find that TFIIH (Core7) is similarly oriented whether interacting with RNAP (TCR) or XPC (GGR), and that the location of lesion downstream of XPD (Core7) in GGR (with XPA's help) appears to be the same as in TCR. These similarities answer the main question in NER as how TFIIH binds to a lesion DNA, scan the lesion strand, promotes and coordinate NER in the presence of a bulky lesion. As the reviewer suggests, in the revised discussion we emphasize that "Although it is unknown how TFIIH is recruited to a stalled transcription site, TFIIH likely interacts with RNA polymerase as in PIC with the lesion between them. After XPA dissociates the CAK module²⁰, Core7 would bind the lesion DNA as if recruited by XPC". In the Abstract we used "would occur" instead of "also occurs" in TCR because TCR has not been observed.

Other issues

1. (Lines 51-53) “How two different lesion-recognition pathways converge and how the XPB and XPD helicases of Core7 move lesion DNA for verification are unclear.”: This was not exactly addressed in this manuscript; the role for TFIIH in TCR is not known or analyzed in this work.

As discussed in the previous paragraph, even in the absence of structures of TCR complexes, the PIC (preinitiation complex of transcription) structures foretell how TFIIH may be recruited for TCR because of the similarity between TFIIH loaded by XPC and XPA onto a lesion strand and TFIIH bound to the RNA polymerase in PIC. Without the structures of GGR complexes, one cannot predict how TFIIH is oriented in TCR. Based on previously published data and our structures, we have revised Fig. 5 and added arrowheads to indicate the movement direction of XPB and XPD on DNA. Kocic’s structure, in which XPC is dissociated and ssDNA bound by XPD, is included in revised Fig. 5.

2. (Lines 56-57) “The DNA lesion is thus positioned outside of Core7, as also occurs with RNA polymerase”: Do authors want to say that after XPA binding during GG-NER the positioning of TFIIH is the same as for TCR relative to the lesion? It is a strange comparison because the mode of lesion recognition is completely different; during TCR the lesion is trapped within RNA polymerase, and it is unknown how or when TFIIH is recruited.

DNA lesions (Cy5 and AP) are positioned outside of Core7 in C7CAD complexes. In the revised manuscript, we have added ED Fig. 1 to compare locations of the lesion relative to the FeS cluster in XPD among C7CD, C7CAD and previously published structures. Although the detail of how TFIIH is recruited by a stalled RNA polymerase is unknown, both XPC (GGR) and RNA polymerase (TCR) bind the DNA downstream of the lesion, and TFIIH (Core7) are placed on the upstream of the lesion by XPC and XPA in GGR and by RNA polymerase in PIC. RNA polymerase can be regarded as “stalled” at a transcription initiation site. Yes, in conclusion we suggest that TFIIH-RNA polymerase interaction is likely to be repeated in TCR as in PIC. In the revised discussion, we have written the following, “TFIIH likely interacts with RNA polymerase as in PIC with the lesion between them” and hope that the meaning is clear.

3. (Lines 57-59) “push and pull the lesion-containing strand”: Unclear what that means. Who is pushing, who is pulling, how does this push-and-pull work?

In the abstract, we wrote simply “XPB and XPD, which track the lesion-containing strand but translocate DNA in opposite directions, push and pull the lesion-containing strand into XPD for verification”. As which pulls and which pushes, it depends on whether the protein complex or the DNA substrate is the frame of reference. The relativity is illustrated in Fig. 4a and 4b. The push-and-pull mechanism is detailed in the last section under the title “C7AD structure and implication for DNA translocation”. In short, in a stationary DNA surveillance complex, after XPD (in Core7) is loaded onto a ssDNA in a small bubble (~ 10 nt), it overpowers the weaker XPB and pulls downstream ssDNA into Core7 for verification. When a bulky lesion stalls the XPD motor, XPB becomes the only working motor and pushes the dsDNA from upstream towards XPD, and XPA separates the duplex, thus enlarging the DNA bubble to over 20 nt in size for XPF and XPG nucleases to make dual incision. “Push” by XPB was used in the original submission, and “pulling” by XPD is added in the revised text.

4. (Lines 76-78) Authors say that XPB enzymatic activity is required for the formation of PIC. This is not correct; PIC can form without any enzymatic activity by TFIIH.

We thank the referee for the correction and have revised the text as following. “Only the ATP-dependent

dsDNA translocase activity of XPB is necessary for transcription initiation ¹⁷. In the transcription pre-initiation complex (PIC) XPD serves as a structural partner.”

5. (Lines 103-104) Authors state the affinity of XPC towards Cy5 DNA without showing any data. How was the affinity determined? The same goes for the statement that core7 and XPA bind with lower affinity.

The data is shown in Fig. 1c by EMSA. The concentrations of DNA and protein in our EMSA assay were calibrated to 10 nM each, and the result is fully shifted DNA by XPC. Since $K_d = [\text{prot}] \times [\text{DNA}] / [\text{Prot-DNA}]$, if K_d were 5 nM, only a half of DNA would be shifted by XPC. With the observed full shift, K_d has to be below 5 nM. In the revised manuscript, DNA and XPC concentrations are indicated in the figure legend, and the following sentence is added in the revised Methods,” K_d of trimeric XPC and lesion DNA was estimated to be < 5 nM because 10 nM of each resulted in nearly full shift of the lesion DNA”.

6. (Lines 113-114) “Previously unknown human trimeric XPC structure was determined at 3.9 Å in the C7CD complex.”: Looking at the density it seems the resolution is exaggerated and varies strongly within the trimeric complex.

After reprocessing the Cy5 data, the resolution of XPC trimer and DNA map is estimated to be 4.1 Å, slightly worse than 3.9 Å. All resolutions are estimated according to the “Gold Standard” of 0.143 Fourier Shell Correlation (FSC) criterion, which is shown in ED Figures 2 and 3 (1 and 2 in the original submission). Resolution variation throughout each complex is shown in panels ED Fig. 2d and 3d. The reported resolution reflects the entire complex and not the worst or the best part of the complex. For example, we find that after reprocessing the XPC map, features of the DNA end and LHN and TGD of XPC are improved, while the map resolution is reduced by 0.2 Å. Revised data processing is shown in revised ED Fig. 2.

7. (Lines 124-125) If DNA is unwound for 7 bp it is not ‘primed’ to form ds-ss junction, but the junctions are already formed.

The unwinding is observed locally 7 bp upstream of the lesion. Therefore, we predicted that the DNA kink and local unwinding by XPA 7 bp upstream of the lesion would lead to unwinding and formation of ~10 nt bubble (between the kink and the lesion) in the presence of ATP and other NER factors. In the revised text, we have clarified the requirement of ATP and other NER factors for the bubble and ssDNA formation.

8. (Lines 130-131) "Dissociation of XPC from DNA may release RAD23 as cellular studies indicate.": I am not sure why this statement is placed here; this was not addressed by the study, and the structure does not reveal anything about it.

We agree and have removed the sentence in the revised text.

9. (Lines 220-221) "Due to severe unwinding...the major groove"...: This sentence is unclear; how are these contacts explained by severe unwinding? How does severe unwinding differ from ‘canonical’ DNA unwinding?

The sentence in question is to explain why with a distance of 35 Å between XPA (α/β domain) and XPC (BHD3), which equals to a full helical turn of B-DNA and should be major groove to major groove or minor groove to minor groove, XPA binds the minor groove and XPC binds the adjacent major groove. We have added “(a full helical turn)” after 35 Å and clarified the sentence in the revised text as following. “But due

to the severe DNA unwinding at the lesion site, the β -hairpin of XPA contacts the minor groove, while the β -hairpin of BHD3 contacts the adjacent major groove”.

10. (Lines 258-272) In the two paragraphs authors describe TFIIH structure that was already described in detail before (Greber, 2019; Kokic, 2019), without properly referencing the previous work. This includes the XPA-p8/p52 interaction that was also noted previously.

We thank the reviewer and cited Kokic’s paper when describing the XPA-p8/p52 interface in the revised text. In Kokic’s structure, XPA model ends before it reaches p8/p52. In the supplemental information, they showed mass spectrometry analysis and the low-resolution volume of the XPA C-terminus reaching p8/p52. In the revised text we have revised the sentence as following, “The linker (aa 234-243) and C-terminal β domain of XPA (aa 244-273) consecutively interact with XPB, p8 and p52 (Fig. 3c) as suggested in previous analyses⁵”.

We are grateful to the published work, particularly Kokic’s, and cited Kokic and Greber papers in the original manuscript (Ref. 19 and 5) as well as papers reporting improved resolution of TFIIH in yeast and mammalian PIC structures (Ref. 6-8). The key difference is that previously reported TFIIH structures cannot be easily compared because of the different setting and pathways, PIC vs. NER. Here the C7CD (U-shaped) and C7CAD (σ -shaped) co-existed on the same DNA and on the same cryoEM grids and thus allowed us to compare structural changes and glean meanings of these changes.

11. (Lines 303-311) “To test this hypothesis...”: Firstly, the assay does not directly test their hypothesis about how XPB and XPD motors are working; it is basically a dual incision assay. Secondly, they did not develop a dual incision assay but just used a fluorescent substrate. This assay is being performed for decades now with same results.

Previous dual incision assays were all carried out with circular and often supercoiled DNAs of over thousands of base pairs, and the dual incision efficiency is rather low, average round 10% of substrate turning into products. We developed the oligo-based dual incision assay with substrates <100 bp and up to 80 % substrate being converted to product. To clarify what we mean by validating the push-and- pull model, we have included the raw data for different upstream and downstream DNA lengths and showed markedly reduced dual incision activities when the upstream DNA is shortened from 62 bp to 52 bp in ED Figure 10. To qualify the development we have revised the text as following, “we developed an in vitro dual incision assay, replacing circular plasmid DNA substrate^{22,56,57} with a linear 94 bp DNA containing either a Cy5 or two consecutive AP sites”.

Reviewer Reports on the First Revision:

Referee #1:

The authors have revised their manuscript on the mechanism of DNA lesion recognition and hand-over in nucleotide excision repair - work that represents an important and timely contribution to the field of DNA repair - according to comments of two reviewers. The authors have addressed my comments satisfactorily. The second reviewer provided additional insightful remarks and observations and should be given the opportunity to require further clarifications if needed.

Referee #2:

In response to the review the authors modified the text of the manuscript, included additional figures, and further processed their data. The responses are well written and to a large extent address the issues well, and I applaud the authors for their effort to further improve their manuscript. However, I am still not convinced on the nature of the C7AD intermediate. Although the authors claim the first two structures are more important, this intermediate is a third of their data, and it is mechanistically interpreted. The fact that the authors do see a few nucleotides of ssDNA entering XPD, but no other parts of the following DNA, is the strongest argument in favor of complex formation on DNA ends with very short overhangs. This sub-population of particles might not be detectable via EMSA experiments and might be further stabilized by crosslinker present during sample preparation for EM. It is unclear what would displace XPC in this complex if the complex was indeed assembled around the lesion site. If no ssDNA (or ATP) is present in their system, as the authors argue, what drives the DNA transfer from XPC to XPD? In my opinion, the possibility of C7AD complex formation on heterogenous DNA ends (meaning very short overhangs) cannot be excluded.

Author Rebuttals to First Revision:

Reviewer 2's comments:

In response to the review the authors modified the text of the manuscript, included additional figures, and further processed their data. The responses are well written and to a large extent address the issues well, and I applaud the authors for their effort to further improve their manuscript.

We are glad that review #2 approves our revision and responses.

However, I am still not convinced on the nature of the C7AD intermediate. Although the authors claim the first two structures are more important, this intermediate is a third of their data, and it is mechanistically interpreted. The fact that the authors do see a few nucleotides of ssDNA entering XPD, but no other parts of the following DNA, is the strongest argument in favor of complex formation on DNA ends with very short overhangs.

As detailed in our previous response and shown in the experimental data, C7AD is a sub-population of C7CAD, in which XPC and the DNA lesion are distinct. In C7AD, the C-terminal 50 residues of XPC are anchored in the Core7 of TFIIH (C7). The remaining XPC and associated DNA lesion are visible in the cryoEM map as a mobile blob and thus disordered, but not dissociated from Core7. These observations indicate that Core7 is bound upstream of the lesion and not at the DNA end.

This sub-population of particles might not be detectable via EMSA experiments and might be further stabilized by crosslinker present during sample preparation for EM.

To structurally observe a macromolecular complex, the complex has to be stable with a defined conformation. We are unaware of any example of a protein-DNA complex that cannot be detected by EMSA assay and yet is observed and recorded by cryoEM. Crosslinking is to stabilize existing complexes during freezing and preparation of cryoEM grids.

In our EMSA assay with lesions embedded in the middle of DNA duplexes, Core7 binds the DNA only in the presence of XPC, which agrees with the observation that XPC is mobile and disordered but remains attached in C7AD.

It is unclear what would displace XPC in this complex if the complex was indeed assembled around the lesion site. If no ssDNA (or ATP) is present in their system, as the authors argue, what drives the DNA transfer from XPC to XPD?

As noted in the revised text and in the previous response, XPC is displaced but not dissociated from Core7 in C7AD. This is because the DNA upstream of the lesion is firmly bound by XPB and XPA as shown in C7CD and C7CAD, but the lesion and downstream DNA is still held by XPC. The 2 nt bound by XPD represent the partition point of DNA between Core7 and mobile XPC. As the reviewer has pointed out, only in the presence of ATP and NER cofactors, can the lesion be completely transferred from XPC to XPD and Core7. We are currently characterizing NER structures after C7AD. We have observed complete dissociation of XPC and ssDNA bound by XPD after incubation with ATP and XPG, XPF and RPA. Therefore, C7AD is on path from C7CAD to Kokic's structure of the artificial ds- and ss-DNA junction bound by Core7 and XPA.

In my opinion, the possibility of C7AD complex formation on heterogenous DNA ends (meaning very short overhangs) cannot be excluded.

The reviewer is correct that we cannot completely exclude this possibility. However, the successful DNA dual incision of NER (as shown in Fig 4C and ED Fig. 10 of our manuscript) supports that most of Core7 must bind upstream of lesions to verify the lesion nature and activate NER.

As we cannot exclude the possibility, the following sentence has been added on p. 6 when C7AD is first described. ““we cannot rule out that some C7AD complexes may have resulted from Core7 and XPA binding to dsDNA ends instead of XPC associated lesions”.