STRUCTURAL BIOLOGY

ISSN 2059-7983

Apurinic/apyrimidinic (AP) endonuclease 1 processing of AP sites with 5′ mismatches

Max S. Fairlamb,‡ Amy M. Whitaker‡ and Bret D. Freudenthal*

Department of Biochemistry and Molecular Biology, University of Kansas Medical Center, 3901 Rainbow Boulevard, Kansas City, KS 66016, USA. *Correspondence e-mail: bfreudenthal@kumc.edu

Received 22 November 2017
Accepted 26 February 2018

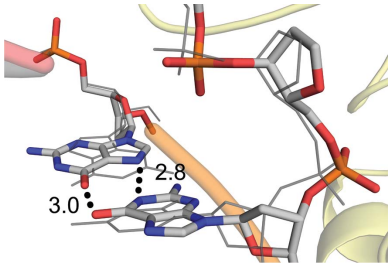
Edited by P. Langan, Oak Ridge National Laboratory, USA

‡ These authors contributed equally.

Keywords: X-ray crystallography; mismatched base pairing; Hoogsteen base pairing; APE1; base-excision repair; apurinic/apyrimidinic endonuclease 1; apurinic/apyrimidinic sites.

PDB references: human APE1, substrate complex with an A·A mismatch adjacent to THF, 6boq; substrate complex with a G·G mismatch adjacent to THF, 6bor; substrate complex with an A·C mismatch adjacent to THF, 6bos; substrate complex with a C·C mismatch adjacent to THF, 6bot; substrate complex with a T·C mismatch adjacent to THF, 6bou; substrate complex with an A·G mismatch adjacent to THF, 6bov; substrate complex with a T·T mismatch adjacent to THF, 6bow

Supporting information: this article has supporting information at journals.iucr.org/d



Despite the DNA duplex being central to biological functions, many intricacies of this molecule, including the dynamic nature of mismatched base pairing, are still unknown. The unique conformations adopted by DNA mismatches can provide insight into the forces at play between nucleotides. Moreover, DNA-binding proteins apply their own individualized steric and electrochemical influences on the nucleotides that they interact with, further altering base-pairing conformations. Here, seven X-ray crystallographic structures of the human nuclease apurinic/apyrimidinic (AP) endonuclease 1 (APE1) in complex with its substrate target flanked by a 5′ mismatch are reported. The structures reveal how APE1 influences the conformations of a variety of different mismatched base pairs. Purine–purine mismatches containing a guanine are stabilized by a rotation of the guanine residue about the N-glycosidic bond to utilize the Hoogsteen edge for hydrogen bonding. Interestingly, no rotation of adenine, the other purine, is observed. Mismatches involving both purine and pyrimidine bases adopt wobble conformations to accommodate the mismatch. Pyrimidine–pyrimidine mismatches also wobble; however, the smaller profile of a pyrimidine base results in a gap between the Watson–Crick faces that is reduced by a C1′–C1′ compression. These results advance our understanding of mismatched base pairing and the influence of a bound protein.

1. Introduction

The faithful replication of the genome during each cell division is essential to prevent genomic mutations induced by mismatched base pairs. These mismatches can arise not only during DNA replication, but also during the repair or bypass of DNA damage *via* specialized, yet relatively error-prone, DNA polymerases (Washington *et al.*, 2010; Freudenthal, Beard & Wilson, 2015). Once established in the genome, mismatches can be repaired either by the mismatch-repair pathway (MMR) or DNA polymerase proofreading mechanisms (Hsieh & Yamane, 2008). In situations where MMR systems are disrupted (*i.e.* Lynch syndrome; Lynch *et al.*, 1966), or the genome is exposed to an excess of DNA-damaging agents, mismatches can accumulate. The resulting genomic instability disrupts cellular homeostasis and promotes cancer-causing mutagenesis (Chatterjee & Walker, 2017).

Canonical base pairs are formed *via* stabilizing hydrogen-bonding interactions between opposing Watson–Crick (WC) faces. In contrast, mismatched nucleotides are contorted into a variety of conformations to accommodate noncanonical base pairing, thus distorting the shape of the DNA helix (Rossetti *et al.*, 2015). Interestingly, mismatch conformations further vary depending on the nature of the protein(s) bound to the DNA, and even functionally related proteins (*i.e.* different DNA polymerases) have been found to elicit distinct base-pairing conformations (Batra *et al.*, 2008, 2016; Johnson & Beese,

2004; Vaisman *et al.*, 2005). These observed differences in base pairing, even within the active sites of similar proteins, hint at the intricate and dynamic nature of the interactions between mismatched bases and DNA-binding proteins. In addition, mismatch-induced structural distortions within a DNA helix can further perturb the activities of critical nucleic acid enzymes responsible for a wide variety of cellular functions (Sassa *et al.*, 2012; Whitaker, Smith *et al.*, 2017; Schermerhorn & Delaney, 2013; Batra *et al.*, 2016). Excluding the enzymes responsible for MMR, which actively seek out the helix-distorting signature of a mismatch, the majority of what is known about the conformation of mismatches in the context of a protein active site comes from polymerase–DNA complex structures (Batra *et al.*, 2016; Johnson & Beese, 2004; Bebenek *et al.*, 2011). Consequently, the molecular-level details of the DNA structural distortion caused by mismatched bases, and the specific effects of this distortion on the activity of other DNA-binding proteins, remains poorly characterized at the atomic level.

Here, we report seven X-ray crystal structures of the essential DNA-repair nuclease human apurinic/apyrimidinic endonuclease 1 (APE1) in complex with DNA substrates containing different mismatched base pairs flanking an abasic, or apurinic/apyrimidinic (AP), site analog. An AP site represents a prevalent DNA lesion, arising at an estimated rate of 10^4 times per cell per day, and a primary target for the APE1 cleavage reaction during base-excision repair (BER; Lindahl, 1993). These structures reveal the unique mismatched base-pairing conformations that occur within the APE1 active site.

2. Methods

2.1. DNA sequences

To generate the 21-mer duplexes for crystallization, the following DNA sequences were used: opposing strand, 5′-GGATCCGTCGA**N**CGCATCAGC-3′; damage-containing strand, 5′-GCTGATGCG**NX**CGACGGATCC-3′. To generate the 30-mer for kinetic studies the following DNA sequences were used: opposing strand, 5′-ATGCGGATCCGTCGA**N**C GCATCAGCGAACG-3′; damage-containing strand labeled with the fluorescein isomer 6-carboxyfluorescein (6-FAM; indicated by an asterisk), 5′-*CGTTCGCTGATGCG**NX**CGA CGGATCCGCAT-3′. **N** represents the mismatched base-pair combination and **X** is the AP-site analog tetrahydrofuran (THF). All sequences were purchased from IDT. The oligonucleotides were separated from other DNA species by electrophoresis on a 16% polyacrylamide gel containing 8 *M* urea in TBE buffer. Purified DNA substrates were annealed in buffer consisting of 50 m*M* Tris, 50 m*M* KCl, and the concentration was determined from the absorbance at 260 nm.

2.2. Expression and purification of APE1

Human wild-type APE1 and a truncated version lacking the flexible 43 N-terminal amino acids (ΔAPE1; Freudenthal, Beard, Cuneo *et al.*, 2015; Mol *et al.*, 2000) were expressed from pET-28a codon-optimized clones purchased from GenScript. All mutagenesis was carried out in either full-length or truncated clones using QuikChange II site-directed mutagenesis (Agilent). APE1 was expressed in One Shot BL21(DE3)pLysS *Escherichia coli* cells (Invitrogen) grown at 37°C, induced at an OD of 0.6 and then grown overnight at 20°C. After harvesting, the cells were lysed at 4°C by sonication in 50 m*M* HEPES pH 7.4, 50 m*M* NaCl and a protease-inhibitor cocktail. The lysate was pelleted at 24 242*g* for 1 h. The resulting supernatant was passed over a HiTrap Heparin HP column (GE Healthcare Life Sciences) equilibrated with lysis buffer. APE1 was eluted from the heparin column with a linear gradient of NaCl up to 1 *M*. APE1 eluting at high salt was buffer-exchanged into 50 m*M* NaCl, loaded onto a POROS HS cation-exchange column (GE Healthcare Life Sciences) and eluted with a linear gradient of NaCl up to 1 *M*. Purified APE1 was subsequently loaded onto a HiPrep 16/60 Sephacryl S-200 HR column (GE Healthcare Life Sciences). The resulting pure fractions were concentrated and stored at −80°C. Final concentrations were determined using a NanoDrop One UV–Vis Spectrophotometer (Thermo Scientific).

2.3. Crystallization and data collection

DNA substrates for ΔAPE1–DNA complex crystals were made by combining 2 m*M* of the two oligonucleotides in a 1:1 ratio and using a PCR thermocycler to heat the mixture for 10 min at 90°C and cool it to 4°C (1°C min^{-1}) to form a 21-mer duplex with a central THF. The annealed DNA was mixed with C138A-ΔAPE1 to achieve a final concentration of 0.56 m*M* DNA and 10–12 mg ml^{-1} protein. The single-amino-acid C138A mutation and truncation of the N-terminal 43 amino acids aid in crystallization (He *et al.*, 2014). ΔAPE1–DNA complexes were crystallized by vapor diffusion. The reservoir solution for crystal formation was 7–14% PEG 20K, 100 m*M* sodium citrate pH 5.0 and 200 m*M* $MgCl_2$. Crystals grew within a week at 20°C. ΔAPE1–DNA crystals were transferred to a cryosolution containing the mother liquor with 20% ethylene glycol. Data were collected at 100 K on a Rigaku MicroMax-007 HF rotating-anode diffractometer equipped with a Dectris PILATUS3 R 200K-A detector system at a wavelength of 1.54 Å. This allowed anomalous data detection after phasing by molecular replacement with high redundancy. Data were processed and scaled with the *HKL*-3000*R* software package (Minor *et al.*, 2006). Initial models were determined by molecular replacement with a modified version of a previously determined ΔAPE1–DNA complex (PDB entry 5dff or 5dfi; Freudenthal, Beard, Cuneo *et al.*, 2015) as a reference. Refinement was carried out with *PHENIX* and model building with *Coot* (Adams *et al.*, 2010; Emsley *et al.*, 2010). Phosphothiolate (PS) linkage-containing substrates were used where indicated in the text and are present as two isomers, *S*p and *R*p. In our crystal structures, we observed both isomers in the active site with equal occupancy. The figures were prepared with *PyMOL* (Schrödinger), and for simplicity only the *R*p conformation is shown.

2.4. APE1 relative product-formation assay

The relative endonuclease activity of APE1 with various proximal mismatched base pairs was determined by analyzing the relative amount of product formation over a period of 30 s. The 30-mer DNA substrates contained a centrally placed mismatch located directly 5′ to an abasic analog (THF). The reactions took place with 5 n*M* APE1 and 500 n*M* annealed DNA substrate in reaction buffer (25 m*M* HEPES, 50 m*M* KCl, 5 m*M* $MgCl_2$, 0.12 mg ml^{-1} BSA) at 37°C. The reactions were initiated by the addition of DNA, and after 30 s the reactions were stopped by the addition of an equal volume of quenching solution (100 m*M* EDTA, 80% deionized formamide, 0.25 mg ml^{-1} bromophenol blue, 0.25 mg ml^{-1} xylene cyanol) to the reaction mixture. The duration of reaction and the concentration of reactants were selected to visualize the full range of activity while maintaining a product formation of less than roughly 50%. Substrate and product DNA were separated on a polyacrylamide gel containing 16% denaturing urea (8 *M*). The 6-FAM-labelled 30-mer ssDNA fragment band (which correlates with substrate) and the 6-FAM-labelled 15-mer ssDNA fragment band (which correlates with product) were imaged using a GE Typhoon 8600 imager in fluorescence mode using a 532 nm excitation laser and a 526 nm short-pass emission filter, and the resulting image was quantified using the *ImageJ* software. The relative product formation was calculated by dividing the product by the sum of the product and the substrate and represents the average of three experiments presented as the mean ± the standard error of the mean (SEM).

3. Results

APE1 is an essential DNA-repair nuclease in the BER pathway (Fig. 1*a*; Li & Wilson, 2014; Whitaker, Schaich *et al.*, 2017). To cleanse the genome of AP sites, APE1 utilizes a base-flipping endonuclease mechanism to incise the DNA phosphodiester backbone at the 5′ side of an AP site (Fig. 1*b*; Mol *et al.*, 2000; Freudenthal, Beard, Cuneo *et al.*, 2015). Rapid cleavage requires proper positioning of the AP site within the APE1 active site (Fig. 1*b*). A flanking mismatched base pair on the 5′ side of an AP site, as opposed to a match, results in at least a fourfold to tenfold reduction in the APE1 cleavage reaction rate (Schermerhorn & Delaney, 2013; Wilson *et al.*, 1995). To investigate the dependence of the activity of APE1 on the nature of the mismatch, we performed APE1 activity assays with a set of oligonucleotide substrates each containing a different mismatched base pair immediately 5′ to a central AP-site analog, tetrahydrofuran (THF). A schematic of the DNA substrates used in this study is presented in Fig. 2(*a*). The site of the abasic analog THF is depicted flipped out of the helix as per the base-flipping mechanism mentioned previously. The relative position of bound APE1 is shown in yellow and the site of impending APE1 cleavage is indicated by a red X. The mismatched base pair is 5′ to the abasic analog and is indicated by $N_O{\cdot}N_A$. N represents either adenine (A), cytosine (C), guanine (G) or thymine (T), with the subscript O indicating the registry opposite the APE1 active site (N_O) and the subscript A representing the 5′ nucleotide immediately adjacent to the THF (N_A). The relative APE1 activity is reflective of the amount of substrate converted to product

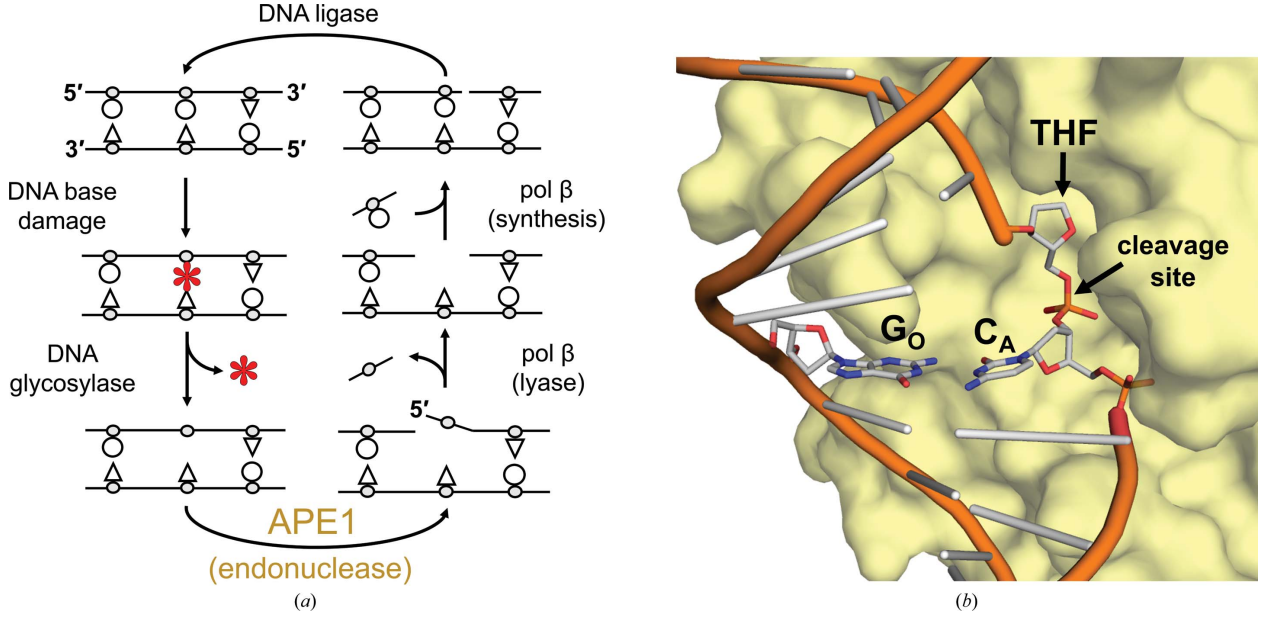


Figure 1
The BER pathway and APE1 substrate binding. (*a*) During BER the damaged base (red asterisk) is removed by a glycosylase, leaving an AP site. The AP site is removed by sequential cleavage of the DNA backbone 5′ to the AP site by APE1 and polymerase (pol) *β* lyase activity. Pol *β* inserts a new nucleotide and the resulting 3′ nick is sealed by DNA ligase. (*b*) Surface representation of ΔAPE1 (yellow; PDB entry 5dff) bound to a THF substrate flipped into the active site with a matched G·C base pair 5′ to the THF. The site of cleavage is indicated and the nomenclature for the opposite (G_O) and adjacent (C_A) base pairs is indicated by subscripts.

during the course of the assay. The resulting levels of substrate and product for each mismatched substrate are shown in the gel depicted in Fig. 2(*b*), which is representative of three replicate assays. The average (assays performed in triplicate) percentage of substrate converted to product is shown below the corresponding mismatches. In general, APE1 activity was inhibited the least by substrates containing pyrimidine–pyrimidine mismatches and the most by purine–purine mismatches. The order of effect, with those mismatches that retain the most activity listed first, was as follows: $T_O{\cdot}C_A$ > $C_O{\cdot}C_A$, $A_O{\cdot}C_A$ > $T_O{\cdot}T_A$ > $A_O{\cdot}G_A$ > $A_O{\cdot}A_A$ > $G_O{\cdot}G_A$.

Table 1
Data-collection and refinement statistics.

Values in parentheses are for the highest resolution shell.

	G·G mismatch	A·G mismatch	A·A mismatch	A·C mismatch	T·T mismatch	T·C mismatch	C·C mismatch
Data collection							
Space group	*P*1	*P*1	*P*1	*P*1	*P*1	*P*1	*P*1
Unit-cell parameters							
a, *b*, *c* (Å)	44.4, 60.2, 73.0	44.3, 60.3, 73.3	44.3, 60.9, 73.2	44.2, 60.7, 72.4	44.2, 61.3, 73.1	44.3, 60.9, 73.3	44.2, 60.9, 73.3
α, *β*, *γ* (°)	82.9, 80.3, 89.1	83.5, 78.4, 88.3	83.4, 78.2, 87.5	83.6, 79.2, 88.3	83.3, 78.2, 86.9	83.3, 78.5, 87.3	83.1, 78.4, 87.2
Resolution (Å)	25–1.85	25–1.98	25–1.96	25–2.31	25–1.60	25–2.55	25–2.32
R_{meas} (%)	66.1 (9.4)	59.7 (7.8)	60.4 (7.2)	64.6 (10.2)	62.9 (6.9)	61.3 (13.0)	69.2 (9.3)
$\langle I/\sigma(I)\rangle$	16.8 (2.0)	21.8 (2.0)	23.6 (2.2)	15.5 (2.1)	26.2 (2.1)	12.2 (1.8)	15.2 (2.1)
$CC_{1/2}$†	0.468	0.801	0.583	0.781	0.857	0.800	0.800
Completeness (%)	99.4 (97.0)	100 (99.8)	99.2 (97.8)	99.8 (99.3)	99.3 (97.3)	99.8 (99.8)	99.9 (99.4)
Multiplicity	4.2 (2.2)	4.3 (2.5)	4.3 (2.4)	3.9 (2.3)	4.0 (2.8)	2.9 (2.7)	4.4 (3.0)
Refinement							
Resolution (Å)	25–1.84	25–1.98	25–1.96	25–2.31	25–1.60	25–2.55	25–2.32
No. of reflections	115113	98881	94749	58993	190771	43274	62321
R_{work}/R_{free} (%)	17.0/20.9	21.1/24.2	22.2/25.5	19.3/23.6	21.4/24.1	20.8/26.8	20.2/24.4
No. of atoms							
Protein	4309	4258	4269	4299	4200	4235	4249
DNA	848	847	876	875	855	854	872
Water	584	310	227	206	560	76	233
B factors (Å^2)							
Protein	19.4	36.3	33.2	34.6	30.2	40.6	38.1
DNA	31.8	52.9	52.4	50.3	42.7	54.0	53.1
Mismatch‡	18.8/26.7	51.4/56.4	50.5/58.1	39.15/46.4	35.5/38.8	44.9/53.7	41.9/54.7
Water	47.9	49.4	41.9	43.2	39.1	39.5	38.2
R.m.s. deviations							
Bond lengths (Å)	0.01	0.01	0.01	0.01	0.01	0.01	0.01
Bond angles (°)	1.44	1.17	0.93	0.83	1.27	0.97	0.98
PDB code	6bor	6bov	6boq	6bos	6bow	6bou	6bot

† For the highest resolution shell. ‡ The reported mismatch combination corresponds to the adjacent and opposing base, respectively.

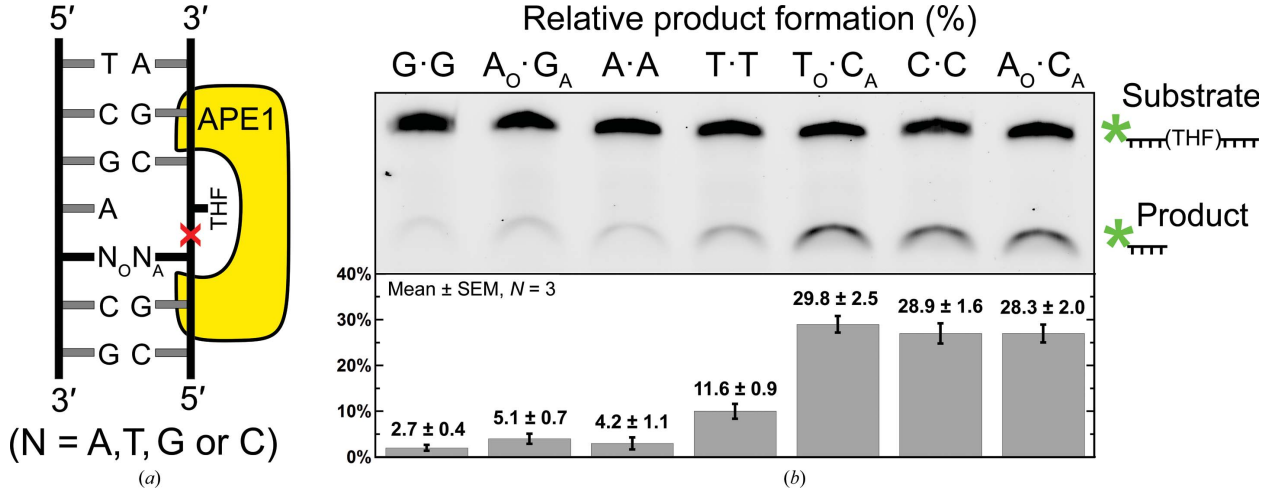


Figure 2
Relative product formation by APE1 with various mismatch combinations adjacent to THF. (*a*) Schematic of the mismatch substrates used in the study, where N_O denotes the nucleotide opposite the APE1 active site and N_A represents the 5′ nucleotide adjacent to THF. (*b*) A representative gel of three replicate assays shows the relative APE1 endonuclease activity with a G·G, $A_O{\cdot}G_A$, A·A, T·T, $T_O{\cdot}C_A$, C·C and $A_O{\cdot}C_A$ mismatch base pair placed directly 5′ to the THF. The substrate and product locations are indicated next to the denaturing gel image, with an asterisk denoting the 6-FAM label used for quantification. The average percentage of product formation is shown as the mean ± SEM, with N = 3.

The APE1 activity assay presented in Fig. 2(*b*) indicates that the rate of the APE1 cleavage reaction varies considerably depending on the specific mismatch flanking the AP site. From this result, we infer that unique, mismatch-specific structural changes within the DNA and/or APE1 active site occur to accommodate electrostatic and steric hindrances resulting from the mismatch. To elucidate the molecular-level details of these structural changes, we obtained X-ray crystal structures of precatalytic complexes of truncated APE1 (lacking the 43 N-terminal amino acids; ΔAPE1) with each 5′-THF mismatch combination (Freudenthal, Beard, Cuneo *et al.*, 2015; Mol *et al.*, 2000).

3.1. Purine–purine mismatches 5′ to a THF lesion

Of the mismatch combinations that were examined, a G·G mismatch 5′ to the AP site showed the most dramatic reduction in APE1 cleavage activity in our activity assays (Fig. 2*b*, lane 1). To determine the structural changes that occur with a G·G mismatch at this position, we obtained a 1.85 Å resolution X-ray crystal structure (Table 1) of the corresponding precatalytic APE1–DNA substrate complex utilizing a catalytically dead variant (E96Q/D210N) of ΔAPE1 (McNeill & Wilson, 2007). We chose to characterize the substrate complex in order to visualize the mismatch conformation prior to the cleavage event, which could result in the relief of structural constraints between the mismatched bases. This structure reveals a rotation of G_O about its N-glycosidic bond into the *syn* conformation, thus allowing it to base-pair with its Hoogsteen edge to the WC face of the opposing *anti*-G_A (Fig. 3*a*). In this conformation, N7 and O6 of the *syn*-G_O Hoogsteen edge are within 2.8 and 3.0 Å of N2 and N1 of *anti*-G_A, respectively. Known canonical proton donors in WC base pairing indicate that the likely proton donors in this structure are *anti*-G_A N2 and *anti*-G_A N1 (Fig. 3*a*, yellow dots). In addition, we observe a C1′–C1′ expansion to 11.2 Å, which is ∼0.7 Å wider than the matched G_O·C_A substrate (10.5 Å; PDB entry 5dgo; Freudenthal, Beard, Cuneo *et al.*, 2015). This results in a 0.9 Å shift of *anti*-G_A towards the minor groove and a 0.8 Å displacement of the 5′ phosphate cleavage site relative to the matched G_O·C_A DNA substrate (Fig. 3*a*). Overall, the structure demonstrates that the electrostatic and steric clashes resulting from a G·G mismatch are preferably accommodated by an N-glycosidic bond rotation of the G located opposite to ΔAPE1, which generates a structural distortion of the DNA backbone relative to a matched G_O·C_A substrate.

We also determined an analogous structure with an A_O·G_A mismatch to 1.95 Å resolution (Table 1). In this structure, *syn*-G_A utilizes its Hoogsteen edge to base-pair with *anti*-A_O. We assume a canonical proton donor and acceptor pair of *anti*-A_O N6 and *syn*-G_A O6 at 3.1 Å (Fig. 3*b*, yellow dot). In addition, two atoms which both typically accept protons in a WC base pair (*anti*-A_O N1 and *syn*-G_A N7) are 2.8 Å apart. The pK_a of adenine N1 has been experimentally determined to be 4.5 when free in solution, but 6.0 in the context of an A^+·G mismatch (Carbonnaux *et al.*, 1991) and between 7.0 and 8.0 in an A^+·C mismatch (Boulard *et al.*, 1992; Moody *et al.*, 2004; Siegfried *et al.*, 2010), thus suggesting that the pK_a could be higher than that in free solution in the case of our reported mismatch structure. If so, a shared proton could exist between

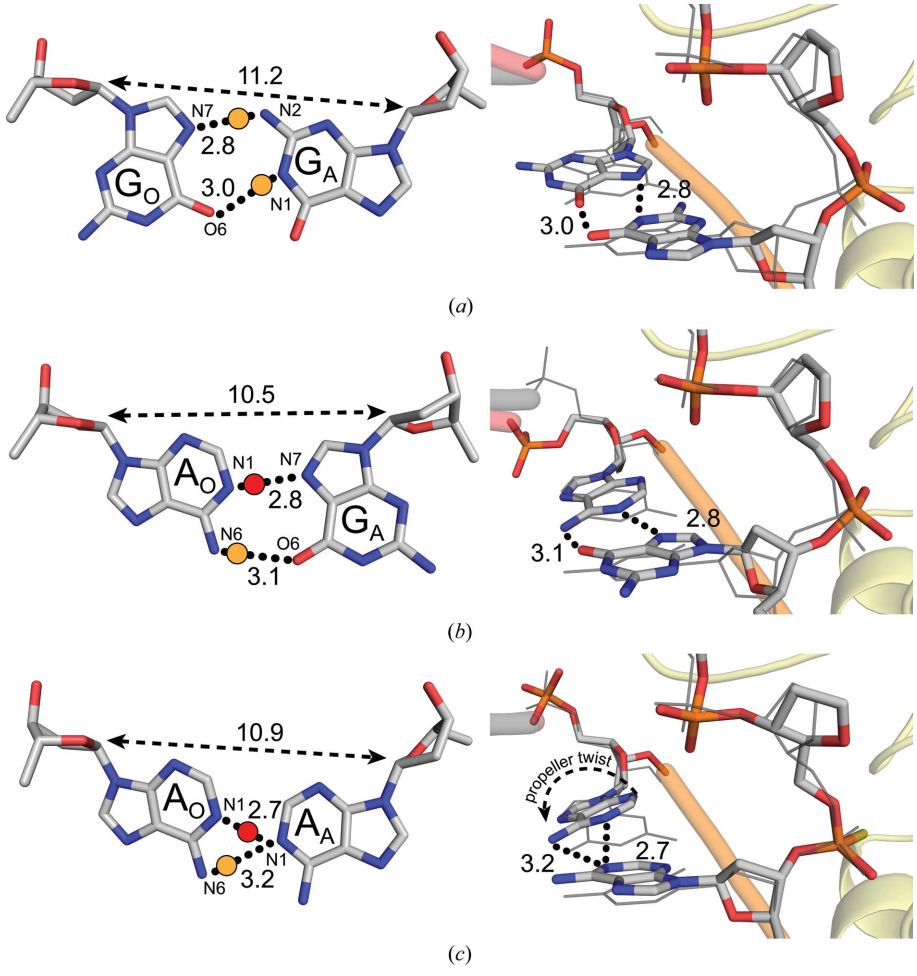


Figure 3
ΔAPE1–DNA complex structures with purine–purine mismatches. A close-up view (left) of the base-pairing interactions and a side view (right) of the superposition of the (*a*) G·G, (*b*) A_O·G_A and (*c*) A·A mismatches with a G_O·C_A reference structure (gray lines; PDB entry 5dfi). Black dotted lines indicate potential hydrogen-bonding interactions with distances. A yellow dot is an expected proton donor with respect to the Watson–Crick donor/acceptor pairs. A red dot denotes a possible proton donor following protonation of the typical proton acceptor. The double arrow with a dashed line denotes the C1′–C1′ distance.

these atoms at pH 5.0 (Fig. 3*b*, red dot). Importantly, the G on the protein side rotates into a Hoogsteen conformation (*syn*-G_A), in contrast to the G·G structure, where the opposite G rotates
(*syn*-G_O). This indicates that in the 5′-THF positon, purine–purine mismatches are preferably accommodated by a *syn*-G conformation over *syn*-A in the ΔAPE1–DNA complex. Of note, we attempted to collect data for both G·G and A_O·G_A precatalytic complexes with wild-type APE1 using a PS linkage-containing substrate, but were unable to obtain sufficiently diffracting crystals. While we do not expect the mutations (E96Q and D210N) to affect the observed base-pairing conformations, the presence of the mutations needs to be considered when interpreting the subtle changes seen in the position of the backbone phosphate within the active site.

To obtain an X-ray crystal structure of ΔAPE1 in complex with a 5′-THF A·A mismatch-containing substrate, we employed a modified DNA substrate containing a PS linkage. This modification substitutes sulfur for a nonbridging O atom, preventing incision (Wilson *et al.*, 1995; Mundle *et al.*, 2009; Freudenthal, Beard, Cuneo *et al.*, 2015). The resulting crystal diffracted to 1.96 Å resolution (Table 1). Unlike the G·G mismatched substrate, both adenines remained in the *anti* conformation, with A_O undergoing a propeller twist to accommodate the mismatch. In this conformation, N1 of A_A is the only atom within hydrogen-bonding distance of the A_O WC face (Fig. 3*c*) and sits 3.2 Å from A_O N6 and 2.7 Å from A_O N1. Although adenine N1 typically accepts protons from an opposing thymine in canonical base pairing, it is possible that one or both of the adenine N1 atoms depicted in this structure are protonated, as described previously. Therefore, the A·A mismatch could have a hydrogen bond between A_A N1 and A_O N6 (Fig. 3*c*, yellow dot) or between A_A N1 and A_O N1 (Fig. 3*c*, red dot), or a mixture of the two. Additionally, the lack of a rotation to utilize the Hoogsteen face suggests that the non-ideal electrostatic and steric forces of the A·A mismatch are not alleviated by Hoogsteen base pairing, as was observed for the G·G and A_O·G_A mismatches bound to ΔAPE1.

3.2. Purine–pyrimidine mismatches 5′ to a THF lesion

A structure containing an A_O·C_A mismatch was solved to a resolution of 2.31 Å (Table 1). To capture a precatalytic substrate complex, we utilized a PS linkage located between the 5′ mismatch and the THF analog as described above. The A_O·C_A structure adopts a wobble conformation to accommodate the electrostatic and steric forces (Fig. 4). This occurs *via* a 0.9 Å shift in A_O into the minor groove of the DNA and towards the ΔAPE1 active site relative to a matched G_O·C_A DNA substrate with canonical WC base pairing. This is consistent with the wobble-base pair conformations observed for A·C mismatches within duplex DNA (Boulard *et al.*, 1992). In our structure, this shift aligns O2 and N3 of *anti*-C_A within hydrogen-bonding distance of N1 (2.7 Å) and N6 (3.1 Å) of *anti*-A_O, respectively (Fig. 4). C_A N3 and A_O N6 are likely to form a proton-donor/acceptor pair considering typical WC hydrogen-bonding configurations (Fig. 4, yellow dot). The C_A O2 and A_O N1 atoms are both typically proton acceptors in WC base pairing, but with the possibility of adenine N1 protonation at pH 5, a hydrogen bond may also be present between these atoms (Fig. 4, red dot; Boulard *et al.*, 1992; Moody *et al.*, 2004; Siegfried *et al.*, 2010). A similar wobble shift conformation was previously observed in a structure with a G_O·T_A mismatch 5′ to a THF in the ΔAPE1 active site (Freudenthal, Beard, Cuneo *et al.*, 2015), suggesting a common theme for pyrimidine–purine mismatches in the context of a ΔAPE1–DNA complex.

3.3. Pyrimidine–pyrimidine mismatches 5′ to a THF lesion

The smaller profile of a pyrimidine nucleotide base, in comparison to a purine, provides a challenge for stable hydrogen-bonding interactions in pyrimidine–pyrimidine mismatches. To obtain a substrate complex of ΔAPE1 with a T·T mismatch 5′ to a THF, we utilized an oligonucleotide containing a PS linkage as described above. A resulting crystal diffracted to a resolution of 1.60 Å (Table 1). The subsequent structure, shown in Fig. 5(*a*), reveals that the T·T WC faces come within hydrogen-bonding distance through a C1′–C1′ compression and wobble shift. The C1′–C1′ distance tightens to 9.0 Å, which is 1.5 Å shorter than the idealized 10.5 Å C1′–C1′ span in G_O·C_A WC base pairing (Fig. 5*a*). The wobble conformation is established when T_O shifts towards the minor groove and T_A shifts towards the major groove. This aligns N3 and O4 of T_O within 3.0 and 2.9 Å of O2 and N3 of T_A, respectively. T_A N3 most likely donates a proton to T_O O4 in addition to T_O N3 donating a proton to T_A O2 (Fig. 5*a*, yellow dots).

Precatalytic structures of ΔAPE1–DNA complexes with either a T_O·C_A or a C·C mismatch located 5′ to the THF were collected to 2.55 and 2.32 Å resolution, respectively, utilizing a PS linkage between the mismatch and THF. Similar to the T·T structure, both of these mismatched base pairs adopt wobble base-pairing interactions (Fig. 5). In the T_O·C_A and C·C structures the gaps between the pyrimidine WC faces are bridged by C1′–C1′ compressions of 1.3 and 1.5 Å, respectively, relative to the G_O·C_A WC base pair (Figs. 5*b* and 5*c*). In the T_O·C_A structure N3 and O4 of T_O sit within 3.3 and 3.0 Å of O2 and N3 of C_A, respectively (Fig. 5*b*). The hydrogen-bonding profile of this structure is unclear, as the likely proton-donor and acceptor pair of T_O N3 and C_A O2 are slightly distant at 3.3 Å (Fig. 5*b*, yellow dot). In addition, T_O O4 and C_A N3 are both typically proton acceptors (Fig. 5*b*); however, similarly to adenine N1, studies have indicated that the pK_a of cytosine N3 increases dramatically in certain electrochemical environments (Bink *et al.*, 2002; Nikolova *et al.*, 2013). Therefore, a hydrogen bond between C_A N3 and the opposing T_O O4 may occur if the local environment perturbs the pK_a of C_A N3 to reach a high enough level for protonation (Fig. 5*b*, red dot). Similarly, in the C·C structure N3 and N4 of C_O are within 3.2 and 3.1 Å of O2 and N3 of C_A (Fig. 5*c*). If C_O N3 is protonated at pH 5 in this particular environment, C_O N3 and C_A O2 could potentially form a proton-donor/

acceptor pair (Fig. 5*c*, red dot). Although C_A N3 protonation is predicted for the T_O·C_A mismatch, C_O N4 typically acts as a proton donor in WC base pairs, suggesting that its pK_a could be lower than that of C_A N3 and that it is more likely to hold a proton. Therefore, a proton-donor/acceptor pair between C_O N4 and C_A N3 is presumed (Fig. 5*c*, yellow dot). In general, these mismatches both adopt similar conformations to the T·T mismatch, suggesting a common wobble shift and C1′–C1′ compression theme in pyrimidine–pyrimidine mismatches within the APE1 active site.

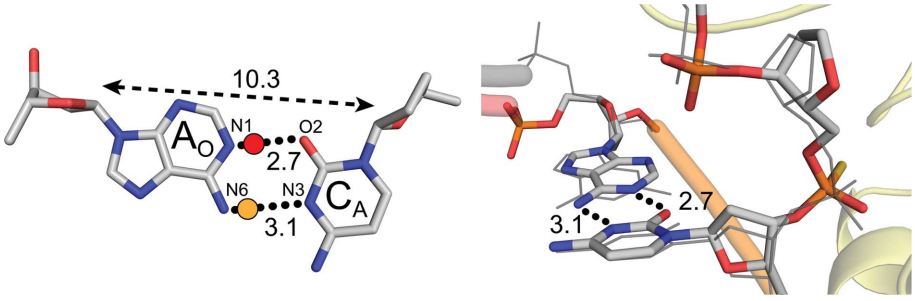


Figure 4
ΔAPE1–DNA complex structure with a purine–pyrimidine A_O·C_A mismatch. A close-up view (left) of the base-pairing interactions and a side view (right) of the superposition with a G_O·C_A reference structure (gray lines; PDB entry 5dfi). Black dotted lines indicate potential hydrogen-bonding interactions with distances. A yellow dot is an expected proton donor with respect to the Watson–Crick donor/acceptor pairs. A red dot denotes a possible proton donor following protonation of the typical proton acceptor. The double arrow with a dashed line denotes the C1′–C1′ distance.

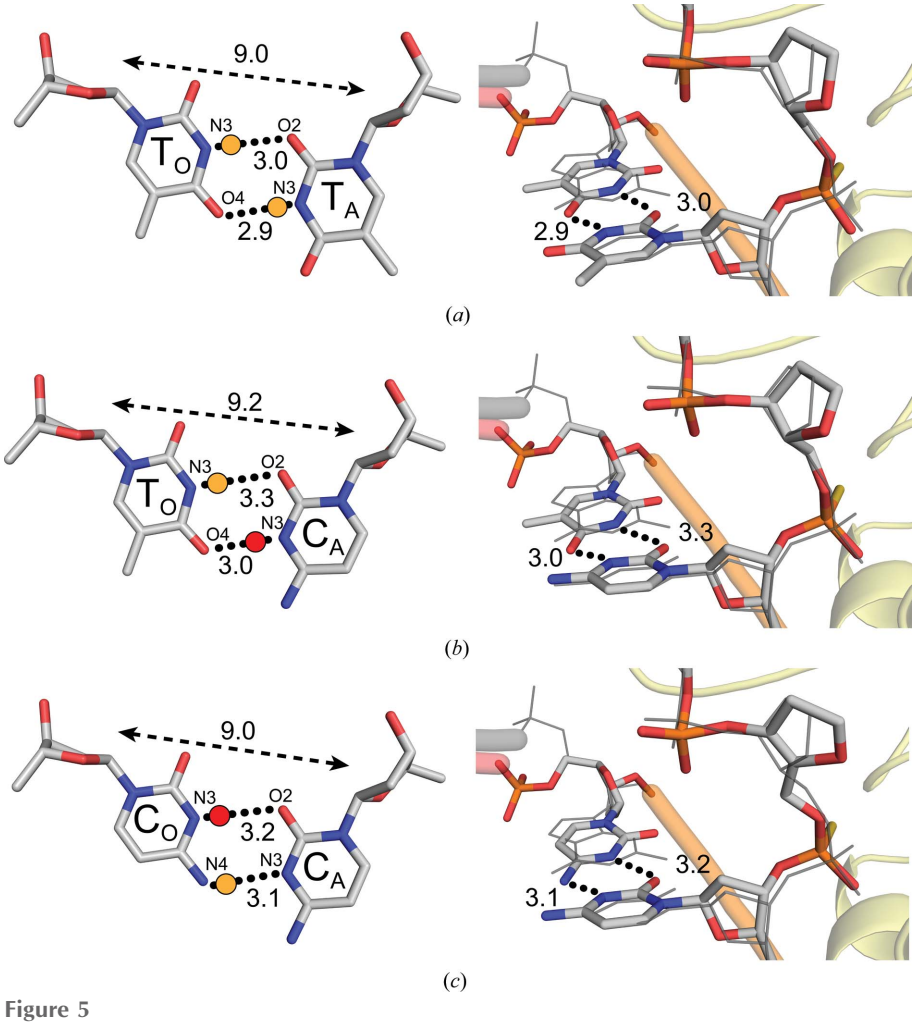


Figure 5
ΔAPE1–DNA complex structures with pyrimidine–pyrimidine mismatches. A close-up (left) view of the base-pairing interactions and a side view (right) of the superposition of the (*a*) T·T, (*b*) T_O·C_A and (*c*) C·C mismatches with a G_O·C_A reference structure (gray lines; PDB entry 5dfi). Black dotted lines indicate potential hydrogen-bonding interactions with distances. A yellow dot is an expected proton donor with respect to the Watson–Crick donor/acceptor pairs. A red dot denotes a possible proton donor following protonation of the typical proton acceptor. The double arrow with a dashed line denotes the C1′–C1′ distance.

4. Discussion

Here, we have characterized mismatched base pairing in the context of a DNA nuclease active site to gain insight into the strategies utilized by nucleotides to accommodate imperfect steric and electrochemical environments. The changes in the APE1 endonuclease activity were found to vary depending on the mismatch combination 5′ to the AP-site analog THF, suggesting that each mismatch induces a unique structural distortion of the DNA helix. We observed a general trend between the nature of the conformation of the mismatch and its effect on the APE1 cleavage reaction, with Hoogsteen base pairing more substantially reducing activity than wobble base pairing in this position. However, an exact link between structural distortion and reduction in activity was not clear, in part because of the challenges in assigning protonation states based on X-ray crystallography.

Several key insights were gained from the analysis of these structures. We observed a preference for a mismatched guanine positioned opposite the APE1 active site (G_O) to flip into a Hoogsteen conformation over a mismatched guanine in the adjacent position (G_A). This rotation about the N-glycosidic bond aligns two probable proton-donor/acceptor pairs between the guanine Hoogsteen face and the opposing guanine WC face (Fig. 3*a*). The *anti*·*syn* conformation of a G·G mismatch has been suggested to be strongly disfavored in duplex DNA (Rossetti *et al.*, 2015); however, this conformation appears to be stable within the APE1 active site. Although G_O prefers to flip

over G_A, the $A_O{\cdot}G_A$ ΔAPE1 complex reveals that G_A will flip to use its Hoogsteen edge over A_O, pointing to a reluctance of adenine to adopt the *syn* conformation. This observation was further supported by the *anti·anti* conformation adopted by the A·A mismatch. Importantly, a rotation of adenine into the *syn* conformation would not result in any obvious proton-donor/acceptor pairs with the opposing adenine WC face, as is the case for G·G and $A_O{\cdot}G_A$ mismatches.

Mismatches containing pyrimidine bases all adopted a wobble conformation, with the nucleotide opposite the APE1 active site (N_O) shifting towards the minor groove and the nucleotide adjacent to the THF (N_A) shifting towards the major groove. The smaller profile of a pyrimidine base, relative to a purine, results in a gap between the WC faces that must be minimized to form stabilizing hydrogen bonds. Wobble base pairing, along with C1′–C1′ compressions of ∼1.5 Å, allow the pyrimidine–pyrimidine mismatches to form the hydrogen-bonding interactions key for stabilization of the base pair. Importantly, the mismatch conformations seen here vary from those seen in the active sites of other enzymes. For example, a G·G mismatch in the active site of DNA polymerase β assumes an *anti·anti* conformation and that in the active site of *Bacillus* fragment DNA polymerase 1 adopts an *anti·syn* conformation, similar to what we see in our APE1 complex structure (Johnson & Beese, 2004; Batra *et al.*, 2016). These observed differences in base-pairing conformations are probably due to the differing electrostatic and steric influences resulting from the structure of the bound protein and highlight the dynamic nature of DNA base-pairing interactions.

Although our high-resolution ΔAPE1–DNA structures provide key insights into mismatched base-pairing properties within the APE1 active site, the precise protonation states of the key atoms remain unknown. Specifically, N3 of cytosine and N1 of adenine are potentially protonated under physiological conditions and in our $A_O{\cdot}C_A$, $A_O{\cdot}G_A$, $T_O{\cdot}C_A$ and C·C mismatch structures. Importantly, adenine N1 and cytosine N3 are within hydrogen-bonding distance of other proton acceptors (G N7, C O2 and T O4) in many of the presented mismatches, further hinting at the possibility of hydrogen bonds between these atoms. Unfortunately, the exact location of H atoms, and therefore hydrogen bonds, cannot easily be determined using X-ray crystallography. The implementation of neutron crystallography would allow the accurate modeling of H atoms, and consequently elucidation of the unknown protonation states (Chaudhuri, 2015; Ho *et al.*, 2004; Moon *et al.*, 2016). Through the presentation of this study at the International Symposium on Diffraction Structural Biology 2016, we intended to shine light on the specific challenges associated with elucidating base-pairing properties, and assigning hydrogen-bonding interactions in general, based on the implied protonation states and X-ray crystallographic data alone.

Acknowledgements

We would like to thank the organizers of the Fifth International Symposium on Diffraction Structural Biology for the opportunity to present our data and receive critical feedback. We also thank Dr Matthew Cuneo (Oak Ridge National Laboratory, Knoxville, Tennessee) for critical feedback on the project and the development of subsequent neutron crystallography projects.

Funding information

The following funding is acknowledged: National Institutes of Environmental Health Sciences of the National Institutes of Health (grant No. R00ES024431 to Bret Freudenthal).

References

Adams, P. D. *et al.* (2010). *Acta Cryst.* D**66**, 213–221.
Batra, V. K., Beard, W. A., Pedersen, L. C. & Wilson, S. H. (2016). *Structure*, **24**, 1863–1875.
Batra, V. K., Beard, W. A., Shock, D. D., Pedersen, L. C. & Wilson, S. H. (2008). *Mol. Cell*, **30**, 315–324.
Bebenek, K., Pedersen, L. C. & Kunkel, T. A. (2011). *Proc. Natl Acad. Sci. USA*, **108**, 1862–1867.
Bink, H. H., Hellendoorn, K., van der Meulen, J. & Pleij, C. W. (2002). *Proc. Natl Acad. Sci. USA*, **99**, 13465–13470.
Boulard, Y., Cognet, J. A., Gabarro-Arpa, J., Le Bret, M., Sowers, L. C. & Fazakerley, G. V. (1992). *Nucleic Acids Res.* **20**, 1933–1941.
Carbonnaux, C., van der Marel, G. A., van Boom, J. H., Guschlbauer, W. & Fazakerley, G. V. (1991). *Biochemistry*, **30**, 5449–5458.
Chatterjee, N. & Walker, G. C. (2017). *Environ. Mol. Mutagen.* **58**, 235–263.
Chaudhuri, B. N. (2015). *Protein Sci.* **24**, 267–276.
Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Cryst.* D**66**, 486–501.
Freudenthal, B. D., Beard, W. A., Cuneo, M. J., Dyrkheeva, N. S. & Wilson, S. H. (2015). *Nature Struct. Mol. Biol.* **22**, 924–931.
Freudenthal, B. D., Beard, W. A. & Wilson, S. H. (2015). *DNA Repair (Amst.)*, **32**, 3–9.
He, H., Chen, Q. & Georgiadis, M. M. (2014). *Biochemistry*, **53**, 6520–6529.
Ho, D. L., Byrnes, W. M., Ma, W.-P., Shi, Y., Callaway, D. J. E. & Bu, Z. (2004). *J. Biol. Chem.* **279**, 39146–39154.
Hsieh, P. & Yamane, K. (2008). *Mech. Ageing Dev.* **129**, 391–407.
Johnson, S. J. & Beese, L. S. (2004). *Cell*, **116**, 803–816.
Li, M. & Wilson, D. M. III (2014). *Antioxid. Redox Signal.* **20**, 678–707.
Lindahl, T. (1993). *Nature (London)*, **362**, 709–715.
Lynch, H. T., Shaw, M. W., Magnuson, C. W., Larsen, A. L. & Krush, A. J. (1966). *Arch. Intern. Med.* **117**, 206–212.
McNeill, D. R. & Wilson, D. M. III (2007). *Mol. Cancer Res.* **5**, 61–70.
Minor, W., Cymborowski, M., Otwinowski, Z. & Chruszcz, M. (2006). *Acta Cryst.* D**62**, 859–866.
Mol, C. D., Izumi, T., Mitra, S. & Tainer, J. A. (2000). *Nature (London)*, **403**, 451–456.
Moody, E. M., Brown, T. S. & Bevilacqua, P. C. (2004). *J. Am. Chem. Soc.* **126**, 10200–10201.
Moon, A. F., Krahn, J. M., Lu, X., Cuneo, M. J. & Pedersen, L. C. (2016). *Nucleic Acids Res.* **44**, 3946–3957.
Mundle, S. T., Delaney, J. C., Essigmann, J. M. & Strauss, P. R. (2009). *Biochemistry*, **48**, 19–26.
Nikolova, E. N., Goh, G. B., Brooks, C. L. III & Al-Hashimi, H. M. (2013). *J. Am. Chem. Soc.* **135**, 6766–6769.
Rossetti, G., Dans, P. D., Gomez-Pinto, I., Ivani, I., Gonzalez, C. & Orozco, M. (2015). *Nucleic Acids Res.* **43**, 4309–4321.
Sassa, A., Beard, W. A., Prasad, R. & Wilson, S. H. (2012). *J. Biol. Chem.* **287**, 36702–36710.

Schermerhorn, K. M. & Delaney, S. (2013). *Biochemistry*, **52**, 7669–7677.
Siegfried, N. A., O'Hare, B. & Bevilacqua, P. C. (2010). *Biochemistry*, **49**, 3225–3236.
Vaisman, A., Ling, H., Woodgate, R. & Yang, W. (2005). *EMBO J.* **24**, 2957–2967.
Washington, M. T., Carlson, K. D., Freudenthal, B. D. & Pryor, J. M. (2010). *Biochim. Biophys. Acta*, **1804**, 1113–1123.
Whitaker, A. M., Schaich, M. A., Smith, M. R., Flynn, T. S. & Freudenthal, B. D. (2017). *Front. Biosci. (Landmark Ed.)*, **22**, 1493–1522.
Whitaker, A. M., Smith, M. R., Schaich, M. A. & Freudenthal, B. D. (2017). *Nucleic Acids Res.* **45**, 6934–6944.
Wilson, D. M. III, Takeshita, M., Grollman, A. P. & Demple, B. (1995). *J. Biol. Chem.* **270**, 16002–16007.