

ARTICLE

Open Access

# Structural and mechanistic insights into the DNA glycosylase AAG-mediated base excision in nucleosome

Lvqin Zheng<sup>1,2</sup>, Bin Tsai<sup>1</sup> and Ning Gao<sup>1</sup>✉

## Abstract

The engagement of a DNA glycosylase with a damaged DNA base marks the initiation of base excision repair. Nucleosome-based packaging of eukaryotic genome obstructs DNA accessibility, and how DNA glycosylases locate the substrate site on nucleosomes is currently unclear. Here, we report cryo-electron microscopy structures of nucleosomes bearing a deoxyinosine (DI) in various geometric positions and structures of them in complex with the DNA glycosylase AAG. The apo nucleosome structures show that the presence of a DI alone perturbs nucleosomal DNA globally, leading to a general weakening of the interface between DNA and the histone core and greater flexibility for the exit/entry of the nucleosomal DNA. AAG makes use of this nucleosomal plasticity and imposes further local deformation of the DNA through formation of the stable enzyme–substrate complex. Mechanistically, local distortion augmentation, translation/rotational register shift and partial opening of the nucleosome are employed by AAG to cope with substrate sites in fully exposed, occluded and completely buried positions, respectively. Our findings reveal the molecular basis for the DI-induced modification on the structural dynamics of the nucleosome and elucidate how the DNA glycosylase AAG accesses damaged sites on the nucleosome with different solution accessibility.

## Introduction

The eukaryotic base excision repair (BER) machinery locates and repairs DNA base damage in chromatin<sup>1,2</sup>. Genomic DNA contains substantial amounts of DNA base damages due to various exogenous damaging agents and spontaneous decomposition reactions — such as deamination of deoxycytidine (DC, cytosine) to deoxyuridine (DU, uracil) and deamination of deoxyadenosine (DA, adenine) to deoxyinosine (DI, hypoxanthine)<sup>3</sup>. In human cells, DNA base damage is primarily detected by damage-specific DNA glycosylases<sup>4,5</sup>. DNA glycosylase recognizes damaged base and catalyzes its excision, leaving an apurinic/apyrimidinic site (AP site) which will be

cleaved by AP site endonuclease APE1 to generate a nick on DNA<sup>6</sup>. Subsequent repair reactions involve AP-lyase activity, DNA polymerase activity and DNA ligase activity. These reactions are executed by XRCC1, DNA polymerase  $\beta$  and DNA ligase III in the short-patch BER sub-pathway, while the long-patch BER sub-pathway requires the replicative DNA polymerase  $\delta/\epsilon$ -PCNA-RFC machinery, flap endonuclease 1 and DNA ligase I<sup>7</sup>.

Alkyladenine DNA glycosylase (AAG, 3-methyladenine DNA glycosylase), as one of the first responders of DNA base damage, can recognize alkylpurines like 3-methyladenine (3-MA) and 7-methylguanine (m<sup>7</sup>G)<sup>8,9</sup>, oxidized adenine 1,*N*<sup>6</sup>-ethenoadenine ( $\epsilon$ A)<sup>10</sup> and deaminated adenine hypoxanthine<sup>6,11–13</sup>. In humans, altered expression and single nucleotide polymorphisms of AAG are associated with microsatellite instability<sup>14</sup>, spontaneous frameshift mutagenesis<sup>15</sup>, and a variety of cancers including osteosarcoma<sup>16</sup>, breast cancer<sup>17,18</sup> and astrocytic tumors<sup>19,20</sup>. In mouse models, Aag-knockout mice are

Correspondence: Ning Gao (gao@pku.edu.cn)

<sup>1</sup>State Key Laboratory of Membrane Biology, Peking-Tsinghua Center for Life Sciences, School of Life Sciences, Peking University, Beijing, China

<sup>2</sup>Academy for Advanced Interdisciplinary Studies, Peking University, Beijing, China

These authors contributed equally: Lvqin Zheng, Bin Tsai

© The Author(s) 2023



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

prone to liver and colorectal cancers<sup>21,22</sup>, while excessive AAG activity in Aag-overexpressing mice causes hepatotoxicity, lethality and other alkylation-induced toxicity<sup>23,24</sup>.

Frequently occurring DNA base damages could be present in all regions of chromatinized eukaryotic genome, including nucleosomal DNA sites where nucleosomes would obstruct BER machinery's accessibility to the damage sites<sup>25</sup>. In a nucleosome, the 147-bp DNA wraps around the octameric core in 1.65 superhelical turns, leaving only a portion of solvent-facing DNA freely accessible<sup>26,27</sup>. Extensive in vitro studies showed that AAG<sup>28,29</sup> along with short-patch BER factors<sup>30–32</sup> could selectively locate and bind to the damage sites in nucleosomes. In general, AAG activity on nucleosome is strongly correlated to solution accessibility of the damaged base, and in vivo studies also confirmed this general correlation<sup>33</sup>, and found that bases at the DNA minor grooves, which interact with the octameric histone core, display higher mutation rates due to low accessibility to repair factors<sup>25</sup>. Structurally, the solution accessibility is predominantly determined by its translational position and rotational orientation on a nucleosome, and solvent-facing damaged bases were indeed excised more efficiently than occluded and embedded damaged bases with medium and low solution accessibility<sup>34</sup>. Another contributing factor is the structural dynamics of nucleosome such as spontaneous opening of nucleosome<sup>35,36</sup>, which could impact on the accessibility of the damaged bases on certain positions. In addition, compared to the other regions of the nucleosomal DNA, the dyad axis is the least accessible to repair factors<sup>34,37</sup>. Overall, the differential AAG activity at various nucleosomal positions could not be solely explained by the solution accessibility predicted from a static nucleosome structure.

While the structural mechanism of base excision enzymes' action on naked DNA duplex is well-understood<sup>4,38–43</sup>, how these proteins overcome nucleosome-imposed obstacle to locate and repair DNA base damage in nucleosome is not fully understood. In the present study, we employed cryo-EM to explain how the BER factor AAG accesses a damaged base introduced to various positions on nucleosomes. Our results show that a single DI nucleotide could induce a global perturbation to the nucleosome structure, and AAG exploits altered nucleosome dynamics and adopts distinct mechanisms to access the substrate site dependent on the geometric positions of the damaged base on the nucleosome.

## Results

### Preparation and structural determination of the AAG–NCP complexes

To capture the state of DNA glycosylase AAG engaging with DNA base damage in nucleosome, we first assembled nucleosome core particles (NCPs) from a 152-base pair (bp) Widom 601 DNA<sup>44</sup> and core histones H2A, H2B, H3 and H4 from *Xenopus laevis* (Supplementary Fig.

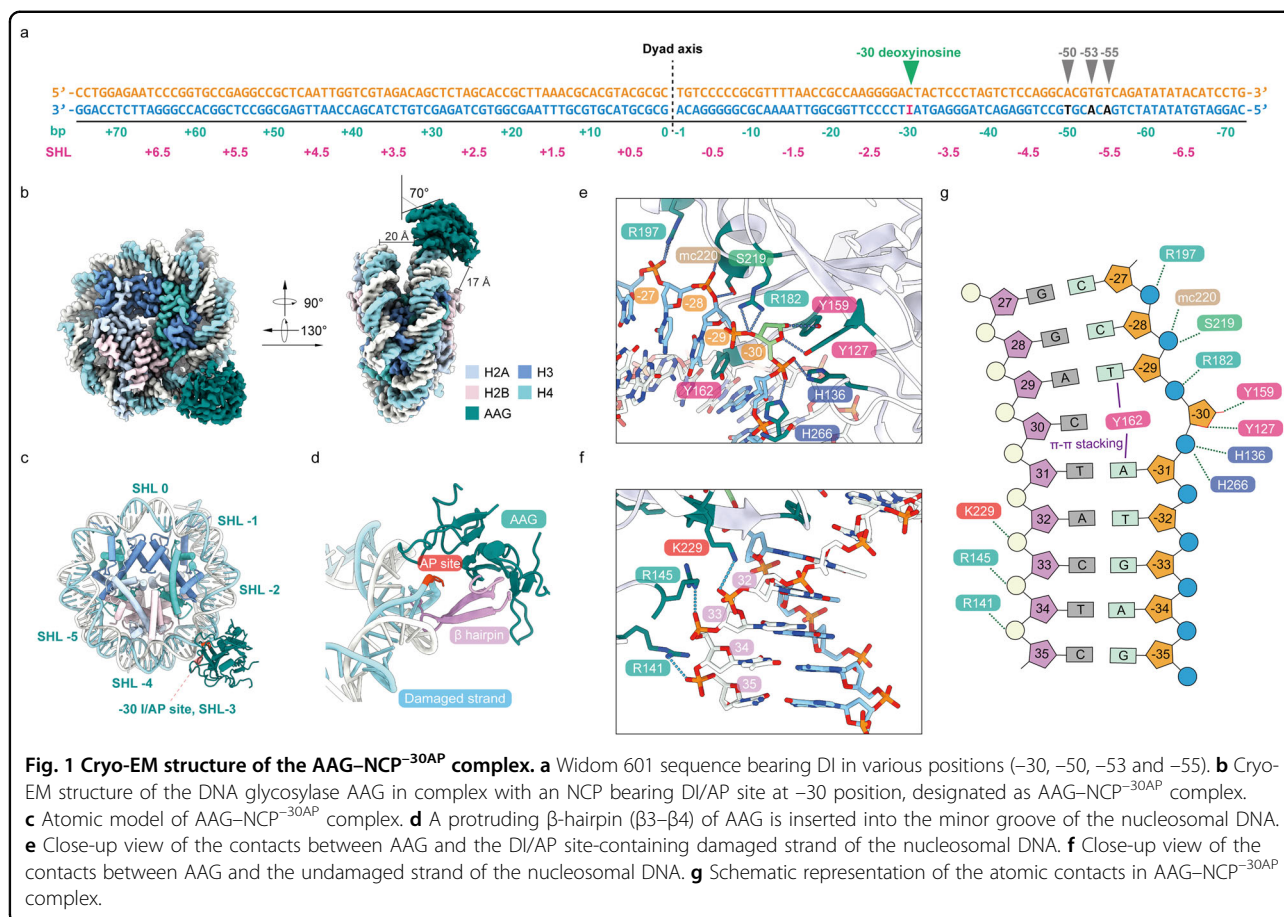
S1a)<sup>45–47</sup>. The bottom strand of Widom 601 DNA (designated as damaged strand) contains a DI to mimic a base damage resulting from deamination of DA (Fig. 1a).

To investigate how AAG acts in positions with different solution accessibility and histone microenvironment, we sampled different superhelical locations (SHLs) and rotational orientations on nucleosome, and a series of single DI-containing DNAs with a DI located in various geometric positions (–30, –50, –53 and –55) were prepared (Fig. 1a). –30 at SHL-3 and –50 at SHL-5 both represent solvent-facing positions with high solution accessibility, whereas –53 and –55 at SHL-5 represent occluded and embedded positions with medium and low solution accessibility, respectively<sup>34</sup>. We assembled these NCPs and incubated them with purified AAG to form AAG–NCP complexes, which were further analyzed by cryo-EM (Supplementary Figs. S2–S5).

### Overall structure of the AAG–NCP<sup>–30AP</sup> complex with a fully exposed damage site

The presence of DI in the –30 position of nucleosomal DNA (NCP<sup>–30I</sup>), which is solvent-facing, triggers a strong engagement of AAG, as shown in the glycerol density gradient centrifugation (Supplementary Fig. S1b, c). Cryo-EM analysis of the AAG–NCP<sup>–30I</sup> sample obtained high-resolution structures of the NCP in both AAG-bound and -unbound states, at resolutions of 2.9 Å and 2.8 Å, respectively (Supplementary Fig. S2). The nucleosomal DNA in the two states and most of the side chains of AAG in the bound state are well resolved (Supplementary Fig. S6a–d). In the map of the free NCP (NCP<sup>–30I</sup>), the DI in the –30 position is well resolved and forms a non-canonical base pair with the DC on the top strand (Supplementary Fig. S6a), reflecting a state for the damaged nucleosome before AAG binding. In contrast, the map of the AAG-bound NCP (AAG–NCP<sup>–30AP</sup>) was identified as a post-catalytic state. The ribose ring of the –30 position is in a flipped-out orientation, and there is no density for hypoxanthine base at –30, indicating that this is a post-catalytic AP site (Supplementary Fig. S6b). In addition, superimposition of a DI onto the catalytic cave comprising residues Y159, Y127, L180, R182, E125 and C167 would result in multiple steric clashes (Supplementary Fig. S6e).

This high-resolution structure of the AAG–NCP<sup>–30AP</sup> complex provides an accurate model explaining the molecular details of the interactions between the damaged NCP and AAG. In general, AAG attaches to the minor groove of the nucleosomal DNA at SHL-3 and is angled at ~70° against the plane parallel to the nucleosomal disc (Fig. 1b, c). The distance between AAG and the closest histone is beyond 17 Å, and no interactions between AAG and octameric histone core were observed (Fig. 1b). In addition to the protruding β-hairpin of AAG, which is



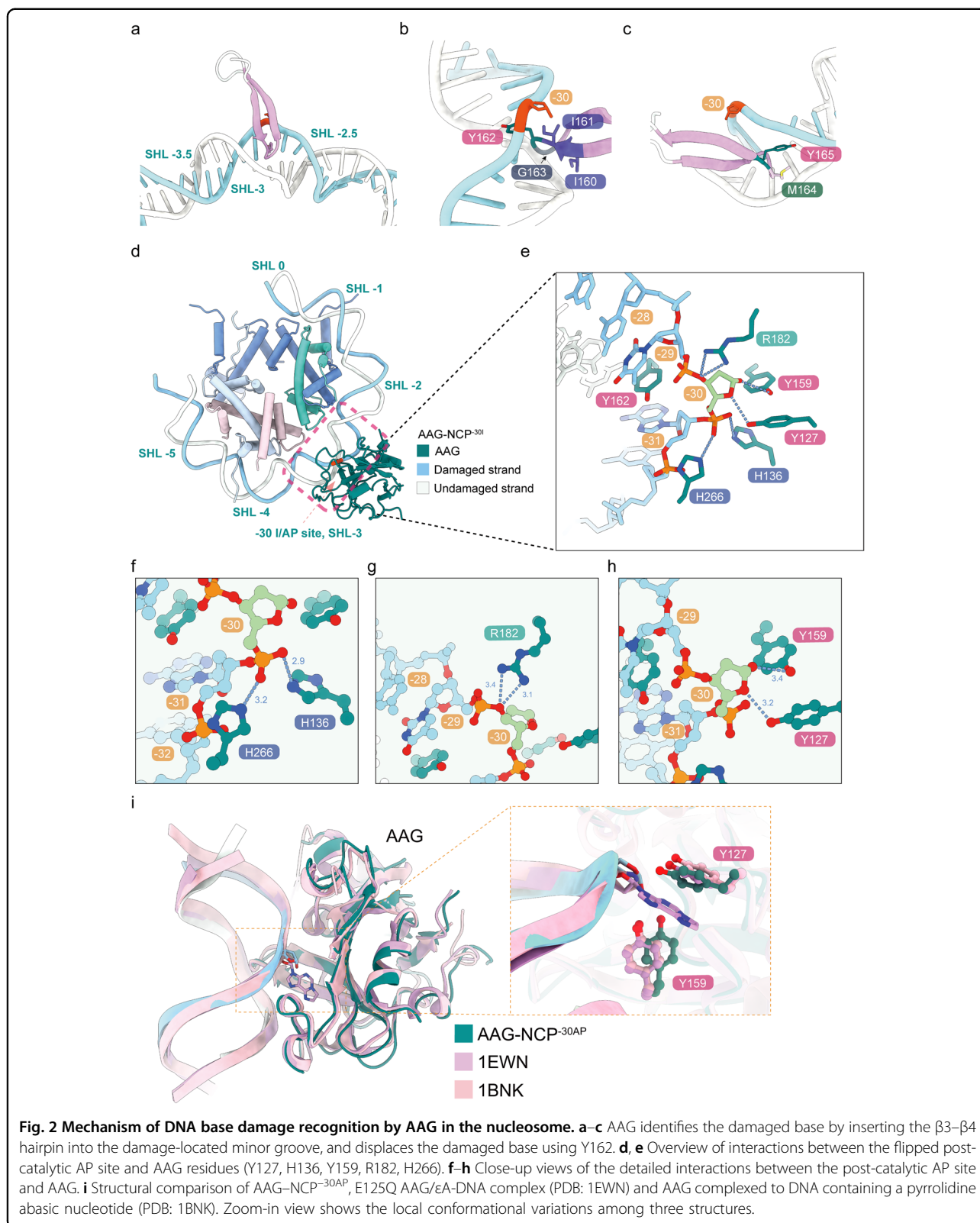
inserted into the minor groove of nucleosomal DNA for direct damage recognition (Fig. 1d), a few other residues also contribute to the stabilization of AAG on the nucleosome. On the damaged strand, R197, S219 and the main chain of K220 display a direct interaction with the DNA backbone (Fig. 1e, g and Supplementary Fig. S6c, d). On the undamaged strand, the attachment of AAG to NCP is strengthened by the interactions between the phosphate backbone and three successive positively charged residues R141, R145 and K229 (Fig. 1f, g and Supplementary Fig. S6c, d).

#### Interaction between AAG and the AP site in the AAG–NCP<sup>-30AP</sup> complex

In the AAG–NCP<sup>-30AP</sup> structure, AAG shadows ~8 bp of nucleosomal DNA near SHL-3 (Fig. 1g). The inserted  $\beta$ -hairpin loop is composed of IYGYM (residues 160–165), which makes extensive contacts with the skewed DNA and serves as the damage-recognition motif (Figs. 1d, 2a and Supplementary Fig. S7a). The residues I160 and I161 establish hydrophobic interactions with both sides of DNA in the minor groove (Fig. 2b). The aromatic residue Y162 stacks with the bases of –31A and –29T on the damaged strand through  $\pi$ – $\pi$  interactions.

Bridging residue G163 bents the terminal loop and links the  $\beta$ 3 and  $\beta$ 4 of AAG (Fig. 2b). The main chain of G163 also contributes to the stabilization of the  $\beta$ -hairpin by interacting with 31T on the undamaged strand. Scaffolding residues M164 and Y165 locate in the center of the minor groove and widen the distance between two DNA backbones (Fig. 2c).

The flipped AP site is stabilized through three sets of interactions (Figs. 1g, 2d, e) and the interacting residues of AAG are highly conserved (Supplementary Fig. S8). The first interacting site consists of two basic amino acids H136 and H266 that coordinate the phosphate backbone of the AP site (Fig. 2e, f). These two histidine residues interact with the phosphate group between –30 and –31 to keep it in a proper conformation. The second site involves a positively charged residue R182 that forms a hydrogen bond with the O3' atom of the AP site (Fig. 2e, g). The third site involves a direct interaction of the AP-site ribose ring with two catalytically important residues, Y127 and Y159 (Fig. 2e, h). Comparison of the AAG–NCP<sup>-30AP</sup> structure with the crystal structure of AAG bound to pyrrolidine-containing free duplex DNA (PDB: 1BNK)<sup>39</sup> revealed a different binding mode for these two residues (Fig. 2i). In the AAG–NCP<sup>-30AP</sup>



structure, Y127 and Y159 interact with the AP site through hydrogen bonds (Fig. 2e, h), whereas the interaction between Y159 and the AP site is absent in the previously reported crystal structure, likely due to the lack of a hydroxyl group in pyrrolidine. In the crystal structure of AAG–DNA complex containing a 1,*N*<sup>6</sup>-ethenoadenine in the pre-catalysis state (PDB: 1EWN)<sup>48</sup>, Y127 and Y159 sandwich the flipped base, and Y127 directly stacks with the modified base (Fig. 2i). Notably, most of these AP site-interacting residues are invariant through evolution (Supplementary Fig. S8), and a previous site-directed random mutagenesis study showed that Y127 could not be substituted and that Y159 was among the group with the lowest mutability<sup>49</sup>.

Overall, AAG uses a similar set of conserved residues to interact with the DNA substrate in both the linear and nucleosomal forms.

#### Distortion of nucleosomal DNA in the NCP<sup>-30I</sup> and AAG–NCP<sup>-30AP</sup> structures

Importantly, the structures of the NCP<sup>-30I</sup> and AAG–NCP<sup>-30AP</sup> demonstrate apparent differences in conformation of nucleosomal DNA when compared with the structures of three representative canonical nucleosomes (PDB: 7OHC, 6ZHX, 6WZ5)<sup>50–52</sup>. These nucleosomes were all derived from *X. laevis* histones and the Widom 601 DNA sequence. The voxel sizes of the published maps were calibrated and the corresponding models were refined in real space before comparison. In the NCP<sup>-30I</sup>, the hypoxanthine base of DI pairs to pyrimidine through weaker hydrogen bonds compared to Watson–Crick base pairs, which destabilizes local DNA and leads to a global perturbation of nucleosomal DNA (Fig. 3a, d, e and Supplementary Fig. S9). Specifically, compared to the root-mean-square deviation (RMSD) values between two canonical nucleosomal DNA backbones (Supplementary Fig. S9c–e), the RMSD values between the damaged-strand DNA of apo NCP<sup>-30I</sup> and that of the canonical NCP are much larger, and increasing displacements are shown towards the nucleosomal DNA ends, especially at the end close to the DI (Fig. 3d, e and Supplementary Fig. S9a, b). Locally, this perturbation is characterized by an outward movement of the DNA from the histone core: the calculated RMSD of the DNA backbone from –25 to –35 (–30 ± 5 positions) is 1.0 Å for both the damaged and intact strands (Fig. 3a).

In the AAG–NCP<sup>-30AP</sup>, the engagement of AAG further distorts the nucleosomal DNA. Previous studies on free DNA substrate indicated that glycosylases require DNA bending at a certain angle to ensure base-flipping and optimal catalysis<sup>53</sup>. Comparing with the NCP<sup>-30I</sup> structure, additional bending and twisting of the nucleosomal DNA around the damage site were observed, which might lead to an optimal DNA bending angle for AAG

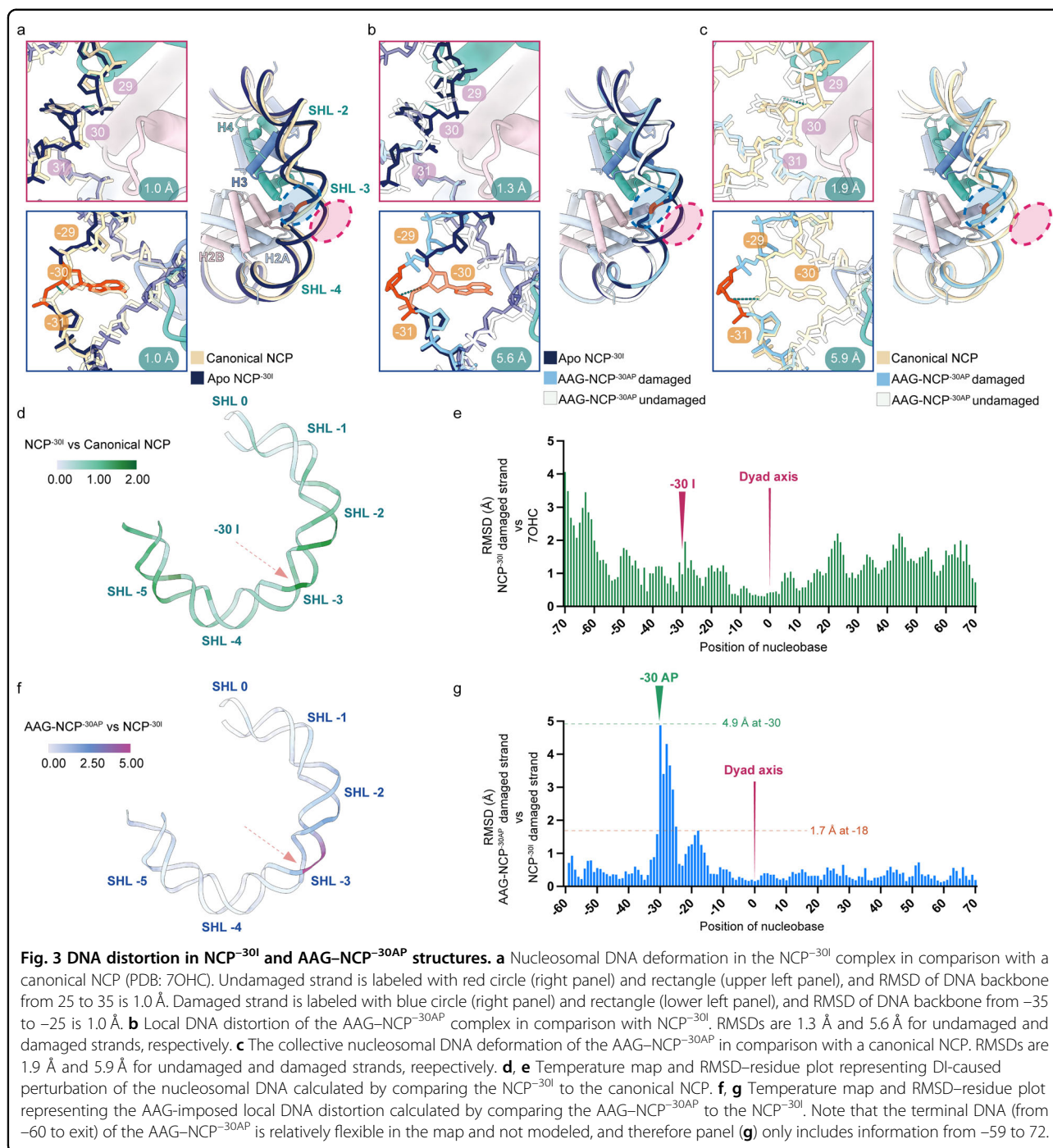
engagement (Fig. 3b, f, g). In general, binding of AAG led to further outward displacement of the DNA around SHL-3, and this distortion was local compared to the global perturbation caused by DI in the NCP<sup>-30I</sup> (Fig. 3f). The distortion out of SHL-3 rapidly decreased, and the maximum was observed at –30 with an RMSD of 4.9 Å (Fig. 3g). In addition, a second distortion peak was seen to be centered around –18 (SHL-2) position (Fig. 3g). The calculated RMSDs around the –30 AP site from –35 to –25 (–30 ± 5 positions) are 5.6 Å and 1.3 Å for the damaged and undamaged strands, respectively (Fig. 3b). A direct comparison of AAG–NCP<sup>-30AP</sup> structure to the canonical NCP confirmed a very dramatic displacement of the damaged strand near SHL-3 (RMSD: 5.9 Å) (Fig. 3c).

Nucleosome imposes restriction to DNA through interactions between nucleosomal DNA and histones<sup>54</sup>. Histone core makes contacts with DNA backbone when the minor groove of nucleosomal DNA faces histone core. Therefore, we analyzed the changes of the interactions between the histone core and the DNA backbone. The presence of DI in NCP<sup>-30I</sup> causes a reduction of the buried surface area from 7141 Å<sup>2</sup> to 6195 Å<sup>2</sup>, indicating that DI alone has significantly destabilized the interaction between the nucleosomal DNA and the octameric histone core (Supplementary Table S1). Further engagement of AAG has a marginal effect on the overall buried surface area (6166 Å<sup>2</sup>), since the dramatic DNA distortion is limited to local region around the damaged site (Fig. 3b).

These results suggest that the damaged base-induced global perturbation of nucleosomal DNA destabilizes the nucleosome and relieves nucleosome-imposed restriction, which likely plays a determinant role in AAG recruitment. Due to the high solution accessibility of the –30 position, additional local DNA distortion induced by AAG alone is sufficient to deform nucleosomal DNA and allow AAG engagement to the damaged site.

#### Distortion of nucleosomal DNA in the NCP<sup>-50I</sup> and AAG–NCP<sup>-50AP</sup> structures

We next examined nucleosomes with a damaged base at a different SHL (SHL-5) to study the effect of translational position. In the AAG–NCP<sup>-50AP</sup> complex, a DI was introduced into the position of –50 at SHL-5, which could be regarded as a roughly equivalent position of –30 in terms of solution accessibility (Fig. 4a, b). Subsequently, the structures of apo NCP<sup>-50I</sup> and AAG–NCP<sup>-50AP</sup> complexes were solved at resolutions of 2.9 Å and 3.0 Å, respectively (Supplementary Figs. S3, S7, S10a, b). Despite the overall high resolution of the AAG–NCP<sup>-50AP</sup> structure, the portion of AAG becomes highly fragmented in the sharpened map, indicating a conformational wobbling of AAG on the nucleosome. Nevertheless, the nucleosomal DNA including the AP site and the inserted residue Tyr162 are well resolved (Supplementary Fig. S10b),

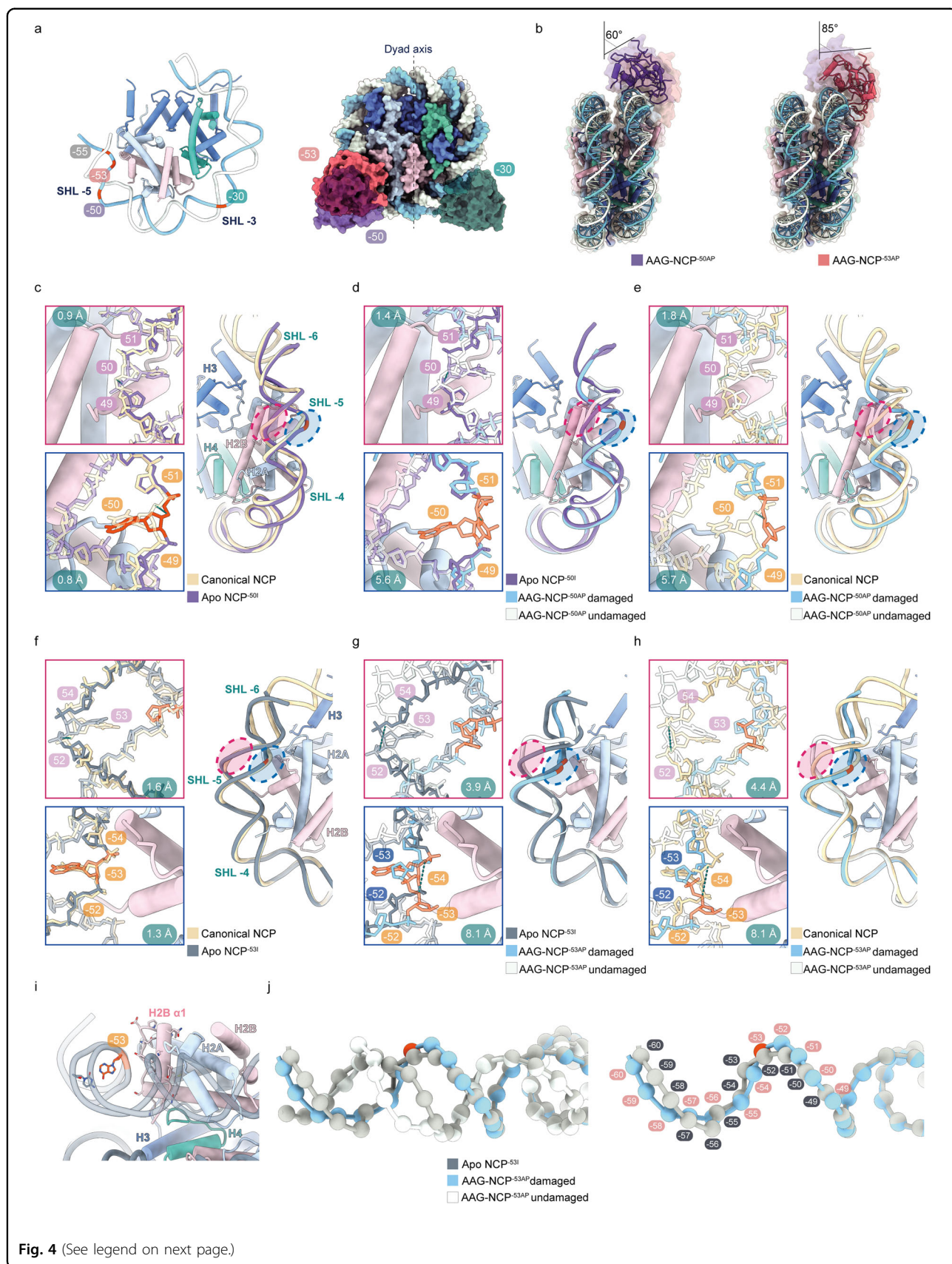


enabling the modeling of the DNA and the assignment of its post-catalytic state. The overall position of AAG was determined by rigid-body fitting.

Comparison of the apo NCP<sup>-50I</sup> to the canonical NCP (PDB: 7OHC)<sup>50</sup> reveals a similar pattern of global perturbation as seen in the structure of the NCP<sup>-30I</sup> (Supplementary Fig. S11a). The presence of DI at -50 causes a noticeable movement of both strands away from the histone core: at -55 to -45 (-50 ± 5 positions) of SHL-5,

the calculated RMSDs for the DNA backbone are 0.8 Å and 0.9 Å for the damaged and undamaged strands, respectively (Fig. 4c).

In the AAG-NCP<sup>-50AP</sup> structure, the engagement of AAG is similar to that in the AAG-NCP<sup>-30AP</sup> in general, although AAG in AAG-NCP<sup>-50AP</sup> is angled at ~60° against the plane parallel to the nucleosome disc (Fig. 4b). Globally, the buried surface area between the DNA and the histone core decreases from 7134 Å<sup>2</sup> to 6248 Å<sup>2</sup> in the



**Fig. 4** (See legend on next page.)

(see figure on previous page)

**Fig. 4 Structural comparison between NCP<sup>-50I</sup>, AAG-NCP<sup>-50AP</sup>, NCP<sup>-53I</sup>, AAG-NCP<sup>-53AP</sup> and the canonical NCP.** **a** Superimposition of the models of the AAG-NCP<sup>-30AP</sup>, AAG-NCP<sup>-50AP</sup> and AAG-NCP<sup>-53AP</sup> to highlight the orientation difference of AAG in the three structures. **b** Comparison of the AAG-NCP<sup>-50AP</sup> and AAG-NCP<sup>-53AP</sup>. **c** The nucleosomal DNA perturbation in the NCP<sup>-50I</sup> in comparison with a canonical NCP (PDB: 7OHC). RMSD of the undamaged DNA backbone from 45 to 55 is 0.9 Å, and RMSD of the undamaged strand from -55 to -45 is 0.8 Å. **d** Local DNA distortion of the AAG-NCP<sup>-50AP</sup> in comparison with the NCP<sup>-50I</sup>. RMSDs are 1.4 Å and 5.6 Å for the undamaged and damaged strands, respectively. **e** The nucleosomal DNA deformation of the AAG-NCP<sup>-50AP</sup> in comparison with a canonical NCP. RMSDs are 1.8 Å and 5.7 Å for the undamaged and damaged strands, respectively. **f** The nucleosomal DNA perturbation in the NCP<sup>-53I</sup> in comparison with a canonical NCP (PDB: 7OHC). RMSD of the undamaged DNA backbone from 48 to 58 is 1.6 Å, and RMSD of the undamaged strand from -58 to -48 is 1.3 Å. **g** In the AAG-NCP<sup>-53AP</sup>, drastic local DNA distortion is accompanied by local register shift. In comparison with the NCP<sup>-53I</sup>, RMSD of the undamaged strand is 3.9 Å, and the RMSD of the damaged strand is 8.1 Å. AP site at -53 is relocated to -52 in relative to the register of the NCP<sup>-53I</sup>. **h** The nucleosomal DNA deformation in the AAG-NCP<sup>-53AP</sup> in comparison with a canonical NCP<sup>-53I</sup>. RMSDs are 4.4 Å and 8.1 Å for the undamaged and damaged strands, respectively. **i** The original DNA register of the NCP<sup>-53I</sup> prohibits base flipping at -53, due to steric clash with histone H2B. **j** The local register shift of the AAG-NCP<sup>-53AP</sup> in comparison to the register of the NCP<sup>-53I</sup>.

presence of DI at -50, and further decreases to 5977 Å<sup>2</sup> after engaging with AAG (Supplementary Table S1). Further DNA distortion was also observed upon AAG engagement, and displayed a predominantly local effect: the maximal distortion was observed at -50 with an RMSD of 3.0 Å (Supplementary Fig. S11b), and the average displacements for the damaged and undamaged strands at -55 to -45 of SHL-5 (-50 ± 5 positions) were 5.6 Å and 1.4 Å, respectively (Fig. 4d). When a canonical NCP was used for comparison, the DNA backbone RMSDs at SHL-5 were 5.7 Å and 1.8 Å for the damaged and undamaged strands, respectively (Fig. 4e).

The degree of local DNA distortion in the AAG-NCP<sup>-50AP</sup> is less than that in the AAG-NCP<sup>-30AP</sup>, but the overall similarity suggests a general mechanism of the engagement of AAG with DNA base damage in solvent-facing positions, which leverages on the global structural perturbation of the nucleosome by DI nucleotide. Since the -50 position is closer to one end of the DNA than the -30 position, the large local DNA distortion around the damage site could have been relaxed by propagating to the end.

#### Structure of the AAG-NCP<sup>-53AP</sup> complex containing an occluded damaged base with medium solution accessibility

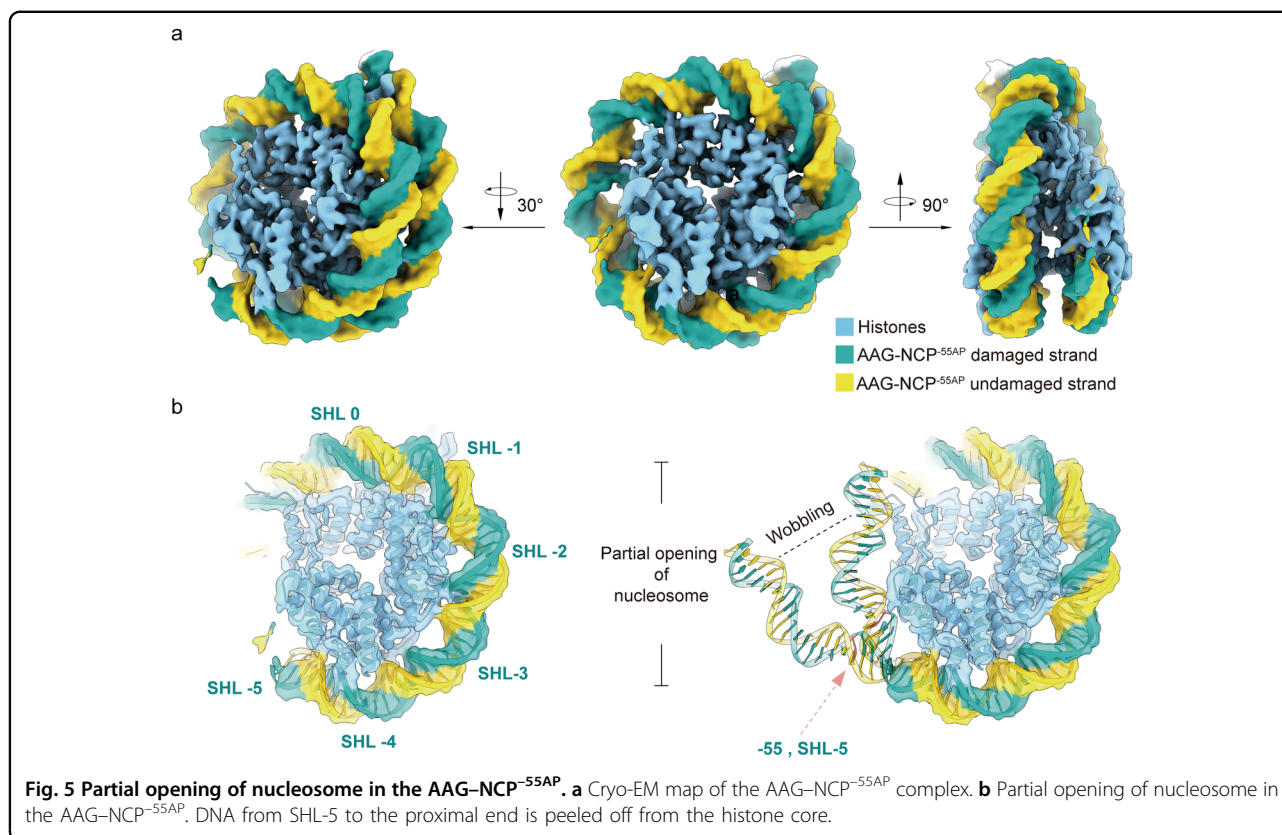
Within an SHL region of the nucleosomal DNA, damaged bases in different rotational orientations should have sharply different solution accessibility, and thus different accessibility to repair proteins. Although the DI in the NCP<sup>-53I</sup> is at an occluded position of SHL-5 with medium solution accessibility (Figs. 1a, 4a), our data show that the AAG-NCP<sup>-53</sup> complex displays a comparable high assembly efficiency as the AAG-NCP<sup>-30</sup> and AAG-NCP<sup>-50</sup> complexes (Supplementary Fig. S4a). We subsequently determined the structures of the NCP<sup>-53I</sup> and AAG-NCP<sup>-53AP</sup> at resolutions of 2.8 Å and 3.1 Å, respectively (Supplementary Figs. S4, S7, S10c, d). Similar to the AAG-NCP<sup>-50AP</sup> complex, AAG in AAG-NCP<sup>-53AP</sup> is flexible, but the nucleosomal DNA

including the AP site and the inserted residue Tyr162 are nicely resolved (Supplementary Fig. S10d). Structural comparison of the NCP<sup>-53I</sup> to the canonical NCP (PDB: 7OHC) reveals a similar global perturbation as in the NCP<sup>-30I</sup> and NCP<sup>-50I</sup> (Supplementary Fig. S11c), featuring a local outward movement of the nucleosomal DNA around the damage site, with the calculated RMSDs of the DNA backbone for the damaged and undamaged strands (-53 ± 5 positions) being 1.3 Å and 1.6 Å, respectively (Fig. 4f).

AAG in AAG-NCP<sup>-53AP</sup> is angled at ~85° against the plane parallel to the nucleosome disc (Fig. 4b). Importantly, in the AAG-NCP<sup>-53AP</sup>, a drastic local DNA distortion was observed (Fig. 4g, h), and the maximal distortion is at -53 with an RMSD of 7.4 Å (Supplementary Fig. S11d). There is a significant widening of the minor groove in the local region of -53 position compared to the apo NCP<sup>-53I</sup> (Fig. 4j), which is also reflected by much larger displacements for both the damaged and undamaged strands: the calculated RMSDs of the DNA backbones from -58 to -48 (-53 ± 5 positions) are 8.1 Å and 3.9 Å for the damaged and undamaged strands, respectively (Fig. 4g). When comparing the AAG-NCP<sup>-53AP</sup> with a canonical NCP (PDB: 7OHC), the RMSDs for the two strands are 8.1 Å and 4.4 Å, respectively (Fig. 4h). Globally, the buried surface area between the DNA and the histone core decreases from 7141 Å<sup>2</sup> to 6075 Å<sup>2</sup> in the presence of DI at -53, and further decreases to 5897 Å<sup>2</sup> after engaging with AAG (Supplementary Table S1).

Intriguingly, in comparison to the local DNA distortions in the AAG-NCP<sup>-30AP</sup> and AAG-NCP<sup>-50AP</sup> (Supplementary Videos S1, S2), an additional translocation of local DNA around -53 was observed (Supplementary Videos S3, S4). The AP site at -53 in the AAG-NCP<sup>-53AP</sup> is relocated to -52 in spatial relation to the translational register of NCP<sup>-53I</sup>. Since base-flipping at -53 on a regular NCP would cause a clash with H2B, this local register shift from -53 to -52 appears to be necessary for AAG to interact with the flipped base (Fig. 4i). It must be noted





that the global translational register is not affected, and this translocation is restricted to the region from the damage site to the near exit end of the nucleosomal DNA.

Therefore, these observations are in line with the previous *in vitro* data that when the  $\epsilon$ A lesion was placed in  $-126$  position ( $-53$  in our structure), AAG was still highly active at this position<sup>34</sup>, indicating that for positions with medium solution accessibility, AAG could induce drastic local DNA distortion that allows local register shift to increase the solution accessibility of the occluded damaged base (Fig. 4).

#### Partial opening of the nucleosome in the AAG-NCP<sup>-55</sup> complex

Different from the previous DI-containing NCPs, the DI in the AAG-NCP<sup>-55AP</sup> is in a completely buried position with low solution accessibility. Therefore, one would expect a very low activity of AAG at this site. To our surprise, the NCP<sup>-551</sup> is still capable of forming a stable complex with AAG, as shown in the glycerol density gradient experiment (Supplementary Fig. S5a). Similarly, cryo-EM analysis revealed two major conformational populations for the AAG-NCP<sup>-55</sup> sample. One is the structure of NCP<sup>-551</sup> determined at  $2.8 \text{ \AA}$  (Supplementary Figs. S5, S7, S10e), and the other is an unusual NCP structure at  $2.9 \text{ \AA}$  resolution, with the terminal nucleosomal DNA highly

flexible, unresolved in the map (from SHL-5 to the proximal end) (Fig. 5a). In this unusual structure, however, density of AAG could not be found. Given the strong association of AAG with the NCPs in our biochemical preparation (Supplementary Fig. S5), it is highly likely that this structure in fact reflects the AAG-NCP<sup>-55AP</sup> complex, in which a drastic conformational change of the DNA at the region of  $-55$  has caused the opening of the terminal DNA (Fig. 5b). This assignment is also supported by the previous data that AAG still retained  $\sim 50\%$  activity towards a damaged base at  $-55$  position<sup>34</sup>.

Similarly, a global perturbation of nucleosomal DNA by DI in NCP<sup>-551</sup> was observed, but the proximal end of the DNA displays a much larger displacement (Supplementary Fig. S11e). The buried surface area between the DNA and the histone core decreases from  $7517 \text{ \AA}^2$  to  $6745 \text{ \AA}^2$  in the presence of DI at  $-55$  (Supplementary Table S1). Locally at  $-60$  to  $-50$  ( $-55 \pm 5$  positions), the  $-55$  DI-induced structural displacements of DNA backbone are  $1.1 \text{ \AA}$  and  $1.3 \text{ \AA}$  for the damaged and undamaged strands, respectively, which are in a similar range to those observed in the NCP<sup>-301</sup>, NCP<sup>-501</sup> and NCP<sup>-531</sup> complexes (Supplementary Fig. S12).

It is of low possibility that this unusual NCP is still in the pre-catalytic state before AAG binding. The dynamic nature of nucleosome could lead to spontaneous unwrapping

of nucleosomal DNA<sup>35–37</sup>, and the significant destabilization of the DNA at the exit observed in the NCP<sup>-551</sup> could further increase the unwrapping possibility. In this scenario, the partial opening of nucleosomal DNA gives a short-lived window for AAG to access the damaged site. The binding of AAG to the detached DNA duplex renders this process less reversible, allowing the formation of a stable AAG–NCP<sup>-55</sup> complex for subsequent excision reaction.

Altogether, the presence of a DI in a completely buried position also leads to a global perturbation of the nucleosomal DNA. When the damaged base is near the exit end of the nucleosomal DNA, the perturbation could increase the possibility of terminal DNA duplex opening to facilitate AAG engagement.

## Discussion

Over the past several decades, immense amount of effort has been devoted to the elucidation of principles governing interactions between DNA-binding proteins and free DNA duplex (reviewed in<sup>55,56</sup>). However, in a more physiological context, how DNA-binding proteins interrogate nucleosomes to detect and read hidden information on chromatin is still not completely clear<sup>57</sup>. Recent structural studies on nucleotide excision repair<sup>58</sup>, retroviral integration<sup>59,60</sup> and pioneer transcription factors<sup>61,62</sup> have revealed that these factors could employ a few different mechanisms, such as nucleosomal DNA register shifting and local DNA distortion, to induce nucleosome deformation to fulfill their molecular functions.

In the present work, we determined a set of cryo-EM structures of the NCPs and AAG–NCP complexes with a damage-mimicking DI nucleotide placed in different representative positions on nucleosome. Although AAG proteins in these complexes are positioned differently on the nucleosomes, the binding mode of AAG on the nucleosome is generally similar to that on free DNA duplex<sup>39,48</sup>, with only limited interactions through several evolutionarily invariant residues with the negatively charged DNA backbone<sup>49</sup> (Fig. 1g). These structures also show a highly conserved mechanism for AAG to recognize the damaged base, involving the insertion of a protruding  $\beta$ -hairpin into the minor groove of the damage site, the replacement of the damaged base by a tyrosine residue, and stabilization of the flipped base/AP site with two aromatic residues (Figs. 1d, 2a–h). These similarities suggest that a crucial and potentially rate-limiting step for base excision reactions could be the recruitment of AAG onto the damage sites.

Interestingly, with pair-wise comparison to the canonical NCP, we discovered that the presence of a single DI nucleotide alone is sufficient to perturb nucleosomal DNA globally (Fig. 3d, e and Supplementary Fig. S11a, c, e), resulting in a collectively reduced

buried surface area between nucleosomal DNA and histone core, by 10%–15% reduction depending on the DI position (Supplementary Table S1). These global perturbations of nucleosomal DNA are generally in a similar pattern regardless of the position of the damaged base, and the DNA deformation caused by DI is most apparent near the exit of nucleosomal DNA (Fig. 3d, e and Supplementary Fig. S11a, c, e).

After formation of the stable AAG–NCP<sup>AP</sup> complex, the binding of AAG induces a dramatic but relatively local distortion on nucleosomal DNA around the damaged base (Fig. 3f, g and Supplementary Fig. S11b, d). Depending on the translational and rotational positions of the damaged site, AAG makes use of distinct mechanisms to get access to the damaged base. For the DI in solvent-facing positions with high solution accessibility, where AAG is highly active<sup>34</sup> and has a direct access to these positions, AAG directly augments the local DNA distortion to recognize the damage site (Supplementary Fig. S13a). For the DI in occluded positions such as –53 with medium accessibility, where AAG activity is expected to be lower comparing to solvent-facing positions<sup>34</sup>, AAG induces drastic local DNA distortion, including both twisting and translocation of DNA to generate a shift of local DNA register by 1 bp to relieve the nucleosome-imposed spatial hindrance for accessing the damaged site (Fig. 4i, j and Supplementary Fig. S13b). As for deeply embedded positions (e.g., –55 in our study), local DNA distortion and limited register shift are insufficient to fully expose the buried base. Our results suggest that globally perturbed nucleosome by DI might be more prone to spontaneous unwrapping<sup>37</sup>. AAG can make use of this feature to capture the detached terminal DNA duplex and render this process less reversible to favor the unwrapping direction. Therefore, in these completely buried sites, a partial opening of the nucleosomal DNA might be a prerequisite for the recruitment and catalysis of AAG (Fig. 5b and Supplementary Fig. S13c).

Altogether, in all these above-mentioned scenarios, the altered structural dynamics of the nucleosome likely plays a major role in recruiting AAG to the damaged sites. The global perturbation of the nucleosomal DNA by DI nucleotide simply reflects the fact that disruption of a single base pair could weaken DNA–histone interaction and alter the conformational landscape of the nucleosome. In a kinetics view, this would result in an increased sampling of the otherwise less possible conformations of the nucleosome. AAG is capable of capturing these transient conformations and forms a stable AAG–NCP complex for the subsequent excision reaction. In summary, our work reveals the effect of a damaged base on nucleosome stability and provides a mechanistic framework for understanding how the DNA glycosylase AAG exploits the structural dynamics of nucleosome to engage with DNA base damage in the nucleosome. In a broader

context, our work also contributes to the general knowledge of how the structural dynamics of nucleosome interplays with DNA-binding proteins to regulate their actions on chromatinized eukaryotic genome.

## Materials and methods

### Construct design and protein expression

The gene encoding the N-terminus-truncated sequence of human AAG (residues 80–298) was cloned into pET22b vector with an N-terminal 6× His tag. Plasmid verified by sequencing was transformed into BL21 (DE3) *Escherichia coli* cells for over-expression. At OD<sub>600</sub> of 0.6, expression of AAG was induced by 0.1 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) at 18 °C for 17 h. Cells were collected by centrifugation (3500 rpm, 10 min, 25 °C) and resuspended in buffer A (20 mM Tris-HCl, pH 7.5, 100 mM NaCl). The suspension was flash-frozen in liquid nitrogen and stored at –80 °C for purification.

### Purification of AAG

All steps were carried out at 4 °C. Frozen cells were thawed and lysed by sonication in buffer A supplemented with 0.1 mM phenylmethanesulfonyl fluoride. The lysate was then supplemented with 0.5 U/mL benzonase nuclease and centrifuged at 18,000 rpm for 40 min. The supernatant was filtered with a 0.45-μm syringe filter, then loaded onto a 5-mL Ni-smart beads 6FF FPLC column (Smart-Lifesciences, Changzhou), and eluted with a linear gradient of imidazole using buffer A and buffer B (20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 M Imidazole). Peak fractions were collected and pooled for dialysis in buffer C (20 mM Tris-HCl, pH 7.5, 100 mM NaCl, 1 mM EDTA, 1 mM DTT) overnight to remove imidazole. Dialyzed sample was centrifuged to remove any possible precipitates, and the supernatant was loaded onto a 5-mL Heparin beads 6FF FPLC column (Smart-Lifesciences, Changzhou) pre-equilibrated with buffer C, and then eluted with a linear gradient of buffer C and buffer D (20 mM Tris-HCl, pH 7.5, 1 M NaCl, 1 mM EDTA, 1 mM DTT). Peak fractions were pooled and concentrated, and then loaded onto a Superdex 75 10/300 GL (Cytiva) column pre-equilibrated in buffer C. Peak fractions were collected and analyzed by SDS-PAGE. The AAG proteins were concentrated to 1 mg/mL and stored at –80 °C in small aliquots.

### Widom 601 DNA preparation

#### Widom 601 top strand sequence

CCTGGAGAATCCCGGTGCCGAGGCCGCTCAATT  
GGTCGTAGACAGCTCTAGCACCGCTTAAACGCAC  
GTACGCGCTGTCCCCGCGTTTTAACCGCCAAGG  
GGATTACTCCCTAGTCTCCAGGCACGTGTCAGA-  
TATATACATCCTGTGCAT

#### Widom 601 –301 bottom strand sequence

ATGCACAGGATGTATATATCTGACACGTGCCTG  
GAGACTAGGGAGTAITCCCCTTGGCGGTTAAAAC  
GCGGGGGACAGCGCGTACGTGCGTTTAAGCGGT  
GCTAGAGCTGTCTACGACCAATTGAGCGGCCTCG  
GCACCGGGATTCTCCAGG

#### Widom 601 –501 bottom strand sequence

ATGCACAGGATGTATATATCTGACACGIGCCTG  
GAGACTAGGGAGTAATCCCCTTGGCGGTTAAAAC  
GCGGGGGACAGCGCGTACGTGCGTTTAAGCGGT  
GCTAGAGCTGTCTACGACCAATTGAGCGGCCTCG  
GCACCGGGATTCTCCAGG

#### Widom 601 –531 bottom strand sequence

ATGCACAGGATGTATATATCTGACICGTGCCTG  
GAGACTAGGGAGTAATCCCCTTGGCGGTTAAAAC  
GCGGGGGACAGCGCGTACGTGCGTTTAAGCGGT  
GCTAGAGCTGTCTACGACCAATTGAGCGGCCTCG  
GCACCGGGATTCTCCAGG

#### Widom 601 –551 bottom strand sequence

ATGCACAGGATGTATATATCTGICACGTGCCTG  
GAGACTAGGGAGTAATCCCCTTGGCGGTTAAAAC  
GCGGGGGACAGCGCGTACGTGCGTTTAAGCGGT  
GCTAGAGCTGTCTACGACCAATTGAGCGGCCTCG  
GCACCGGGATTCTCCAGG

### DNA preparation

The Widom 601 sequence<sup>44</sup> was inserted into pET-51b vector, and the resulting plasmid was used as PCR template. The plasmid was transformed into Trans1-T1 *E. coli* cells and isolated using Plasmid Mini Kit I (Omega Bio-Tek). Damaged base-containing Widom 601 DNA was prepared by PCR with DI-containing primers. The primers used in this study were listed in Supplementary Table S2. A typical purification procedure required 10 mL of PCR reaction product. PCR reaction product was loaded onto a 5-mL Q beads 6FF FPLC column (Smart-Lifesciences, Changzhou) pre-equilibrated with 20 mM Tris-HCl (pH 8.0), and then eluted with linear gradient of NaCl (from 0 to 2 M, 20 mM Tris-HCl, pH 8.0). Peak fractions containing Widom 601 DNA were collected and concentrated using 10 K AmiconUltra-15 centrifugal filter unit (Merck).

Concentrated solution was supplemented with 1/10 volume of 3 M sodium acetate and 2 volumes of 100% ethanol to precipitate DNA. After 1 h of freezing at –40 °C, the mixture was centrifuged at 15000 rpm at 4 °C for 10 min, and the DNA pellet was washed with ice-cold 70% ethanol and dried in air. The resulting DNA pellet was stored at –80 °C for nucleosome reconstitution.

### Histone octamer preparation

Methods for expression and purification of *X. laevis* histones were adapted from previous studies<sup>45,47</sup>. In brief, H2A from *X. laevis* was cloned into pET-22b vector, which contained a 6× His tag and a TEV site at the N-terminus of H2A. H2B, H3 and H4 were cloned into pET-3a vectors.

Histone expression was induced with IPTG at 37 °C, and histone-expressing bacterial cells from 1 L H2A, 4 L H2B, 2 L H3 and 2 L H4 of bacterial cultures were mixed to achieve stoichiometry of 1:1:1:1. Mixed cells were lysed by sonication, and inclusion body was recovered by centrifugation and dissolved in 20 mM acetate, pH 5.2, 8 M guanidine hydrochloride, 10 mM DTT. Supernatants containing denatured histones were recovered by centrifugation, and transferred into dialysis bags for histone octamer refolding. Histone octamer was dialyzed against a buffer (20 mM Tris, pH 8.0, 2 M NaCl, and 2 mM β-mercaptoethanol) at 4 °C for three times, with each time more than 8 h. After dialysis, supernatant containing refolded histone octamer was centrifugated, loaded onto a 5-mL Ni-smart beads 6FF FPLC column (Smart-Life-sciences, Changzhou), and eluted with a linear gradient of imidazole mixed by (20 mM Tris-HCl, 2 M NaCl, 2 mM β-mercaptoethanol) and (20 mM Tris-HCl, 2 M NaCl, 2 mM β-mercaptoethanol, 500 mM imidazole). Peak fractions were collected and concentrated, and then loaded onto Superdex 200 10/300 GL (Cytiva) in buffer E (20 mM Tris-HCl, pH 7.5, 2 M NaCl, 1 mM EDTA, 1 mM DTT). Peak fractions were analyzed by SDS-PAGE, and fractions containing only histone octamer were pooled, concentrated and stored at −80 °C in small aliquots.

### Nucleosome assembly

Nucleosome assembly was performed as previously described<sup>46,63</sup> with some modifications. Briefly, the caps of 1.5-mL Eppendorf tube were used to make dialysis button. Histone octamer and DI-containing Widom 601 DNA were mixed at a molar ratio of 1:1 in buffer E, and incubated for at least 0.5 h on ice. The mixture was then transferred into dialysis buttons sealed with dialysis membrane. Dialysis buttons were transferred into a dialysis bag filled with buffer E. The dialysis bag was dialyzed against buffer F (20 mM Tris-HCl, pH 7.5, 50 mM NaCl, 1 mM EDTA, 1 mM DTT) for 17 h at 4 °C. To complete NCP assembly, the dialysis buttons were dialyzed against fresh buffer F for additional 4 h. Assembled nucleosome was verified by 6% TBE-PAGE and used for AAG–NCP complex assembly.

### Preparation of AAG–NCP complex

All AAG–NCP complexes were prepared with the following method. Nucleosome was mixed with AAG protein at a molar ration of 1:50 in buffer C. The complex was

stabilized by GraFix<sup>64</sup> using TLS-55 rotor, the mixture was loaded onto a glycerol gradient of 20 mM HEPES, pH 7.5, 100 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 20% glycerol (v/v) and 20 mM HEPES, pH 7.5, 100 mM NaCl, 0.1 mM EDTA, 1 mM DTT, 40% glycerol (v/v), 0.05% glutaraldehyde. The sample was centrifuged at 200,000× *g* for 17 h at 4 °C, and fractionated into 100-μL aliquots manually. 10 μL of 1 M Tris-HCl (pH 7.5) was added to each aliquot to quench cross-linking reaction. Each fraction was examined for AAG–NCP complex assembly by 6% TBE-PAGE. Fractions containing AAG–NCP complex were pooled and concentrated, and the glycerol removal and buffer exchange were achieved by ultrafiltration using 50 K AmiconUltra-0.5 centrifugal filter unit. A final buffer of 20 mM Tris-HCl, pH 7.5, 50 mM NaCl, 0.1 mM EDTA, 1 mM DTT was used for cryo-EM sample preparation.

### Cryo-EM sample preparation and data collection

AAG–NCP complexes concentrated to ~1.0 mg/mL were used for cryo-grid preparation. Quantifoil holey carbon Au R1.2/1.3 grids were glow-discharged, and 4 μL of the sample was applied to glow-discharged grids in Vitrobot Mark IV at 8 °C and 100% humidity. After waiting for 10 s, the grids were plunged into the liquid ethane for vitrification. Grids were screened on a Talos Arctica (ThermoFisher) 200 kV TEM equipped with a K2 detector (Gatan). Data collection was performed with an FEI Titan Krios G2 TEM operated at 300 kV with a Gatan K2 (GIF) direct electron detector. Automated data collection was done using SerialEM<sup>65</sup>. Data were collected with nominal magnification of 130,000× (corresponding to a calibrated pixel size of 1.052 Å), with defocus range of −1.2 μm to −1.7 μm. For each movie stack, a total of 32 frames were collected at a dose rate of 8 e<sup>−</sup>Å<sup>−2</sup>s<sup>−1</sup> for 8 s.

### Image processing

For each dataset, the movie stacks were first subjected to beam-induced motion correction, electron-dose weighting and two-fold binning using MotionCor2<sup>66</sup>. Contrast transfer function (CTF) parameters of dose-weighted motion-corrected micrographs were estimated by program Gctf<sup>67</sup>. Around 400 particles were manually picked to generate an initial set of 2D template for automatic particle picking. Two rounds of reference-free 2D classification (25 iterations) were applied after particle auto-picking, and bad 2D classes were discarded. The initial 3D reference was produced using RELION3.1<sup>68</sup>. All subsequent processing steps were performed using RELION 3.1 unless otherwise stated. The image processing workflow was summarized in Supplementary Figs. S2–S5.

For the AAG–NCP<sup>−30</sup> dataset, a total of 3814 raw micrographs were acquired, and 1,421,601 particles were auto-picked. After 2D classification, 1,249,347 particles

were kept and subjected to two rounds of 3D classification with no symmetry imposed, which generated two major states. One is a free NCP (189,350 particles), called apo state. The other is the AAG-bound state (126,881 particles), called post-catalytic state (confirmed in the model building). The particles from these two states were re-extracted and re-centered using a box size of 200 pixels for 3D refinement. Two individual soft-edged masks were used for high-resolution refinement, leading to two density maps at resolutions of 3.0 Å (apo state) and 3.1 Å (post-catalytic state) (gold-standard Fourier shell correlation 0.143 criteria). Application of CTF refinement and Bayesian polishing boosted the resolutions to 2.8 and 2.9 Å (Supplementary Fig. S2), respectively. To further improve the local density of the AAG region in the AAG-bound state, another round of mask-based 3D refinement was performed. Final density maps were corrected for the modulation transfer function of the K2 Summit detector. The map sharpening was carried out by both RELION and DeepEMhancer<sup>69</sup>. The B factors used in RELION were automatically estimated through the post-processing procedure and the local resolution maps were generated using ResMap<sup>70</sup> in RELION.

For the AAG–NCP<sup>-50</sup> dataset, 1,583,682 particles were selected from 3918 raw images. After cleaning up the dataset by two rounds of 2D classification and one round of 3D classification, 460,662 particles were left for further processing. For the apo state, another round of 3D classification was applied and one class showing clear features was chosen and subjected to 3D refinement. Finally, the resolution of the apo state was pushed to 2.9 Å. For the post-catalytic state, all particles from good classes of the first round 3D classification were merged, and a small soft spherical mask centered at the extra density near SHL-5 of the nucleosome was generated to perform a focused 3D classification (skip alignment). A class (73,954 particles) exhibiting strong additional density was selected and further refined. The CTF refinement and Bayesian polishing procedures were subsequently applied to improve the resolution of the final map to 3.0 Å (Supplementary Fig. S3).

For the AAG–NCP<sup>-53</sup> dataset, 4374 raw images were obtained and 1,711,509 particles were auto-picked. The data processing steps of AAG–NCP<sup>-53AP</sup> were similar to those of the AAG–NCP<sup>-50</sup> dataset. A set of 129,507 and 98,691 particles were used to generate the final density maps for the apo state and AAG-bound state at 2.8 Å and 3.1 Å, respectively (Supplementary Fig. S4).

For the AAG–NCP<sup>-55</sup> dataset, 1,706,627 particles were auto-picked from 4363 raw micrographs. Similarly, the final map of the apo state was determined at 2.8 Å resolution, and the final map of the post-catalytic state was determined at 2.9 Å resolution (Supplementary Fig. S5).

### Model building and refinement

Cryo-EM structure of a canonical nucleosome (PDB: 7OHC)<sup>50</sup> was used as the initial template for modeling. For the NCP<sup>-30I</sup>, the crystal structure was first docked into the density map using UCSF ChimeraX<sup>71,72</sup>, and structures of histones and DNA were adjusted manually using Coot<sup>73</sup>. The –30A was replaced with DI. For the AAG–NCP<sup>-30AP</sup>, the structures of the nucleosome (PDB: 7OHC) and AAG (PDB: 1BNK)<sup>39</sup> were similarly modeled, and the –30A was replaced with AP site based on local density. These atomic models were refined in real space using Phenix<sup>74</sup> with the geometry and secondary structure restraints applied. The refined structures were re-checked in Coot to adjust the side chain to proper locations. Final atomic models were evaluated using Molprobit<sup>75</sup>. The same modeling procedures were performed for other maps. Although the AAG regions in the AAG–NCP<sup>-50AP</sup> and AAG–NCP<sup>-53AP</sup> maps were fragmented in the final sharpened maps, the nucleosomal DNA and the inserted residue Tyr162 were clearly resolved. Therefore, the atomic model of AAG was docked into these two density maps by rigid body fitting. The statistics of data collection and model validation were summarized in Supplementary Table S3.

### Structural analysis

Three published NCP structures (EMD-12900, PDB: 7OHC<sup>50</sup>; EMD-11220, PDB: 6ZHX<sup>51</sup>; EMD-21970, PDB: 6WZ5<sup>52</sup>) were used as a reference. Initial structural analysis indicated a scaling factor between the published map and our maps. Therefore, we further calibrated the voxel sizes of the published map and our maps using the crystal structure of a canonical NCP (PDB: 1KX5)<sup>76</sup> as a standard. Based on the calibration, the voxel size of the reference density map (EMD-12900) should be 1.06 Å, and the other two references were not modified (calibrated scaling factors less than 1%). Subsequently, the reference model (PDB: 7OHC) was adjusted by real-space refinement against the calibrated reference map using Phenix. The resulting refined model was used as the atomic model of the canonical NCP for structural analysis. Structural comparisons were performed using UCSF ChimeraX, and atom models were aligned by histone H3 using the matchmaker function. RMSD of nucleosomal DNA was calculated based on atoms from DNA backbone (C1', C2', C3', C4', C5', O3', O4', O5', P, OP1, OP2). Buried surface area between the nucleosomal DNA and the octameric histone core was determined using buried area function in UCSF ChimeraX.

### Acknowledgements

We thank Prof. Qing Li (Peking University) for sharing plasmids of *X. laevis* histones. We thank the Core Facilities of Peking University School of Life Sciences for assistance with negative-staining electron microscopy; the Cryo-EM Platform and the Electron Microscopy Laboratory of Peking University for

cryo-EM data collection; the High-performance Computing Platform of Peking University for help with computation. This work was supported by the Ministry of Science and Technology of China (2019YFA0508904 to N.G.), and the Qidong-SLS Innovation Fund to N.G.

#### Author contributions

N.G. supervised the project. B.T. conceived the project; B.T. and L.Z. designed experiments; L.Z. and B.T. performed all the experiments; L.Z. and B.T. performed EM analysis; B.T., L.Z. and N.G. wrote the manuscript.

#### Data availability

Atomic coordinates of the associated structures have been deposited to the Protein Data Bank (PDB) with the following accession codes 7XFC (NCP<sup>-30I</sup>), 7XFH (AAG-NCP<sup>-30AP</sup>), 7XFI (NCP<sup>-50I</sup>), 7XFJ (AAG-NCP<sup>-50AP</sup>), 7XFL (NCP<sup>-53I</sup>), 7XFM (AAG-NCP<sup>-53AP</sup>), 7XFN (NCP<sup>-55I</sup>) and 7XNP (AAG-NCP<sup>-55AP</sup>), respectively. The corresponding maps have been deposited to the Electron Microscopy Data Bank (EMDB) under the following accession numbers EMD-33171 (NCP<sup>-30I</sup>), EMD-33172 (AAG-NCP<sup>-30AP</sup>), EMD-33173 (NCP<sup>-50I</sup>), EMD-33174 (AAG-NCP<sup>-50AP</sup>), EMD-33175 (NCP<sup>-53I</sup>), EMD-33176 (AAG-NCP<sup>-53AP</sup>), EMD-33177 (NCP<sup>-55I</sup>) and EMD-33322 (AAG-NCP<sup>-55AP</sup>), respectively. All data are available from the corresponding author.

#### Competing interests

The authors declare no competing interests.

#### Publisher's note

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

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41421-023-00560-0>.

Received: 21 July 2022 Accepted: 6 May 2023

Published online: 20 June 2023

#### References

- Barnes, D. E. & Lindahl, T. Repair and genetic consequences of endogenous DNA base damage in mammalian cells. *Annu. Rev. Genet.* **38**, 445–476 (2004).
- Nilsen, H., Lindahl, T. & Verreault, A. DNA base excision repair of uracil residues in reconstituted nucleosome core particles. *EMBO J.* **21**, 5943–5952 (2002).
- Lindahl, T. The intrinsic fragility of DNA (Nobel Lecture). *Angew. Chem. Int. Ed. Engl.* **55**, 8528–8534 (2016).
- Beard, W. A., Horton, J. K., Prasad, R. & Wilson, S. H. Eukaryotic base excision repair: new approaches shine light on mechanism. *Annu. Rev. Biochem.* **88**, 137–162 (2019).
- Wood, R. D. DNA repair in eukaryotes. *Annu. Rev. Biochem.* **65**, 135–167 (1996).
- Saparbaev, M. & Laval, J. Excision of hypoxanthine from DNA containing dIMP residues by the *Escherichia coli*, yeast, rat, and human alkylpurine DNA glycosylases. *Proc. Natl. Acad. Sci. USA* **91**, 5873–5877 (1994).
- Lindahl, T. Recognition and processing of damaged DNA. *J. Cell Sci. Suppl.* **19**, 73–77 (1995).
- Gallagher, P. E. & Brent, T. P. Partial purification and characterization of 3-methyladenine-DNA glycosylase from human placenta. *Biochemistry* **21**, 6404–6409 (1982).
- O'Connor, T. R. Purification and characterization of human 3-methyladenine-DNA glycosylase. *Nucleic Acids Res.* **1**, 5561–5569 (1993).
- Singer, B. et al. Both purified human 1,N<sup>6</sup>-ethenoadenine-binding protein and purified human 3-methyladenine-DNA glycosylase act on 1,N<sup>6</sup>-ethenoadenine and 3-methyladenine. *Proc. Natl. Acad. Sci. USA* **89**, 9386–9390 (1992).
- O'Brien, P. J. & Ellenberger, T. Dissecting the broad substrate specificity of human 3-methyladenine-DNA glycosylase. *J. Biol. Chem.* **279**, 9750–9757 (2004).
- Hang, B., Singer, B., Margison, G. P. & Elder, R. H. Targeted deletion of alkylpurine-DNA-N-glycosylase in mice eliminates repair of 1,N<sup>6</sup>-ethenoadenine and hypoxanthine but not of 3,N<sup>4</sup>-ethenocytosine or 8-oxoguanine. *Proc. Natl. Acad. Sci. USA* **94**, 12869–12874 (1997).
- Miao, F., Bouziane, M. & O'Connor, T. R. Interaction of the recombinant human methylpurine-DNA glycosylase (MPG protein) with oligodeoxyribonucleotides containing either hypoxanthine or abasic sites. *Nucleic Acids Res.* **26**, 4034–4041 (1998).
- Hofseth, L. J. et al. The adaptive imbalance in base excision-repair enzymes generates microsatellite instability in chronic inflammation. *J. Clin. Invest.* **112**, 1887–1894 (2003).
- Klapacz, J. et al. Frameshift mutagenesis and microsatellite instability induced by human alkyladenine DNA glycosylase. *Mol. Cell* **37**, 843–853 (2010).
- Mirabello, L. et al. A comprehensive candidate gene approach identifies genetic variation associated with osteosarcoma. *BMC Cancer* **11**, 209 (2011).
- Cerda, S. R., Turk, P. W., Thor, A. D. & Weitzman, S. A. Altered expression of the DNA repair protein, N-methylpurine-DNA glycosylase (MPG) in breast cancer. *FEBS Lett.* **431**, 12–18 (1998).
- Curtis, C. D., Thorngren, D. L. & Nardulli, A. M. Immunohistochemical analysis of oxidative stress and DNA repair proteins in normal mammary and breast cancer tissues. *BMC Cancer* **10**, 9 (2010).
- Kim, N. K. et al. Expression of the DNA repair enzyme, N-methylpurine-DNA glycosylase (MPG) in astrocytic tumors. *Anticancer Res.* **23**, 1417–1423 (2003).
- Fu, D., Calvo, J. A. & Samson, L. D. Balancing repair and tolerance of DNA damage caused by alkylating agents. *Nat. Rev. Cancer* **12**, 104–120 (2012).
- Wirtz, S. et al. Both base excision repair and O<sup>6</sup>-methylguanine-DNA methyltransferase protect against methylation-induced colon carcinogenesis. *Carcinogenesis* **31**, 2111–2117 (2010).
- Meira, L. B. et al. DNA damage induced by chronic inflammation contributes to colon carcinogenesis in mice. *J. Clin. Invest.* **118**, 2516–2525 (2008).
- Calvo, J. A. et al. Aag DNA glycosylase promotes alkylation-induced tissue damage mediated by Parp1. *PLoS Genet.* **9**, e1003413 (2013).
- Kay, J. E. et al. Excision of mutagenic replication-blocking lesions suppresses cancer but promotes cytotoxicity and lethality in nitrosamine-exposed mice. *Cell Rep.* **34**, 108864 (2021).
- Pich, O. et al. Somatic and germline mutation periodicity follow the orientation of the DNA minor groove around nucleosomes. *Cell* **175**, 1074–1087.e18 (2018).
- Andrews, A. J. & Luger, K. Nucleosome structure(s) and stability: variations on a theme. *Annu. Rev. Biophys.* **40**, 99–117 (2011).
- Luger, K., Mäder, A. W., Richmond, R. K., Sargent, D. F. & Richmond, T. J. Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251–260 (1997).
- Olmon, E. D. & Delaney, S. Differential ability of five DNA glycosylases to recognize and repair damage on nucleosomal DNA. *ACS Chem. Biol.* **12**, 692–701 (2017).
- Caffrey, P. J., Kher, R., Bian, K., Li, D. & Delaney, S. Comparison of the base excision and direct reversal repair pathways for correcting 1,N<sup>6</sup>-ethenoadenine in strongly positioned nucleosome core particles. *Chem. Res. Toxicol.* **33**, 1888–1896 (2020).
- Cannan, W. J., Rashid, I., Tomkinson, A. E., Wallace, S. S. & Pederson, D. S. The human ligase IIIa-XRCC1 protein complex performs DNA Nick repair after transient unwrapping of nucleosomal DNA. *J. Biol. Chem.* **292**, 5227–5238 (2017).
- Hinz, J. M., Mao, P., McNeill, D. R. & Wilson, D. M. 3rd Reduced nuclease activity of apurinic/apyrimidinic endonuclease (APE1) variants on nucleosomes: IDENTIFICATION OF ACCESS RESIDUES. *J. Biol. Chem.* **290**, 21067–21075 (2015).
- Schlaeger, E. J., van Telgen, H. J., Klempner, K. H. & Knippers, R. Association of DNA polymerase with nucleosomes from mammalian cell chromatin. *Eur. J. Biochem.* **84**, 95–102 (1978).
- Mao, P. et al. Genome-wide maps of alkylation damage, repair, and mutagenesis in yeast reveal mechanisms of mutational heterogeneity. *Genome Res.* **27**, 1674–1684 (2017).
- Kennedy, E. E., Li, C. & Delaney, S. Global repair profile of human alkyladenine DNA glycosylase on nucleosomes reveals DNA packaging effects. *ACS Chem. Biol.* **14**, 1687–1692 (2019).
- Buning, R. & van Noort, J. Single-pair FRET experiments on nucleosome conformational dynamics. *Biochimie* **92**, 1729–1740 (2010).
- Polach, K. J. & Widom, J. Restriction enzymes as probes of nucleosome stability and dynamics. *Methods Enzymol.* **304**, 278–298 (1999).
- Fierz, B. & Poirier, M. G. Biophysics of chromatin dynamics. *Annu. Rev. Biophys.* **48**, 321–345 (2019).
- Cotner-Gohara, E. et al. Human DNA ligase III recognizes DNA ends by dynamic switching between two DNA-bound states. *Biochemistry* **49**, 6165–6176 (2010).
- Lau, A. Y., Schärer, O. D., Samson, L., Verdine, G. L. & Ellenberger, T. Crystal structure of a human alkylbase-DNA repair enzyme complexed to DNA: mechanisms for nucleotide flipping and base excision. *Cell* **95**, 249–258 (1998).

40. Mol, C. D., Izumi, T., Mitra, S. & Tainer, J. A. DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination [corrected]. *Nature* **403**, 451–456 (2000).
41. Sawaya, M. R., Prasad, R., Wilson, S. H., Kraut, J. & Pelletier, H. Crystal structures of human DNA polymerase beta complexed with gapped and nicked DNA: evidence for an induced fit mechanism. *Biochemistry* **36**, 11205–11215 (1997).
42. Brooks, S. C., Adhikary, S., Rubinson, E. H. & Eichman, B. F. Recent advances in the structural mechanisms of DNA glycosylases. *Biochim. Biophys. Acta* **1834**, 247–271 (2013).
43. Krokan, H. E. & Bjørås, M. Base excision repair. *Cold Spring Harb. Perspect. Biol.* **5**, a012583 (2013).
44. Lowary, P. T. & Widom, J. New DNA sequence rules for high affinity binding to histone octamer and sequence-directed nucleosome positioning. *J. Mol. Biol.* **276**, 19–42 (1998).
45. Luger, K., Rechsteiner, T. J. & Richmond, T. J. Expression and purification of recombinant histones and nucleosome reconstitution. *Methods Mol. Biol.* **119**, 1–16 (1999).
46. Thåström, A., Lowary, P. T. & Widom, J. Measurement of histone-DNA interaction free energy in nucleosomes. *Methods* **33**, 33–44 (2004).
47. Lee, Y. T., Gibbons, G., Lee, S. Y., Nikolovska-Coleska, Z. & Dou, Y. One-pot refolding of core histones from bacterial inclusion bodies allows rapid reconstitution of histone octamer. *Protein Expr. Purif.* **110**, 89–94 (2015).
48. Lau, A. Y., Wyatt, M. D., Glassner, B. J., Samson, L. D. & Ellenberger, T. Molecular basis for discriminating between normal and damaged bases by the human alkyladenine glycosylase, AAG. *Proc. Natl. Acad. Sci. USA* **97**, 13573–13578 (2000).
49. Chen, C. Y. et al. Substrate binding pocket residues of human alkyladenine-DNA glycosylase critical for methylating agent survival. *DNA Repair* **7**, 1731–1745 (2008).
50. Wang, H., Xiong, L. & Cramer, P. Structures and implications of TBP-nucleosome complexes. *Proc. Natl. Acad. Sci. USA* **118**, e2108859118 (2021).
51. Lehmann, L. C. et al. Mechanistic insights into regulation of the ALC1 remodeler by the nucleosome acidic patch. *Cell Rep.* **33**, 108529 (2020).
52. Bilokapic, S., Suskiewicz, M. J., Ahel, I. & Halic, M. Bridging of DNA breaks activates PARP2-HPF1 to modify chromatin. *Nature* **585**, 609–613 (2020).
53. Stivers, J. T. & Jiang, Y. L. A mechanistic perspective on the chemistry of DNA repair glycosylases. *Chem. Rev.* **103**, 2729–2759 (2003).
54. Luger, K. & Richmond, T. J. DNA binding within the nucleosome core. *Curr. Opin. Struct. Biol.* **8**, 33–40 (1998).
55. Hudson, W. H. & Ortlund, E. A. The structure, function and evolution of proteins that bind DNA and RNA. *Nat. Rev. Mol. Cell Biol.* **15**, 749–760 (2014).
56. Rohs, R. et al. Origins of specificity in protein-DNA recognition. *Annu. Rev. Biochem.* **79**, 233–269 (2010).
57. Michael, A. K. & Thomä, N. H. Reading the chromatinized genome. *Cell* **184**, 3599–3611 (2021).
58. Matsumoto, S. et al. DNA damage detection in nucleosomes involves DNA register shifting. *Nature* **571**, 79–84 (2019).
59. Maskell, D. P. et al. Structural basis for retroviral integration into nucleosomes. *Nature* **523**, 366–369 (2015).
60. Wilson, M. D. et al. Retroviral integration into nucleosomes through DNA looping and sliding along the histone octamer. *Nat. Commun.* **10**, 4189 (2019).
61. Dodonova, S. O., Zhu, F., Dienemann, C., Taipale, J. & Cramer, P. Nucleosome-bound SOX2 and SOX11 structures elucidate pioneer factor function. *Nature* **580**, 669–672 (2020).
62. Michael, A. K. et al. Mechanisms of OCT4-SOX2 motif readout on nucleosomes. *Science* **368**, 1460–1465 (2020).
63. Xu, T. H. et al. Structure of nucleosome-bound DNA methyltransferases DNMT3A and DNMT3B. *Nature* **586**, 151–155 (2020).
64. Stark, H. GraFix: stabilization of fragile macromolecular complexes for single particle cryo-EM. *Methods Enzymol.* **481**, 109–126 (2010).
65. Mastrorade, D. N. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* **152**, 36–51 (2005).
66. Zheng, S. Q. et al. MotionCor2: anisotropic correction of beam-induced motion for improved cryo-electron microscopy. *Nat. Methods* **14**, 331–332 (2017).
67. Zhang, K. Gctf: Real-time CTF determination and correction. *J. Struct. Biol.* **193**, 1–12 (2016).
68. Zivanov, J. et al. New tools for automated high-resolution cryo-EM structure determination in RELION-3. *Elife* **7**, e42166 (2018).
69. Sanchez-Garcia, R. et al. DeepEMhancer: a deep learning solution for cryo-EM volume post-processing. *Commun. Biol.* **4**, 874 (2021).
70. Kucukelbir, A., Sigworth, F. J. & Tagare, H. D. Quantifying the local resolution of cryo-EM density maps. *Nat. Methods* **11**, 63–65 (2014).
71. Pettersen, E. F. et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci.* **30**, 70–82 (2021).
72. Goddard, T. D. et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci.* **27**, 14–25 (2018).
73. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
74. Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 213–221 (2010).
75. Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 12–21 (2010).
76. Davey, C. A., Sargent, D. F., Luger, K., Maeder, A. W. & Richmond, T. J. Solvent mediated interactions in the structure of the nucleosome core particle at 1.9 Å resolution. *J. Mol. Biol.* **319**, 1097–1113 (2002).