THE EMBO JOURNAL

Manuscript EMBO-2017-97833

Selective base excision repair of DNA damage by the non-base-flipping DNA glycosylase AlkC

Rongxin Shi, Elwood A. Mullins, Xing-Xing Shen, Kori T. Lay, Philip K. Yuen, Sheila S. David, Antonis Rokas & Brandt F. Eichman

Corresponding author: Brandt F. Eichman, Vanderbilt University

Review timeline:

Submission date:	22 July 2017
Editorial Decision:	28 August 2017
Revision received:	11 September 2017
Accepted:	22 September 2017

Editor: Hartmut Vodermaier

Transaction Report:

(Note: With the exception of the correction of typographical or spelling errors that could be a source of ambiguity, letters and reports are not edited. The original formatting of letters and referee reports may not be reflected in this compilation.)

1st Editorial Decision 28 August 2017

Thank you for submitting your manuscript on AlkC glycosylase structure for our editorial consideration. We have now received the comments of three expert referees, copied below for your information. As you will see, all referees consider your results interesting and potentially important, but they also raise a number of issues and queries that would need to be satisfactorily answered prior to eventual publication. In this regard, it would appear particularly important to clarify an apparent conformational variability and its relevance for interpreting the presented 'active conformation' structure (see referees 2 and 3); as well as to follow up on the significance of the Ig domain (and its lack in the AlkCa family) through additional mutagenesis and modeling approaches as requested by referees 1 and 3.

Should you be able to adequately address these key point, as well as the various more specific comments raised in all three reports, then we should be happy to consider a revised manuscript further for publication in The EMBO Journal. However, please bear in mind that it is our policy to allow only a single round of major revision, making it important to carefully respond to all points raised during this round. Should you have any additional questions/comments regarding the referee reports or the revision requirements, please therefore do not hesitate to get in touch with me ahead of resubmission. If needed, we might also extend the revision period, during which publication of any competing work elsewhere would have no negative impact on our final assessment of your own study.

Thank you again for the opportunity to consider this work for The EMBO Journal! I look forward to your revision.

REFEREE REPORTS

Referee #1:

Many glycosylases are known to catalyze the cleavage of an N-glycosidic bond through a base-flipping mechanism, independent of the substrate specificity or damage recognition. However, in previous studies, the authors showed that AlkD with a HEAT-like repeat fold excises the small or bulky bases from DNA via a non-base-flipping mechanism, demonstrating the diversity of base excision mechanism by glycosylases. In this study, authors showed another example of a non-base-flipping mechanism by AlkC glycosylase. In contrast to AlkD, AlkC exhibited narrow substrate-specificity as it can remove small alkylated-bases. Interestingly, there are two classes of AlkC glycosylases. The AlkCa group is formed with a HLR domain, whereas the AlkCb group consists of the HLR domain and the Ig domain at N- and C-terminal, respectively. Here, Shi et al determined the structure of AlkCb and provided insights into how AlkC selects and excises methylated bases from DNA. AlkC is unique in that it recognizes DNA by 60 degree bending without a base-flipping, and both HLR and Ig domains play important roles in this process. Ig is not only important for DNA recognition but also contributes to protein stability. Also, the modeling analysis to understand the substrate specificity is quite reasonable. Overall, this is an interesting manuscript that provides important insights into the mechanism of glycosylases in BER. This work further extends the importance of a non-base-flipping mechanism in BER by glycosylases and a role of Ig fold in DNA recognition in bacterial DNA repair proteins. Biochemical experiments support the structural data in this well-written manuscript. I have a few comments and it would be very helpful if authors explain these questions prior to publication.

Comments
1. Authors report the THF-bound structure in supplemental figure S7. While the 1aR-bound structure can be considered as a catalytic analogue, the THF-bound structure may be a non-catalytic but a product-analogue. In that sense, it is interesting that the partner base of 1aR is slipped in the 1aR-bound structure, whereas the THF nucleotide is slipped (although no close up view of this region is provided, see S7d). Unfortunately, there is no discussion on this structure in the main-text. Authors concluded that the biological significance of the THF-bound structure is uncertain and maybe that's why they did not describe this structure at all in the main-text. However, what is the basis of their conclusion that this structure is biologically irrelavant? Is there any evidence that this structure does not represent the PfAlkC-product complex? Furthermore, the THF-bound structures showed two different DNA structures (60 degree- and 85- degree bending), which indicate the flexible nature of DNA near the lesion. Authors may comment on this. I assume that the DNA does not make any contact with symmetry-related molecules. But if any, authors should describe them. I also strongly suggest that more detailed description of the THF-bound AlkC structure are needed in the text.

2. The importance of the Ig domain raises a question on the AlkCa family members that do not have Ig, yet exhibit similar activity relative to AlkCb. Although authors suggested that weak sequence conservation in the N-terminal helical bundle and alphaE, some species from two different groups share similarity. Can the authors model AlkCa or make some additional comments on this issue?

Referee #2:

Shi et al document novel crystal structures and biochemical analysis of AlkC DNA complexes. The mechanisms of repair of alkylation DNA damage remains of high interest due to the common use of DNA alkylating agents in cancer therapy, and the gene-environment impacts of DNA alkylation.The AlkC class of cationic methylbase glycosylases have not been previously characterized at the molecular in complex with DNA. Thus the study is a significant novel advance, and this work provides a clear testable basis for understanding new activities by this clade of AlkC bacterial DNA glycosylases. Overall this is a very interesting and well written manuscript. The data are of high quality and this work is likely of interest to diverse readership of EMBO journal, not limited to but

including those studying DNA transactions, DNA repair, DNA alkylation, genome integrity, cancer therapeutics etc.

I have minor comments:

1. Abstract: "normally processed" , used twice. What is normally? Consider word choice.

2. The introduction is very comprehensive. While interesting, it is at times off point. Consider refocusing the introduction.

3. How do the DNA binding contacts of the Ig fold domain compare to other Ig fold DNA interactions. For instance those made by NF-kB and p53 which are referenced, or Ndt80 (not referenced). A figure of an overlay or side-by-side comparison could be informative here.

4. Figure S7 looks to document interesting conformational differences in the asymmetric unit of the THF structure. However, this is not discussed in the text? Perhaps this discussion was omitted by accident?

5. In figure 4b - the modeled damaged cationic nucleobase appears to participate in cation-pi interactions with W164. This might be worth commenting on, even though this is a model?

6. page 12 "Like its cousin..." consider word choice.

7. What are the relevant distances between Glu121, the proposed catalytic water and the 1aR moiety? Having these parameters in the context of a chemical structure of 1aR would also be informative.

8. What parameters were used to assign the sodium atom (Figure s4a)? Does the geometry satisfy validation tools such as the check my metal server (i.e. coordination geometry, distances etc)?

9. The authors should consider use of modern data collection and reduction statistics (CC* or CC1/2) as this would extend the resolution of their structures. The authors cut the data at very high I/sigI. 5.8 sigma and 5.3 sigma. Is there a reason why, or did the data simply hit the edge of the detector and was limited by a suboptimal data collection strategy?

Referee #3:

The manuscript presents a phylogenetic, biochemical and structural investigation of the AlkC and AlkD glycosylases, and the results reveal some novel features of AlkC distinct from the structurally similar AlkD glycosylase:

(i) AlkC has concomitant recognition of the alkylated base in both the minor and major groove, thanks to the Ig-like and HLR domains wrapping around the DNA (but only the AlkC-beta subfamily contains the Ig-like domain)

(ii) AlkC has narrower substrate specificity and the presented structure supports a molecular explanation for this discrimination

(iii) AlkC has in vitro activity for 3mC and 1mA which are normally repaired by AlkB

The study presents a mutational analysis to reveal key residues involved in base excision and AlkC selectivity.

The presented structure shows some novel features like the Ig-like domain in the AlkC-beta subclass, the alpha-IJ loop that penetrates the minor groove of DNA, and the 60 degree bent DNA.

Questions and comments:

How can the authors be confident that the 1aR structure represents an active conformation of the DNA substrate and not a post-catalytic low energy state different from the conformation at the time of base excision? The structure of the AlkC in complex with the closely related THF DNA substrate is quite different as shown in Fig S7. Do the authors have any plausible explanation why a substitution
of a few atoms (from 1aR to THF in the same DNA context) results in such large structural changes in the complex?

In the abstract, the autohrs refer to a "structure of a catalytic intermediate". What is the intermediate? Aren't the structures models of product analoges (THF and 1aR instead of true AP-sites)?

The AlkC-alpha family has glycosylase activity despite the lack of the Ig-like domain. The working hypothesis is that the alpha-E helix and N-terminal helical bundle might compensate for the lack of the Ig-like domain in the AlkC-alpha class. Is this possible to test experimentally by mutagenesis? The Ig-like domain is obviously important for DNA binding in the AlkC-beta class.

The structure reveals some other protein-DNA contacts such as Thr337 -- Cyt5D. Could this be DNA sequence specific?

In Fig S4, we see that the co-crystallized crystallization agent pentaerythriton propoxylate (PEP) wraps two "arms" around the opposing Thy, which is outside of the helical stack. Is it possible that the PEP locks Thy in this position? It is noteworthy that the THF complex was obtained with completely different crystalizing agent and that the structure reveals a quite different position of the AP-site and opposing Thy conformations. Could the 1aR structure be a result of the capture of a AlkC-DNA-PEP complex? Is the rotation and orientation of the 1aR "ribose" ring also limited by steric contact with PEP (difficult to interpret from Fig S4b).

Were other metal ions than Na+ concidered during structural modeling?

How large and varied was the oligo library for crystal screening? It would be very useful with more details on this.

The text in the box in Fig S7 should be moved to the main text as this is important information related to the interpretation of the structural models. In the figure: give a short description of what the curved arrows illustrate.

Add a structural formula for 7mG and YTMA in Fig S2a and b, similar to Fig 6b and c.

Any explanation for the smear of the YTMA product in Fig 2S?

In Fig S5, it seems that the 1aR rings in AlkC and AlkD are partly flipped and not non-flipped as compared with B-DNA on the far right. Also in Fig 4a, it seems that the 1aR site is displaced into the minor groove.

Table S1 - add atoms of Na+ and EP in the list of number of atoms in the structure

Page 5, end of 1. paragraph. Authors claim AlkD excises "diverse .. lesions, including those normally processed by NER". Is it not only one bulky lesion, the YTMA that has been tested on AlkD?

1st Revision - authors' response 11 September 2017

Response to Reviewers

We would like to thank the reviewers for their time and thoughtful critiques of our manuscript, and are pleased with the overall positive responses. We have taken these suggestions and have revised the manuscript to address each point raised. Consequently, the new manuscript clarifies issues related to the catalytic relevance of the THF structure and the lack of the Ig-domain in the AlkCα proteins.

Referee #1:

1. Authors report the THF-bound structure in supplemental figure S7. While the 1aR-bound structure can be considered as a catalytic analogue, the THF-bound structure may be a non-catalytic but a product-analogue. In that sense, it is interesting that the partner base of 1aR is slipped in the 1aR-bound structure, whereas the THF nucleotide is slipped (although no close up view of this region is provided, see S7d). Unfortunately, there is no discussion on this structure in the main-text. Authors concluded that the biological significance of the THF-bound structure is uncertain and maybe that's why they did not describe this structure at all in the main-text. However, what is the basis of their conclusion that this structure is biologically irrelavant? Is there any evidence that this structure does not represent the PfAlkC-product complex?

Response: Our primary rationale for why the THF structure does not represent the enzyme-product complex is that the THF moiety, in contrast to 1aR, sits well outside of the active site and does not contact any of the catalytic residues. This was not articulated in the original manuscript, and we have now clarified this point by moving the description of the THF structure back to the main text and by adding a close-up view of the THF structure to Fig EV5 (old Figure S7).

We originally had decided to relegate the THF structure to the supplement since we thought it deviated from the catalytic mechanism, but in doing so we did not explain ourselves properly. We agree that the THF structure is interesting from the standpoint of conformational flexibility and overall similarity to the 1aR structure, and it highlights a key difference between base-flipping and non-base-flipping enzymes even though it does not help to describe the mechanism of base excision by AlkC. Abasic sites (natural and THF) are extremely dynamic and can force the DNA to adopt a range of structures. The more traditional base-flipping glycosylases insert a “plug” residue into the DNA helix where the missing base resided. The plug residue stabilizes both the enzyme-substrate and the enzyme-product complex, and limits the conformations that the abasic site can adopt by holding it in the active site. Thus, abasic site analogs often form specific structures that mimic the *bona fide* product complex to those particular enzymes. In contrast, AlkC and AlkD do not use a plug residue, and thus do not stabilize the abasic site conformation to the same extent. In support of this idea, we have observed in a number of published [Rubinson et al (2010) *Nature* 468: 406] and unpublished structures that AlkD can bind to abasic site DNA in a large number of conformations that are not informative to catalysis. We have added a short discussion on the consequences of lacking a plug residue in the main text (p.13).

Furthermore, the THF-bound structures showed two different DNA structures (60 degree- and 85-degree bending), which indicate the flexible nature of DNA near the lesion. Authors may comment on this. I assume that the DNA does not make any contact with symmetry-related molecules. But if any, authors should describe them. I also strongly suggest that more detailed description of the THF-bound AlkC structure are needed in the text.

Response: The protein-DNA complex stacks in a head-to-head arrangement such that the 5′-overhanging A1 bases from adjacent molecules are stacked on one another. We have included the symmetry-related DNA contacts within the new description of the THF complex on p.13.

2. The importance of the Ig domain raises a question on the AlkCa family members that do not have Ig, yet exhibit similar activity relative to AlkCb. Although authors suggested that weak sequence conservation in the N-terminal helical bundle and alphaE, some species from two different groups share similarity. Can the authors model AlkCa or make some additional comments on this issue?

Response: We now include two homology models of BcAlkCα (Fig EV3), both of which show how extension of helix E could interact with either the minor groove immediately adjacent to the lesion or the backbone in a similar manner as the N-terminal helical bundle. Both interactions would serve to stabilize the DNA kink in the absence of an Ig-domain. This point has been added to the main text on p.10.

Referee #2:

1. Abstract: "normally processed", used twice. What is normally? Consider word choice.
Response: “Normally processed by” has been changed to “more commonly associated with”. “Normally demethylated by” has been changed to “which are also repaired by AlkB-catalyzed oxidative demethylation.”

2. The introduction is very comprehensive. While interesting, it is at times off point. Consider refocusing the introduction.
Response: We have rewritten the second paragraph and removed several off-topic sentences pertaining to repair of O^6- and O^4-alkylbases and structural details about eukaryotic alkylpurine glycosylases.

3. How do the DNA binding contacts of the Ig fold domain compare to other Ig fold DNA interactions. For instance those made by NF-kB and p53 which are referenced, or Ndt80 (not referenced). A figure of an overlay or side-by-side comparison could be informative here.
Response: We have added Fig EV1 to illustrate structural differences between AlkC and other DNA binding Ig-like domains.

4. Figure S7 looks to document interesting conformational differences in the asymmetric unit of the THF structure. However, this is not discussed in the text? Perhaps this discussion was omitted by accident?
Response: We have moved the discussion of the THF structure to the main text on p. 12-13.

5. In figure 4b - the modeled damaged cationic nucleobase appears to participate in cation-pi interactions with W164. This might be worth commenting on, even though this is a model?
Response: We have added text on p. 11 to point out the putative cation-π interaction.

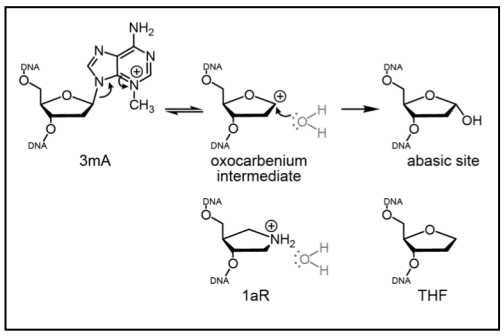
6. page 12 "Like its cousin..." consider word choice.
Response: The qualifier has been removed and now reads “Like AlkD, …”

7. What are the relevant distances between Glu121, the proposed catalytic water and the 1aR moiety? Having these parameters in the context of a chemical structure of 1aR would also be informative.
Response: We have added the distances to Fig 4A and also include a schematic of the chemical interactions within the active site (Fig 4B).

8. What parameters were used to assign the sodium atom (Figure s4a)? Does the geometry satisfy validation tools such as the check my metal server (i.e. coordination geometry, distances etc)?
Response: The sodium was assigned based on metal-ligand distances and coordination geometry as defined in Harding (2002) Acta Cryst. D58:872 and Harding (2001) Acta Cryst D57:401, and has now been validated using the CheckMyMetal webserver. We have added text to this effect in the Methods section and have also included the interatomic distances to Fig EV4.

9. The authors should consider use of modern data collection and reduction statistics (CC* or CC1/2) as this would extend the resolution of their structures. The authors cut the data at very high I/sigI. 5.8 sigma and 5.3 sigma. Is there a reason why, or did the data simply hit the edge of the detector and was limited by a suboptimal data collection strategy?
Response: We appreciate this suggestion and in fact did take both CC1/2 and CC* into account when scaling the data. However, the diffraction (completeness and intensity) for both crystals fell off sharply beyond the specified resolution limits we imposed. This decrease was intrinsic to the crystals and not a result of poor data collection. The few remaining reflections that were beyond these limits gave Rsym values well above and CC values well below statistically significant values.

Referee #3:
How can the authors be confident that the 1aR structure represents an active conformation of the DNA substrate and not a post-catalytic low energy state different from the conformation at the time of base excision?



Response: The positively charged 1aR is a mimetic of the positively charged oxocarbenium intermediate formed upon dissociation of the 3mA base from the ribose ring, but before attack of the nucleophilic water [Drohat and Maiti (2014) *Org. Biomol. Chem*. 12:8367]. This analog has been well-validated as an oxocarbenium ion intermediate in mechanistically similar enzymes that work on nucleosides [Schramm (2011) *Annual review of biochemistry* 80: 703] and other glycosylases [e.g., Hollis et al (2000) *EMBO J* 19:758; Chu et al (2011) *Bioorg Med Chem Lett* 21:4969]. The oxocarbenium is a high-energy intermediate that is stabilized by the active site and must be positioned correctly for attack by a water. The water in our 1aR structure is 2.6 Å away from the 1aR N1 (which would be the cationic C1′ in the oxocarbenium ribose ring) and has optimal geometry for attack of this position. Modeling also showed that the pocket created between the protein and the DNA in this structure is ideally suited for 3mA (Fig. 4C), suggesting the conformation of the DNA is similar to that of DNA in an enzyme-substrate complex. We have articulated this in the main text and added the figure above to Fig. S7.

The structure of the AlkC in complex with the closely related THF DNA substrate is quite different as shown in Fig S7. Do the authors have any plausible explanation why a substitution of a few atoms (from 1aR to THF in the same DNA context) results in such large structural changes in the complex?
Response: The main difference between these two abasic analogs is that 1aR bears a formal positive charge and THF is uncharged. This charge is likely to affect specific binding to the electron-rich active site. Moreover, in the absence of a plug residue used by all base flipping enzymes to stabilize the enzyme-DNA complex, the dynamic THF abasic site is free to adopt a wider range of conformations. Please see our response to Referee #1, point 1 for a full description of this phenomenon.

In the abstract, the authors refer to a "structure of a catalytic intermediate". What is the intermediate? Aren't the structures models of product analoges (THF and 1aR instead of true AP-sites)?
Response: No, the 1aR more closely approximates the oxocarbenium intermediate as described above.

The AlkC-alpha family has glycosylase activity despite the lack of the Ig-like domain. The working hypothesis is that the alpha-E helix and N-terminal helical bundle might compensate for the lack of the Ig-like domain in the AlkC-alpha class. Is this possible to test experimentally by mutagenesis?
Response: We feel that abrogating activity by mutating the AlkCα structural elements in question would not sufficiently show that these motifs compensate for DNA binding activity. In order to do the mutagenesis correctly, we would need to make a gain-of-function mutant by converting the HLR domain in the AlkCβ Ig-delta construct into a functional unit like the AlkCα clade. We have already spent 4 months attempting this with several mutants and have so far been unsuccessful. Thus, this is a major undertaking in itself as there is no easy way to predict if compensatory changes are necessary to stabilize the engineered AlkCα protein. Even if successful, this will be a minor advance to the current work and falls outside of the current scope of this paper. We intend to pursue this as a stand-alone project that will address this topic from several angles, including a structure of AlkCα.

The structure reveals some other protein-DNA contacts such as Thr337 -- Cyt5D. Could this be DNA sequence specific?
Response: No, we do not expect this to be sequence specific. Threonine can be either a hydrogen bond donor or acceptor, and thus can interact with any base in the major groove.

In Fig S4, we see that the co-crystallized crystallization agent pentaerythriton propoxylate (PEP) wraps two "arms" around the opposing Thy, which is outside of the helical stack. Is it possible that the PEP locks Thy in this position? It is noteworthy that the THF complex was obtained with completely different crystalizing agent and that the structure reveals a quite different position of the AP-site and opposing Thy conformations. Could the 1aR structure be a result of the capture of an

AlkC-DNA-PEP complex? Is the rotation and orientation of the 1aR "ribose" ring also limited by steric contact with PEP (difficult to interpret from Fig S4b).
Response: To clarify, the PEP arms are not wrapped around the opposing thymine; one arm projects into the DNA kink and contacts the 1aR and the flanking base pairs, two arms project outward to solvent, and the fourth arm next to the thymine is not present. The hydroxyl group from this missing arm forms a hydrogen bond to the thymine base. Thus, the position of the thymine is certainly affected by the presence of the PEP (as well as by coordination to the sodium ion). However, even in the absence of the PEP, this thymidine would be displaced into the minor groove as a result of the kink in the DNA. That is, the thymine might be pushed farther out by PEP, but it would have to be pushed toward the minor groove because of the bend angle of the DNA. Indeed, this thymine is displaced in the THF structure (see Fig EV5F). On the 1aR side, it is difficult to know the extent to which, if any, the PEP perturbs the 1aR position. If anything, the PEP limits rotation of the 1aR back *toward* the DNA, as PEP sits between the DNA stack and the 1aR. We have added text to this effect to the legend of Fig EV4. Because the majority of contacts to the 1aR are made by the catalytic residues, and because the conformation of the kinked DNA in the THF structure (which lacks the PEP molecule) is very similar to the 1aR structure, we are confident that the PEP has only minimal impact on the structure. The most important point, which we have worked to clarify in the revised manuscript, is that this void in the active site would be occupied by the damaged nucleobase, both before and after cleavage from the DNA. We tried to trap a ternary AlkC/1aR-DNA/3mA base complex, and while we were able to reduce the concentration of PEP required for crystallization by adding the free 3mA base, we were unsuccessful in fully displacing the PEP from the crystals. It is worth noting that crystallization of a catalytically relevant product complex between AlkD, which also uses a non-base-flipping mechanism, and THF-DNA required the presence of free 3mA base to stabilize the void in the duplex [Mullins et al (2015) *Nature* 527: 168]. The absence of free 3mA base, or a different intercalating molecule, produced structures that were not catalytically informative [Rubinson et al (2010) *Nature* 468: 406 and unpublished]. We have added a discussion of this to the main text on p. 13.

Were other metal ions than Na+ considered during structural modeling?
Response: We considered other metals, but we assigned this as Na+ because of (1) its coordination geometry is consistent with sodium (which we have now verified using the CheckMyMetal webserver), (2) sodium was the only cation present in both the mother liquor and the protein buffer, and (3) our protein buffer contained EDTA which would sequester divalent ions.

How large and varied was the oligo library for crystal screening? It would be very useful with more details on this.
Response: Against each of the seven AlkC orthologs, we tested eight DNA duplexes ranging in length from 8-15 nucleotides and containing a THF•T. We have added this information to the Methods section.

The text in the box in Fig S7 should be moved to the main text as this is important information related to the interpretation of the structural models. In the figure: give a short description of what the curved arrows illustrate.
Response: We have moved the description of the THF structure to the main text and added a description of the curved arrows to the legend of Fig EV5 (old Fig. S7).

Add a structural formula for 7mG and YTMA in Fig S2a and b, similar to Fig 6b and c.
Response: The structures of 7mG and YTMA have been added to Appendix Fig S2.

Any explanation for the smear of the YTMA product in Fig 2S?
Response: We attribute the smearing to an incompletely denatured 12-mer/25-mer product. The sequence required for the YTMA experiments is extremely GC-rich and is different from those used in the 7mG, 1mA, 3mC, 3mT, and 1mG cleavage assays. In fact, 11 out of 12 nucleotides in the cleaved YTMA strand are GCs and thus the cleaved product has a higher melting temperature than the products in the other gels. Unlike the YTMA experiments published recently, in which samples were heated for longer periods of time prior to running the gels, here the YTMA products were treated in the same manner as the products from the methylated lesions. We have added text to this effect to the legend of Appendix Fig S2.

In Fig S5, it seems that the 1aR rings in AlkC and AlkD are partly flipped and not non-flipped as compared with B-DNA on the far right. Also in Fig 4a, it seems that the 1aR site is displaced into the minor groove.
Response: We agree that the abasic sites in AlkC and AlkD are partially displaced into the minor groove relative B-DNA, but they are not rotated 180° around the backbone as they are in the base-flipping enzymes. The more important point, however, is that we define non-base-flipping by the fact that the nucleobase in the AlkD structure is partially stacked with the flanking nucleotides. Our AlkC structures argue that this would also be the case for this enzyme.

Table S1 - add atoms of Na+ and EP in the list of number of atoms in the structure
Response: The Na+ and PEP atoms are included in the number of Solvent atoms in the table. For clarity, we prefer not to specifically list each solvent molecule on its own line in the table since there are several solvent molecules present in these structures. Thus, we have added a footnote to the table to specify that the values for solvent include molecules other than water.

Page 5, end of 1. paragraph. Authors claim AlkD excises "diverse .. lesions, including those normally processed by NER". Is it not only one bulky lesion, the YTMA that has been tested on AlkD?
Response: No—we showed previously that AlkD can excise cationic pyridyloxobutyl (POB) adducts of dG and dC [Rubinson et al (2010) *Nature* 468:406]. We have revised the text in the introduction to reflect this.

2nd Editorial Decision 22 September 2017

Thank you for submitting your final revised manuscript for our consideration. In light of the positive re-review by two of the original referees (see comments below), I am pleased to inform you that we have now accepted it for publication in The EMBO Journal.

Thank you again for this contribution to The EMBO Journal and congratulations on a successful publication! Please consider us again in the future for your most exciting work.

REFEREE REPORTS

Referee #1:

In this revised manuscript, authors have clearly resolved issues addressed by this reviewer. I believe that this is a well-written manuscript that will contribute to the community of DNA repair. I recommend this paper to be published in EMBO J.

Referee #2:

The authors have addressed all points with this revision.

EMBO PRESS

YOU MUST COMPLETE ALL CELLS WITH A PINK BACKGROUND ↓

PLEASE NOTE THAT THIS CHECKLIST WILL BE PUBLISHED ALONGSIDE YOUR PAPER

Corresponding Author Name: Brandt F. Eichman
Journal Submitted to: EMBO J
Manuscript Number: EMBOJ-2017-97833

Reporting Checklist For Life Sciences Articles (Rev. June 2017)

This checklist is used to ensure good reporting standards and to improve the reproducibility of published results. These guidelines are consistent with the Principles and Guidelines for Reporting Preclinical Research issued by the NIH in 2014. Please follow the journal's authorship guidelines in preparing your manuscript.

A- Figures

1. Data

The data shown in figures should satisfy the following conditions:

- the data were obtained and processed according to the field's best practice and are presented to reflect the results of the experiments in an accurate and unbiased manner.
- figure panels include only data points, measurements or observations that can be compared to each other in a scientifically meaningful way.
- graphs include clearly labeled error bars for independent experiments and sample sizes. Unless justified, error bars should not be shown for technical replicates.
- if n< 5, the individual data points from each experiment should be plotted and any statistical test employed should be justified
- Source Data should be included to report the data underlying graphs. Please follow the guidelines set out in the author ship guidelines on Data Presentation.

2. Captions

Each figure caption should contain the following information, for each panel where they are relevant:

- a specification of the experimental system investigated (eg cell line, species name).
- the assay(s) and method(s) used to carry out the reported observations and measurements
- an explicit mention of the biological and chemical entity(ies) that are being measured.
- an explicit mention of the biological and chemical entity(ies) that are altered/varied/perturbed in a controlled manner.

- the exact sample size (n) for each experimental group/condition, given as a number, not a range;
- a description of the sample collection allowing the reader to understand whether the samples represent technical or biological replicates (including how many animals, litters, cultures, etc.).
- a statement of how many times the experiment shown was independently replicated in the laboratory.
- definitions of statistical methods and measures:
  - common tests, such as t-test (please specify whether paired vs. unpaired), simple $\chi 2$ tests, Wilcoxon and Mann-Whitney tests, can be unambiguously identified by name only, but more complex techniques should be described in the methods section;
  - are tests one-sided or two-sided?
  - are there adjustments for multiple comparisons?
  - exact statistical test results, e.g., P values = x but not P values < x;
  - definition of 'center values' as median or average;
  - definition of error bars as s.d. or s.e.m.

Any descriptions too long for the figure legend should be included in the methods section and/or with the source data.

In the pink boxes below, please ensure that the answers to the following questions are reported in the manuscript itself. Every question should be answered. If the question is not relevant to your research, please write NA (non applicable). We encourage you to include a specific subsection in the methods section for statistics, reagents, animal models and human subjects.

B- Statistics and general methods

Please fill out these boxes ↓ (Do not worry if you cannot see all your text once you press return)

1.a. How was the sample size chosen to ensure adequate power to detect a pre-specified effect size?	experiments performed in triplicate
1.b. For animal studies, include a statement about sample size estimate even if no statistical methods were used.	n/a
2. Describe inclusion/exclusion criteria if samples or animals were excluded from the analysis. Were the criteria pre-established?	no data excluded
3. Were any steps taken to minimize the effects of subjective bias when allocating animals/samples to treatment (e.g. randomization procedure)? If yes, please describe.	no
For animal studies, include a statement about randomization even if no randomization was used.	n/a
4.a. Were any steps taken to minimize the effects of subjective bias during group allocation or/and when assessing results (e.g. blinding of the investigator)? If yes please describe.	no
4.b. For animal studies, include a statement about blinding even if no blinding was done	n/a
5. For every figure, are statistical tests justified as appropriate?	yes
Do the data meet the assumptions of the tests (e.g., normal distribution)? Describe any methods used to assess it.	yes
Is there an estimate of variation within each group of data?	no
Is the variance similar between the groups that are being statistically compared?	n/a

C- Reagents

USEFUL LINKS FOR COMPLETING THIS FORM

http://www.antibodypedia.com
http://1degreebio.org
http://www.equator-network.org/reporting-guidelines/improving-bioscience-research-repo

http://grants.nih.gov/grants/olaw/olaw.htm
http://www.mrc.ac.uk/Ourresearch/Ethicsresearchguidance/Useofanimals/index.htm
http://ClinicalTrials.gov
http://www.consort-statement.org
http://www.consort-statement.org/checklists/view/32-consort/66-title

http://www.equator-network.org/reporting-guidelines/reporting-recommendations-for-tun

http://datadryad.org

http://figshare.com

http://www.ncbi.nlm.nih.gov/gap

http://www.ebi.ac.uk/ega

http://biomodels.net/

http://biomodels.net/miriam/
http://jjj.biochem.sun.ac.za
http://oba.od.nih.gov/biosecurity/biosecurity_documents.html
http://www.selectagents.gov/

6. To show that antibodies were profiled for use in the system under study (assay and species), provide a citation, catalog number and/or clone number, supplementary information or reference to an antibody validation profile. e.g., Antibodypedia (see link list at top right), 1DegreeBio (see link list at top right).	n/a
7. Identify the source of cell lines and report if they were recently authenticated (e.g., by STR profiling) and tested for mycoplasma contamination.	n/a

* for all hyperlinks, please see the table at the top right of the document

D- Animal Models

8. Report species, strain, gender, age of animals and genetic modification status where applicable. Please detail housing and husbandry conditions and the source of animals.	n/a
9. For experiments involving live vertebrates, include a statement of compliance with ethical regulations and identify the committee(s) approving the experiments.	n/a
10. We recommend consulting the ARRIVE guidelines (see link list at top right) (PLoS Biol. 8(6), e1000412, 2010) to ensure that other relevant aspects of animal studies are adequately reported. See author guidelines, under 'Reporting Guidelines'. See also: NIH (see link list at top right) and MRC (see link list at top right) recommendations. Please confirm compliance.	n/a

E- Human Subjects

11. Identify the committee(s) approving the study protocol.	n/a
12. Include a statement confirming that informed consent was obtained from all subjects and that the experiments conformed to the principles set out in the WMA Declaration of Helsinki and the Department of Health and Human Services Belmont Report.	n/a
13. For publication of patient photos, include a statement confirming that consent to publish was obtained.	n/a
14. Report any restrictions on the availability (and/or on the use) of human data or samples.	n/a
15. Report the clinical trial registration number (at ClinicalTrials.gov or equivalent), where applicable.	n/a
16. For phase II and III randomized controlled trials, please refer to the CONSORT flow diagram (see link list at top right) and submit the CONSORT checklist (see link list at top right) with your submission. See author guidelines, under 'Reporting Guidelines'. Please confirm you have submitted this list.	n/a
17. For tumor marker prognostic studies, we recommend that you follow the REMARK reporting guidelines (see link list at top right). See author guidelines, under 'Reporting Guidelines'. Please confirm you have followed these guidelines.	n/a

F- Data Accessibility

18: Provide a "Data Availability" section at the end of the Materials & Methods, listing the accession codes for data generated in this study and deposited in a public database (e.g. RNA-Seq data: Gene Expression Omnibus GSE39462, Proteomics data: PRIDE PXD000208 etc.) Please refer to our author guidelines for 'Data Deposition'. Data deposition in a public repository is mandatory for: a. Protein, DNA and RNA sequences b. Macromolecular structures c. Crystallographic data for small molecules d. Functional genomics data e. Proteomics and molecular interactions	Data Availability section included.
19. Deposition is strongly recommended for any datasets that are central and integral to the study; please consider the journal's data policy. If no structured public repository exists for a given data type, we encourage the provision of datasets in the manuscript as a Supplementary Document (see author guidelines under 'Expanded View' or in unstructured repositories such as Dryad (see link list at top right) or Figshare (see link list at top right).	n/a
20. Access to human clinical and genomic datasets should be provided with as few restrictions as possible while respecting ethical obligations to the patients and relevant medical and legal issues. If practically possible and compatible with the individual consent agreement used in the study, such data should be deposited in one of the major public access-controlled repositories such as dbGAP (see link list at top right) or EGA (see link list at top right).	n/a
21. Computational models that are central and integral to a study should be shared without restrictions and provided in a machine-readable form. The relevant accession numbers or links should be provided. When possible, standardized format (SBML, CellML) should be used instead of scripts (e.g. MATLAB). Authors are strongly encouraged to follow the MIRIAM guidelines (see link list at top right) and deposit their model in a public database such as Biomodels (see link list at top right) or JWS Online (see link list at top right). If computer source code is provided with the paper, it should be deposited in a public repository or included in supplementary information.	n/a

G- Dual use research of concern

22. Could your study fall under dual use research restrictions? Please check biosecurity documents (see link list at top right) and list of select agents and toxins (APHIS/CDC) (see link list at top right). According to our biosecurity guidelines, provide a statement only if it could.	n/a