



**HAL**  
open science

## Abasic site–peptide cross-links are blocking lesions repaired by AP endonucleases

Anna V Yudkina, Nikita A Bulgakov, Daria V Kim, Svetlana V Baranova, Alexander A Ishchenko, Murat K Saparbaev, Vladimir V Koval, Dmitry O Zharkov

► **To cite this version:**

Anna V Yudkina, Nikita A Bulgakov, Daria V Kim, Svetlana V Baranova, Alexander A Ishchenko, et al.. Abasic site–peptide cross-links are blocking lesions repaired by AP endonucleases. *Nucleic Acids Research*, 2023, 51 (12), pp.6321-6336. 10.1093/nar/gkad423 . hal-04289451

**HAL Id: hal-04289451**

**<https://hal.science/hal-04289451>**

Submitted on 16 Nov 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License



**Some supplementary files may need to be viewed online via your Referee Centre at <http://mc.manuscriptcentral.com/nar>. If the figures are small, you can view the original files in your Referee Centre.**

**Abasic site-peptide cross-links are blocking lesions repaired by AP endonucleases**

Journal:	<i>Nucleic Acids Research</i>
Manuscript ID	NAR-03976-Q-2022.R1
Manuscript Type:	1 Standard Manuscript
Key Words:	DNA-peptide cross-links, AP sites, DNA polymerases, AP endonucleases, base excision repair

SCHOLARONE™  
Manuscripts

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## DATA AVAILABILITY

Does the manuscript use or report the following? If so, please provide details in a Data Availability statement below and in the manuscript.	
<p><b>New genome expression or sequencing data (ChIP-seq, RNA-seq...)</b></p> <ul style="list-style-type: none"> <li>- Must comply with <a href="#">ENCODE</a> Guidelines.</li> <li>- All datasets must be validated via biological replicates.</li> <li>- Must deposit data in <a href="#">GEO</a> or an equivalent publicly available depository and provide accession numbers, private tokens, reviewer login details and/or private URLs for Referees.</li> <li>- Excluding RNA-Seq, data must be viewable on the <a href="#">UCSC</a> (eukaryotes) or other suitable genome browsers; must provide genome browser session links (even if GEO entries are publicly available).</li> </ul>	<b>No</b>
<p><b>Novel nucleic acid sequences</b></p> <ul style="list-style-type: none"> <li>- Must deposit in <a href="#">EMBL</a> / <a href="#">GenBank</a> / <a href="#">DDBJ</a>.</li> <li>- Must provide sequence names and accession numbers.</li> </ul>	<b>No</b>
<p><b>Illumina-type sequencing data</b></p> <ul style="list-style-type: none"> <li>- Must submit data to <a href="#">BioProject/SRA</a>, <a href="#">ArrayExpress</a> or <a href="#">GEO</a>.</li> <li>- Must provide link for reviewers (BioProject/SRA), login details (ArrayExpress) or accession numbers and private tokens (GEO).</li> </ul>	<b>No</b>
<p><b>Novel protein sequences</b></p> <ul style="list-style-type: none"> <li>- Must deposit <a href="#">UniProt</a> using the interactive tool SPIN.</li> <li>- Must provide sequence names and accession number.</li> </ul>	<b>No</b>
<p><b>Novel molecular structures</b> determined by X-ray crystallography, NMR and/or CryoEM/EM</p> <ul style="list-style-type: none"> <li>- Must deposit to a member site of the Worldwide Protein Data Bank (RCSB <a href="#">PDB</a>, <a href="#">PDBe</a>, <a href="#">PDBj</a>) and provide the accession numbers.</li> <li>- If structures are unreleased (i.e. status HPUB), <b>MUST</b> upload: <ul style="list-style-type: none"> <li>- the validation reports (.pdf)</li> <li>- molecular coordinates (.pdb or .mmcif).</li> <li>- one of the following: <ul style="list-style-type: none"> <li>• X-ray data (.mtz, .cif)</li> <li>• NMR restraints and chemical shift files (.mr, .tbl or .str)</li> <li>• CryoEM map files (.map).</li> </ul> </li> </ul> </li> </ul>	<b>No</b>
<p><b>Novel molecular models</b> based on SAXS, computational modeling, or other combinations of strategies that are generally not appropriate for deposition in the PDB</p> <ul style="list-style-type: none"> <li>- Must deposit coordinates and all underlying data in appropriate databases (including but not limited to the <a href="#">Small Angle Scattering Database</a> and <a href="#">PDB-Dev</a>).</li> <li>- Must report on validation of the structure against experimental data (if available) or report on statistical validation of the structure by model quality assessment programs. If applicable, these should be uploaded as a Data file.</li> </ul>	<b>No</b>
<p><b>Molecular behaviour studies</b> derived from biological NMR spectroscopy data (not necessarily leading to new structures)</p>	<b>No</b>

- Must deposit NMR spectral data, including assigned chemical shifts, coupling constants, relaxation parameters (T1, T2, and NOE values), dipolar couplings, in <a href="#">BMRB</a> .	
<b>Novel nucleic acids structure</b> - Must deposit to <a href="#">NDB</a> (via <a href="#">PDB</a> if possible) and provide accession numbers.	<b>No</b>
<b>Structures of nucleosides, nucleotides, other small molecules</b> - Must deposit in the Cambridge Crystallographic Data Centre ( <a href="#">CCDC</a> ) and provide the structure identifiers.	<b>No</b>
<b>Mass spectrometry proteomics</b> - Must deposit to <a href="#">ProteomeXchange</a> consortium and provide Dataset Identifier and reviewer account details. If appropriate, data and corresponding details can also be deposited in the <a href="#">Panorama</a> repository for targeted mass spec assays and workflows.	<b>No</b>
<b>Microarray data</b> - Must comply with the <a href="#">MIAME</a> Guidelines - Must deposit the data to <a href="#">GEO</a> or <a href="#">Array Express</a> , and provide accession numbers and private tokens (GEO) or login details (ArrayExpress).	<b>No</b>
<b>Quantitative PCR</b> - Must comply with the <a href="#">MIQE</a> Guidelines. - Details should be supplied in Materials and Methods section of manuscript.	<b>No</b>
<b>Synthetic nucleic acid oligonucleotides including siRNAs or shRNAs</b> - The manuscript should include controls to rule out off-target effects, such as use of multiple siRNA/shRNAs or inclusion of cDNA rescue data. - Manuscript should provide exact sequences, exact details of chemical modifications at any position, and source of reagent or precise methods for creation. These can be included in the main text or in Supplementary Material.	<b>Yes</b>
<b>Software and source codes</b> - Must deposit in <a href="#">FigShare</a> and provide link to code and/or DOI or upload source code as Data file.	<b>No</b>
<b>Gel images, micrographs, graphs, and tables</b> - Optionally, may deposit in a general-purpose repository such as <a href="#">Zenodo</a> or <a href="#">Dryad</a> . If applicable, provide access details.	<b>Yes</b>

**REFEREES – you will find data deposition details below**

All data are provided in the paper and the supplementary materials.

## KEY POINTS

(3 bullet points summarizing the manuscript's contribution to the field)

“Abasic (AP) site” peptide cross-links arise in the proteolytic repair of AP site “protein cross-links

1  
2  
3  
4  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

â€ AP siteâ€ peptide cross-links are blocking and follow the A-rule or induce primer misalignment

â€ AP siteâ€ peptide cross-links can be repaired by the TIM barrel superfamily AP endonucleases

## Abasic site–peptide cross-links are blocking lesions repaired by AP endonucleases

Anna V. Yudkina<sup>1,2,\*</sup>, Nikita A. Bulgakov<sup>1</sup>, Daria V. Kim<sup>1,2</sup>, Svetlana V. Baranova<sup>1</sup>, Alexander A. Ishchenko<sup>3</sup>, Murat K. Saparbaev<sup>3</sup>, Vladimir V. Koval<sup>1</sup> and Dmitry O. Zharkov<sup>1,2,\*</sup>

<sup>1</sup> SB RAS Institute of Chemical Biology and Fundamental Medicine, Novosibirsk, 630090, Russia

<sup>2</sup> Department of Natural Sciences, Novosibirsk State University, Novosibirsk, 630090, Russia

<sup>3</sup> Groupe “Mechanisms of DNA Repair and Carcinogenesis”, Equipe Labellisée LIGUE 2016, CNRS UMR9019, Université Paris-Saclay, Gustave Roussy Cancer Campus, F-94805 Villejuif, France

\* To whom correspondence should be addressed. Tel: +7 383 363 5187; Fax: +7 383 363 5128; Email: dzharkov@niboch.nsc.ru. Correspondence may also be addressed to Anna Yudkina. Tel: +7 383 363 5188; Fax: +7 383 363 5128; Email: ayudkina@niboch.nsc.ru.

### KEY POINTS

- Abasic (AP) site–peptide cross-links arise in the proteolytic repair of AP site–protein cross-links
- AP site–peptide cross-links are blocking and follow the A-rule or induce primer misalignment
- AP site–peptide cross-links can be repaired by the TIM barrel superfamily AP endonucleases

### ABSTRACT

Apurinic/aprimidinic (AP) sites are abundant DNA lesions arising from spontaneous hydrolysis of the N-glycosidic bond and as base excision repair (BER) intermediates. AP sites and their derivatives readily trap DNA-bound proteins, resulting in DNA–protein cross-links. Those are subject to proteolysis but the fate of the resulting AP–peptide cross-links (APPXLs) is unclear. Here we report two *in vitro* models of APPXLs synthesized by cross-linking of DNA glycosylases Fpg and OGG1 to DNA followed by trypsinolysis. The reaction with Fpg produces a 10-mer peptide cross-linked through its N-terminus, while OGG1 yields a 23-mer peptide attached through an internal lysine. Both adducts strongly blocked Klenow fragment, phage RB69 polymerase, *Saccharolobus solfataricus* Dpo4, and African swine fever virus PolX. In the residual lesion bypass, mostly dAMP and dGMP were incorporated by Klenow and RB69 polymerases, while Dpo4 and PolX used primer/template misalignment. Of AP endonucleases involved in BER, *E. coli* endonuclease IV and its yeast homolog Apn1p efficiently hydrolyzed both adducts. In contrast, *E. coli* exonuclease III and human APE1 showed little activity on APPXL substrates. Our data suggest that APPXLs produced by proteolysis of AP site-trapped proteins may be removed by the BER pathway, at least in bacterial and yeast cells.

## INTRODUCTION

Apurinic/aprimidinic (AP) sites occur by hydrolysis of the *N*-glycosidic bond in deoxyribonucleotides and are among the most abundant spontaneous and induced DNA lesions in the cell (1,2). They are also formed as intermediates in the base excision repair pathway, often being more genotoxic and mutagenic than the original lesion and thus requiring tight control over their processing (3,4).

Chemically, AP sites are a family of lesions, which includes the much-studied aldehydic AP site as well as its derivatives oxidized at C1' (deoxyribolactone, DRL) (5-7), C2' (C2-AP) (8), C4' (C4-AP) or C5' position (5'-(2-phosphoryl-1,4-dioxobutane), DOB) (9,10).

Normally, AP sites are quickly removed from the genome by the base excision DNA repair (BER) system. However, there are important differences in the processing of different types of AP sites. The repair of aldehydic AP sites follows the classical short-patch BER branch, which, in human cells, involves the AP endonuclease (APE1) followed by polymerase and deoxyribophosphate lyase (dRPase) activities of DNA polymerase  $\beta$  (POL $\beta$ ) (11). At the same time, C2-AP is not a substrate for the dRPase activity of POL $\beta$  and is removed through the long-patch branch of BER: the DNA flap containing the lesion at the 5'-end is displaced during the repair DNA synthesis and hydrolyzed by FEN1 nuclease (8,12). DRL repair is even more problematic: this lesion forms DNA-protein cross-links (DPXLs) with Lys residues in the active sites of DNA glycosylase Nth and POL $\beta$  (13,14), and the  $\beta$ -elimination product of DRL can similarly trap DNA glycosylases Fpg and NEIL1 (14). The presence of a 1,4-dicarbonyl moiety in C4-AP and DOB also promotes covalent capture of POL $\beta$  and DNA polymerase  $\lambda$  (15,16). The irreversible inhibition of the main DNA repair enzymes leads to significant cytotoxicity of modified AP sites (7).

Both aldehydic and oxidized AP sites readily form DPXLs. The covalent capture of many DNA-binding proteins by AP sites has been reported, including histones (6,17,18), integrase (19), DNA ligase (20), Ku antigen (21), as well as RNA-binding proteins (22,23), chaperones (22) and enzymes of carbohydrate metabolism (22,24). Recently, a unique mechanism for the repair of AP sites in single-stranded DNA was discovered (25-27), based on the formation of cross-links between the AP site and the HMCES protein (in human cells) or YedK protein (in *E. coli*), followed by switching to the proteasome-dependent pathway of DPXL repair. Thus, DPXL with AP sites are highly relevant adducts *in vivo*.

DPXLs are not limited to cross-links with AP sites and appear through a multitude of mechanisms including reactions with aldehydes and heavy metals, UV and ionizing radiation, and failed action of enzymes such as topoisomerases or C5-methyltransferases (28,29). Thus, a special pathway dedicated to DPXL repair has evolved. It is tightly coupled to replication and involves proteolytic degradation of the protein part of the cross-link (30-33). However, at the moment it remains poorly understood how the last peptide remnant conjugated with DNA is finally removed. It was suggested that this role can be played by nucleotide excision repair (NER), which normally protects the genome from bulky adducts (34-38). Also, Fpg/Nei-like DNA glycosylases are able to excise oxidized purine residues cross-linked to peptides (38). Recombination repair does not appear to be involved in the removal of peptide conjugates from DNA, although it may play a role in tolerance to larger DPXLs (31,36). It should be noted that all these data were obtained on cross-links chemically different from

1  
2  
3 those formed by AP sites, and peptide adducts of this type remain uncharacterized with respect to  
4 their repair and their effect on replication. While the studied peptide adducts with nucleobases usually  
5 retain some base-pairing ability, the bulky nature and non-instructing properties of AP site-peptide  
6 cross-links make these lesions more problematic for the cell. Here we report a model of stable,  
7 chemically determined AP site-peptide cross-links (APPXLs), conjugated through either a terminal or  
8 an internal position in the peptide, and show that they block DNA polymerases and can be repaired by  
9 AP endonucleases.  
10  
11  
12

## 13 14 **MATERIAL AND METHODS**

### 15 16 **Oligonucleotides and enzymes**

17  
18 *E. coli* formamidopyrimidine-DNA glycosylase (Fpg) (39), endonuclease VIII (Nei) (40) and  
19 endonuclease IV (Nfo) (41), human AP endonuclease (APE1) (42), 8-oxoguanine-DNA glycosylase  
20 (OGG1) (43), endonuclease VIII-like protein 1 (NEIL1) (44) and endonuclease VIII-like protein 2  
21 (NEIL2) (45), yeast AP endonuclease (Apn1p) (41), 3'→5'-exonuclease-deficient (*exo*<sup>-</sup>) Klenow  
22 fragment of *E. coli* DNA polymerase I (KF) (46) and *exo*<sup>-</sup> bacteriophage RB69 DNA polymerase  
23 (RBpol) (47) were overexpressed and purified essentially as described. Cloning and purification of  
24 DNA polymerase X from African swine fever virus (ASFV PolX) is described below. *E. coli*  
25 exonuclease III (Xth) was purchased from Roche (Basel, Switzerland), *E. coli* uracil-DNA glycosylase  
26 (Ung) and exonuclease-proficient KF, from SibEnzyme (Novosibirsk, Russia), full-length *E. coli* DNA  
27 polymerase I (Pol I) and *Saccharolobus (Sulfolobus) solfataricus* DNA polymerase IV (Dpo4), from  
28 New England Biolabs (Ipswich, MA) and trypsin, from Samson-Med (St. Petersburg, Russia). Purity of  
29 the DNA polymerases and AP endonucleases, except for those obtained commercially, is illustrated in  
30 Supplementary Figure S1. Oligonucleotides (Supplementary Table S1) were synthesized in-house  
31 from commercially available phosphoramidites (Glen Research, Sterling, VA).  
32  
33  
34  
35  
36  
37  
38

### 39 40 **ASFV PolX cloning and purification**

41  
42 The coding frame of PolX optimized for expression in *E. coli* was synthesized by Gene Universal  
43 (Newark, DE, USA) and confirmed by Sanger sequencing. The insert was subcloned into the pET-24b  
44 vector (Merck, Darmstadt, Germany) at NdeI/XhoI sites. The plasmid was subsequently introduced  
45 into the *E. coli* Rossetta 2(DE3) strain (Merck). One liter of LB medium supplemented with 100 µg/ml  
46 of kanamycin was inoculated with 5 ml of an overnight culture. The cells were grown with vigorous  
47 shaking at 37°C to A<sub>600</sub> = 0.6, isopropyl-β-D-thiogalactopyranoside was added to 1 mM, and the  
48 growth was continued for 4 h at 37°C. The cells were harvested by centrifugation at 12,000×g at 4°C  
49 for 20 min and stored at -72°C. Before the purification, the pellet was thawed on ice in 40 ml of  
50 Buffer A consisting of 20 mM sodium phosphate (pH 7.5), 1 mM ethylenediaminetetraacetic acid  
51 (EDTA), 1 mM dithiothreitol (DTT) and supplemented with 500 mM NaCl and 1 mM  
52 phenylmethylsulfonyl fluoride. The cells were sonicated, and the lysate was clarified by centrifugation  
53 at 12,000×g at 4°C for 30 min. The supernatant was filtered through a 0.45-µm polyvinylidene  
54 difluoride membrane, diluted with four volumes of Buffer A and loaded onto a 5-ml SP Sepharose  
55  
56  
57  
58  
59  
60

1  
2  
3 HiTrap column (GE Healthcare, Chicago, IL) equilibrated in Buffer A plus 100 mM NaCl. The column  
4 was washed with Buffer A supplemented with 100 mM NaCl, and PolX was eluted with a NaCl  
5 gradient at ~650 mM. The fractions containing PolX were pooled, diluted with six volumes of Buffer A  
6 and loaded onto a 5-ml heparine Sepharose HiTrap column (GE Healthcare) equilibrated in Buffer A  
7 plus 100 mM NaCl. The column was washed with the same buffer, and PolX was eluted with a NaCl  
8 gradient at ~730 mM. The fractions containing PolX were pooled, diluted with six volumes of Buffer A  
9 and additionally purified on a 1-ml MonoS column (GE Healthcare) previously equilibrated in Buffer A  
10 plus 100 mM NaCl. PolX was eluted with a NaCl gradient at ~600 mM. The fractions containing >90%  
11 homogeneous protein were pooled, dialyzed against the buffer containing 20 mM sodium phosphate  
12 (pH 7.5), 400 mM NaCl, 1 mM EDTA, 1 mM DTT, and 50% glycerol, and stored at -20°C.  
13  
14  
15  
16  
17

### 18 **Synthesis of model peptide conjugates with AP sites**

19  
20  
21 DNA cross-links to Fpg or OGG1 proteins were obtained as described (48-50). Briefly, ~2 nmol of the  
22 oligonucleotide duplex (40oG//28down or 28oG//21comp; Supplementary Table S1) was treated with  
23 3–6 nmol of the enzyme in 50 mM sodium phosphate (pH 6.8), 1 mM EDTA, 1 mM DTT and 100 mM  
24 NaBH<sub>4</sub> in a total volume of 40 µl. The reaction was allowed to proceed at 37°C for 2 h. The resulting  
25 DPXL was denatured at 95°C for 10 min, cooled, and pH of the solution was adjusted to ~8.0 by  
26 adding Tris base. The mixture was supplemented with 2 µg of trypsin and incubated for 4 h at 37°C.  
27 The oligonucleotide–peptide conjugate was separated by electrophoresis in 20% polyacrylamide/7 M  
28 urea, excised under UV illumination, desalted by reverse-phase chromatography on Isolute C18  
29 sorbent (Biotage, Uppsala, Sweden) and characterized by MALDI MS. If necessary, the  
30 oligonucleotides for the following experiments were 5'-labeled using γ[<sup>32</sup>P]ATP and annealed to a 1.5-  
31 fold molar excess of the complementary strand.  
32  
33  
34  
35  
36

### 37 **DNA polymerase assays**

38  
39 For the standing-start assay, the substrates were assembled from the 40-mer template strand  
40 containing an APPXL and the <sup>32</sup>P-labeled primer (13pri; Supplementary Table S1) or the 28-mer  
41 template strand containing an APPXL and the <sup>32</sup>P-labeled primer (16pri; Supplementary Table S1).  
42 The reaction mixture (10 µl) contained 20 nM pre-annealed substrate, DNA polymerase (100 nM KF  
43 or RBpol, 400 nM PolX, 0.1 U/µl Dpo4), dNTPs (either 200 µM individual dATP, dCTP, dGTP, or  
44 dTTP, or a mixture of all four dNTPs each at 200 µM) in the reaction buffer (50 mM Tris–HCl (pH 7.5),  
45 5 mM MgCl<sub>2</sub> and 1 mM DTT for KF, RBpol and PolX; 20 mM Tris–HCl (pH 8.8), 10 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>,  
46 10 mM KCl, 2 mM MgSO<sub>4</sub>, 0.1% Triton X-100 for Dpo4). The reaction was allowed to proceed at 37°C  
47 for 30 min, quenched with an equal volume of the loading solution (80% formamide, 20 mM Na-  
48 EDTA, 0.1% xylene cyanol, 0.1% bromophenol blue), and heated for 1 min at 95°C. The reaction  
49 products were resolved by 20% denaturing PAGE and visualized on a Typhoon FLA 9500 imager (GE  
50 Healthcare). The bands were quantified using Quantity One v4.6.8 (Bio-Rad Laboratories, Hercules,  
51 CA). For the running-start assay, the substrates were assembled from the 40-mer template strand  
52 containing an APPXL and the <sup>32</sup>P-labeled primer (11pri; Supplementary Table S1). The substrates  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 were taken at 50 nM, and the enzymes, at 50 nM (KF or RBpol), 200 nM (PolX) or 0.1 U/μl (Dpo4);  
4 aliquots were taken at 2, 5, and 30 min; otherwise the reactions were run and processed as above.  
5  
6

### 7 **AP endonuclease assay**

8  
9 The substrates were assembled from the <sup>32</sup>P-labeled 40-mer strand containing an APPXL and the  
10 complementary 40-mer strand (40compC; Supplementary Table S1). The reaction mixture (10 μl)  
11 contained 50 nM substrate prepared as above and 10–1000 nM Xth, APE1, Apn1p or Nfo in the  
12 respective buffer optimal for BER or NIR (Supplementary Table S2). As a control, instead of the  
13 APPXL, we used the same oligonucleotide construct with an aldehydic AP site, freshly prepared by  
14 treatment of the U-containing substrate with Ung at 37°C for 20 min. The endonuclease reaction was  
15 allowed to proceed at 37°C for 30 min, treated and analyzed as above.  
16  
17  
18  
19

### 20 **AP endonuclease kinetics**

21  
22 The reaction mixture (10 μl) contained increasing concentrations of the <sup>32</sup>P-labeled 40-mer APPXL  
23 substrates or control substrates 23OHU:23comp or 20THF:20comp (1–500 nM) in the BER or NIR  
24 buffer. Concentrations of AP endonucleases were optimized for each substrate and buffer to produce  
25 ≤20% cleavage and were in the 5 pM – 0.5 nM range. The reaction was allowed to proceed for 10 min  
26 at 37°C, treated and analyzed as above. The data were fitted by nonlinear regression to the  
27 Michaelis–Menten equation using Sigmaplot v11.0 (Systat Software, Chicago, IL). The reported  
28 values are averages of 3–5 independent experiments.  
29  
30  
31  
32

### 33 **AP endonuclease / DNA polymerase assay**

34  
35 The reaction mixture (10 μl) contained 200 nM <sup>32</sup>P-labeled duplex 40-mer APPXL substrate and 50–  
36 150 nM Nfo in the BER buffer (Supplementary Table S2). After 10 min, the reaction mixture was  
37 supplemented with dNTPs (200 μM each), 5 mM MgCl<sub>2</sub>, and 50–250 nM KF. The reaction was  
38 allowed to proceed at 37°C for 2, 5, or 30 min, treated and analyzed as above.  
39  
40  
41

### 42 **DNA glycosylase assay**

43  
44 The reaction mixture (10 μl) contained 10 nM labeled duplex substrate (see above), 1 μM DNA  
45 glycosylase (*E. coli* Fpg or Nei, human OGG1, NEIL1, or NEIL2), 50 mM Tris–HCl (pH 7.5), 100 mM  
46 NaCl, 1 mM EDTA and 1 mM DTT. The reaction was allowed to proceed at 37°C for 30 min, treated  
47 and analyzed as above.  
48  
49  
50

## 51 **RESULTS**

### 52 **Synthesis and characterization of DNA–peptide cross-links with AP site**

53  
54 Since AP sites can form cross-links to a variety of proteins, which can be further processed to  
55 APPXLs, it is usually assumed that the repair of such adducts does not depend on the exact nature of  
56 the cross-linked peptide. Hence, to obtain model APPXLs, we made use of two enzymes, *E. coli* Fpg  
57 and human OGG1 DNA glycosylases, that form transient covalent intermediates with AP sites during  
58  
59  
60

1  
2  
3 the catalysis. To excise the damaged base (8-oxoguanine; 8-oxoG) from DNA, Fpg and OGG1 carry  
4 out a nucleophilic attack by an amino group at the C1' atom of the target nucleotide forming a labile  
5 Schiff base, which may be nearly quantitatively reduced by NaBH<sub>4</sub> or NaBH<sub>3</sub>CN to a stable conjugate  
6 (Figure 1). This reaction was successfully used to crystallize covalent Fpg–DNA and OGG1–DNA  
7 complexes (39,51) and to study the interaction of full-length DPXLs with the replication and repair  
8 machinery (36,48-50). Thus, we first prepared DPXLs by treatment of 8-oxoG-containing duplexes  
9 with Fpg or OGG1 in the presence of NaBH<sub>4</sub> and subjected them to trypsinolysis (Supplementary  
10 Figures S2A, S2B). Fpg and OGG1 belong to different structural superfamilies and use different  
11 nucleophilic residues for catalysis: N-terminal Pro1 in Fpg and Lys249 in OGG1. Hence, after  
12 trypsinolysis, the resulting peptide conjugates are expected to have different structures (Figure 1).  
13 The Fpg–DNA conjugate gives an adduct with the decapeptide (P)ELPEVETSR (the residue that  
14 forms the cross-link is in parentheses), in which the peptide forms a bond with DNA through its N-  
15 terminus. On the other hand, trypsinolysis of the OGG1 conjugate should give an adduct with the 23-  
16 mer peptide ALCILPGVGT(K)VADCICLMALDK, cross-linked through an internal position of the  
17 peptide. Based on the visual resemblance (Figure 1), we will further call these AP site–peptide cross-  
18 links APPXL-I (terminally adducted, linear peptide) and APPXL-Y (an adduct through an internal  
19 residue with two peptide branches).

20  
21 APPXL-I and APPXL-Y were purified by denaturing polyacrylamide gel electrophoresis followed by  
22 reverse-phase chromatography and characterized by MALDI-TOF mass spectrometry  
23 (Supplementary Figures S2C, S2D). The yield of the homogeneous stable APPXLs was ~5–15% of  
24 the initial 8-oxoG-containing oligonucleotide. In addition to the cross-link with the full-length modified  
25 strand, we have observed some products that likely corresponded to products of strand cleavage by  
26 Fpg/OGG1, which were easily separated from the major adduct by electrophoresis (Supplementary  
27 Figures S2A, S2B). Despite the long oligonucleotide parts, of the cross-links, their masses were better  
28 resolved by MALDI mass spectrometry in a positive mode and were within a few Da of the expected  
29 values. No peaks of the starting material (expected [M+H<sup>+</sup>], 12138.9 Da) was evident, indicating good  
30 separation of the APPXLs by electrophoresis. In the APPXL-I mass spectrum, we observed two peaks  
31 corresponding to the expected molecular weight of the adduct of the 40-mer oligonucleotide with the  
32 (P)ELPEVETSR peptide (observed [M+H<sup>+</sup>], 13131.0 Da, expected, 13127.9 Da) and  
33 (P)ELPEVETSR peptide (observed [M+H<sup>+</sup>], 13285.8 Da, expected, 13284.0 Da), most likely  
34 resulting from incomplete trypsin cleavage at two adjacent Arg residues (Supplementary Figure S2C).  
35 As heterogeneity of cross-linked peptides may be naturally expected from proteolysis of DPXLs in the  
36 cell, we have proceeded to characterize the biochemical properties of our adducts.

### 51 **AP site–peptide cross-links block DNA polymerases**

52  
53 To investigate the ability of DNA polymerases to bypass APPXLs, we annealed the cross-linked  
54 template oligonucleotide with a labeled primer with its 3'-terminus placed 12 bases upstream of the  
55 cross-link site ("running start" conditions, see the substrate schemes in Figures 2A, 2B, 3A). This  
56 distance exceeds the footprints of the full-length proteins from which the cross-linked peptides were  
57 derived, and thus the peptide was not expected to prevent DNA polymerase binding. We also tested  
58  
59  
60

1  
2  
3 the ability of DNA polymerases to carry out synthesis through the APPXL in the context of DNA  
4 duplex when the polymerase has to displace the downstream strand (Figures 2C, 2D, 3B). As  
5 APPXLs are non-instructive, both APPXL-I and APPXL-Y are expected to be strongly blocking.  
6 Therefore, we used a natural freshly prepared AP site as a control to determine which events would  
7 result from DNA polymerases blocking due to the lack of a base or due to the nature of cross-linked  
8 peptide. Non-damaged primer–template or primer–downstream strand–template constructs were  
9 used as polymerase activity controls.  
10  
11  
12

13 Four DNA polymerases, representing a range of structural and functional features, were  
14 investigated. As we were interested in the general ability of DNA polymerases to bypass APPXLs, we  
15 opted to use easily accessible representative members of various structural families originating from  
16 bacteria, archaea, and viruses, rather than human enzymes. The Klenow fragment of *E. coli* DNA  
17 polymerase I (KF) belongs to Family A of DNA polymerases, acting in DNA repair and replication. The  
18 homologous human enzymes are DNA polymerases  $\gamma$ ,  $\theta$ , and  $\nu$ . Bacteriophage RB69 DNA  
19 polymerase (RBpol) is a high-fidelity replicative polymerase from Family B and is often taken as a  
20 convenient model of structurally related human replicative DNA polymerases  $\alpha$ ,  $\delta$ , and  $\epsilon$ , and a  
21 translesion DNA polymerase  $\zeta$ . In both cases, we have used 3'→5'-exonuclease-deficient versions of  
22 the polymerases to prevent primer degradation; exonuclease-proficient KF was also brought in for  
23 comparison. DNA polymerase X (PolX) from African swine fever virus (ASFV) belongs to structural  
24 Family X, mostly engaged in DNA repair and encompassing human DNA polymerases  $\beta$ ,  $\lambda$ , and  $\mu$ .  
25 Finally, DNA polymerase IV from *Saccharolobus* (formerly *Sulfolobus*) *solfataricus* (Dpo4) is a  
26 representative of Family Y translesion polymerases, of which human cells possess DNA  
27 polymerases  $\eta$ ,  $\iota$ , and  $\kappa$ .  
28  
29  
30  
31  
32  
33  
34

35 Under the running start conditions, the blocking properties of APPXL-I were comparable with the  
36 properties of the natural AP site: KF demonstrated slightly better bypass of APPXL-I compared to AP  
37 site in comparison with RBpol, which was slightly more efficient on AP site-containing substrates.  
38 After 30 min, KF was able to fully elongate 6.8% of the primer if APPXL-I was in the template, and  
39 only 4.6% if an AP site was in the template whereas RB69 demonstrated 6.6% bypass of APPXL-I  
40 and 11% of the AP site (Figure 2A). Moreover, two strong pause points were evident one nucleotide  
41 before the cross-linking site and directly opposite the cross-link (23- and 24-nt long products; marked  
42 by grey arrows in Figures 2 and 3), indicating low ability of DNA polymerases both to incorporate a  
43 dNMP opposite APPXL-I and to extend the primer after such incorporation. The absence of any  
44 additional pause points suggests that the bulky peptide does not hamper moving DNA polymerase  
45 access to the cross-link site. The aldehyde AP site produced a single pause point one nucleotide  
46 before the lesion. Exonuclease-proficient KF demonstrated the same behavior as KF  $\text{exo}^-$  except for  
47 noticeable primer degradation at longer reaction times (Supplementary Figure S3). Thus, APPXL-I,  
48 while strongly blocking, still allows better incorporation of a dNMP than the smaller non-instructive AP  
49 site. PolX, which is known to be nearly fully blocked by AP sites (52), demonstrated no bypass of the  
50 APPXL-I as well as the natural AP site, showing a single pause point corresponding to the 23-mer  
51 synthesis product (Figure 3A). Blocking of PolX may be due to its distributive mode of synthesis,  
52 which facilitates DNA release and complicates polymerase binding near the blocking lesion.  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 Translesion DNA polymerase Dpo4, as expected, demonstrated the strongest ability to bypass the AP  
4 site (Dpo4 was able to fully elongate ~28% of the primer) and APPXL-I (~30% of the bypass),  
5 producing the same two pause points in the latter case (Figure 3A). The presence of a downstream  
6 strand complicated the bypass of the blocking lesion, yet the synthesis pause points and the relative  
7 efficiency of lesion bypass seen with the primer–template system were preserved (Figure 2C,  
8 Figure 3B).  
9

10  
11 Strong polymerase inhibition and two pronounced pause points for KF, RBpol and Dpo4 were also  
12 evident for substrates containing the branched APPXL-Y adduct (Figures 2B, 2D, 3, Supplementary  
13 Figure S3). The bulkier APPXL-Y turned out to be even more blocking than APPXL-I. The pause point  
14 corresponding to the 24-nt product for KF and RBpol was weaker than the corresponding pause for  
15 APPXL-I, and the full-length product decreased significantly for KF and Dpo4.  
16

17  
18 We have also addressed polymerase-blocking ability of AP site–peptide cross-links present in an  
19 internal position of the non-template downstream strand. However, only for PolX was a minor pause  
20 point observed, while KF and RBpol displaced the cross-linked strand without effort, and Dpo4, while  
21 having lower strand displacement activity, did not pause at the cross-link (Supplementary Figure S4).  
22

23  
24 Interestingly, when different substrates were compared, we systematically observed primer  
25 utilization efficiency decreasing in the order AP site > APPXL-I > APPXL-Y. Possibly, the peptide part  
26 may partially interfere with DNA binding by the polymerase even when the primer end is 12 nt away  
27 from the cross-link site, for example, by obstructing non-specific DNA binding and facilitated diffusion  
28 used by polymerases to locate the primer end (53). The only exception was Dpo4 with APPXL-Y in  
29 the downstream strand, in which case the bulky peptide adduct might destabilize the duplex and  
30 improve the poor strand displacement ability of Dpo4.  
31  
32

### 33 34 35 **AP site–peptide cross-links mostly direct A and G incorporation and can induce primer–** 36 **template misalignment** 37

38  
39 Since we observed some dNMP insertion opposite APPXLs, we sought to explore the spectra of the  
40 incorporated nucleotides. To do this, we moved the primer to place its 3'-end immediately next to the  
41 site of the lesion and presented the polymerases with the individual dNMPs as well as with their  
42 mixture ("standing start" conditions, Figure 4). Here, KF utilized the APPXL-I adducted template much  
43 better than the APPXL-Y adduct, whereas for RBpol the efficiency was somewhat better for the  
44 APPXL-Y. Notably, in all cases the primer elongation was significantly lower than under the running  
45 start conditions, most likely because the bulky peptide interferes with DNA polymerases' binding to  
46 the primer/template junction. Nevertheless, the polymerase preferences for dNMP incorporation  
47 opposite APPXLs were clear. Both KF and RBpol predominantly incorporated dAMP, and, to a lesser  
48 extent, dGMP (Figures 4A, 4B). We observed the preference of dNMP incorporation in the order  
49 A > G > C > T, which is consistent with the "A rule", the preference pattern for better dAMP  
50 incorporation opposite the AP site (54-56). dGMP incorporation could possibly be explained by  
51 template/primer slippage since the next base after the lesion in the template is C. To address this  
52 possibility, we repeated the experiment with the substrate of another sequence featuring G as the  
53 next base (Supplementary Figure S5), for which template/primer slippage would guide the  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 incorporation of dCMP. However, the same general dNMP incorporation pattern A > G > C > T for KF  
4 and Rbpol was observed (Supplementary Figures S5A, S5B), confirming that these preferences  
5 characterize the miscoding properties of the peptide adducts rather than the tendency for template  
6 misalignment. The order of preference of dNTP incorporation by the Family Y Dpo4 was G >> A > C, T  
7 (Figure 4C); however, when the substrate was switched, the preference for dNTP usage completely to  
8 C >> A > G, T (Supplementary Figure S5C), indicating that Dpo4 use +1 nucleotide template  
9 misalignment mechanism rather than the “A rule” during the synthesis on the AP-peptide APPXL-  
10 containing substrates. Dpo4 tended to misincorporate several nucleotides, in agreement with its  
11 reported low fidelity characteristic of Family Y DNA polymerases (57). PolX also bypassed the peptide  
12 adduct by +1 slippage mechanism, albeit with a low efficiency: the enzyme preferred G > A, T, C on the  
13 substrate presenting C as the +1 template nucleotide while changing the specificity to C > A >> T, G on  
14 the substrate presenting G at this position (Figure 4D, Supplementary Figure 5D).

### 21 **AP site–peptide cross-links can be processed by AP endonucleases**

22  
23 Aldehydic AP sites and many of their natural and synthetic modifications are removed from DNA  
24 through the BER pathway initiated by AP endonucleases. These enzymes belong to two major  
25 unrelated structural groups: the exonuclease–endonuclease–phosphatase (EEP) superfamily and the  
26 triosephosphate isomerase (TIM) barrel superfamily. Therefore, we have tested representative  
27 members of each group – *E. coli* exonuclease III (Xth) and the main human AP endo/exonuclease 1  
28 (APE1) from the EEP superfamily and *E. coli* endonuclease IV (Nfo) and yeast endonuclease Apn1p  
29 from the TIM barrel superfamily – for their ability to cleave oligonucleotide substrates containing  
30 APPXL-I or APPXL-Y. Notably, all these enzymes but Xth, in addition to their AP endonuclease  
31 activity, were shown to possess the “nucleotide incision repair” (NIR) activity that cleaves DNA 5' to  
32 many lesions of different chemical nature including quite bulky ones such as 1,*N*<sup>6</sup>-etheno-dA, 3,*N*<sup>4</sup>-  
33 etheno-dC, and 3,*N*<sup>4</sup>-benzetheno-dC (41,58-62). Since the pH, buffer and ionic composition optimal  
34 for BER and NIR are enzyme-specific, we tested the ability of these AP endonucleases to cleave  
35 APPXL-containing substrates under both conditions (Supplementary Table S2).

36  
37 APE1 efficiently hydrolyzed the natural AP site but did not yield any detectable product of the  
38 APPXL-I substrate cleavage (Figure 5A). However, the APPXL-Y substrate was modestly hydrolyzed  
39 by APE1, especially under NIR conditions (Figure 6A). This result was rather surprising, since the  
40 bulkier APPXL-Y adduct was expected to have larger inhibitory effect, and suggested that the  
41 chemical nature of the peptide adduct may be important for its recognition by AP endonucleases.  
42 Another member of the EEP superfamily, Xth, also showed low activity on substrates of both types  
43 (Figure 5C, 6C). The degradation product observed at the highest enzyme concentrations migrates  
44 slower than the expected APPXL cleavage product and is likely due to the robust 3'→5'-exonuclease  
45 activity of Xth evident on the AP substrate. In contrast to APE1 and Xth, both Nfo and Apn1p  
46 efficiently processed the substrates containing APPXL-I or APPXL-Y adducts, cleaving the DNA at the  
47 cross-link site and showing some additional exonucleolytic degradation (Figure 5B, 5D, 6B, 6D). The  
48 cleavage was evident under either BER or NIR conditions. Thus, both TIM barrel AP endonucleases  
49 cleaved APPXL-containing DNA significantly better than did EEP superfamily AP endonucleases.  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 To compare the efficiency of AP endonucleases processing APPXLs with their established  
4 substrates in a quantitative way, we have measured the steady-state kinetic parameters ( $K_M$ ,  $k_{cat}$ , and  
5 the catalytic efficiency  $k_{cat}/K_M$ ) values using duplex oligonucleotides containing APPXL-I, APPXL-Y,  
6 (3-hydroxytetrahydrofuran-2-yl)methyl phosphate (THF, an AP site analog resistant to  $\beta$ -elimination  
7 and often used as a model AP site in BER studies) or 5-hydroxy-2'-deoxyuridine (OHU, a typical NIR  
8 substrate) (61,62). As shown in Table 1 and Supplementary Figure S6, the apparent  $k_{cat}/K_M$  values  
9 confirm APPXL-Y is a better substrate compared to APPXL-I for all studied AP endonucleases. For  
10 the TIM barrel enzymes, APPXL-Y cleavage activity was comparable with (~0.5-fold for Apn1p) or  
11 even significantly higher than (~13-fold for Nfo) the OHU cleavage, strongly suggesting that these AP  
12 endonucleases may be biologically relevant for processing of AP site-peptide adducts. Moreover,  
13 even under BER conditions the efficiency of APPXL-Y cleavage by Nfo was comparable with THF  
14 cleavage (THF favored only ~2-fold). More efficient processing of APPXL-Y than of APPXL-I may also  
15 be of biological importance, as peptides covalently bound to the AP site through an internal residue  
16 are statistically more likely to arise in the cell after protein conjugation followed by proteolysis.

17 We also tested the ability of several bifunctional DNA glycosylases capable of AP site cleavage,  
18 namely *E. coli* Fpg and Nei, human OGG1, NEIL1 and NEIL2, to cleave the APPXL adducts.  
19 However, none of them was able to process cross-links of either type to any significant degree  
20 (Supplementary Figure S7).

### 21 **Strand with a hanging 5'-terminal peptide cross-link can be displaced by DNA polymerases**

22 AP endonucleases nick DNA 5' of the lesion producing a hanging 5'-terminal 2'-deoxyribosephosphate  
23 (or another lesion remnant if the substrate was something else than an aldehydic AP site) and a free  
24 3'-OH, which then serves as a primer for a DNA polymerase. To see whether the nick with a 5'-  
25 terminal peptide adduct can be processed in the same way, we have first treated the APPXL-I or  
26 APPXL-Y substrate with Nfo and then added KF and all four dNTPs (Figure 7A, 7B). Although in both  
27 cases a pause was observed after the incorporation of one nucleotide, eventually the full-length  
28 primer extension product was synthesized. Thus, even if the peptide part of the cross-link interferes  
29 with polymerase binding or the early steps of primer extension, once the flap is >1 nt there is  
30 apparently no obstacles for further strand displacement.

31 Furthermore, we have reconstituted the full BER cycle with APPXL-I (annealed opposite to G;  
32 40compG template strand), Nfo, full-length *E. coli* DNA polymerase I, and DNA ligase. Unlike KF,  
33 Pol I degrades the strand ahead by a combined exo/endonucleolytic action as it extends the primer  
34 (63). To prevent full degradation of the downstream strand, we have omitted dATP and dGTP from  
35 the reaction mixture, which should cause the polymerase to stop after incorporating 14 nt, yielding a  
36 30-mer product (Figure 7C). Nfo cleaved the duplex at the adduct, producing the 16-mer with the 3'-  
37 OH end, which served as an entry point for Pol I (Figure 7C). In the absence of dNTPs, the 3'→5'  
38 exonuclease activity of Pol I degraded the primer, whereas addition of dCTP and dTTP allowed  
39 primer extension to the position of the first pyrimidine in the template. Finally, if DNA ligase was  
40 present, the product had a higher electrophoretic mobility than the DNA-peptide adduct, migrating  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 approximately at the same level as an undamaged 40-mer (Figure 7C), which clearly indicated that  
4 the repair took place.

5  
6 These observations support the possibility of APPXL removal via the long-patch BER subpathway  
7 (Figure 7D). Obviously, short-patch repair is also possible if a non-stabilized 5'-terminal conjugate can  
8 be processed by a dRP lyase.  
9

## 10 11 **DISCUSSION**

12  
13 For many years, AP sites were considered a fairly well studied type of DNA damage, but recent  
14 discoveries regarding new features of their enzymatic processing and biologically relevant reactions  
15 of AP sites with non-specific DNA-bound proteins (18,25-27) put these lesions in an unexpected  
16 perspective. There is mounting evidence that native and oxidized AP sites in the cell can covalently  
17 capture proteins to form highly toxic DNA-protein cross-links. Several breakthrough works in recent  
18 years have outlined the general repair scheme for DNA damage of this type. DPXLs mainly serve as  
19 substrates for a specialized repair pathway, which is initiated by proteolytic degradation of the protein  
20 part of the cross-link by the proteasome or specialized replication-associated proteases (SPRTN1,  
21 FAM111A, DDI1, or others yet to be discovered) to a small peptide (30-33). What happens after that,  
22 however, remains poorly understood. The studied models of naturally occurring DNA-peptide cross-  
23 links include peptide conjugates with exocyclic amino groups of G and A modified through (i) a  $\gamma$ -  
24 hydroxypropano bridge (a product of acrolein-induced DNA damage) (35,64-66) or (ii) through a C2  
25 linker derived from alkylation of Cys residues by 1,2-dibromoethane (67), and (iii) conjugates at the  
26 aldehyde group of 5-formylcytosine (68-70). In addition, unnatural adducts that are readily obtained  
27 synthetically, such as products of azide-alkyl addition to the functionalized C5 position of thymine (71)  
28 and cross-links to 7-deaza-7-(2-oxoethyl)guanine (72), were used as model systems.

29  
30 The major difference between peptide cross-links through nucleobases and through AP sites is the  
31 inability of the latter to form complementary bonds. Until now, there have been no reports in the  
32 literature regarding the levels of APPXLs, their effect on DNA polymerases or the ways to repair such  
33 residual damage. However, recent data on the reactivity of aldehyde and oxidized AP sites suggest  
34 that the number of cross-links with the AP site can be comparable to the number of cross-links with  
35 bases, or even exceed them. Moreover, some conjugates with oxidized AP sites are much more  
36 stable than reversible imine conjugates with aldehydic AP sites. In this work, we developed a model to  
37 generate cross-links of AP sites with peptides of different structure based on the reduction of the  
38 Schiff base intermediate formed by DNA glycosylases Fpg and OGG1 followed by trypsinolysis. Fpg  
39 carries an N-terminal nucleophilic amino group and, after proteolysis, leaves behind a decapeptide  
40 attached via its end (APPXL-I), while OGG1 uses the internal Lys249 for catalysis, and upon  
41 trypsinolysis yields a longer peptide remnant, in which the cross-link is formed at an internal position  
42 (APPXL-Y). This peptide adduct may be of greater biological relevance because DPXLs in the cell are  
43 statistically much more likely to be attached within the polypeptide chain, and protease cleavage  
44 immediately next to the cross-link point may be hindered by DNA. Although the diversity of proteins  
45 capable of cross-linking to DNA dictates the universality of the mechanisms for APPXL repair, the size  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 and properties of the peptide remnant might be expected to affect the APPXL interactions with DNA  
4 polymerases and the repair machinery.  
5

6 A key aspect of APPXL biochemistry determining their biological consequences is the ability to  
7 block DNA polymerases or guide incorporation of particular dNMPs. Although no single DNA–peptide  
8 adduct was studied with a fully representative set of DNA polymerases, *in vitro* experiments with  
9 peptides conjugated through nucleobases show that bypass of such lesions depends both on their  
10 size and the preservation of complementarity. DNA polymerases bind DNA in the minor groove and  
11 leave the major groove mostly accessible, so bulky moieties occupying the major groove may be  
12 better tolerated. For example, 5-formylcytosine adducts or peptides attached to the C5 position of dU  
13 through click chemistry, which protrude to the major groove and do not interfere with base pairing, are  
14 quite easily bypassed by both replicative and translesion human DNA polymerases, although longer  
15 peptides may present a problem (e. g., 10-mer peptides were bypassed, while 23-mers were not)  
16 (70,71,73). In another report, even a large 31-mer peptide conjugated to 5-formylcytosine was  
17 bypassed by translesion DNA polymerases  $\kappa$  and  $\eta$  (68). Similarly, conjugates through position C7 of  
18 7-deazaG are bypassed by translesion polymerases in a size-dependent manner (74). Peptides  
19 adducted through the exocyclic amino group of A, which may either eliminate or keep base pairing  
20 depending on the orientation of the peptide, completely block *E. coli* replicative DNA polymerase III  
21 but demonstrate low-to-full bypass by *E. coli* DNA polymerases I, II, IV and V, human translesion DNA  
22 polymerases  $\eta$ ,  $\iota$ ,  $\kappa$ , and  $\nu$ , and replicative yeast DNA polymerase  $\delta$  (65-67). On the other hand,  
23 peptide adducts with the exocyclic group of G, which either disrupt base pairing or protrude into the  
24 minor groove, are strongly blocking for most polymerases and were efficiently bypassed only by  
25 Family Y translesion DNA polymerases (*E. coli* Pol IV and human Pol  $\kappa$ ) (64,65). In the cell,  
26 mutagenic bypass of peptide adducts requires translesion DNA polymerases (69,70,75).  
27  
28  
29  
30  
31  
32  
33  
34  
35

36 Bypass of natural AP sites and their modified varieties is often described by the “A-rule”, which  
37 states that dAMP insertion is preferred opposite a non-instructive lesion. Interestingly, the structural  
38 reasons underlying this preference are different between DNA polymerases. For example, the  
39 structure of the Stoffel fragment of Taq DNA polymerase (which corresponds to KF in the *E. coli*  
40 enzyme) shows templating by a conservative Tyr residue inserted in place of a template nucleotide  
41 and acting as a thymine mimic (76,77). RBpol, on the other hand, relies on the favorable stacking  
42 between the incoming dATP and the extended planar system formed by the terminal primer  
43 nucleotide and its pairing partner in the absence of the template nucleobase (78,79). However, in the  
44 structures of both polymerases with AP-DNA, C1' of the AP site residues is buried at the protein–DNA  
45 interface, and the template strand would require extensive conformational rearrangement for a bypass  
46 if a large peptide is attached to this position. The steric clash could be relieved by flipping the  
47 template nucleotide around its flanking phosphates, as often observed in structures of Families X and  
48 Y DNA polymerases stabilizing a misaligned primer–template junction (80-82). In the “open” state of  
49 the enzyme–DNA complex, Family A DNA polymerases, including KF, also rotate the downstream  
50 template nucleotides outside of the helix, and it has been suggested that this movement helps prevent  
51 the template bases from misalignment in the absence of an incoming dNTP (83). As we do not see  
52 preferential incorporation of a dNMP corresponding to +1 template nucleotide, characteristic of the  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 primer–template misalignment bypass mechanism, we hypothesize that favorable stacking can guide  
4 dAMP and dGMP insertion opposite extrahelical APPXLs in the rare bypass events. On the other  
5 hand, structures of Dpo4 bypassing an abasic site, demonstrate extensive conformational flexibility  
6 allowing for primer/template misalignment and generation of mismatches, +1 and –1 frameshifts (84).  
7 In the structure with misalignment between the primer end and the +1 position in the template (which  
8 was observed in our experiments), the AP site is looped out of the helix into a gap between the  
9 fingers and the little finger domains, providing plenty of space for accommodation of a cross-linked  
10 peptide. No structure of PolX in the frameshift mode is available but POL $\beta$ , a prototypical member of  
11 Family X, can also loop out a template nucleotide to match the primer end to the +1 position in the  
12 template (82).

13  
14 As we used representative DNA polymerases of non-human origin in our study, it remains to be  
15 seen which DNA polymerases may participate in APPXL bypass in human cells. Based on the *in vitro*  
16 insertion specificity data, it has long been held that natural AP sites are bypassed by DNA  
17 polymerase  $\eta$  (POL $\eta$ ) (85,86). On the other hand, AP site bypass in POL $\eta$ -deficient human cells is  
18 almost as efficient as in wild-type cells (87). An alternative model was proposed, based on yeast  
19 mutation spectra and *in vitro* biochemistry of human and yeast DNA polymerase  $\zeta$  (POL $\zeta$ ), that the  
20 POL $\zeta$ /REV1 complex acting both as an inserter and an extender polymerase provides mutagenic  
21 bypass of AP sites (reviewed in (88)). However, mutation spectra induced by AP sites in yeast and  
22 human cells seem to be different (89), and human REV1 also interacts with DNA polymerases  $\eta$ ,  $\kappa$ ,  
23 and  $\iota$  (88), so the nature of the inserter polymerase in human cells exposed to AP sites or their  
24 derivatives, including APPXLs, needs further investigation.

25  
26 The high proficiency of TIM barrel AP endonucleases in nicking DNA at APPXLs was rather  
27 unexpected, since even the known substrates for the NIR activity of these enzymes are much less  
28 bulky. However, inspection of the structures of DNA-bound Nfo (90,91) reveals that C1' is solvent-  
29 accessible in both enzyme–substrate and enzyme–product complexes (Supplementary Figure S8)  
30 and the attached peptide could possibly be accommodated in the major groove of the DNA molecule,  
31 which is sharply bent upon Nfo binding. The structure of yeast Apn1p has not been solved but a  
32 similar C1' exposure is observed in a DNA-bound African swine fever virus AP endonuclease (92),  
33 another TIM barrel superfamily member. At the same time, the C1' atom of the target nucleotide is  
34 fully buried in the complex of APE1 with AP site-containing DNA, the nicked AP product, or 3'-terminal  
35 mismatched dNMPs that are substrates for the exonuclease activity of APE1 (93-95). The structure of  
36 DNA-bound *E. coli* Xth is not available but in the structurally related AP endonuclease from *Neisseria*  
37 *meningitidis* the C1' atom is also inaccessible (96). Thus, structural reasons may explain the lack of  
38 cleavage of the APPXL substrates by Xth and the mediocre cleavage by APE1 despite its NIR activity  
39 being capable of processing bulky etheno adducts (59,62). The remaining activity of APE1 on APPXL-  
40 Y may be due to a more pronounced duplex destabilization by the larger peptide, which would make  
41 easier the access to the phosphate of the damaged nucleotide without fully moving the peptide part  
42 into the AP site-recognition pocket.

43  
44 Overall, our results suggest that, at least in bacteria and budding yeast, APPXL may be repaired  
45 by the BER pathway. After the initial nick by Nfo/Apn1p, the repair could proceed along the long-patch  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 branch displacing or degrading a flap decorated with the 5'-terminal adduct. Alternatively, it is possible  
4 that 2'-deoxyribo-5'-phosphate lyases, such as RecJ or Fpg in *E. coli* (97,98) and Trf4 in yeast (99), or  
5 exonucleases resolving topoisomerase II-type adducts (100,101), can catalyze excision of 5'-terminal  
6 imine-peptide AP adducts, or that such conjugates undergo spontaneous  $\beta$ -elimination (102,103),  
7 allowing the repair to continue through the short-patch subpathway. As for human cells, given that  
8 under NIR conditions APE1 processed APPXL-Y and OHU at about the same order of magnitude,  
9 BER could also contribute, at least partly, to APPXL repair. Possible alternatives include NER  
10 (104,105), homologous recombination (106), or direct adduct uncoupling by HMCES-SRAP proteins  
11 (107). Interestingly, Nfo/Apn1p homologs are present in fish and amphibians, raising a possibility that  
12 BER of APPXL may be evolutionarily conserved and occur even in vertebrates.  
13  
14  
15  
16  
17  
18  
19

## 20 **DATA AVAILABILITY**

21  
22 All data are provided in the paper and the supplementary materials.  
23  
24  
25

## 26 **SUPPLEMENTARY DATA**

27  
28  
29 Supplementary Data are available at NAR online.  
30  
31  
32

## 33 **ACKNOWLEDGEMENT**

34  
35 Mass spectrometry was performed at the SB RAS Institute of Chemical Biology and Fundamental  
36 Medicine Joint Center for Genomic, Proteomic and Metabolomics Studies. We thank Dr. Alexander A.  
37 Chernonosov for helpful discussion.  
38  
39  
40  
41  
42

## 43 **FUNDING**

44  
45 This work was supported by the Russian Science Foundation (grant 21-74-00061 to A.V.Y., all  
46 biochemical experiments) and Russian Ministry of Higher Education and Science (grant  
47 121031300056-8 to D.O.Z.). D.V.K. is supported by a graduate student fellowship from the Russian  
48 Foundation for Basic Research (No. 20-34-90092).  
49  
50  
51  
52  
53

## 54 **CONFLICT OF INTEREST**

55  
56 None declared.  
57  
58  
59  
60

## REFERENCES

1. Lindahl, T. and Nyberg, B. (1972) Rate of depurination of native deoxyribonucleic acid. *Biochemistry*, **11**, 3610-3618.
2. Atamna, H., Cheung, I. and Ames, B.N. (2000) A method for detecting abasic sites in living cells: Age-dependent changes in base excision repair. *Proc. Natl Acad. Sci. U.S.A.*, **97**, 686-691.
3. Boiteux, S. and Guillet, M. (2004) Abasic sites in DNA: Repair and biological consequences in *Saccharomyces cerevisiae*. *DNA Repair*, **3**, 1-12.
4. Thompson, P.S. and Cortez, D. (2020) New insights into abasic site repair and tolerance. *DNA Repair*, **90**, 102866.
5. DeMott, M.S., Beyret, E., Wong, D., Bales, B.C., Hwang, J.-T., Greenberg, M.M. and Demple, B. (2002) Covalent trapping of human DNA polymerase  $\beta$  by the oxidative DNA lesion 2-deoxyribonolactone. *J. Biol. Chem.*, **277**, 7637-7640.
6. Sczepanski, J.T., Wong, R.S., McKnight, J.N., Bowman, G.D. and Greenberg, M.M. (2010) Rapid DNA-protein cross-linking and strand scission by an abasic site in a nucleosome core particle. *Proc. Natl Acad. Sci. U.S.A.*, **107**, 22475-22480.
7. Greenberg, M.M. (2014) Abasic and oxidized abasic site reactivity in DNA: Enzyme inhibition, cross-linking, and nucleosome catalyzed reactions. *Acc. Chem. Res.*, **47**, 646-655.
8. Greenberg, M.M., Weledji, Y.N., Kroeger, K.M. and Kim, J. (2004) In vitro replication and repair of DNA containing a C2'-oxidized abasic site. *Biochemistry*, **43**, 15217-15222.
9. Guan, L. and Greenberg, M.M. (2009) DNA interstrand cross-link formation by the 1,4-dioxobutane abasic lesion. *J. Am. Chem. Soc.*, **131**, 15225-15231.
10. Ghosh, S. and Greenberg, M.M. (2015) Correlation of thermal stability and structural distortion of DNA interstrand cross-links produced from oxidized abasic sites with their selective formation and repair. *Biochemistry*, **54**, 6274-6283.
11. Fortini, P. and Dogliotti, E. (2007) Base damage and single-strand break repair: Mechanisms and functional significance of short- and long-patch repair subpathways. *DNA Repair*, **6**, 398-409.
12. Wong, R.S., Sczepanski, J.T. and Greenberg, M.M. (2010) Excision of a lyase-resistant oxidized abasic lesion from DNA. *Chem. Res. Toxicol.*, **23**, 766-770.
13. Hashimoto, M., Greenberg, M.M., Kow, Y.W., Hwang, J.-T. and Cunningham, R.P. (2001) The 2-deoxyribonolactone lesion produced in DNA by neocarzinostatin and other damaging agents forms cross-links with the base-excision repair enzyme endonuclease III. *J. Am. Chem. Soc.*, **123**, 3161-3162.
14. Kroeger, K.M., Hashimoto, M., Kow, Y.W. and Greenberg, M.M. (2003) Cross-linking of 2-deoxyribonolactone and its  $\beta$ -elimination product by base excision repair enzymes. *Biochemistry*, **42**, 2449-2455.
15. Guan, L. and Greenberg, M.M. (2010) Irreversible inhibition of DNA polymerase  $\beta$  by an oxidized abasic lesion. *J. Am. Chem. Soc.*, **132**, 5004-5005.

- 1
- 2
- 3 16. Guan, L., Bebenek, K., Kunkel, T.A. and Greenberg, M.M. (2010) Inhibition of short patch and
- 4 long patch base excision repair by an oxidized abasic site. *Biochemistry*, **49**, 9904-9910.
- 5
- 6 17. Sczepanski, J.T., Zhou, C. and Greenberg, M.M. (2013) Nucleosome core particle-catalyzed
- 7 strand scission at abasic sites. *Biochemistry*, **52**, 2157-2164.
- 8
- 9 18. Ren, M., Shang, M., Wang, H., Xi, Z. and Zhou, C. (2021) Histones participate in base
- 10 excision repair of 8-oxodGuo by transiently cross-linking with active repair intermediates in
- 11 nucleosome core particles. *Nucleic Acids Res.*, **49**, 257-268.
- 12
- 13 19. Mazumder, A., Neamati, N., Pilon, A.A., Sunder, S. and Pommier, Y. (1996) Chemical
- 14 trapping of ternary complexes of human immunodeficiency virus type 1 integrase, divalent
- 15 metal, and DNA substrates containing an abasic site: Implications for the role of lysine 136 in
- 16 DNA binding. *J. Biol. Chem.*, **271**, 27330-27338.
- 17
- 18 20. Bogenhagen, D.F. and Pinz, K.G. (1998) The action of DNA ligase at abasic sites in DNA. *J.*
- 19 *Biol. Chem.*, **273**, 7888-7893.
- 20
- 21 21. Kosova, A.A., Khodyreva, S.N. and Lavrik, O.I. (2016) Ku antigen displays the AP lyase
- 22 activity on a certain type of duplex DNA. *Biochim. Biophys. Acta*, **1864**, 1244-1252.
- 23
- 24 22. Rieger, R.A., Zaika, E.I., Xie, W., Johnson, F., Grollman, A.P., Iden, C.R. and Zharkov, D.O.
- 25 (2006) Proteomic approach to identification of proteins reactive for abasic sites in DNA. *Mol.*
- 26 *Cell. Proteomics*, **5**, 858-867.
- 27
- 28 23. Grosheva, A.S., Zharkov, D.O., Stahl, J., Gopanenko, A.V., Tupikin, A.E., Kabilov, M.R.,
- 29 Graifer, D.M. and Karpova, G.G. (2017) Recognition but no repair of abasic site in single-
- 30 stranded DNA by human ribosomal uS3 protein residing within intact 40S subunit. *Nucleic*
- 31 *Acids Res.*, **45**, 3833-3843.
- 32
- 33 24. Kosova, A.A., Khodyreva, S.N. and Lavrik, O.I. (2015) Glyceraldehyde-3-phosphate
- 34 dehydrogenase (GAPDH) interacts with apurinic/apyrimidinic sites in DNA. *Mutat. Res.*, **779**,
- 35 46-57.
- 36
- 37 25. Halabelian, L., Ravichandran, M., Li, Y., Zeng, H., Rao, A., Aravind, L. and Arrowsmith, C.H.
- 38 (2019) Structural basis of HMCES interactions with abasic DNA and multivalent substrate
- 39 recognition. *Nat. Struct. Mol. Biol.*, **26**, 607-612.
- 40
- 41 26. Mohni, K.N., Wessel, S.R., Zhao, R., Wojciechowski, A.C., Luzwick, J.W., Layden, H.,
- 42 Eichman, B.F., Thompson, P.S., Mehta, K.P.M. and Cortez, D. (2019) HMCES maintains
- 43 genome integrity by shielding abasic sites in single-strand DNA. *Cell*, **176**, 144-153.
- 44
- 45 27. Thompson, P.S., Amidon, K.M., Mohni, K.N., Cortez, D. and Eichman, B.F. (2019) Protection
- 46 of abasic sites during DNA replication by a stable thiazolidine protein-DNA cross-link. *Nat.*
- 47 *Struct. Mol. Biol.*, **26**, 613-618.
- 48
- 49 28. Barker, S., Weinfeld, M. and Murray, D. (2005) DNA-protein crosslinks: Their induction,
- 50 repair, and biological consequences. *Mutat. Res.*, **589**, 111-135.
- 51
- 52 29. Pommier, Y., Leo, E., Zhang, H. and Marchand, C. (2010) DNA topoisomerases and their
- 53 poisoning by anticancer and antibacterial drugs. *Chem. Biol.*, **17**, 421-433.
- 54
- 55 30. Duxin, J.P., Dewar, J.M., Yardimci, H. and Walter, J.C. (2014) Repair of a DNA-protein
- 56 crosslink by replication-coupled proteolysis. *Cell*, **159**, 346-357.
- 57
- 58
- 59
- 60

- 1
- 2
- 3 31. Stinglee, J., Schwarz, M.S., Bloemeke, N., Wolf, P.G. and Jentsch, S. (2014) A DNA-
- 4 dependent protease involved in DNA-protein crosslink repair. *Cell*, **158**, 327-338.
- 5
- 6 32. Vaz, B., Popovic, M., Newman, J.A., Fielden, J., Aitkenhead, H., Halder, S., Singh, A.N.,
- 7 Vendrell, I., Fischer, R., Torrecilla, I. *et al.* (2016) Metalloprotease SPRTN/DVC1 orchestrates
- 8 replication-coupled DNA-protein crosslink repair. *Mol. Cell*, **64**, 704-719.
- 9
- 10 33. Larsen, N.B., Gao, A.O., Sparks, J.L., Gallina, I., Wu, R.A., Mann, M., Räsche, M., Walter,
- 11 J.C. and Duxin, J.P. (2019) Replication-coupled DNA-protein crosslink repair by SPRTN and
- 12 the proteasome in *Xenopus* egg extracts. *Mol. Cell*, **73**, 574-588.e577.
- 13
- 14 34. Quievryn, G. and Zhitkovich, A. (2000) Loss of DNA–protein crosslinks from formaldehyde-
- 15 exposed cells occurs through spontaneous hydrolysis and an active repair process linked to
- 16 proteasome function. *Carcinogenesis*, **21**, 1573-1580.
- 17
- 18 35. Minko, I.G., Kurtz, A.J., Croteau, D.L., Van Houten, B., Harris, T.M. and Lloyd, R.S. (2005)
- 19 Initiation of repair of DNA–polypeptide cross-links by the UvrABC nuclease. *Biochemistry*, **44**,
- 20 3000-3009.
- 21
- 22 36. Nakano, T., Morishita, S., Katafuchi, A., Matsubara, M., Horikawa, Y., Terato, H., Salem,
- 23 A.M.H., Izumi, S., Pack, S.P., Makino, K. *et al.* (2007) Nucleotide excision repair and
- 24 homologous recombination systems commit differentially to the repair of DNA-protein
- 25 crosslinks. *Mol. Cell*, **28**, 147-158.
- 26
- 27 37. de Graaf, B., Clore, A. and McCullough, A.K. (2009) Cellular pathways for DNA repair and
- 28 damage tolerance of formaldehyde-induced DNA-protein crosslinks. *DNA Repair*, **8**, 1207-
- 29 1214.
- 30
- 31 38. McKibbin, P.L., Fleming, A.M., Towheed, M.A., Van Houten, B., Burrows, C.J. and David,
- 32 S.S. (2013) Repair of hydantoin lesions and their amine adducts in DNA by base and
- 33 nucleotide excision repair. *J. Am. Chem. Soc.*, **135**, 13851-13861.
- 34
- 35 39. Gilboa, R., Zharkov, D.O., Golan, G., Fernandes, A.S., Gerchman, S.E., Matz, E., Kycia, J.H.,
- 36 Grollman, A.P. and Shoham, G. (2002) Structure of formamidopyrimidine-DNA glycosylase
- 37 covalently complexed to DNA. *J. Biol. Chem.*, **277**, 19811-19816.
- 38
- 39 40. Rieger, R.A., McTigue, M.M., Kycia, J.H., Gerchman, S.E., Grollman, A.P. and Iden, C.R.
- 40 (2000) Characterization of a cross-linked DNA-endonuclease VIII repair complex by
- 41 electrospray ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.*, **11**, 505-515.
- 42
- 43 41. Ishchenko, A.A., Sanz, G., Privezentzev, C.V., Maksimenko, A.V. and Sapparbaev, M. (2003)
- 44 Characterisation of new substrate specificities of *Escherichia coli* and *Saccharomyces*
- 45 *cerevisiae* AP endonucleases. *Nucleic Acids Res.*, **31**, 6344-6353.
- 46
- 47 42. Sidorenko, V.S., Nevinsky, G.A. and Zharkov, D.O. (2007) Mechanism of interaction between
- 48 human 8-oxoguanine-DNA glycosylase and AP endonuclease. *DNA Repair*, **6**, 317-328.
- 49
- 50 43. Kuznetsov, N.A., Koval, V.V., Zharkov, D.O., Nevinsky, G.A., Douglas, K.T. and Fedorova,
- 51 O.S. (2005) Kinetics of substrate recognition and cleavage by human 8-oxoguanine-DNA
- 52 glycosylase. *Nucleic Acids Res.*, **33**, 3919-3931.
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60

- 1  
2  
3 44. Katafuchi, A., Nakano, T., Masaoka, A., Terato, H., Iwai, S., Hanaoka, F. and Ide, H. (2004)  
4 Differential specificity of human and *Escherichia coli* endonuclease III and VIII homologues for  
5 oxidative base lesions. *J. Biol. Chem.*, **279**, 14464-14471.  
6  
7 45. Kakhkharova, Z.I., Zharkov, D.O. and Grin, I.R. (2022) A low-activity polymorphic variant of  
8 human NEIL2 DNA glycosylase. *Int. J. Mol. Sci.*, **23**, 2212.  
9  
10 46. Miller, H. and Grollman, A.P. (1997) Kinetics of DNA polymerase I (Klenow fragment exo<sup>-</sup>)  
11 activity on damaged DNA templates: Effect of proximal and distal template damage on DNA  
12 synthesis. *Biochemistry*, **36**, 15336-15342.  
13  
14 47. Freisinger, E., Grollman, A.P., Miller, H. and Kisker, C. (2004) Lesion (in)tolerance reveals  
15 insights into DNA replication fidelity. *EMBO J.*, **23**, 1494-1505.  
16  
17 48. Yudkina, A.V., Dvornikova, A.P. and Zharkov, D.O. (2018) Variable termination sites of DNA  
18 polymerases encountering a DNA–protein cross-link. *PLoS ONE*, **13**, e0198480.  
19  
20 49. Boldinova, E.O., Yudkina, A.V., Shilkin, E.S., Gagarinskaya, D.I., Baranovskiy, A.G., Tahirov,  
21 T.H., Zharkov, D.O. and Makarova, A.V. (2021) Translesion activity of PrimPol on DNA with  
22 cisplatin and DNA–protein cross-links. *Sci. Rep.*, **11**, 17588.  
23  
24 50. Yudkina, A.V., Shilkin, E.S., Makarova, A.V. and Zharkov, D.O. (2022) Stalling of eukaryotic  
25 translesion DNA polymerases at DNA-protein cross-links. *Genes*, **13**, 166.  
26  
27 51. Fromme, J.C., Bruner, S.D., Yang, W., Karplus, M. and Verdine, G.L. (2003) Product-assisted  
28 catalysis in base-excision DNA repair. *Nat. Struct. Biol.*, **10**, 204-211.  
29  
30 52. Kumar, S., Lamarche, B.J. and Tsai, M.-D. (2007) Use of damaged DNA and dNTP  
31 substrates by the error-prone DNA polymerase X from african swine fever virus. *Biochemistry*,  
32 **46**, 3814-3825.  
33  
34 53. Howard, M.J. and Wilson, S.H. (2018) DNA scanning by base excision repair enzymes and  
35 implications for pathway coordination. *DNA Repair*, **71**, 101-107.  
36  
37 54. Goodman, M.F., Cai, H., Bloom, L.B. and Eritja, R. (1994) Nucleotide insertion and primer  
38 extension at abasic template sites in different sequence contexts. *Ann. N. Y. Acad. Sci.*, **726**,  
39 132-142.  
40  
41 55. Paz-Elizur, T., Takeshita, M. and Livneh, Z. (1997) Mechanism of bypass synthesis through  
42 an abasic site analog by DNA polymerase I. *Biochemistry*, **36**, 1766-1773.  
43  
44 56. Taylor, J.-S. (2002) New structural and mechanistic insight into the A-rule and the  
45 instructional and non-instructional behavior of DNA photoproducts and other lesions. *Mutat.*  
46 *Res.*, **510**, 55-70.  
47  
48 57. Kokoska, R.J., Bebenek, K., Boudsocq, F., Woodgate, R. and Kunkel, T.A. (2002) Low fidelity  
49 DNA synthesis by a  $\gamma$  family DNA polymerase due to misalignment in the active site. *J. Biol.*  
50 *Chem.*, **277**, 19633-19638.  
51  
52 58. Ide, H., Tedzuka, K., Shimzu, H., Kimura, Y., Purmal, A.A., Wallace, S.S. and Kow, Y.W.  
53 (1994)  $\alpha$ -Deoxyadenosine, a major anoxic radiolysis product of adenine in DNA, is a  
54 substrate for *Escherichia coli* endonuclease IV. *Biochemistry*, **33**, 7842-7847.  
55  
56 59. Hang, B., Chenna, A., Fraenkel-Conrat, H. and Singer, B. (1996) An unusual mechanism for  
57 the major human apurinic/apyrimidinic (AP) endonuclease involving 5' cleavage of DNA  
58  
59  
60

- 1  
2  
3 containing a benzene-derived exocyclic adduct in the absence of an AP site. *Proc. Natl Acad. Sci. U.S.A.*, **93**, 13737-13741.
- 4  
5  
6 60. Ischenko, A.A. and Saparbaev, M.K. (2002) Alternative nucleotide incision repair pathway for  
7 oxidative DNA damage. *Nature*, **415**, 183-187.
- 8  
9 61. Gros, L., Ishchenko, A.A., Ide, H., Elder, R.H. and Saparbaev, M.K. (2004) The major human  
10 AP endonuclease (Ape1) is involved in the nucleotide incision repair pathway. *Nucleic Acids*  
11 *Res.*, **32**, 73-81.
- 12  
13 62. Prorok, P., Saint-Pierre, C., Gasparutto, D., Fedorova, O.S., Ishchenko, A.A., Leh, H., Buckle,  
14 M., Tudek, B. and Saparbaev, M. (2012) Highly mutagenic exocyclic DNA adducts are  
15 substrates for the human nucleotide incision repair pathway. *PLoS ONE*, **7**, e51776.
- 16  
17 63. Lyamichev, V., Brow, M.A.D. and Dahlberg, J.E. (1993) Structure-specific endonucleolytic  
18 cleavage of nucleic acids by eubacterial DNA polymerases. *Science*, **260**, 778-783.
- 19  
20 64. Minko, I.G., Yamanaka, K., Kozekov, I.D., Kozekova, A., Indiani, C., O'Donnell, M.E., Jiang,  
21 Q., Goodman, M.F., Rizzo, C.J. and Lloyd, R.S. (2008) Replication bypass of the acrolein-  
22 mediated deoxyguanine DNA-peptide cross-links by DNA polymerases of the DinB family.  
23 *Chem. Res. Toxicol.*, **21**, 1983-1990.
- 24  
25 65. Yamanaka, K., Minko, I.G., Takata, K.-i., Kolbanovskiy, A., Kozekov, I.D., Wood, R.D., Rizzo,  
26 C.J. and Lloyd, R.S. (2010) Novel enzymatic function of DNA polymerase  $\nu$  in translesion  
27 DNA synthesis past major groove DNA-peptide and DNA-DNA cross-links. *Chem. Res.*  
28 *Toxicol.*, **23**, 689-695.
- 29  
30 66. Yamanaka, K., Minko, I.G., Finkel, S.E., Goodman, M.F. and Lloyd, R.S. (2011) Role of high-  
31 fidelity *Escherichia coli* DNA polymerase I in replication bypass of a deoxyadenosine DNA-  
32 peptide cross-link. *J. Bacteriol.*, **193**, 3815-3821.
- 33  
34 67. Ghodke, P.P., Gonzalez-Vasquez, G., Wang, H., Johnson, K.M., Sedgeman, C.A. and  
35 Guengerich, F.P. (2021) Enzymatic bypass of an  $N^6$ -deoxyadenosine DNA-ethylene  
36 dibromide-peptide crosslink by translesion DNA polymerases. *J. Biol. Chem.*, **296**, 100444.
- 37  
38 68. Ji, S., Fu, I., Naldiga, S., Shao, H., Basu, A.K., Broyde, S. and Tretyakova, N.Y. (2018) 5-  
39 Formylcytosine mediated DNA-protein cross-links block DNA replication and induce  
40 mutations in human cells. *Nucleic Acids Res.*, **46**, 6455-6469.
- 41  
42 69. Ji, S., Park, D., Kropachev, K., Essawy, M., Geacintov, N.E. and Tretyakova, N.Y. (2019) 5-  
43 Formylcytosine-induced DNA-peptide cross-links reduce transcription efficiency, but do not  
44 cause transcription errors in human cells. *J. Biol. Chem.*, **294**, 18387-18397.
- 45  
46 70. Naldiga, S., Ji, S., Thomforde, J., Nicolae, C.M., Lee, M., Zhang, Z., Moldovan, G.-L.,  
47 Tretyakova, N.Y. and Basu, A.K. (2019) Error-prone replication of a 5-formylcytosine-  
48 mediated DNA-peptide cross-link in human cells. *J. Biol. Chem.*, **294**, 10619-10627.
- 49  
50 71. Wickramaratne, S., Boldry, E.J., Buehler, C., Wang, Y.-C., Distefano, M.D. and Tretyakova,  
51 N.Y. (2015) Error-prone translesion synthesis past DNA-peptide cross-links conjugated to the  
52 major groove of DNA via C5 of thymidine. *J. Biol. Chem.*, **290**, 775-787.
- 53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 72. Pande, P., Ji, S., Mukherjee, S., Schärer, O.D., Tretyakova, N.Y. and Basu, A.K. (2017)  
4 Mutagenicity of a model DNA-peptide cross-link in human cells: Roles of translesion synthesis  
5 DNA polymerases. *Chem. Res. Toxicol.*, **30**, 669-677.  
6  
7 73. Yeo, J.E., Wickramaratne, S., Khatwani, S., Wang, Y.-C., Vervacke, J., Distefano, M.D. and  
8 Tretyakova, N.Y. (2014) Synthesis of site-specific DNA–protein conjugates and their effects  
9 on DNA replication. *ACS Chem. Biol.*, **9**, 1860-1868.  
10  
11 74. Wickramaratne, S., Ji, S., Mukherjee, S., Su, Y., Pence, M.G., Lior-Hoffmann, L., Fu, I.,  
12 Broyde, S., Guengerich, F.P., Distefano, M. *et al.* (2016) Bypass of DNA-protein cross-links  
13 conjugated to the 7-deazaguanine position of DNA by translesion synthesis polymerases. *J.*  
14 *Biol. Chem.*, **291**, 23589-23603.  
15  
16 75. Ji, S., Thomforde, J., Rogers, C., Fu, I., Broyde, S. and Tretyakova, N.Y. (2019)  
17 Transcriptional bypass of DNA–protein and DNA–peptide conjugates by T7 RNA polymerase.  
18 *ACS Chem. Biol.*, **14**, 2564-2575.  
19  
20 76. Obeid, S., Blatter, N., Kranaster, R., Schnur, A., Diederichs, K., Welte, W. and Marx, A.  
21 (2010) Replication through an abasic DNA lesion: Structural basis for adenine selectivity.  
22 *EMBO J.*, **29**, 1738-1747.  
23  
24 77. Obeid, S., Welte, W., Diederichs, K. and Marx, A. (2012) Amino acid templating mechanisms  
25 in selection of nucleotides opposite abasic sites by a family A DNA polymerase. *J. Biol.*  
26 *Chem.*, **287**, 14099-14108.  
27  
28 78. Zahn, K.E., Belrhali, H., Wallace, S.S. and Doublé, S. (2007) Caught bending the A-rule:  
29 Crystal structures of translesion DNA synthesis with a non-natural nucleotide. *Biochemistry*,  
30 **46**, 10551-10561.  
31  
32 79. Xia, S., Vashishtha, A., Bulkley, D., Eom, S.H., Wang, J. and Konigsberg, W.H. (2012)  
33 Contribution of partial charge interactions and base stacking to the efficiency of primer  
34 extension at and beyond abasic sites in DNA. *Biochemistry*, **51**, 4922-4931.  
35  
36 80. Garcia-Diaz, M., Bebenek, K., Krahn, J.M., Pedersen, L.C. and Kunkel, T.A. (2006) Structural  
37 analysis of strand misalignment during DNA synthesis by a human DNA polymerase. *Cell*,  
38 **124**, 331-342.  
39  
40 81. Bauer, J., Xing, G., Yagi, H., Sayer, J.M., Jerina, D.M. and Ling, H. (2007) A structural gap in  
41 Dpo4 supports mutagenic bypass of a major benzo[a]pyrene dG adduct in DNA through  
42 template misalignment. *Proc. Natl Acad. Sci. U.S.A.*, **104**, 14905-14910.  
43  
44 82. Howard, M.J., Cavanaugh, N.A., Batra, V.K., Shock, D.D., Beard, W.A. and Wilson, S.H.  
45 (2020) DNA polymerase  $\beta$  nucleotide-stabilized template misalignment fidelity depends on  
46 local sequence context. *J. Biol. Chem.*, **295**, 529-538.  
47  
48 83. Patel, P.H., Suzuki, M., Adman, E., Shinkai, A. and Loeb, L.A. (2001) Prokaryotic DNA  
49 polymerase I: Evolution, structure, and "base flipping" mechanism for nucleotide selection. *J.*  
50 *Mol. Biol.*, **308**, 823-837.  
51  
52 84. Ling, H., Boudsocq, F., Woodgate, R. and Yang, W. (2004) Snapshots of replication through  
53 an abasic lesion: Structural basis for base substitutions and frameshifts. *Mol. Cell*, **13**, 751-  
54 762.  
55  
56  
57  
58  
59  
60

- 1  
2  
3 85. Zhang, Y., Yuan, F., Wu, X., Rechkoblit, O., Taylor, J.-S., Geacintov, N.E. and Wang, Z.  
4 (2000) Error-prone lesion bypass by human DNA polymerase  $\eta$ . *Nucleic Acids Res.*, **28**,  
5 4717-4724.  
6  
7 86. Choi, J.-Y., Lim, S., Kim, E.-J., Jo, A. and Guengerich, F.P. (2010) Translesion synthesis  
8 across abasic lesions by human B-family and Y-family DNA polymerases  $\alpha$ ,  $\delta$ ,  $\eta$ ,  $\iota$ ,  $\kappa$ , and  
9 REV1. *J. Mol. Biol.*, **404**, 34-44.  
10  
11 87. Avkin, S., Adar, S., Blander, G. and Livneh, Z. (2002) Quantitative measurement of  
12 translesion replication in human cells: Evidence for bypass of abasic sites by a replicative  
13 DNA polymerase. *Proc. Natl Acad. Sci. U.S.A.*, **99**, 3764-3769.  
14  
15 88. Makarova, A.V. and Burgers, P.M. (2015) Eukaryotic DNA polymerase  $\zeta$ . *DNA Repair*, **29**, 47-  
16 55.  
17  
18 89. Weerasooriya, S., Jasti, V.P. and Basu, A.K. (2014) Replicative bypass of abasic site in  
19 *Escherichia coli* and human cells: Similarities and differences. *PLoS ONE*, **9**, e107915.  
20  
21 90. Hosfield, D.J., Guan, Y., Haas, B.J., Cunningham, R.P. and Tainer, J.A. (1999) Structure of  
22 the DNA repair enzyme endonuclease IV and its DNA complex: Double-nucleotide flipping at  
23 abasic sites and three-metal-ion catalysis. *Cell*, **98**, 397-408.  
24  
25 91. Garcin, E.D., Hosfield, D.J., Desai, S.A., Haas, B.J., Björas, M., Cunningham, R.P. and  
26 Tainer, J.A. (2008) DNA apurinic-apyrimidinic site binding and excision by endonuclease IV.  
27 *Nat. Struct. Mol. Biol.*, **15**, 515-522.  
28  
29 92. Chen, Y., Chen, X., Huang, Q., Shao, Z., Gao, Y., Li, Y., Yang, C., Liu, H., Li, J., Wang, Q. *et al.*  
30 (2020) A unique DNA-binding mode of African swine fever virus AP endonuclease. *Cell*  
31 *Discov.*, **6**, 13.  
32  
33 93. Mol, C.D., Izumi, T., Mitra, S. and Tainer, J.A. (2000) DNA-bound structures and mutants  
34 reveal abasic DNA binding by APE1 and DNA repair coordination. *Nature*, **403**, 451-456.  
35  
36 94. Freudenthal, B.D., Beard, W.A., Cuneo, M.J., Dyrkheeva, N.S. and Wilson, S.H. (2015)  
37 Capturing snapshots of APE1 processing DNA damage. *Nat. Struct. Mol. Biol.*, **22**, 924-931.  
38  
39 95. Whitaker, A.M., Flynn, T.S. and Freudenthal, B.D. (2018) Molecular snapshots of APE1  
40 proofreading mismatches and removing DNA damage. *Nat. Commun.*, **9**, 399.  
41  
42 96. Lu, D., Silhan, J., MacDonald, J.T., Carpenter, E.P., Jensen, K., Tang, C.M., Baldwin, G.S.  
43 and Freemont, P.S. (2012) Structural basis for the recognition and cleavage of abasic DNA in  
44 *Neisseria meningitidis*. *Proc. Natl Acad. Sci. U.S.A.*, **109**, 16852-16857.  
45  
46 97. Graves, R.J., Felzenszwalb, I., Laval, J. and O'Connor, T.R. (1992) Excision of 5'-terminal  
47 deoxyribose phosphate from damaged DNA is catalyzed by the Fpg protein of *Escherichia*  
48 *coli*. *J. Biol. Chem.*, **267**, 14429-14435.  
49  
50 98. Dianov, G., Sedgwick, B., Daly, G., Olsson, M., Lovett, S. and Lindahl, T. (1994) Release of  
51 5'-terminal deoxyribose-phosphate residues from incised abasic sites in DNA by the  
52 *Escherichia coli* RecJ protein. *Nucleic Acids Res.*, **22**, 993-998.  
53  
54 99. Gellon, L., Carson, D.R., Carson, J.P. and Demple, B. (2008) Intrinsic 5'-deoxyribose-5-  
55 phosphate lyase activity in *Saccharomyces cerevisiae* Trf4 protein with a possible role in base  
56 excision DNA repair. *DNA Repair*, **7**, 187-198.  
57  
58  
59  
60

- 1  
2  
3 100. Neale, M.J., Pan, J. and Keeney, S. (2005) Endonucleolytic processing of covalent protein-  
4 linked DNA double-strand breaks. *Nature*, **436**, 1053-1057.  
5  
6 101. Huang, S.-Y.N., Michaels, S.A., Mitchell, B.B., Majdalani, N., Vanden Broeck, A., Canela, A.,  
7 Tse-Dinh, Y.-C., Lamour, V. and Pommier, Y. (2021) Exonuclease VII repairs quinolone-  
8 induced damage by resolving DNA gyrase cleavage complexes. *Sci. Adv.*, **7**, eabe0384.  
9  
10 102. Jha, J.S., Nel, C., Haldar, T., Peters, D., Housh, K. and Gates, K.S. (2022) Products  
11 generated by amine-catalyzed strand cleavage at apurinic/apyrimidinic sites in DNA: New  
12 insights from a biomimetic nucleoside model system. *Chem. Res. Toxicol.*, **35**, 203-217.  
13  
14 103. Haldar, T., Jha, J.S., Yang, Z., Nel, C., Housh, K., Cassidy, O.J. and Gates, K.S. (2022)  
15 Unexpected complexity in the products arising from NaOH-, heat-, amine-, and glycosylase-  
16 induced strand cleavage at an abasic site in DNA. *Chem. Res. Toxicol.*, **35**, 218-232.  
17  
18 104. Reardon, J.T. and Sancar, A. (2006) Repair of DNA–polypeptide crosslinks by human  
19 excision nuclease. *Proc. Natl Acad. Sci. U.S.A.*, **103**, 4056-4061.  
20  
21 105. Baker, D.J., Wuenschell, G., Xia, L., Termini, J., Bates, S.E., Riggs, A.D. and O'Connor, T.R.  
22 (2007) Nucleotide excision repair eliminates unique DNA-protein cross-links from mammalian  
23 cells. *J. Biol. Chem.*, **282**, 22592-22604.  
24  
25 106. Nakano, T., Katafuchi, A., Matsubara, M., Terato, H., Tsuboi, T., Masuda, T., Tatsumoto, T.,  
26 Pack, S.P., Makino, K., Croteau, D.L. *et al.* (2009) Homologous recombination but not  
27 nucleotide excision repair plays a pivotal role in tolerance of DNA-protein cross-links in  
28 mammalian cells. *J. Biol. Chem.*, **284**, 27065-27076.  
29  
30 107. Paulin, K.A., Cortez, D. and Eichman, B.F. (2022) The SOS response-associated peptidase  
31 (SRAP) domain of YedK catalyzes ring opening of abasic sites and reversal of its DNA-  
32 protein crosslink. *J. Biol. Chem.*, **298**, 102307.  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## TABLE AND FIGURES LEGENDS

**Table 1.** Steady-state kinetics of cleavage of APPXL substrates by AP endonucleases.

**Figure 1.** Scheme of the DNA–peptide cross-link preparation. 8-OxoG-containing DNA duplexes were treated with bifunctional DNA glycosylases Fpg (top) or OGG1 (bottom) to produce a Schiff base conjugate with the nascent aldehydic AP site through a terminal amino group in case of Fpg or an internal amino group in case of OGG1. The enzyme–DNA intermediates were stabilized with NaBH<sub>4</sub> to yield covalent DNA-protein cross-links (DPXLs). Then, DPXL were subjected to full trypsinolysis to obtain AP-peptide cross-links.

**Figure 2.** Stalling of KF and RBpol by AP–peptide cross-links. A, APPXL-I, primer–template; B, APPXL-Y, primer–template; C, APPXL-I, primer–downstream strand–template; D, APPXL-Y, primer–downstream strand–template. The <sup>32</sup>P-labeled primer is shown in red, the downstream strand is in lowercase. Grey arrows mark the sites of synthesis termination at the lesion. Reaction time and the nature of the lesion are indicated under the gel images. In all panels, lanes 1–3 are size markers corresponding to the primer (11 nt), full-size product (40 nt), and primer extended to the cross-link site (23 nt). “–”, no enzyme added; K1, undamaged primer–template, 30 min; K2, undamaged primer–template–downstream strand, 30 min. The inset shows KF pause sites at a lower band intensity.

**Figure 3.** Stalling of Dpo4 and PolX by AP–peptide cross-links. (A), primer–template; (B), primer–downstream strand–template. The nature of the polymerase and the cross-link is indicated over and below the gel images, respectively. The <sup>32</sup>P-labeled primer is shown in red, the downstream strand is in lowercase. Grey arrows mark the sites of synthesis termination at the lesion. Reaction time and the nature of the lesion are indicated under the gel images. In all panels, lanes 1–3 are size markers corresponding to the primer (11 nt), full-size product (40 nt), and primer extended to the cross-link site (23 nt). “–”, no enzyme added; K1, undamaged primer–template, 30 min; K2, undamaged primer–template–downstream strand, 30 min. No bypass was observed for PolX.

**Figure 4.** Incorporation of dNMPs opposite APPXLs by KF (A), RBpol (B), Dpo4 (C) and PolX (D). The nature of the substrate and dNTP is indicated below the gel images. The <sup>32</sup>P-labeled primer is shown in red. “–”, no enzyme or dNTPs added; N, all dNTPs added; K1, undamaged primer–template, all dNTPs, 30 min.

**Figure 5.** Cleavage of APPXL-I by AP endonucleases: APE1 (A), Apn1p (B), Xth (C), and Nfo (D). For panels A–C: lane 1, APPXL-I, no enzyme; lanes 2–5, APPXL-I cleavage under the BER conditions; lanes 6–9, APPXL-I cleavage under the NIR conditions; lanes 10–13, AP site cleavage under the BER conditions; lanes 14–17, AP site cleavage under the NIR conditions; lane 18, AP site treated with NaOH; lane 19, AP site, no enzyme. Panel D: lanes 1–4, APPXL-I cleavage under the BER conditions; lanes 5–8, APPXL-I cleavage under the NIR conditions; lanes 9–12, AP site cleavage under the BER conditions; lane 13, AP site treated with NaOH; lane 14, AP site, no enzyme. BER and NIR conditions for the individual AP endonucleases are specified in Supplementary Table S2.

1  
2  
3 **Figure 6.** Cleavage of APPXL-Y by AP endonucleases: APE1 (A), Apn1p (B), Xth (C), and Nfo (D).  
4 For panels A–C: lane 1, APPXL-Y, no enzyme; lanes 2–5, APPXL-Y cleavage under the BER  
5 conditions; lanes 6–9, APPXL-Y cleavage under the NIR conditions; lanes 10–13, AP site cleavage  
6 under the BER conditions; lanes 14–17, AP site cleavage under the NIR conditions; lane 18, AP site  
7 untreated (A) or treated with NaOH (B, C). Panel D: lane 1, APPXL-Y, no enzyme; lanes 2–5, APPXL-  
8 Y cleavage under the BER conditions; lanes 6–9, APPXL-Y cleavage under the NIR conditions;  
9 lanes 10–13, AP site cleavage under the BER conditions. BER and NIR conditions for the individual  
10 AP endonucleases are specified in Supplementary Table S2.  
11  
12  
13  
14

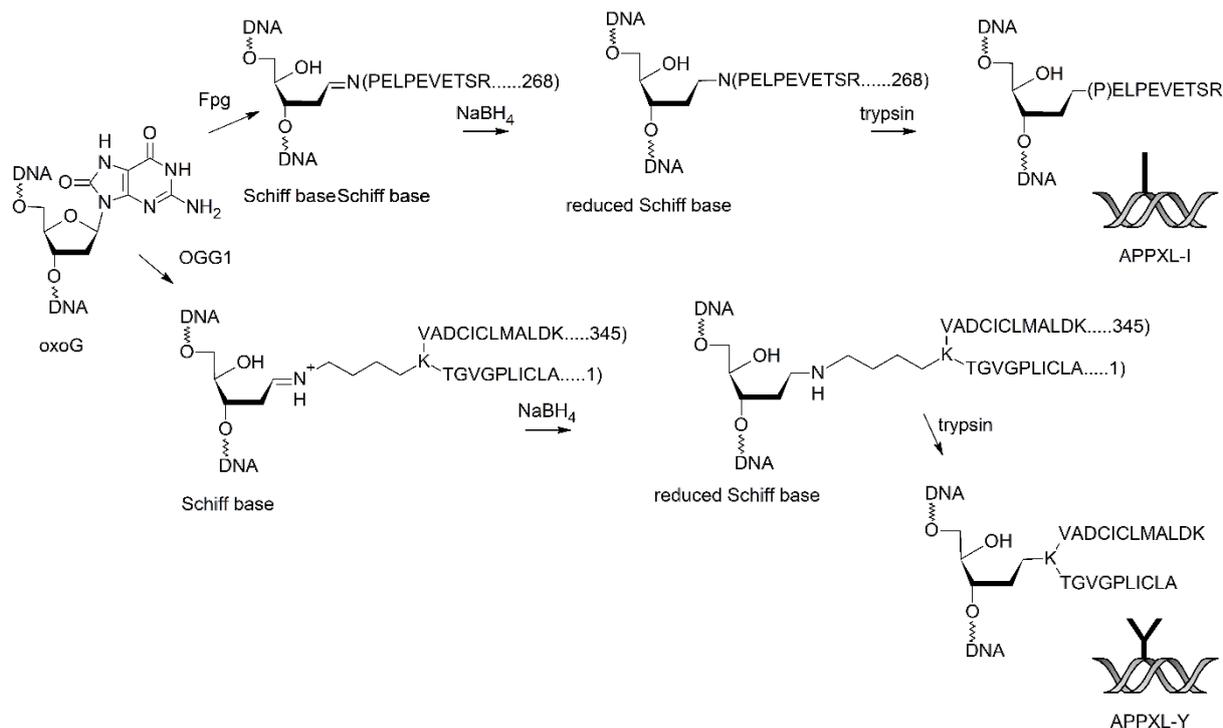
15 **Figure 7.** Base excision repair continuation after the cleavage of AP site–peptide cross-links. A, B,  
16 Extension by KF of the primer with the displacement of the downstream strand bearing a 5'-terminal  
17 cross-link after Nfo cleavage of APPXL-I (A) and APPXL-Y (B). Lane 1, cross-linked strand; lane 2,  
18 APPXL treated with Nfo; lanes 3–5, APPXL treated with Nfo followed by KF extension for the  
19 indicated time; lane 6, size marker corresponding to the unmodified strand. C, Reconstitution of the  
20 full BER cycle of APPXL-I with Nfo, Pol I, and DNA ligase. Lane 1 and 2, unmodified and cross-linked  
21 strands, respectively; lane 3, substrate cleaved by Nfo; lane 4, substrate cleaved by Nfo in the  
22 presence of Pol I without dNTPs; lane 5; substrate cleaved by Nfo and extended by Pol I; lane 6,  
23 substrate cleaved by Nfo, extended by Pol I and ligated. D General scheme of possible BER-  
24 dependent repair of AP site–peptide cross-links.  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Table 1.** Steady-state kinetics of cleavage of APPXL substrates by AP endonucleases.

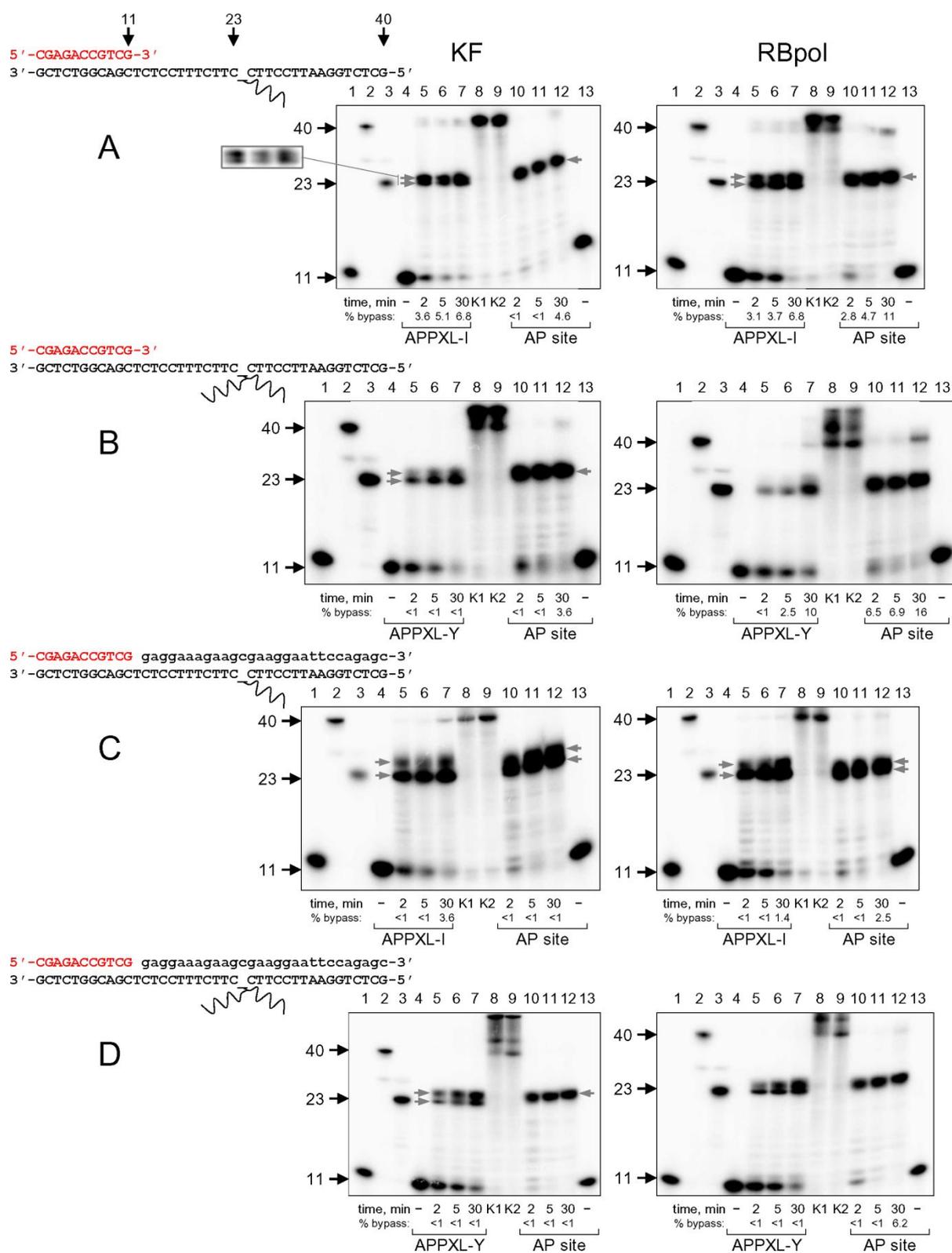
Enzyme	Substrate	Conditions	$K_M$ , nM	$k_{cat}$ , $s^{-1}$ ( $\times 10^{-5}$ )	$k_{cat}/K_M$ , $nM^{-1}\cdot s^{-1}$ ( $\times 10^{-8}$ )
Apn1p	APPXL-I	BER	190 $\pm$ 60	2.0 $\pm$ 0.3	10 $\pm$ 3
Apn1p	APPXL-I	NIR	21 $\pm$ 8	1.2 $\pm$ 0.1	58 $\pm$ 24
Apn1p	APPXL-Y	BER	110 $\pm$ 30	47 $\pm$ 5	430 $\pm$ 130
Apn1p	APPXL-Y	NIR	200 $\pm$ 80	65 $\pm$ 13	330 $\pm$ 150
Apn1p	THF	BER	8.8 $\pm$ 3.0	160 $\pm$ 20	18600 $\pm$ 6700
Apn1p	OHU	NIR	61 $\pm$ 21	38 $\pm$ 4	620 $\pm$ 220
Nfo	APPXL-I	BER	n/s <sup>1</sup>	n/s	5.2 $\pm$ 1.0
Nfo	APPXL-I	NIR	130 $\pm$ 50	6.5 $\pm$ 1.0	50 $\pm$ 20
Nfo	APPXL-Y	BER	n/s	n/s	200 $\pm$ 30
Nfo	APPXL-Y	NIR	61 $\pm$ 17	260 $\pm$ 20	4200 $\pm$ 1200
Nfo	THF	BER	150 $\pm$ 60.0	66 $\pm$ 12	430 $\pm$ 180
Nfo	OHU	NIR	140 $\pm$ 40.0	46 $\pm$ 5	320 $\pm$ 90
APE1	APPXL-I	BER	n/c <sup>2</sup>	n/c	n/c
APE1	APPXL-I	NIR	n/c	n/c	n/c
APE1	APPXL-Y	BER	n/c	n/c	n/c
APE1	APPXL-Y	NIR	44 $\pm$ 19	1.6 $\pm$ 0.2	37 $\pm$ 16
APE1	THF	BER	21 $\pm$ 7	580 $\pm$ 70	28200 $\pm$ 10000
APE1	OHU	NIR	170 $\pm$ 80	19 $\pm$ 4	110 $\pm$ 50

<sup>1</sup> n/s, not saturated, saturation with substrate could not be achieved;  $k_{cat}/K_M$  determined from the initial phase of reaction velocity vs substrate concentration plots.

<sup>2</sup> n/c, not cleaved, cleavage absent or too low to reliably estimate kinetic constants.

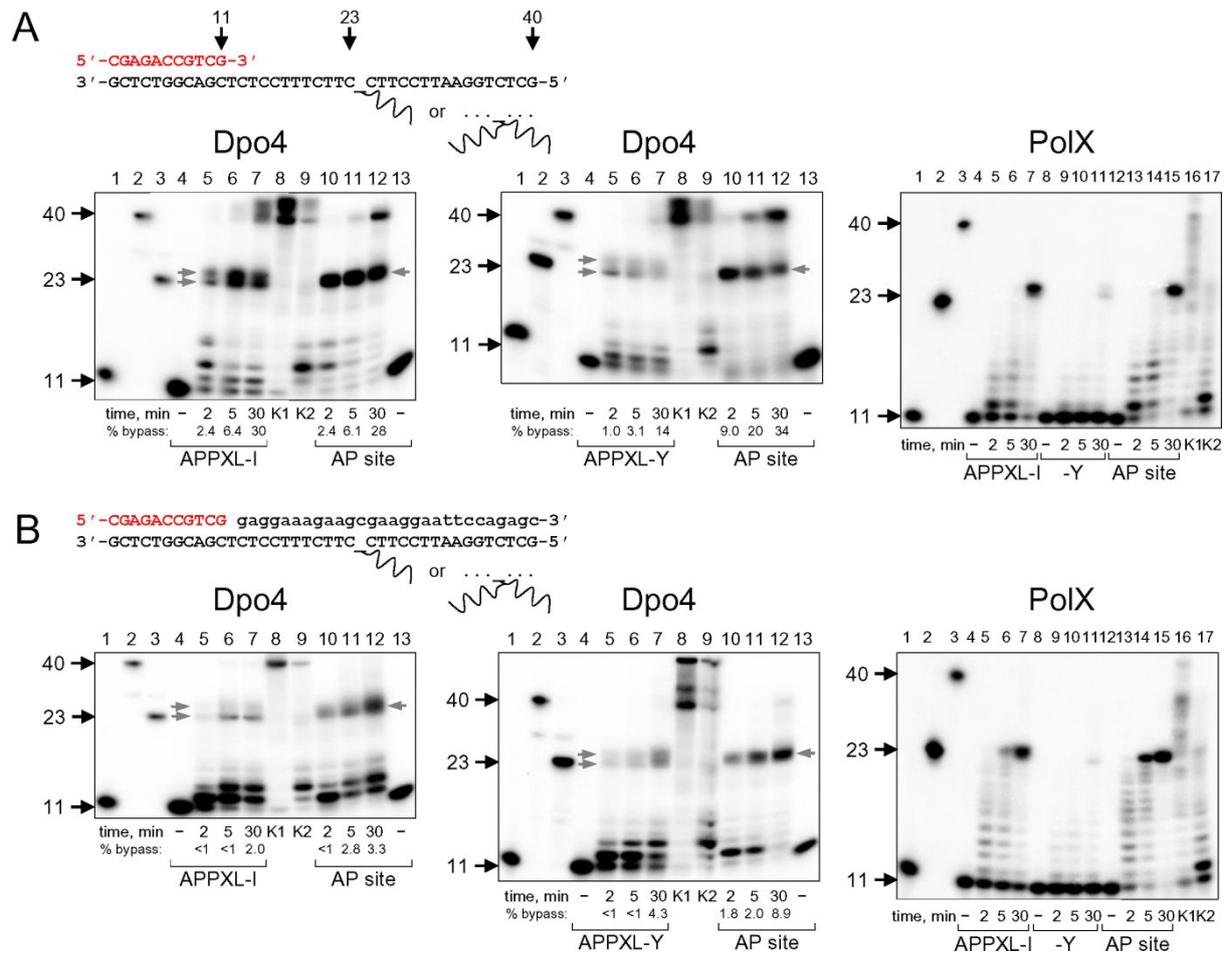


**Figure 1.** Scheme of the DNA–peptide cross-link preparation. 8-OxoG-containing DNA duplexes were treated with bifunctional DNA glycosylases Fpg (top) or OGG1 (bottom) to produce a Schiff base conjugate with the nascent aldehydic AP site through a terminal amino group in case of Fpg or an internal amino group in case of OGG1. The enzyme–DNA intermediates were stabilized with  $\text{NaBH}_4$  to yield covalent DNA–protein cross-links (DPXLs). Then, DPXL were subjected to full trypsinolysis to obtain AP-peptide cross-links.

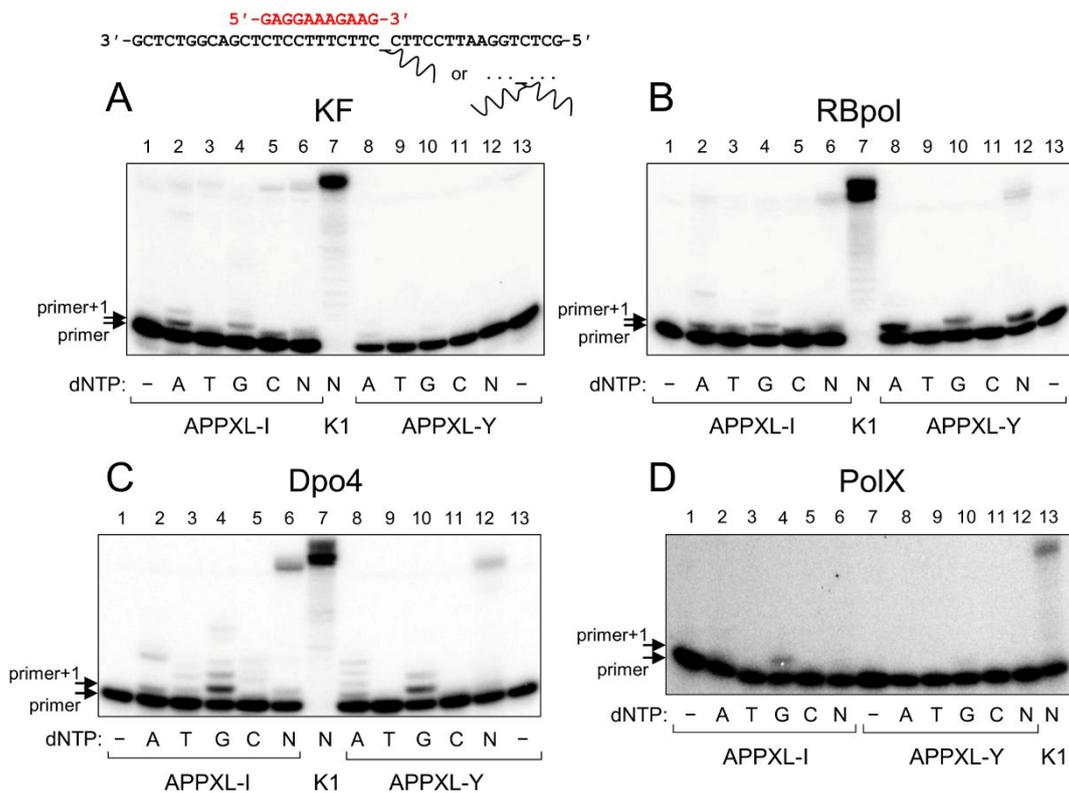


**Figure 2.** Stalling of KF and RBpol by AP-peptide cross-links. A, APPXL-I, primer-template; B, APPXL-Y, primer-template; C, APPXL-I, primer-downstream strand-template; D, APPXL-Y, primer-downstream strand-template. The  $^{32}\text{P}$ -labeled primer is shown in red, the downstream strand is in lowercase. Grey arrows mark the sites of synthesis termination at the lesion. Reaction time and the nature of the lesion are indicated under the gel images. In all panels, lanes 1–3 are size markers

1  
2  
3 corresponding to the primer (11 nt), full-size product (40 nt), and primer extended to the cross-link site  
4 (23 nt). "-", no enzyme added; K1, undamaged primer–template, 30 min; K2, undamaged primer–  
5 template–downstream strand, 30 min. The inset shows KF pause sites at a lower band intensity.  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

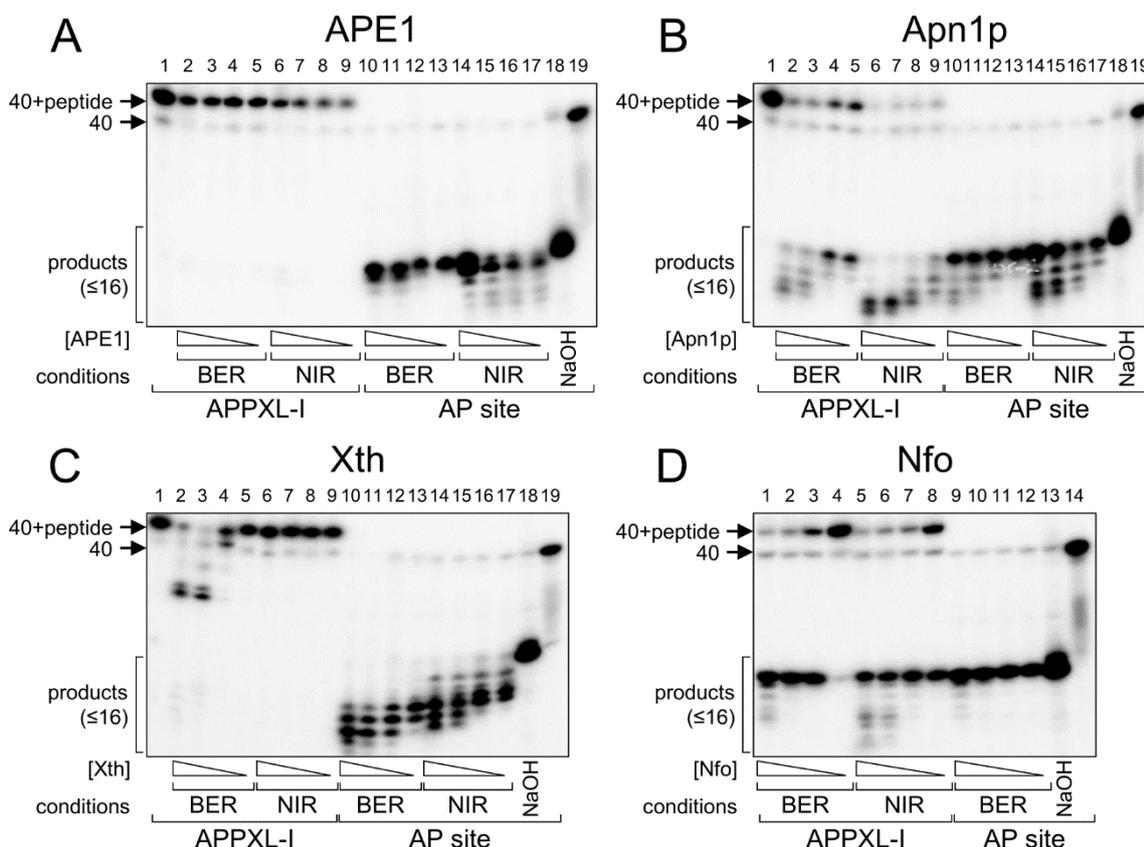


**Figure 3.** Stalling of Dpo4 and PolX by AP-peptide cross-links. (A), primer-template; (B), primer-downstream strand-template. The nature of the polymerase and the cross-link is indicated over and below the gel images, respectively. The  $^{32}\text{P}$ -labeled primer is shown in red, the downstream strand is in lowercase. Grey arrows mark the sites of synthesis termination at the lesion. Reaction time and the nature of the lesion are indicated under the gel images. In all panels, lanes 1–3 are size markers corresponding to the primer (11 nt), full-size product (40 nt), and primer extended to the cross-link site (23 nt). “–”, no enzyme added; K1, undamaged primer-template, 30 min; K2, undamaged primer-template-downstream strand, 30 min. No bypass was observed for PolX.

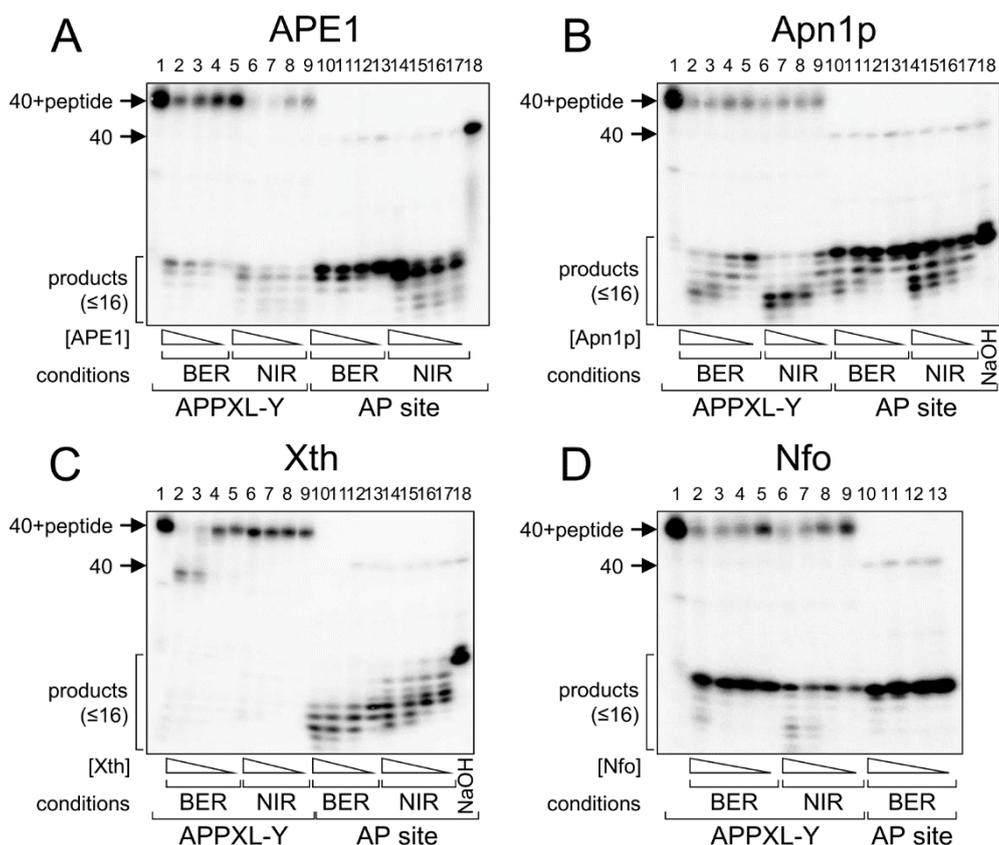


29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Figure 4.** Incorporation of dNMPs opposite APPXLs by KF (A), RBpol (B), Dpo4 (C) and PolX (D). The nature of the substrate and dNTP is indicated below the gel images. The <sup>32</sup>P-labeled primer is shown in red. "-", no enzyme or dNTPs added; N, all dNTPs added; K1, undamaged primer-template, all dNTPs, 30 min.

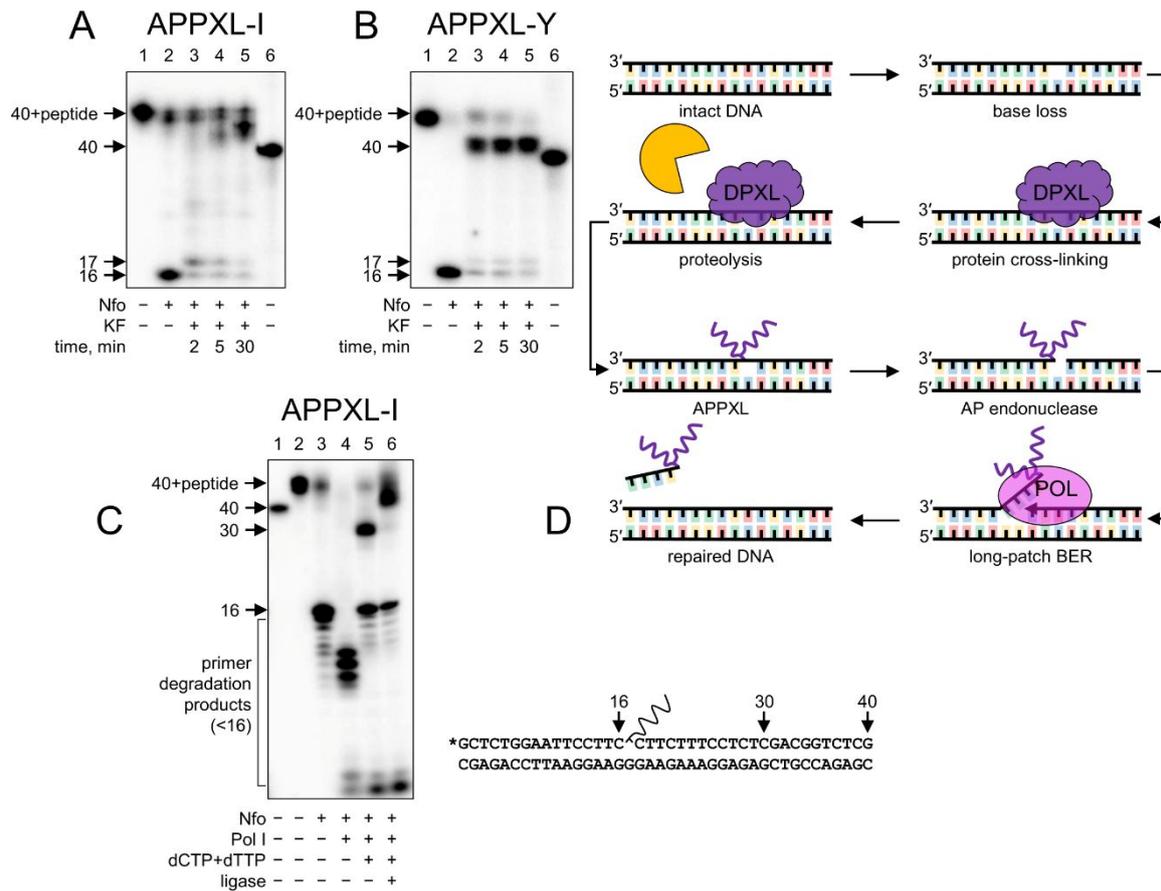


**Figure 5.** Cleavage of APPXL-I by AP endonucleases: APE1 (A), Apn1p (B), Xth (C), and Nfo (D). For panels A–C: lane 1, APPXL-I, no enzyme; lanes 2–5, APPXL-I cleavage under the BER conditions; lanes 6–9, APPXL-I cleavage under the NIR conditions; lanes 10–13, AP site cleavage under the BER conditions; lanes 14–17, AP site cleavage under the NIR conditions; lane 18, AP site treated with NaOH; lane 19, AP site, no enzyme. Panel D: lanes 1–4, APPXL-I cleavage under the BER conditions; lanes 5–8, APPXL-I cleavage under the NIR conditions; lanes 9–12, AP site cleavage under the BER conditions; lane 13, AP site treated with NaOH; lane 14, AP site, no enzyme. BER and NIR conditions for the individual AP endonucleases are specified in Supplementary Table S2.



**Figure 6.** Cleavage of APPXL-Y by AP endonucleases: APE1 (A), Apn1p (B), Xth (C), and Nfo (D).

For panels A–C: lane 1, APPXL-Y, no enzyme; lanes 2–5, APPXL-Y cleavage under the BER conditions; lanes 6–9, APPXL-Y cleavage under the NIR conditions; lanes 10–13, AP site cleavage under the BER conditions; lanes 14–17, AP site cleavage under the NIR conditions; lane 18, AP site untreated (A) or treated with NaOH (B, C). Panel D: lane 1, APPXL-Y, no enzyme; lanes 2–5, APPXL-Y cleavage under the BER conditions; lanes 6–9, APPXL-Y cleavage under the NIR conditions; lanes 10–13, AP site cleavage under the BER conditions. BER and NIR conditions for the individual AP endonucleases are specified in Supplementary Table S2.



**Figure 7.** Base excision repair continuation after the cleavage of AP site-peptide cross-links. A, B, Extension by KF of the primer with the displacement of the downstream strand bearing a 5'-terminal cross-link after Nfo cleavage of APPXL-I (A) and APPXL-Y (B). Lane 1, cross-linked strand; lane 2, APPXL treated with Nfo; lanes 3–5, APPXL treated with Nfo followed by KF extension for the indicated time; lane 6, size marker corresponding to the unmodified strand. C, Reconstitution of the full BER cycle of APPXL-I with Nfo, Pol I, and DNA ligase. Lane 1 and 2, unmodified and cross-linked strands, respectively; lane 3, substrate cleaved by Nfo; lane 4, substrate cleaved by Nfo in the presence of Pol I without dNTPs; lane 5, substrate cleaved by Nfo and extended by Pol I; lane 6, substrate cleaved by Nfo, extended by Pol I and ligated. D General scheme of possible BER-dependent repair of AP site-peptide cross-links.

## Abasic site–peptide cross-links are blocking lesions repaired by AP endonucleases

Anna V. Yudkina<sup>1,2,\*</sup>, Nikita A. Bulgakov<sup>1</sup>, **Daria V. Kim<sup>1,2</sup>**, Svetlana V. Baranova<sup>1</sup>, Alexander A. Ishchenko<sup>3</sup>, Murat K. Saparbaev<sup>3</sup>, Vladimir V. Koval<sup>1</sup> and Dmitry O. Zharkov<sup>1,2,\*</sup>

<sup>1</sup> SB RAS Institute of Chemical Biology and Fundamental Medicine, Novosibirsk, 630090, Russia

<sup>2</sup> Department of Natural Sciences, Novosibirsk State University, Novosibirsk, 630090, Russia

<sup>3</sup> Groupe “Mechanisms of DNA Repair and Carcinogenesis”, Equipe Labellisée LIGUE 2016, CNRS UMR9019, Université Paris-Saclay, Gustave Roussy Cancer Campus, F-94805 Villejuif, France

\* To whom correspondence should be addressed. Tel: +7 383 363 5187; Fax: +7 383 363 5128; Email: dzharkov@niboch.nsc.ru. Correspondence may also be addressed to Anna Yudkina. Tel: +7 383 363 5188; Fax: +7 383 363 5128; Email: ayudkina@niboch.nsc.ru.

### KEY POINTS

- Abasic (AP) site–peptide cross-links arise in the proteolytic repair of AP site–protein cross-links
- AP site–peptide cross-links are blocking and follow the A-rule or induce primer misalignment
- AP site–peptide cross-links can be repaired by the TIM barrel superfamily AP endonucleases

### ABSTRACT

Apurinic/aprimidinic (AP) sites are abundant DNA lesions arising from spontaneous hydrolysis of the N-glycosidic bond and as base excision repair (BER) intermediates. AP sites and their derivatives readily trap DNA-bound proteins, resulting in DNA–protein cross-links. Those are subject to proteolysis but the fate of the resulting AP–peptide cross-links (APPXLs) is unclear. Here we report two *in vitro* models of APPXLs **synthesized** by cross-linking of DNA glycosylases Fpg and OGG1 to DNA followed by trypsinolysis. The reaction with Fpg produces a 10-mer peptide cross-linked through its N-terminus, while OGG1 yields a 23-mer peptide attached through an internal lysine. Both adducts strongly blocked Klenow fragment, phage RB69 polymerase, ***Saccharolobus solfataricus* Dpo4, and African swine fever virus PolX. In the residual lesion bypass, mostly dAMP and dGMP were incorporated by Klenow and RB69 polymerases, while Dpo4 and PolX used primer/template misalignment.** Of AP endonucleases involved in BER, *E. coli* endonuclease IV and its yeast homolog Apn1p efficiently hydrolyzed both adducts. In contrast, *E. coli* exonuclease III and human APE1 showed little activity on APPXL substrates. Our data suggest that APPXLs produced by proteolysis of AP site-trapped proteins may be removed by the BER pathway, at least in bacterial and yeast cells.

## INTRODUCTION

Apurinic/aprimidinic (AP) sites occur by hydrolysis of the *N*-glycosidic bond in deoxyribonucleotides and are among the most abundant spontaneous and induced DNA lesions in the cell (1,2). They are also formed as intermediates in the base excision repair pathway, often being more genotoxic and mutagenic than the original lesion and thus requiring tight control over their processing (3,4).

Chemically, AP sites are a family of lesions, which includes the much-studied aldehydic AP site as well as its derivatives oxidized at C1' (deoxyribolactone, DRL) (5-7), C2' (C2-AP) (8), C4' (C4-AP) or C5' position (5'-(2-phosphoryl-1,4-dioxobutane), DOB) (9,10).

Normally, AP sites are quickly removed from the genome by the base excision DNA repair (BER) system. However, there are important differences in the processing of different types of AP sites. The repair of aldehydic AP sites follows the classical short-patch BER branch, which, in human cells, involves the AP endonuclease (APE1) followed by polymerase and deoxyribosephosphate lyase (dRPase) activities of DNA polymerase  $\beta$  (POL $\beta$ ) (11). At the same time, C2-AP is not a substrate for the dRPase activity of POL $\beta$  and is removed through the long-patch branch of BER: the DNA flap containing the lesion at the 5'-end is displaced during the repair DNA synthesis and hydrolyzed by FEN1 nuclease (8,12). DRL repair is even more problematic: this lesion forms DNA-protein cross-links (DPXLs) with Lys residues in the active sites of DNA glycosylase Nth and POL $\beta$  (13,14), and the  $\beta$ -elimination product of DRL can similarly trap DNA glycosylases Fpg and NEIL1 (14). The presence of a 1,4-dicarbonyl moiety in C4-AP and DOB also promotes covalent capture of POL $\beta$  and DNA polymerase  $\lambda$  (15,16). The irreversible inhibition of the main DNA repair enzymes leads to significant cytotoxicity of modified AP sites (7).

Both aldehydic and oxidized AP sites readily form DPXLs. The covalent capture of many DNA-binding proteins by AP sites has been reported, including histones (6,17,18), integrase (19), DNA ligase (20), Ku antigen (21), as well as RNA-binding proteins (22,23), chaperones (22) and enzymes of carbohydrate metabolism (22,24). Recently, a unique mechanism for the repair of AP sites in single-stranded DNA was discovered (25-27), based on the formation of cross-links between the AP site and the HMCES protein (in human cells) or YedK protein (in *E. coli*), followed by switching to the proteasome-dependent pathway of DPXL repair. **Thus, DPXL with AP sites are highly relevant adducts *in vivo*.**

DPXLs are not limited to cross-links with AP sites and appear through a multitude of mechanisms including reactions with aldehydes and heavy metals, UV and ionizing radiation, and failed action of enzymes such as topoisomerases or C5-methyltransferases (28,29). Thus, a special pathway dedicated to DPXL repair has evolved. It is tightly coupled to replication and involves proteolytic degradation of the protein part of the cross-link (30-33). However, at the moment it remains poorly understood how the last peptide remnant conjugated with DNA is finally removed. It was suggested that this role can be played by nucleotide excision repair (NER), which normally protects the genome from bulky adducts (34-38). Also, Fpg/Nei-like DNA glycosylases are able to excise oxidized purine residues cross-linked to peptides (38). Recombination repair does not appear to be involved in the removal of peptide conjugates from DNA, although it may play a role in tolerance to larger DPXLs (31,36). It should be noted that all these data were obtained on cross-links chemically different from

1  
2  
3 those formed by AP sites, and peptide adducts of this type remain uncharacterized with respect to  
4 their repair and their effect on replication. While the studied peptide adducts with nucleobases usually  
5 retain some base-pairing ability, the bulky nature and non-instructing properties of AP site-peptide  
6 cross-links make these lesions more problematic for the cell. Here we report a model of stable,  
7 chemically determined AP site-peptide cross-links (APPXLs), conjugated through either a terminal or  
8 an internal position in the peptide, and show that they block DNA polymerases and can be repaired by  
9 AP endonucleases.  
10  
11  
12

## 13 MATERIAL AND METHODS

### 14 Oligonucleotides and enzymes

15  
16  
17  
18 *E. coli* formamidopyrimidine–DNA glycosylase (Fpg) (39), endonuclease VIII (Nei) (40) and  
19 endonuclease IV (Nfo) (41), human AP endonuclease (APE1) (42), 8-oxoguanine–DNA glycosylase  
20 (OGG1) (43), endonuclease VIII-like protein 1 (NEIL1) (44) and endonuclease VIII-like protein 2  
21 (NEIL2) (45), yeast AP endonuclease (Apn1p) (41), 3'→5'-exonuclease-deficient (*exo*<sup>-</sup>) Klenow  
22 fragment of *E. coli* DNA polymerase I (KF) (46) and *exo*<sup>-</sup> bacteriophage RB69 DNA polymerase  
23 (RBpol) (47) were overexpressed and purified essentially as described. Cloning and purification of  
24 DNA polymerase X from African swine fever virus (ASFV PolX) is described below. *E. coli*  
25 exonuclease III (Xth) was purchased from Roche (Basel, Switzerland), *E. coli* uracil–DNA glycosylase  
26 (Ung) and exonuclease-proficient KF, from SibEnzyme (Novosibirsk, Russia), full-length *E. coli* DNA  
27 polymerase I (Pol I) and *Saccharolobus (Sulfolobus) solfataricus* DNA polymerase IV (Dpo4), from  
28 New England Biolabs (Ipswich, MA) and trypsin, from Samson-Med (St. Petersburg, Russia). Purity of  
29 the DNA polymerases and AP endonucleases, except for those obtained commercially, is illustrated in  
30 Supplementary Figure S1. Oligonucleotides (Supplementary Table S1) were synthesized in-house  
31 from commercially available phosphoramidites (Glen Research, Sterling, VA).  
32  
33  
34  
35  
36  
37  
38

### 39 ASFV PolX cloning and purification

40  
41  
42 The coding frame of PolX optimized for expression in *E. coli* was synthesized by Gene Universal  
43 (Newark, DE, USA) and confirmed by Sanger sequencing. The insert was subcloned into the pET-24b  
44 vector (Merck, Darmstadt, Germany) at NdeI/XhoI sites. The plasmid was subsequently introduced  
45 into the *E. coli* Rossetta 2(DE3) strain (Merck). One liter of LB medium supplemented with 100 µg/ml  
46 of kanamycin was inoculated with 5 ml of an overnight culture. The cells were grown with vigorous  
47 shaking at 37°C to A<sub>600</sub> = 0.6, isopropyl-β-D-thiogalactopyranoside was added to 1 mM, and the  
48 growth was continued for 4 h at 37°C. The cells were harvested by centrifugation at 12,000×g at 4°C  
49 for 20 min and stored at –72°C. Before the purification, the pellet was thawed on ice in 40 ml of  
50 Buffer A consisting of 20 mM sodium phosphate (pH 7.5), 1 mM ethylenediaminetetraacetic acid  
51 (EDTA), 1 mM dithiothreitol (DTT) and supplemented with 500 mM NaCl and 1 mM  
52 phenylmethylsulfonyl fluoride. The cells were sonicated, and the lysate was clarified by centrifugation  
53 at 12,000×g at 4°C for 30 min. The supernatant was filtered through a 0.45-µm polyvinylidene  
54 difluoride membrane, diluted with four volumes of Buffer A and loaded onto a 5-ml SP Sepharose  
55  
56  
57  
58  
59  
60

1  
2  
3 HiTrap column (GE Healthcare, Chicago, IL) equilibrated in Buffer A plus 100 mM NaCl. The column  
4 was washed with Buffer A supplemented with 100 mM NaCl, and PolX was eluted with a NaCl  
5 gradient at ~650 mM. The fractions containing PolX were pooled, diluted with six volumes of Buffer A  
6 and loaded onto a 5-ml heparine Sepharose HiTrap column (GE Healthcare) equilibrated in Buffer A  
7 plus 100 mM NaCl. The column was washed with the same buffer, and PolX was eluted with a NaCl  
8 gradient at ~730 mM. The fractions containing PolX were pooled, diluted with six volumes of Buffer A  
9 and additionally purified on a 1-ml MonoS column (GE Healthcare) previously equilibrated in Buffer A  
10 plus 100 mM NaCl. PolX was eluted with a NaCl gradient at ~600 mM. The fractions containing >90%  
11 homogeneous protein were pooled, dialyzed against the buffer containing 20 mM sodium phosphate  
12 (pH 7.5), 400 mM NaCl, 1 mM EDTA, 1 mM DTT, and 50% glycerol, and stored at -20°C.  
13  
14  
15  
16  
17

### 18 **Synthesis of model peptide conjugates with AP sites**

19  
20  
21 DNA cross-links to Fpg or OGG1 proteins were obtained as described (48-50). Briefly, ~2 nmol of the  
22 oligonucleotide duplex (40oG//28down or 28oG//21comp; Supplementary Table S1) was treated with  
23 3–6 nmol of the enzyme in 50 mM sodium phosphate (pH 6.8), 1 mM EDTA, 1 mM DTT and 100 mM  
24 NaBH<sub>4</sub> in a total volume of 40 µl. The reaction was allowed to proceed at 37°C for 2 h. The resulting  
25 DPXL was denatured at 95°C for 10 min, cooled, and pH of the solution was adjusted to ~8.0 by  
26 adding Tris base. The mixture was supplemented with 2 µg of trypsin and incubated for 4 h at 37°C.  
27 The oligonucleotide–peptide conjugate was separated by electrophoresis in 20% polyacrylamide/7 M  
28 urea, excised under UV illumination, desalted by reverse-phase chromatography on Isolute C18  
29 sorbent (Biotage, Uppsala, Sweden) and characterized by MALDI MS. If necessary, the  
30 oligonucleotides for the following experiments were 5'-labeled using γ[<sup>32</sup>P]ATP and annealed to a 1.5-  
31 fold molar excess of the complementary strand.  
32  
33  
34  
35  
36

### 37 **DNA polymerase assays**

38  
39 For the standing-start assay, the substrates were assembled from the 40-mer template strand  
40 containing an APPXL and the <sup>32</sup>P-labeled primer (13pri; Supplementary Table S1) or the 28-mer  
41 template strand containing an APPXL and the <sup>32</sup>P-labeled primer (16pri; Supplementary Table S1).  
42 The reaction mixture (10 µl) contained 20 nM pre-annealed substrate, DNA polymerase (100 nM KF  
43 or RBpol, 400 nM PolX, 0.1 U/µl Dpo4), dNTPs (either 200 µM individual dATP, dCTP, dGTP, or  
44 dTTP, or a mixture of all four dNTPs each at 200 µM) in the reaction buffer (50 mM Tris–HCl (pH 7.5),  
45 5 mM MgCl<sub>2</sub> and 1 mM DTT for KF, RBpol and PolX; 20 mM Tris–HCl (pH 8.8), 10 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>,  
46 10 mM KCl, 2 mM MgSO<sub>4</sub>, 0.1% Triton X-100 for Dpo4). The reaction was allowed to proceed at 37°C  
47 for 30 min, quenched with an equal volume of the loading solution (80% formamide, 20 mM Na-  
48 EDTA, 0.1% xylene cyanol, 0.1% bromophenol blue), and heated for 1 min at 95°C. The reaction  
49 products were resolved by 20% denaturing PAGE and visualized on a Typhoon FLA 9500 imager (GE  
50 Healthcare). The bands were quantified using Quantity One v4.6.8 (Bio-Rad Laboratories, Hercules,  
51 CA). For the running-start assay, the substrates were assembled from the 40-mer template strand  
52 containing an APPXL and the <sup>32</sup>P-labeled primer (11pri; Supplementary Table S1). **The substrates**  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 were taken at 50 nM, and the enzymes, at 50 nM (KF or RBpol), 200 nM (PolX) or 0.1 U/μl (Dpo4);  
4 aliquots were taken at 2, 5, and 30 min; otherwise the reactions were run and processed as above.  
5  
6

### 7 **AP endonuclease assay**

8  
9 The substrates were assembled from the <sup>32</sup>P-labeled 40-mer strand containing an APPXL and the  
10 complementary 40-mer strand (40compC; Supplementary Table S1). The reaction mixture (10 μl)  
11 contained 50 nM substrate prepared as above and 10–1000 nM Xth, APE1, Apn1p or Nfo in the  
12 respective buffer optimal for BER or NIR (Supplementary Table S2). As a control, instead of the  
13 APPXL, we used the same oligonucleotide construct with an aldehydic AP site, freshly prepared by  
14 treatment of the U-containing substrate with Ung at 37°C for 20 min. The endonuclease reaction was  
15 allowed to proceed at 37°C for 30 min, treated and analyzed as above.  
16  
17  
18  
19

### 20 **AP endonuclease kinetics**

21  
22 The reaction mixture (10 μl) contained increasing concentrations of the <sup>32</sup>P-labeled 40-mer APPXL  
23 substrates or control substrates 23OHU:23comp or 20THF:20comp (1–500 nM) in the BER or NIR  
24 buffer. Concentrations of AP endonucleases were optimized for each substrate and buffer to produce  
25 ≤20% cleavage and were in the 5 pM – 0.5 nM range. The reaction was allowed to proceed for 10 min  
26 at 37°C, treated and analyzed as above. The data were fitted by nonlinear regression to the  
27 Michaelis–Menten equation using Sigmaplot v11.0 (Systat Software, Chicago, IL). The reported  
28 values are averages of 3–5 independent experiments.  
29  
30  
31  
32

### 33 **AP endonuclease / DNA polymerase assay**

34  
35 The reaction mixture (10 μl) contained 200 nM <sup>32</sup>P-labeled duplex 40-mer APPXL substrate and 50–  
36 150 nM Nfo in the BER buffer (Supplementary Table S2). After 10 min, the reaction mixture was  
37 supplemented with dNTPs (200 μM each), 5 mM MgCl<sub>2</sub>, and 50–250 nM KF. The reaction was  
38 allowed to proceed at 37°C for 2, 5, or 30 min, treated and analyzed as above.  
39  
40  
41

### 42 **DNA glycosylase assay**

43  
44 The reaction mixture (10 μl) contained 10 nM labeled duplex substrate (see above), 1 μM DNA  
45 glycosylase (*E. coli* Fpg or Nei, human OGG1, NEIL1, or NEIL2), 50 mM Tris–HCl (pH 7.5), 100 mM  
46 NaCl, 1 mM EDTA and 1 mM DTT. The reaction was allowed to proceed at 37°C for 30 min, treated  
47 and analyzed as above.  
48  
49  
50

## 51 **RESULTS**

### 52 **Synthesis and characterization of DNA–peptide cross-links with AP site**

53  
54 Since AP sites can form cross-links to a variety of proteins, which can be further processed to  
55 APPXLs, it is usually assumed that the repair of such adducts does not depend on the exact nature of  
56 the cross-linked peptide. Hence, to obtain model APPXLs, we made use of two enzymes, *E. coli* Fpg  
57 and human OGG1 DNA glycosylases, that form transient covalent intermediates with AP sites during  
58  
59  
60

1  
2  
3 the catalysis. To excise the damaged base (8-oxoguanine; 8-oxoG) from DNA, Fpg and OGG1 carry  
4 out a nucleophilic attack by an amino group at the C1' atom of the target nucleotide forming a labile  
5 Schiff base, which may be nearly quantitatively reduced by NaBH<sub>4</sub> or NaBH<sub>3</sub>CN to a stable conjugate  
6 (Figure 1). This reaction was successfully used to crystallize covalent Fpg–DNA and OGG1–DNA  
7 complexes (39,51) and to study the interaction of full-length DPXLs with the replication and repair  
8 machinery (36,48-50). Thus, we first prepared DPXLs by treatment of 8-oxoG-containing duplexes  
9 with Fpg or OGG1 in the presence of NaBH<sub>4</sub> and subjected them to trypsinolysis (Supplementary  
10 Figures S2A, S2B). Fpg and OGG1 belong to different structural superfamilies and use different  
11 nucleophilic residues for catalysis: N-terminal Pro1 in Fpg and Lys249 in OGG1. Hence, after  
12 trypsinolysis, the resulting peptide conjugates are expected to have different structures (Figure 1).  
13 The Fpg–DNA conjugate gives an adduct with the decapeptide (P)ELPEVETSR (the residue that  
14 forms the cross-link is in parentheses), in which the peptide forms a bond with DNA through its N-  
15 terminus. On the other hand, trypsinolysis of the OGG1 conjugate should give an adduct with the 23-  
16 mer peptide ALCILPGVGT(K)VADCICLMALDK, cross-linked through an internal position of the  
17 peptide. Based on the visual resemblance (Figure 1), we will further call these AP site–peptide cross-  
18 links APPXL-I (terminally adducted, linear peptide) and APPXL-Y (an adduct through an internal  
19 residue with two peptide branches).

20  
21 APPXL-I and APPXL-Y were purified by denaturing polyacrylamide gel electrophoresis followed by  
22 reverse-phase chromatography and characterized by MALDI-TOF mass spectrometry  
23 (Supplementary Figures S2C, S2D). The yield of the homogeneous stable APPXLs was ~5–15% of  
24 the initial 8-oxoG-containing oligonucleotide. In addition to the cross-link with the full-length modified  
25 strand, we have observed some products that likely corresponded to products of strand cleavage by  
26 Fpg/OGG1, which were easily separated from the major adduct by electrophoresis (Supplementary  
27 Figures S2A, S2B). Despite the long oligonucleotide parts, of the cross-links, their masses were better  
28 resolved by MALDI mass spectrometry in a positive mode and were within a few Da of the expected  
29 values. No peaks of the starting material (expected [M+H<sup>+</sup>], 12138.9 Da) was evident, indicating good  
30 separation of the APPXLs by electrophoresis. In the APPXL-I mass spectrum, we observed two peaks  
31 corresponding to the expected molecular weight of the adduct of the 40-mer oligonucleotide with the  
32 (P)ELPEVETSR peptide (observed [M+H<sup>+</sup>], 13131.0 Da, expected, 13127.9 Da) and  
33 (P)ELPEVETSR peptide (observed [M+H<sup>+</sup>], 13285.8 Da, expected, 13284.0 Da), most likely  
34 resulting from incomplete trypsin cleavage at two adjacent Arg residues (Supplementary Figure S2C).  
35 As heterogeneity of cross-linked peptides may be naturally expected from proteolysis of DPXLs in the  
36 cell, we have proceeded to characterize the biochemical properties of our adducts.

### 51 AP site–peptide cross-links block DNA polymerases

52  
53 To investigate the ability of DNA polymerases to bypass APPXLs, we annealed the cross-linked  
54 template oligonucleotide with a labeled primer with its 3'-terminus placed 12 bases upstream of the  
55 cross-link site ("running start" conditions, see the substrate schemes in Figures 2A, 2B, 3A). This  
56 distance exceeds the footprints of the full-length proteins from which the cross-linked peptides were  
57 derived, and thus the peptide was not expected to prevent DNA polymerase binding. We also tested  
58  
59  
60

1  
2  
3 the ability of DNA polymerases to carry out synthesis through the APPXL in the context of DNA  
4 duplex when the polymerase has to displace the downstream strand (Figures 2C, 2D, 3B). As  
5 APPXLs are non-instructive, both APPXL-I and APPXL-Y are expected to be strongly blocking.  
6 Therefore, we used a natural freshly prepared AP site as a control to determine which events would  
7 result from DNA polymerases blocking due to the lack of a base or due to the nature of cross-linked  
8 peptide. Non-damaged primer–template or primer–downstream strand–template constructs were  
9 used as polymerase activity controls.  
10  
11  
12

13 Four DNA polymerases, representing a range of structural and functional features, were  
14 investigated. As we were interested in the general ability of DNA polymerases to bypass APPXLs, we  
15 opted to use easily accessible representative members of various structural families originating from  
16 bacteria, archaea, and viruses, rather than human enzymes. The Klenow fragment of *E. coli* DNA  
17 polymerase I (KF) belongs to Family A of DNA polymerases, acting in DNA repair and replication. The  
18 homologous human enzymes are DNA polymerases  $\gamma$ ,  $\theta$ , and  $\nu$ . Bacteriophage RB69 DNA  
19 polymerase (RBpol) is a high-fidelity replicative polymerase from Family B and is often taken as a  
20 convenient model of structurally related human replicative DNA polymerases  $\alpha$ ,  $\delta$ , and  $\epsilon$ , and a  
21 translesion DNA polymerase  $\zeta$ . In both cases, we have used 3'→5'-exonuclease-deficient versions of  
22 the polymerases to prevent primer degradation; exonuclease-proficient KF was also brought in for  
23 comparison. DNA polymerase X (PolX) from African swine fever virus (ASFV) belongs to structural  
24 Family X, mostly engaged in DNA repair and encompassing human DNA polymerases  $\beta$ ,  $\lambda$ , and  $\mu$ .  
25 Finally, DNA polymerase IV from *Saccharolobus* (formerly *Sulfolobus*) *solfataricus* (Dpo4) is a  
26 representative of Family Y translesion polymerases, of which human cells possess DNA  
27 polymerases  $\eta$ ,  $\iota$ , and  $\kappa$ .  
28  
29  
30  
31  
32  
33  
34

35 Under the running start conditions, the blocking properties of APPXL-I were comparable with the  
36 properties of the natural AP site: KF demonstrated slightly better bypass of APPXL-I compared to AP  
37 site in comparison with RBpol, which was slightly more efficient on AP site-containing substrates.  
38 After 30 min, KF was able to fully elongate 6.8% of the primer if APPXL-I was in the template, and  
39 only 4.6% if an AP site was in the template whereas RB69 demonstrated 6.6% bypass of APPXL-I  
40 and 11% of the AP site (Figure 2A). Moreover, two strong pause points were evident one nucleotide  
41 before the cross-linking site and directly opposite the cross-link (23- and 24-nt long products; marked  
42 by grey arrows in Figures 2 and 3), indicating low ability of DNA polymerases both to incorporate a  
43 dNMP opposite APPXL-I and to extend the primer after such incorporation. The absence of any  
44 additional pause points suggests that the bulky peptide does not hamper moving DNA polymerase  
45 access to the cross-link site. The aldehyde AP site produced a single pause point one nucleotide  
46 before the lesion. Exonuclease-proficient KF demonstrated the same behavior as KF  $\text{exo}^-$  except for  
47 noticeable primer degradation at longer reaction times (Supplementary Figure S3). Thus, APPXL-I,  
48 while strongly blocking, still allows better incorporation of a dNMP than the smaller non-instructive AP  
49 site. PolX, which is known to be nearly fully blocked by AP sites (52), demonstrated no bypass of the  
50 APPXL-I as well as the natural AP site, showing a single pause point corresponding to the 23-mer  
51 synthesis product (Figure 3A). Blocking of PolX may be due to its distributive mode of synthesis,  
52 which facilitates DNA release and complicates polymerase binding near the blocking lesion.  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 Translesion DNA polymerase Dpo4, as expected, demonstrated the strongest ability to bypass the AP  
4 site (Dpo4 was able to fully elongate ~28% of the primer) and APPXL-I (~30% of the bypass),  
5 producing the same two pause points in the latter case (Figure 3A). The presence of a downstream  
6 strand complicated the bypass of the blocking lesion, yet the synthesis pause points and the relative  
7 efficiency of lesion bypass seen with the primer–template system were preserved (Figure 2C,  
8 Figure 3B).

9  
10  
11 Strong polymerase inhibition and two pronounced pause points for KF, RBpol and Dpo4 were also  
12 evident for substrates containing the branched APPXL-Y adduct (Figures 2B, 2D, 3, Supplementary  
13 Figure S3). The bulkier APPXL-Y turned out to be even more blocking than APPXL-I. The pause point  
14 corresponding to the 24-nt product for KF and RBpol was weaker than the corresponding pause for  
15 APPXL-I, and the full-length product decreased significantly for KF and Dpo4.

16  
17 We have also addressed polymerase-blocking ability of AP site–peptide cross-links present in an  
18 internal position of the non-template downstream strand. However, only for PolX was a minor pause  
19 point observed, while KF and RBpol displaced the cross-linked strand without effort, and Dpo4, while  
20 having lower strand displacement activity, did not pause at the cross-link (Supplementary Figure S4).

21  
22 Interestingly, when different substrates were compared, we systematically observed primer  
23 utilization efficiency decreasing in the order AP site > APPXL-I > APPXL-Y. Possibly, the peptide part  
24 may partially interfere with DNA binding by the polymerase even when the primer end is 12 nt away  
25 from the cross-link site, for example, by obstructing non-specific DNA binding and facilitated diffusion  
26 used by polymerases to locate the primer end (53). The only exception was Dpo4 with APPXL-Y in  
27 the downstream strand, in which case the bulky peptide adduct might destabilize the duplex and  
28 improve the poor strand displacement ability of Dpo4.

### 29 30 31 32 33 34 35 36 37 **AP site–peptide cross-links mostly direct A and G incorporation and can induce primer– 38 template misalignment**

39  
40 Since we observed some dNMP insertion opposite APPXLs, we sought to explore the spectra of the  
41 incorporated nucleotides. To do this, we moved the primer to place its 3'-end immediately next to the  
42 site of the lesion and presented the polymerases with the individual dNMPs as well as with their  
43 mixture ("standing start" conditions, Figure 4). Here, KF utilized the APPXL-I adducted template much  
44 better than the APPXL-Y adduct, whereas for RBpol the efficiency was somewhat better for the  
45 APPXL-Y. Notably, in all cases the primer elongation was significantly lower than under the running  
46 start conditions, most likely because the bulky peptide interferes with DNA polymerases' binding to  
47 the primer/template junction. Nevertheless, the polymerase preferences for dNMP incorporation  
48 opposite APPXLs were clear. Both KF and RBpol predominantly incorporated dAMP, and, to a lesser  
49 extent, dGMP (Figures 4A, 4B). We observed the preference of dNMP incorporation in the order  
50 A > G > C > T, which is consistent with the "A rule", the preference pattern for better dAMP  
51 incorporation opposite the AP site (54-56). dGMP incorporation could possibly be explained by  
52 template/primer slippage since the next base after the lesion in the template is C. To address this  
53 possibility, we repeated the experiment with the substrate of another sequence featuring G as the  
54 next base (Supplementary Figure S5), for which template/primer slippage would guide the  
55  
56  
57  
58  
59  
60

1  
2  
3 incorporation of dCMP. However, the same general dNMP incorporation pattern A > G > C > T for KF  
4 and RBpol was observed (Supplementary Figures S5A, S5B), confirming that these preferences  
5 characterize the miscoding properties of the peptide adducts rather than the tendency for template  
6 misalignment. The order of preference of dNTP incorporation by the Family Y Dpo4 was G >> A > C, T  
7 (Figure 4C); however, when the substrate was switched, the preference for dNTP usage completely to  
8 C >> A > G, T (Supplementary Figure S5C), indicating that Dpo4 use +1 nucleotide template  
9 misalignment mechanism rather than the “A rule” during the synthesis on the AP-peptide APPXL-  
10 containing substrates. Dpo4 tended to misincorporate several nucleotides, in agreement with its  
11 reported low fidelity characteristic of Family Y DNA polymerases (57). PolX also bypassed the peptide  
12 adduct by +1 slippage mechanism, albeit with a low efficiency: the enzyme preferred G > A, T, C on the  
13 substrate presenting C as the +1 template nucleotide while changing the specificity to C > A >> T, G on  
14 the substrate presenting G at this position (Figure 4D, Supplementary Figure 5D).

### 21 AP site–peptide cross-links can be processed by AP endonucleases

22  
23 Aldehydic AP sites and many of their natural and synthetic modifications are removed from DNA  
24 through the BER pathway initiated by AP endonucleases. These enzymes belong to two major  
25 unrelated structural groups: the exonuclease–endonuclease–phosphatase (EEP) superfamily and the  
26 triosephosphate isomerase (TIM) barrel superfamily. Therefore, we have tested representative  
27 members of each group – *E. coli* exonuclease III (Xth) and the main human AP endo/exonuclease 1  
28 (APE1) from the EEP superfamily and *E. coli* endonuclease IV (Nfo) and yeast endonuclease Apn1p  
29 from the TIM barrel superfamily – for their ability to cleave oligonucleotide substrates containing  
30 APPXL-I or APPXL-Y. Notably, all these enzymes but Xth, in addition to their AP endonuclease  
31 activity, were shown to possess the “nucleotide incision repair” (NIR) activity that cleaves DNA 5' to  
32 many lesions of different chemical nature including quite bulky ones such as 1,N<sup>6</sup>-etheno-dA, 3,N<sup>4</sup>-  
33 etheno-dC, and 3,N<sup>4</sup>-benzetheno-dC (41,58-62). Since the pH, buffer and ionic composition optimal  
34 for BER and NIR are enzyme-specific, we tested the ability of these AP endonucleases to cleave  
35 APPXL-containing substrates under both conditions (Supplementary Table S2).

36  
37 APE1 efficiently hydrolyzed the natural AP site but did not yield any detectable product of the  
38 APPXL-I substrate cleavage (Figure 5A). However, the APPXL-Y substrate was modestly hydrolyzed  
39 by APE1, especially under NIR conditions (Figure 6A). This result was rather surprising, since the  
40 bulkier APPXL-Y adduct was expected to have larger inhibitory effect, and suggested that the  
41 chemical nature of the peptide adduct may be important for its recognition by AP endonucleases.  
42 Another member of the EEP superfamily, Xth, also showed low activity on substrates of both types  
43 (Figure 5C, 6C). The degradation product observed at the highest enzyme concentrations migrates  
44 slower than the expected APPXL cleavage product and is likely due to the robust 3'→5'-exonuclease  
45 activity of Xth evident on the AP substrate. In contrast to APE1 and Xth, both Nfo and Apn1p  
46 efficiently processed the substrates containing APPXL-I or APPXL-Y adducts, cleaving the DNA at the  
47 cross-link site and showing some additional exonucleolytic degradation (Figure 5B, 5D, 6B, 6D). The  
48 cleavage was evident under either BER or NIR conditions. Thus, both TIM barrel AP endonucleases  
49 cleaved APPXL-containing DNA significantly better than did EEP superfamily AP endonucleases.  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

To compare the efficiency of AP endonucleases processing APPXLs with their established substrates in a quantitative way, we have measured the steady-state kinetic parameters ( $K_M$ ,  $k_{cat}$ , and the catalytic efficiency  $k_{cat}/K_M$ ) values using duplex oligonucleotides containing APPXL-I, APPXL-Y, (3-hydroxytetrahydrofuran-2-yl)methyl phosphate (THF, an AP site analog resistant to  $\beta$ -elimination and often used as a model AP site in BER studies) or 5-hydroxy-2'-deoxyuridine (OHU, a typical NIR substrate) (61,62). As shown in Table 1 and Supplementary Figure S6, the apparent  $k_{cat}/K_M$  values confirm APPXL-Y is a better substrate compared to APPXL-I for all studied AP endonucleases. For the TIM barrel enzymes, APPXL-Y cleavage activity was comparable with (~0.5-fold for Apn1p) or even significantly higher than (~13-fold for Nfo) the OHU cleavage, strongly suggesting that these AP endonucleases may be biologically relevant for processing of AP site-peptide adducts. Moreover, even under BER conditions the efficiency of APPXL-Y cleavage by Nfo was comparable with THF cleavage (THF favored only ~2-fold). More efficient processing of APPXL-Y than of APPXL-I may also be of biological importance, as peptides covalently bound to the AP site through an internal residue are statistically more likely to arise in the cell after protein conjugation followed by proteolysis.

We also tested the ability of several bifunctional DNA glycosylases capable of AP site cleavage, namely *E. coli* Fpg and Nei, human OGG1, NEIL1 and NEIL2, to cleave the APPXL adducts. However, none of them was able to process cross-links of either type to any significant degree (Supplementary Figure S7).

### **Strand with a hanging 5'-terminal peptide cross-link can be displaced by DNA polymerases**

AP endonucleases nick DNA 5' of the lesion producing a hanging 5'-terminal 2'-deoxyribose phosphate (or another lesion remnant if the substrate was something else than an aldehydic AP site) and a free 3'-OH, which then serves as a primer for a DNA polymerase. To see whether the nick with a 5'-terminal peptide adduct can be processed in the same way, we have first treated the APPXL-I or APPXL-Y substrate with Nfo and then added KF and all four dNTPs (Figure 7A, 7B). Although in both cases a pause was observed after the incorporation of one nucleotide, eventually the full-length primer extension product was synthesized. Thus, even if the peptide part of the cross-link interferes with polymerase binding or the early steps of primer extension, once the flap is >1 nt there is apparently no obstacles for further strand displacement.

Furthermore, we have reconstituted the full BER cycle with APPXL-I (annealed opposite to G; 40compG template strand), Nfo, full-length *E. coli* DNA polymerase I, and DNA ligase. Unlike KF, Pol I degrades the strand ahead by a combined exo/endonucleolytic action as it extends the primer (63). To prevent full degradation of the downstream strand, we have omitted dATP and dGTP from the reaction mixture, which should cause the polymerase to stop after incorporating 14 nt, yielding a 30-mer product (Figure 7C). Nfo cleaved the duplex at the adduct, producing the 16-mer with the 3'-OH end, which served as an entry point for Pol I (Figure 7C). In the absence of dNTPs, the 3'→5' exonuclease activity of Pol I degraded the primer, whereas addition of dCTP and dTTP allowed primer extension to the position of the first pyrimidine in the template. Finally, if DNA ligase was present, the product had a higher electrophoretic mobility than the DNA-peptide adduct, migrating

1  
2  
3 approximately at the same level as an undamaged 40-mer (Figure 7C), which clearly indicated that  
4 the repair took place.  
5

6 These observations support the possibility of APPXL removal via the long-patch BER subpathway  
7 (Figure 7D). Obviously, short-patch repair is also possible if a non-stabilized 5'-terminal conjugate can  
8 be processed by a dRP lyase.  
9

## 10 11 **DISCUSSION**

12  
13 For many years, AP sites were considered a fairly well studied type of DNA damage, but recent  
14 discoveries regarding new features of their enzymatic processing and biologically relevant reactions  
15 of AP sites with non-specific DNA-bound proteins (18,25-27) put these lesions in an unexpected  
16 perspective. There is mounting evidence that native and oxidized AP sites in the cell can covalently  
17 capture proteins to form highly toxic DNA-protein cross-links. Several breakthrough works in recent  
18 years have outlined the general repair scheme for DNA damage of this type. DPXLs mainly serve as  
19 substrates for a specialized repair pathway, which is initiated by proteolytic degradation of the protein  
20 part of the cross-link by the proteasome or specialized replication-associated proteases (SPRTN1,  
21 FAM111A, DDI1, or others yet to be discovered) to a small peptide (30-33). What happens after that,  
22 however, remains poorly understood. The studied models of naturally occurring DNA-peptide cross-  
23 links include peptide conjugates with exocyclic amino groups of G and A modified through (i) a  $\gamma$ -  
24 hydroxypropano bridge (a product of acrolein-induced DNA damage) (35,64-66) or (ii) through a C2  
25 linker derived from alkylation of Cys residues by 1,2-dibromoethane (67), and (iii) conjugates at the  
26 aldehyde group of 5-formylcytosine (68-70). In addition, unnatural adducts that are readily obtained  
27 synthetically, such as products of azide-alkyl addition to the functionalized C5 position of thymine (71)  
28 and cross-links to 7-deaza-7-(2-oxoethyl)guanine (72), were used as model systems.  
29  
30  
31  
32  
33  
34  
35  
36

37 The major difference between peptide cross-links through nucleobases and through AP sites is the  
38 inability of the latter to form complementary bonds. Until now, there have been no reports in the  
39 literature regarding the levels of APPXLs, their effect on DNA polymerases or the ways to repair such  
40 residual damage. However, recent data on the reactivity of aldehyde and oxidized AP sites suggest  
41 that the number of cross-links with the AP site can be comparable to the number of cross-links with  
42 bases, or even exceed them. Moreover, some conjugates with oxidized AP sites are much more  
43 stable than reversible imine conjugates with aldehydic AP sites. In this work, we developed a model to  
44 generate cross-links of AP sites with peptides of different structure based on the reduction of the  
45 Schiff base intermediate formed by DNA glycosylases Fpg and OGG1 followed by trypsinolysis. Fpg  
46 carries an N-terminal nucleophilic amino group and, after proteolysis, leaves behind a decapeptide  
47 attached via its end (APPXL-I), while OGG1 uses the internal Lys249 for catalysis, and upon  
48 trypsinolysis yields a longer peptide remnant, in which the cross-link is formed at an internal position  
49 (APPXL-Y). This peptide adduct may be of greater biological relevance because DPXLs in the cell are  
50 statistically much more likely to be attached within the polypeptide chain, and protease cleavage  
51 immediately next to the cross-link point may be hindered by DNA. Although the diversity of proteins  
52 capable of cross-linking to DNA dictates the universality of the mechanisms for APPXL repair, the size  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 and properties of the peptide remnant might be expected to affect the APPXL interactions with DNA  
4 polymerases and the repair machinery.  
5

6 A key aspect of APPXL biochemistry determining their biological consequences is the ability to  
7 block DNA polymerases or guide incorporation of particular dNMPs. Although no single DNA–peptide  
8 adduct was studied with a fully representative set of DNA polymerases, *in vitro* experiments with  
9 peptides conjugated through nucleobases show that bypass of such lesions depends both on their  
10 size and the preservation of complementarity. DNA polymerases bind DNA in the minor groove and  
11 leave the major groove mostly accessible, so bulky moieties occupying the major groove may be  
12 better tolerated. For example, 5-formylcytosine adducts or peptides attached to the C5 position of dU  
13 through click chemistry, which protrude to the major groove and do not interfere with base pairing, are  
14 quite easily bypassed by both replicative and translesion human DNA polymerases, although longer  
15 peptides may present a problem (e. g., 10-mer peptides were bypassed, while 23-mers were not)  
16 (70,71,73). In another report, even a large 31-mer peptide conjugated to 5-formylcytosine was  
17 bypassed by translesion DNA polymerases  $\kappa$  and  $\eta$  (68). Similarly, conjugates through position C7 of  
18 7-deazaG are bypassed by translesion polymerases in a size-dependent manner (74). Peptides  
19 adducted through the exocyclic amino group of A, which may either eliminate or keep base pairing  
20 depending on the orientation of the peptide, completely block *E. coli* replicative DNA polymerase III  
21 but demonstrate low-to-full bypass by *E. coli* DNA polymerases I, II, IV and V, human translesion DNA  
22 polymerases  $\eta$ ,  $\iota$ ,  $\kappa$ , and  $\nu$ , and replicative yeast DNA polymerase  $\delta$  (65-67). On the other hand,  
23 peptide adducts with the exocyclic group of G, which either disrupt base pairing or protrude into the  
24 minor groove, are strongly blocking for most polymerases and were efficiently bypassed only by  
25 Family Y translesion DNA polymerases (*E. coli* Pol IV and human Pol  $\kappa$ ) (64,65). In the cell,  
26 mutagenic bypass of peptide adducts requires translesion DNA polymerases (69,70,75).  
27  
28  
29  
30  
31  
32  
33  
34  
35

36 Bypass of natural AP sites and their modified varieties is often described by the “A-rule”, which  
37 states that dAMP insertion is preferred opposite a non-instructive lesion. Interestingly, the structural  
38 reasons underlying this preference are different between DNA polymerases. For example, the  
39 structure of the Stoffel fragment of Taq DNA polymerase (which corresponds to KF in the *E. coli*  
40 enzyme) shows templating by a conservative Tyr residue inserted in place of a template nucleotide  
41 and acting as a thymine mimic (76,77). RBpol, on the other hand, relies on the favorable stacking  
42 between the incoming dATP and the extended planar system formed by the terminal primer  
43 nucleotide and its pairing partner in the absence of the template nucleobase (78,79). However, in the  
44 structures of both polymerases with AP-DNA, C1' of the AP site residues is buried at the protein–DNA  
45 interface, and the template strand would require extensive conformational rearrangement for a bypass  
46 if a large peptide is attached to this position. The steric clash could be relieved by flipping the  
47 template nucleotide around its flanking phosphates, as often observed in structures of Families X and  
48 Y DNA polymerases stabilizing a misaligned primer–template junction (80-82). In the “open” state of  
49 the enzyme–DNA complex, Family A DNA polymerases, including KF, also rotate the downstream  
50 template nucleotides outside of the helix, and it has been suggested that this movement helps prevent  
51 the template bases from misalignment in the absence of an incoming dNTP (83). As we do not see  
52 preferential incorporation of a dNMP corresponding to +1 template nucleotide, characteristic of the  
53  
54  
55  
56  
57  
58  
59  
60

1  
2  
3 primer–template misalignment bypass mechanism, we hypothesize that favorable stacking can guide  
4 dAMP and dGMP insertion opposite extrahelical APPXLs in the rare bypass events. **On the other**  
5 **hand, structures of Dpo4 bypassing an abasic site, demonstrate extensive conformational flexibility**  
6 **allowing for primer/template misalignment and generation of mismatches, +1 and –1 frameshifts (84).**  
7 **In the structure with misalignment between the primer end and the +1 position in the template (which**  
8 **was observed in our experiments), the AP site is looped out of the helix into a gap between the**  
9 **fingers and the little finger domains, providing plenty of space for accommodation of a cross-linked**  
10 **peptide. No structure of PolX in the frameshift mode is available but POL $\beta$ , a prototypical member of**  
11 **Family X, can also loop out a template nucleotide to match the primer end to the +1 position in the**  
12 **template (82).**

13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
As we used representative DNA polymerases of non-human origin in our study, it remains to be  
seen which DNA polymerases may participate in APPXL bypass in human cells. Based on the *in vitro*  
insertion specificity data, it has long been held that natural AP sites are bypassed by DNA  
polymerase  $\eta$  (POL $\eta$ ) (85,86). On the other hand, AP site bypass in POL $\eta$ -deficient human cells is  
almost as efficient as in wild-type cells (87). An alternative model was proposed, based on yeast  
mutation spectra and *in vitro* biochemistry of human and yeast DNA polymerase  $\zeta$  (POL $\zeta$ ), that the  
POL $\zeta$ /REV1 complex acting both as an inserter and an extender polymerase provides mutagenic  
bypass of AP sites (reviewed in (88)). However, mutation spectra induced by AP sites in yeast and  
human cells seem to be different (89), and human REV1 also interacts with DNA polymerases  $\eta$ ,  $\kappa$ ,  
and  $\iota$  (88), so the nature of the inserter polymerase in human cells exposed to AP sites or their  
derivatives, including APPXLs, needs further investigation.

The high proficiency of TIM barrel AP endonucleases in nicking DNA at APPXLs was rather  
unexpected, since even the known substrates for the NIR activity of these enzymes are much less  
bulky. However, inspection of the structures of DNA-bound Nfo (90,91) reveals that C1' is solvent-  
accessible in both enzyme–substrate and enzyme–product complexes (Supplementary Figure S8)  
and the attached peptide could possibly be accommodated in the major groove of the DNA molecule,  
which is sharply bent upon Nfo binding. The structure of yeast Apn1p has not been solved but a  
similar C1' exposure is observed in a DNA-bound African swine fever virus AP endonuclease (92),  
another TIM barrel superfamily member. At the same time, the C1' atom of the target nucleotide is  
fully buried in the complex of APE1 with AP site-containing DNA, the nicked AP product, or 3'-terminal  
mismatched dNMPs that are substrates for the exonuclease activity of APE1 (93-95). The structure of  
DNA-bound *E. coli* Xth is not available but in the structurally related AP endonuclease from *Neisseria*  
*meningitidis* the C1' atom is also inaccessible (96). Thus, structural reasons may explain the lack of  
cleavage of the APPXL substrates by Xth and the mediocre cleavage by APE1 despite its NIR activity  
being capable of processing bulky etheno adducts (59,62). The remaining activity of APE1 on APPXL-  
Y may be due to a more pronounced duplex destabilization by the larger peptide, which would make  
easier the access to the phosphate of the damaged nucleotide without fully moving the peptide part  
into the AP site-recognition pocket.

Overall, our results suggest that, at least in bacteria and budding yeast, APPXL may be repaired  
by the BER pathway. After the initial nick by Nfo/Apn1p, the repair could proceed along the long-patch

1  
2  
3 branch displacing or degrading a flap decorated with the 5'-terminal adduct. Alternatively, it is possible  
4 that 2'-deoxyribo-5'-phosphate lyases, such as RecJ or Fpg in *E. coli* (97,98) and Trf4 in yeast (99), or  
5 exonucleases resolving topoisomerase II-type adducts (100,101), can catalyze excision of 5'-terminal  
6 imine-peptide AP adducts, or that such conjugates undergo spontaneous  $\beta$ -elimination (102,103),  
7 allowing the repair to continue through the short-patch subpathway. As for human cells, given that  
8 under NIR conditions APE1 processed APPXL-Y and OHU at about the same order of magnitude,  
9 BER could also contribute, at least partly, to APPXL repair. Possible alternatives include NER  
10 (104,105), homologous recombination (106), or direct adduct uncoupling by HMCES-SRAP proteins  
11 (107). Interestingly, Nfo/Apn1p homologs are present in fish and amphibians, raising a possibility that  
12 BER of APPXL may be evolutionarily conserved and occur even in vertebrates.  
13  
14  
15  
16  
17  
18  
19

## 20 **DATA AVAILABILITY**

21  
22 All data are provided in the paper and the supplementary materials.  
23  
24  
25

## 26 **SUPPLEMENTARY DATA**

27  
28  
29 Supplementary Data are available at NAR online.  
30  
31  
32

## 33 **ACKNOWLEDGEMENT**

34  
35 Mass spectrometry was performed at the SB RAS Institute of Chemical Biology and Fundamental  
36 Medicine Joint Center for Genomic, Proteomic and Metabolomics Studies. We thank Dr. Alexander A.  
37 Chernonosov for helpful discussion.  
38  
39  
40  
41

## 42 **FUNDING**

43  
44 This work was supported by the Russian Science Foundation (grant 21-74-00061 to A.V.Y., all  
45 biochemical experiments) and Russian Ministry of Higher Education and Science (grant  
46 121031300056-8 to D.O.Z.). D.V.K. is supported by a graduate student fellowship from the Russian  
47 Foundation for Basic Research (No. 20-34-90092).  
48  
49  
50  
51

## 52 **CONFLICT OF INTEREST**

53  
54  
55  
56 None declared.  
57  
58  
59  
60

## REFERENCES

1. Lindahl, T. and Nyberg, B. (1972) Rate of depurination of native deoxyribonucleic acid. *Biochemistry*, **11**, 3610-3618.
2. Atamna, H., Cheung, I. and Ames, B.N. (2000) A method for detecting abasic sites in living cells: Age-dependent changes in base excision repair. *Proc. Natl Acad. Sci. U.S.A.*, **97**, 686-691.
3. Boiteux, S. and Guillet, M. (2004) Abasic sites in DNA: Repair and biological consequences in *Saccharomyces cerevisiae*. *DNA Repair*, **3**, 1-12.
4. Thompson, P.S. and Cortez, D. (2020) New insights into abasic site repair and tolerance. *DNA Repair*, **90**, 102866.
5. DeMott, M.S., Beyret, E., Wong, D., Bales, B.C., Hwang, J.-T., Greenberg, M.M. and Demple, B. (2002) Covalent trapping of human DNA polymerase  $\beta$  by the oxidative DNA lesion 2-deoxyribonolactone. *J. Biol. Chem.*, **277**, 7637-7640.
6. Sczepanski, J.T., Wong, R.S., McKnight, J.N., Bowman, G.D. and Greenberg, M.M. (2010) Rapid DNA-protein cross-linking and strand scission by an abasic site in a nucleosome core particle. *Proc. Natl Acad. Sci. U.S.A.*, **107**, 22475-22480.
7. Greenberg, M.M. (2014) Abasic and oxidized abasic site reactivity in DNA: Enzyme inhibition, cross-linking, and nucleosome catalyzed reactions. *Acc. Chem. Res.*, **47**, 646-655.
8. Greenberg, M.M., Weledji, Y.N., Kroeger, K.M. and Kim, J. (2004) In vitro replication and repair of DNA containing a C2'-oxidized abasic site. *Biochemistry*, **43**, 15217-15222.
9. Guan, L. and Greenberg, M.M. (2009) DNA interstrand cross-link formation by the 1,4-dioxobutane abasic lesion. *J. Am. Chem. Soc.*, **131**, 15225-15231.
10. Ghosh, S. and Greenberg, M.M. (2015) Correlation of thermal stability and structural distortion of DNA interstrand cross-links produced from oxidized abasic sites with their selective formation and repair. *Biochemistry*, **54**, 6274-6283.
11. Fortini, P. and Dogliotti, E. (2007) Base damage and single-strand break repair: Mechanisms and functional significance of short- and long-patch repair subpathways. *DNA Repair*, **6**, 398-409.
12. Wong, R.S., Sczepanski, J.T. and Greenberg, M.M. (2010) Excision of a lyase-resistant oxidized abasic lesion from DNA. *Chem. Res. Toxicol.*, **23**, 766-770.
13. Hashimoto, M., Greenberg, M.M., Kow, Y.W., Hwang, J.-T. and Cunningham, R.P. (2001) The 2-deoxyribonolactone lesion produced in DNA by neocarzinostatin and other damaging agents forms cross-links with the base-excision repair enzyme endonuclease III. *J. Am. Chem. Soc.*, **123**, 3161-3162.
14. Kroeger, K.M., Hashimoto, M., Kow, Y.W. and Greenberg, M.M. (2003) Cross-linking of 2-deoxyribonolactone and its  $\beta$ -elimination product by base excision repair enzymes. *Biochemistry*, **42**, 2449-2455.
15. Guan, L. and Greenberg, M.M. (2010) Irreversible inhibition of DNA polymerase  $\beta$  by an oxidized abasic lesion. *J. Am. Chem. Soc.*, **132**, 5004-5005.

16. Guan, L., Bebenek, K., Kunkel, T.A. and Greenberg, M.M. (2010) Inhibition of short patch and long patch base excision repair by an oxidized abasic site. *Biochemistry*, **49**, 9904-9910.
17. Sczepanski, J.T., Zhou, C. and Greenberg, M.M. (2013) Nucleosome core particle-catalyzed strand scission at abasic sites. *Biochemistry*, **52**, 2157-2164.
18. Ren, M., Shang, M., Wang, H., Xi, Z. and Zhou, C. (2021) Histones participate in base excision repair of 8-oxodGuo by transiently cross-linking with active repair intermediates in nucleosome core particles. *Nucleic Acids Res.*, **49**, 257-268.
19. Mazumder, A., Neamati, N., Pilon, A.A., Sunder, S. and Pommier, Y. (1996) Chemical trapping of ternary complexes of human immunodeficiency virus type 1 integrase, divalent metal, and DNA substrates containing an abasic site: Implications for the role of lysine 136 in DNA binding. *J. Biol. Chem.*, **271**, 27330-27338.
20. Bogenhagen, D.F. and Pinz, K.G. (1998) The action of DNA ligase at abasic sites in DNA. *J. Biol. Chem.*, **273**, 7888-7893.
21. Kosova, A.A., Khodyreva, S.N. and Lavrik, O.I. (2016) Ku antigen displays the AP lyase activity on a certain type of duplex DNA. *Biochim. Biophys. Acta*, **1864**, 1244-1252.
22. Rieger, R.A., Zaika, E.I., Xie, W., Johnson, F., Grollman, A.P., Iden, C.R. and Zharkov, D.O. (2006) Proteomic approach to identification of proteins reactive for abasic sites in DNA. *Mol. Cell. Proteomics*, **5**, 858-867.
23. Grosheva, A.S., Zharkov, D.O., Stahl, J., Gopanenko, A.V., Tupikin, A.E., Kabilov, M.R., Graifer, D.M. and Karpova, G.G. (2017) Recognition but no repair of abasic site in single-stranded DNA by human ribosomal uS3 protein residing within intact 40S subunit. *Nucleic Acids Res.*, **45**, 3833-3843.
24. Kosova, A.A., Khodyreva, S.N. and Lavrik, O.I. (2015) Glyceraldehyde-3-phosphate dehydrogenase (GAPDH) interacts with apurinic/apyrimidinic sites in DNA. *Mutat. Res.*, **779**, 46-57.
25. Halabelian, L., Ravichandran, M., Li, Y., Zeng, H., Rao, A., Aravind, L. and Arrowsmith, C.H. (2019) Structural basis of HMCES interactions with abasic DNA and multivalent substrate recognition. *Nat. Struct. Mol. Biol.*, **26**, 607-612.
26. Mohni, K.N., Wessel, S.R., Zhao, R., Wojciechowski, A.C., Luzwick, J.W., Layden, H., Eichman, B.F., Thompson, P.S., Mehta, K.P.M. and Cortez, D. (2019) HMCES maintains genome integrity by shielding abasic sites in single-strand DNA. *Cell*, **176**, 144-153.
27. Thompson, P.S., Amidon, K.M., Mohni, K.N., Cortez, D. and Eichman, B.F. (2019) Protection of abasic sites during DNA replication by a stable thiazolidine protein-DNA cross-link. *Nat. Struct. Mol. Biol.*, **26**, 613-618.
28. Barker, S., Weinfeld, M. and Murray, D. (2005) DNA-protein crosslinks: Their induction, repair, and biological consequences. *Mutat. Res.*, **589**, 111-135.
29. Pommier, Y., Leo, E., Zhang, H. and Marchand, C. (2010) DNA topoisomerases and their poisoning by anticancer and antibacterial drugs. *Chem. Biol.*, **17**, 421-433.
30. Duxin, J.P., Dewar, J.M., Yardimci, H. and Walter, J.C. (2014) Repair of a DNA-protein crosslink by replication-coupled proteolysis. *Cell*, **159**, 346-357.

- 1
- 2
- 3 31. Stinglele, J., Schwarz, M.S., Bloemeke, N., Wolf, P.G. and Jentsch, S. (2014) A DNA-
- 4 dependent protease involved in DNA-protein crosslink repair. *Cell*, **158**, 327-338.
- 5
- 6 32. Vaz, B., Popovic, M., Newman, J.A., Fielden, J., Aitkenhead, H., Halder, S., Singh, A.N.,
- 7 Vendrell, I., Fischer, R., Torrecilla, I. *et al.* (2016) Metalloprotease SPRTN/DVC1 orchestrates
- 8 replication-coupled DNA-protein crosslink repair. *Mol. Cell*, **64**, 704-719.
- 9
- 10 33. Larsen, N.B., Gao, A.O., Sparks, J.L., Gallina, I., Wu, R.A., Mann, M., Räschle, M., Walter,
- 11 J.C. and Duxin, J.P. (2019) Replication-coupled DNA-protein crosslink repair by SPRTN and
- 12 the proteasome in *Xenopus* egg extracts. *Mol. Cell*, **73**, 574-588.e577.
- 13
- 14 34. Quievryn, G. and Zhitkovich, A. (2000) Loss of DNA–protein crosslinks from formaldehyde-
- 15 exposed cells occurs through spontaneous hydrolysis and an active repair process linked to
- 16 proteasome function. *Carcinogenesis*, **21**, 1573-1580.
- 17
- 18 35. Minko, I.G., Kurtz, A.J., Croteau, D.L., Van Houten, B., Harris, T.M. and Lloyd, R.S. (2005)
- 19 Initiation of repair of DNA–polypeptide cross-links by the UvrABC nuclease. *Biochemistry*, **44**,
- 20 3000-3009.
- 21
- 22 36. Nakano, T., Morishita, S., Katafuchi, A., Matsubara, M., Horikawa, Y., Terato, H., Salem,
- 23 A.M.H., Izumi, S., Pack, S.P., Makino, K. *et al.* (2007) Nucleotide excision repair and
- 24 homologous recombination systems commit differentially to the repair of DNA-protein
- 25 crosslinks. *Mol. Cell*, **28**, 147-158.
- 26
- 27 37. de Graaf, B., Clore, A. and McCullough, A.K. (2009) Cellular pathways for DNA repair and
- 28 damage tolerance of formaldehyde-induced DNA-protein crosslinks. *DNA Repair*, **8**, 1207-
- 29 1214.
- 30
- 31 38. McKibbin, P.L., Fleming, A.M., Towheed, M.A., Van Houten, B., Burrows, C.J. and David,
- 32 S.S. (2013) Repair of hydantoin lesions and their amine adducts in DNA by base and
- 33 nucleotide excision repair. *J. Am. Chem. Soc.*, **135**, 13851-13861.
- 34
- 35 39. Gilboa, R., Zharkov, D.O., Golan, G., Fernandes, A.S., Gerchman, S.E., Matz, E., Kycia, J.H.,
- 36 Grollman, A.P. and Shoham, G. (2002) Structure of formamidopyrimidine-DNA glycosylase
- 37 covalently complexed to DNA. *J. Biol. Chem.*, **277**, 19811-19816.
- 38
- 39 40. Rieger, R.A., McTigue, M.M., Kycia, J.H., Gerchman, S.E., Grollman, A.P. and Iden, C.R.
- 40 (2000) Characterization of a cross-linked DNA-endonuclease VIII repair complex by
- 41 electrospray ionization mass spectrometry. *J. Am. Soc. Mass Spectrom.*, **11**, 505-515.
- 42
- 43 41. Ishchenko, A.A., Sanz, G., Privezentzev, C.V., Maksimenko, A.V. and Sapparbaev, M. (2003)
- 44 Characterisation of new substrate specificities of *Escherichia coli* and *Saccharomyces*
- 45 *cerevisiae* AP endonucleases. *Nucleic Acids Res.*, **31**, 6344-6353.
- 46
- 47 42. Sidorenko, V.S., Nevinsky, G.A. and Zharkov, D.O. (2007) Mechanism of interaction between
- 48 human 8-oxoguanine-DNA glycosylase and AP endonuclease. *DNA Repair*, **6**, 317-328.
- 49
- 50 43. Kuznetsov, N.A., Koval, V.V., Zharkov, D.O., Nevinsky, G.A., Douglas, K.T. and Fedorova,
- 51 O.S. (2005) Kinetics of substrate recognition and cleavage by human 8-oxoguanine-DNA
- 52 glycosylase. *Nucleic Acids Res.*, **33**, 3919-3931.
- 53
- 54
- 55
- 56
- 57
- 58
- 59
- 60

- 1  
2  
3 44. Katafuchi, A., Nakano, T., Masaoka, A., Terato, H., Iwai, S., Hanaoka, F. and Ide, H. (2004)  
4 Differential specificity of human and *Escherichia coli* endonuclease III and VIII homologues for  
5 oxidative base lesions. *J. Biol. Chem.*, **279**, 14464-14471.  
6  
7 45. Kakhkharova, Z.I., Zharkov, D.O. and Grin, I.R. (2022) A low-activity polymorphic variant of  
8 human NEIL2 DNA glycosylase. *Int. J. Mol. Sci.*, **23**, 2212.  
9  
10 46. Miller, H. and Grollman, A.P. (1997) Kinetics of DNA polymerase I (Klenow fragment exo<sup>-</sup>)  
11 activity on damaged DNA templates: Effect of proximal and distal template damage on DNA  
12 synthesis. *Biochemistry*, **36**, 15336-15342.  
13  
14 47. Freisinger, E., Grollman, A.P., Miller, H. and Kisker, C. (2004) Lesion (in)tolerance reveals  
15 insights into DNA replication fidelity. *EMBO J.*, **23**, 1494-1505.  
16  
17 48. Yudkina, A.V., Dvornikova, A.P. and Zharkov, D.O. (2018) Variable termination sites of DNA  
18 polymerases encountering a DNA–protein cross-link. *PLoS ONE*, **13**, e0198480.  
19  
20 49. Boldinova, E.O., Yudkina, A.V., Shilkin, E.S., Gagarińska, D.I., Baranovskiy, A.G., Tahirov,  
21 T.H., Zharkov, D.O. and Makarova, A.V. (2021) Translesion activity of PrimPol on DNA with  
22 cisplatin and DNA–protein cross-links. *Sci. Rep.*, **11**, 17588.  
23  
24 50. Yudkina, A.V., Shilkin, E.S., Makarova, A.V. and Zharkov, D.O. (2022) Stalling of eukaryotic  
25 translesion DNA polymerases at DNA-protein cross-links. *Genes*, **13**, 166.  
26  
27 51. Fromme, J.C., Bruner, S.D., Yang, W., Karplus, M. and Verdine, G.L. (2003) Product-assisted  
28 catalysis in base-excision DNA repair. *Nat. Struct. Biol.*, **10**, 204-211.  
29  
30 52. Kumar, S., Lamarche, B.J. and Tsai, M.-D. (2007) Use of damaged DNA and dNTP  
31 substrates by the error-prone DNA polymerase X from african swine fever virus. *Biochemistry*,  
32 **46**, 3814-3825.  
33  
34 53. Howard, M.J. and Wilson, S.H. (2018) DNA scanning by base excision repair enzymes and  
35 implications for pathway coordination. *DNA Repair*, **71**, 101-107.  
36  
37 54. Goodman, M.F., Cai, H., Bloom, L.B. and Eritja, R. (1994) Nucleotide insertion and primer  
38 extension at abasic template sites in different sequence contexts. *Ann. N. Y. Acad. Sci.*, **726**,  
39 132-142.  
40  
41 55. Paz-Elizur, T., Takeshita, M. and Livneh, Z. (1997) Mechanism of bypass synthesis through  
42 an abasic site analog by DNA polymerase I. *Biochemistry*, **36**, 1766-1773.  
43  
44 56. Taylor, J.-S. (2002) New structural and mechanistic insight into the A-rule and the  
45 instructional and non-instructional behavior of DNA photoproducts and other lesions. *Mutat.*  
46 *Res.*, **510**, 55-70.  
47  
48 57. Kokoska, R.J., Bebenek, K., Boudsocq, F., Woodgate, R. and Kunkel, T.A. (2002) Low fidelity  
49 DNA synthesis by a  $\gamma$  family DNA polymerase due to misalignment in the active site. *J. Biol.*  
50 *Chem.*, **277**, 19633-19638.  
51  
52 58. Ide, H., Tedzuka, K., Shimzu, H., Kimura, Y., Purmal, A.A., Wallace, S.S. and Kow, Y.W.  
53 (1994)  $\alpha$ -Deoxyadenosine, a major anoxic radiolysis product of adenine in DNA, is a  
54 substrate for *Escherichia coli* endonuclease IV. *Biochemistry*, **33**, 7842-7847.  
55  
56 59. Hang, B., Chenna, A., Fraenkel-Conrat, H. and Singer, B. (1996) An unusual mechanism for  
57 the major human apurinic/aprimidinic (AP) endonuclease involving 5' cleavage of DNA  
58  
59  
60

- 1  
2  
3 containing a benzene-derived exocyclic adduct in the absence of an AP site. *Proc. Natl Acad. Sci. U.S.A.*, **93**, 13737-13741.
- 4  
5  
6 60. Ischenko, A.A. and Saparbaev, M.K. (2002) Alternative nucleotide incision repair pathway for  
7 oxidative DNA damage. *Nature*, **415**, 183-187.
- 8  
9 61. Gros, L., Ishchenko, A.A., Ide, H., Elder, R.H. and Saparbaev, M.K. (2004) The major human  
10 AP endonuclease (Ape1) is involved in the nucleotide incision repair pathway. *Nucleic Acids*  
11 *Res.*, **32**, 73-81.
- 12  
13 62. Prorok, P., Saint-Pierre, C., Gasparutto, D., Fedorova, O.S., Ishchenko, A.A., Leh, H., Buckle,  
14 M., Tudek, B. and Saparbaev, M. (2012) Highly mutagenic exocyclic DNA adducts are  
15 substrates for the human nucleotide incision repair pathway. *PLoS ONE*, **7**, e51776.
- 16  
17 63. Lyamichev, V., Brow, M.A.D. and Dahlberg, J.E. (1993) Structure-specific endonucleolytic  
18 cleavage of nucleic acids by eubacterial DNA polymerases. *Science*, **260**, 778-783.
- 19  
20 64. Minko, I.G., Yamanaka, K., Kozekov, I.D., Kozekova, A., Indiani, C., O'Donnell, M.E., Jiang,  
21 Q., Goodman, M.F., Rizzo, C.J. and Lloyd, R.S. (2008) Replication bypass of the acrolein-  
22 mediated deoxyguanine DNA-peptide cross-links by DNA polymerases of the DinB family.  
23 *Chem. Res. Toxicol.*, **21**, 1983-1990.
- 24  
25 65. Yamanaka, K., Minko, I.G., Takata, K.-i., Kolbanovskiy, A., Kozekov, I.D., Wood, R.D., Rizzo,  
26 C.J. and Lloyd, R.S. (2010) Novel enzymatic function of DNA polymerase  $\nu$  in translesion  
27 DNA synthesis past major groove DNA-peptide and DNA-DNA cross-links. *Chem. Res.*  
28 *Toxicol.*, **23**, 689-695.
- 29  
30 66. Yamanaka, K., Minko, I.G., Finkel, S.E., Goodman, M.F. and Lloyd, R.S. (2011) Role of high-  
31 fidelity *Escherichia coli* DNA polymerase I in replication bypass of a deoxyadenosine DNA-  
32 peptide cross-link. *J. Bacteriol.*, **193**, 3815-3821.
- 33  
34 67. Ghodke, P.P., Gonzalez-Vasquez, G., Wang, H., Johnson, K.M., Sedgeman, C.A. and  
35 Guengerich, F.P. (2021) Enzymatic bypass of an  $N^6$ -deoxyadenosine DNA-ethylene  
36 dibromide-peptide crosslink by translesion DNA polymerases. *J. Biol. Chem.*, **296**, 100444.
- 37  
38 68. Ji, S., Fu, I., Naldiga, S., Shao, H., Basu, A.K., Broyde, S. and Tretyakova, N.Y. (2018) 5-  
39 Formylcytosine mediated DNA-protein cross-links block DNA replication and induce  
40 mutations in human cells. *Nucleic Acids Res.*, **46**, 6455-6469.
- 41  
42 69. Ji, S., Park, D., Kropachev, K., Essawy, M., Geacintov, N.E. and Tretyakova, N.Y. (2019) 5-  
43 Formylcytosine-induced DNA-peptide cross-links reduce transcription efficiency, but do not  
44 cause transcription errors in human cells. *J. Biol. Chem.*, **294**, 18387-18397.
- 45  
46 70. Naldiga, S., Ji, S., Thomforde, J., Nicolae, C.M., Lee, M., Zhang, Z., Moldovan, G.-L.,  
47 Tretyakova, N.Y. and Basu, A.K. (2019) Error-prone replication of a 5-formylcytosine-  
48 mediated DNA-peptide cross-link in human cells. *J. Biol. Chem.*, **294**, 10619-10627.
- 49  
50 71. Wickramaratne, S., Boldry, E.J., Buehler, C., Wang, Y.-C., Distefano, M.D. and Tretyakova,  
51 N.Y. (2015) Error-prone translesion synthesis past DNA-peptide cross-links conjugated to the  
52 major groove of DNA via C5 of thymidine. *J. Biol. Chem.*, **290**, 775-787.
- 53  
54  
55  
56  
57  
58  
59  
60

- 1  
2  
3 72. Pande, P., Ji, S., Mukherjee, S., Schärer, O.D., Tretyakova, N.Y. and Basu, A.K. (2017)  
4 Mutagenicity of a model DNA-peptide cross-link in human cells: Roles of translesion synthesis  
5 DNA polymerases. *Chem. Res. Toxicol.*, **30**, 669-677.  
6  
7 73. Yeo, J.E., Wickramaratne, S., Khatwani, S., Wang, Y.-C., Vervacke, J., Distefano, M.D. and  
8 Tretyakova, N.Y. (2014) Synthesis of site-specific DNA–protein conjugates and their effects  
9 on DNA replication. *ACS Chem. Biol.*, **9**, 1860-1868.  
10  
11 74. Wickramaratne, S., Ji, S., Mukherjee, S., Su, Y., Pence, M.G., Lior-Hoffmann, L., Fu, I.,  
12 Broyde, S., Guengerich, F.P., Distefano, M. *et al.* (2016) Bypass of DNA-protein cross-links  
13 conjugated to the 7-deazaguanine position of DNA by translesion synthesis polymerases. *J.*  
14 *Biol. Chem.*, **291**, 23589-23603.  
15  
16 75. Ji, S., Thomforde, J., Rogers, C., Fu, I., Broyde, S. and Tretyakova, N.Y. (2019)  
17 Transcriptional bypass of DNA–protein and DNA–peptide conjugates by T7 RNA polymerase.  
18 *ACS Chem. Biol.*, **14**, 2564-2575.  
19  
20 76. Obeid, S., Blatter, N., Kranaster, R., Schnur, A., Diederichs, K., Welte, W. and Marx, A.  
21 (2010) Replication through an abasic DNA lesion: Structural basis for adenine selectivity.  
22 *EMBO J.*, **29**, 1738-1747.  
23  
24 77. Obeid, S., Welte, W., Diederichs, K. and Marx, A. (2012) Amino acid templating mechanisms  
25 in selection of nucleotides opposite abasic sites by a family A DNA polymerase. *J. Biol.*  
26 *Chem.*, **287**, 14099-14108.  
27  
28 78. Zahn, K.E., Belrhali, H., Wallace, S.S. and Doublé, S. (2007) Caught bending the A-rule:  
29 Crystal structures of translesion DNA synthesis with a non-natural nucleotide. *Biochemistry*,  
30 **46**, 10551-10561.  
31  
32 79. Xia, S., Vashishtha, A., Bulkley, D., Eom, S.H., Wang, J. and Konigsberg, W.H. (2012)  
33 Contribution of partial charge interactions and base stacking to the efficiency of primer  
34 extension at and beyond abasic sites in DNA. *Biochemistry*, **51**, 4922-4931.  
35  
36 80. Garcia-Diaz, M., Bebenek, K., Krahn, J.M., Pedersen, L.C. and Kunkel, T.A. (2006) Structural  
37 analysis of strand misalignment during DNA synthesis by a human DNA polymerase. *Cell*,  
38 **124**, 331-342.  
39  
40 81. Bauer, J., Xing, G., Yagi, H., Sayer, J.M., Jerina, D.M. and Ling, H. (2007) A structural gap in  
41 Dpo4 supports mutagenic bypass of a major benzo[a]pyrene dG adduct in DNA through  
42 template misalignment. *Proc. Natl Acad. Sci. U.S.A.*, **104**, 14905-14910.  
43  
44 82. Howard, M.J., Cavanaugh, N.A., Batra, V.K., Shock, D.D., Beard, W.A. and Wilson, S.H.  
45 (2020) DNA polymerase  $\beta$  nucleotide-stabilized template misalignment fidelity depends on  
46 local sequence context. *J. Biol. Chem.*, **295**, 529-538.  
47  
48 83. Patel, P.H., Suzuki, M., Adman, E., Shinkai, A. and Loeb, L.A. (2001) Prokaryotic DNA  
49 polymerase I: Evolution, structure, and "base flipping" mechanism for nucleotide selection. *J.*  
50 *Mol. Biol.*, **308**, 823-837.  
51  
52 84. Ling, H., Boudsocq, F., Woodgate, R. and Yang, W. (2004) Snapshots of replication through  
53 an abasic lesion: Structural basis for base substitutions and frameshifts. *Mol. Cell*, **13**, 751-  
54 762.  
55  
56  
57  
58  
59  
60

- 1  
2  
3 85. Zhang, Y., Yuan, F., Wu, X., Rechkoblit, O., Taylor, J.-S., Geacintov, N.E. and Wang, Z.  
4 (2000) Error-prone lesion bypass by human DNA polymerase  $\eta$ . *Nucleic Acids Res.*, **28**,  
5 4717-4724.  
6  
7 86. Choi, J.-Y., Lim, S., Kim, E.-J., Jo, A. and Guengerich, F.P. (2010) Translesion synthesis  
8 across abasic lesions by human B-family and Y-family DNA polymerases  $\alpha$ ,  $\delta$ ,  $\eta$ ,  $\iota$ ,  $\kappa$ , and  
9 REV1. *J. Mol. Biol.*, **404**, 34-44.  
10  
11 87. Avkin, S., Adar, S., Blander, G. and Livneh, Z. (2002) Quantitative measurement of  
12 translesion replication in human cells: Evidence for bypass of abasic sites by a replicative  
13 DNA polymerase. *Proc. Natl Acad. Sci. U.S.A.*, **99**, 3764-3769.  
14  
15 88. Makarova, A.V. and Burgers, P.M. (2015) Eukaryotic DNA polymerase  $\zeta$ . *DNA Repair*, **29**, 47-  
16 55.  
17  
18 89. Weerasooriya, S., Jasti, V.P. and Basu, A.K. (2014) Replicative bypass of abasic site in  
19 *Escherichia coli* and human cells: Similarities and differences. *PLoS ONE*, **9**, e107915.  
20  
21 90. Hosfield, D.J., Guan, Y., Haas, B.J., Cunningham, R.P. and Tainer, J.A. (1999) Structure of  
22 the DNA repair enzyme endonuclease IV and its DNA complex: Double-nucleotide flipping at  
23 abasic sites and three-metal-ion catalysis. *Cell*, **98**, 397-408.  
24  
25 91. Garcin, E.D., Hosfield, D.J., Desai, S.A., Haas, B.J., Björas, M., Cunningham, R.P. and  
26 Tainer, J.A. (2008) DNA apurinic-apyrimidinic site binding and excision by endonuclease IV.  
27 *Nat. Struct. Mol. Biol.*, **15**, 515-522.  
28  
29 92. Chen, Y., Chen, X., Huang, Q., Shao, Z., Gao, Y., Li, Y., Yang, C., Liu, H., Li, J., Wang, Q. *et al.*  
30 (2020) A unique DNA-binding mode of African swine fever virus AP endonuclease. *Cell*  
31 *Discov.*, **6**, 13.  
32  
33 93. Mol, C.D., Izumi, T., Mitra, S. and Tainer, J.A. (2000) DNA-bound structures and mutants  
34 reveal abasic DNA binding by APE1 and DNA repair coordination. *Nature*, **403**, 451-456.  
35  
36 94. Freudenthal, B.D., Beard, W.A., Cuneo, M.J., Dyrkheeva, N.S. and Wilson, S.H. (2015)  
37 Capturing snapshots of APE1 processing DNA damage. *Nat. Struct. Mol. Biol.*, **22**, 924-931.  
38  
39 95. Whitaker, A.M., Flynn, T.S. and Freudenthal, B.D. (2018) Molecular snapshots of APE1  
40 proofreading mismatches and removing DNA damage. *Nat. Commun.*, **9**, 399.  
41  
42 96. Lu, D., Silhan, J., MacDonald, J.T., Carpenter, E.P., Jensen, K., Tang, C.M., Baldwin, G.S.  
43 and Freemont, P.S. (2012) Structural basis for the recognition and cleavage of abasic DNA in  
44 *Neisseria meningitidis*. *Proc. Natl Acad. Sci. U.S.A.*, **109**, 16852-16857.  
45  
46 97. Graves, R.J., Felzenszwalb, I., Laval, J. and O'Connor, T.R. (1992) Excision of 5'-terminal  
47 deoxyribose phosphate from damaged DNA is catalyzed by the Fpg protein of *Escherichia*  
48 *coli*. *J. Biol. Chem.*, **267**, 14429-14435.  
49  
50 98. Dianov, G., Sedgwick, B., Daly, G., Olsson, M., Lovett, S. and Lindahl, T. (1994) Release of  
51 5'-terminal deoxyribose-phosphate residues from incised abasic sites in DNA by the  
52 *Escherichia coli* RecJ protein. *Nucleic Acids Res.*, **22**, 993-998.  
53  
54 99. Gellon, L., Carson, D.R., Carson, J.P. and Demple, B. (2008) Intrinsic 5'-deoxyribose-5-  
55 phosphate lyase activity in *Saccharomyces cerevisiae* Trf4 protein with a possible role in base  
56 excision DNA repair. *DNA Repair*, **7**, 187-198.  
57  
58  
59  
60

- 1  
2  
3 100. Neale, M.J., Pan, J. and Keeney, S. (2005) Endonucleolytic processing of covalent protein-  
4 linked DNA double-strand breaks. *Nature*, **436**, 1053-1057.
- 5  
6 101. Huang, S.-Y.N., Michaels, S.A., Mitchell, B.B., Majdalani, N., Vanden Broeck, A., Canela, A.,  
7 Tse-Dinh, Y.-C., Lamour, V. and Pommier, Y. (2021) Exonuclease VII repairs quinolone-  
8 induced damage by resolving DNA gyrase cleavage complexes. *Sci. Adv.*, **7**, eabe0384.
- 9  
10 102. Jha, J.S., Nel, C., Haldar, T., Peters, D., Housh, K. and Gates, K.S. (2022) Products  
11 generated by amine-catalyzed strand cleavage at apurinic/aprimidinic sites in DNA: New  
12 insights from a biomimetic nucleoside model system. *Chem. Res. Toxicol.*, **35**, 203-217.
- 13  
14 103. Haldar, T., Jha, J.S., Yang, Z., Nel, C., Housh, K., Cassidy, O.J. and Gates, K.S. (2022)  
15 Unexpected complexity in the products arising from NaOH-, heat-, amine-, and glycosylase-  
16 induced strand cleavage at an abasic site in DNA. *Chem. Res. Toxicol.*, **35**, 218-232.
- 17  
18 104. Reardon, J.T. and Sancar, A. (2006) Repair of DNA–polypeptide crosslinks by human  
19 excision nuclease. *Proc. Natl Acad. Sci. U.S.A.*, **103**, 4056-4061.
- 20  
21 105. Baker, D.J., Wuenschell, G., Xia, L., Termini, J., Bates, S.E., Riggs, A.D. and O'Connor, T.R.  
22 (2007) Nucleotide excision repair eliminates unique DNA-protein cross-links from mammalian  
23 cells. *J. Biol. Chem.*, **282**, 22592-22604.
- 24  
25 106. Nakano, T., Katafuchi, A., Matsubara, M., Terato, H., Tsuboi, T., Masuda, T., Tatsumoto, T.,  
26 Pack, S.P., Makino, K., Croteau, D.L. *et al.* (2009) Homologous recombination but not  
27 nucleotide excision repair plays a pivotal role in tolerance of DNA-protein cross-links in  
28 mammalian cells. *J. Biol. Chem.*, **284**, 27065-27076.
- 29  
30 107. Paulin, K.A., Cortez, D. and Eichman, B.F. (2022) The SOS response-associated peptidase  
31 (SRAP) domain of YedK catalyzes ring opening of abasic sites and reversal of its DNA-  
32 protein crosslink. *J. Biol. Chem.*, **298**, 102307.
- 33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

## TABLE AND FIGURES LEGENDS

**Table 1.** Steady-state kinetics of cleavage of APPXL substrates by AP endonucleases.

**Figure 1.** Scheme of the DNA–peptide cross-link preparation. 8-OxoG-containing DNA duplexes were treated with bifunctional DNA glycosylases Fpg (top) or OGG1 (bottom) to produce a Schiff base conjugate with the nascent aldehydic AP site through a terminal amino group in case of Fpg or an internal amino group in case of OGG1. The enzyme–DNA intermediates were stabilized with NaBH<sub>4</sub> to yield covalent DNA-protein cross-links (DPXLs). Then, DPXL were subjected to full trypsinolysis to obtain AP-peptide cross-links.

**Figure 2.** Stalling of KF and RBpol by AP–peptide cross-links. A, APPXL-I, primer–template; B, APPXL-Y, primer–template; C, APPXL-I, primer–downstream strand–template; D, APPXL-Y, primer–downstream strand–template. The <sup>32</sup>P-labeled primer is shown in red, the downstream strand is in lowercase. Grey arrows mark the sites of synthesis termination at the lesion. Reaction time and the nature of the lesion are indicated under the gel images. In all panels, lanes 1–3 are size markers corresponding to the primer (11 nt), full-size product (40 nt), and primer extended to the cross-link site (23 nt). “–”, no enzyme added; K1, undamaged primer–template, 30 min; K2, undamaged primer–template–downstream strand, 30 min. The inset shows KF pause sites at a lower band intensity.

**Figure 3.** Stalling of Dpo4 and PolX by AP–peptide cross-links. (A), primer–template; (B), primer–downstream strand–template. The nature of the polymerase and the cross-link is indicated over and below the gel images, respectively. The <sup>32</sup>P-labeled primer is shown in red, the downstream strand is in lowercase. Grey arrows mark the sites of synthesis termination at the lesion. Reaction time and the nature of the lesion are indicated under the gel images. In all panels, lanes 1–3 are size markers corresponding to the primer (11 nt), full-size product (40 nt), and primer extended to the cross-link site (23 nt). “–”, no enzyme added; K1, undamaged primer–template, 30 min; K2, undamaged primer–template–downstream strand, 30 min. No bypass was observed for PolX.

**Figure 4.** Incorporation of dNMPs opposite APPXLs by KF (A), RBpol (B), Dpo4 (C) and PolX (D). The nature of the substrate and dNTP is indicated below the gel images. The <sup>32</sup>P-labeled primer is shown in red. “–”, no enzyme or dNTPs added; N, all dNTPs added; K1, undamaged primer–template, all dNTPs, 30 min.

**Figure 5.** Cleavage of APPXL-I by AP endonucleases: APE1 (A), Apn1p (B), Xth (C), and Nfo (D). For panels A–C: lane 1, APPXL-I, no enzyme; lanes 2–5, APPXL-I cleavage under the BER conditions; lanes 6–9, APPXL-I cleavage under the NIR conditions; lanes 10–13, AP site cleavage under the BER conditions; lanes 14–17, AP site cleavage under the NIR conditions; lane 18, AP site treated with NaOH; lane 19, AP site, no enzyme. Panel D: lanes 1–4, APPXL-I cleavage under the BER conditions; lanes 5–8, APPXL-I cleavage under the NIR conditions; lanes 9–12, AP site cleavage under the BER conditions; lane 13, AP site treated with NaOH; lane 14, AP site, no enzyme. BER and NIR conditions for the individual AP endonucleases are specified in Supplementary Table S2.

1  
2  
3 **Figure 6.** Cleavage of APPXL-Y by AP endonucleases: APE1 (A), Apn1p (B), Xth (C), and Nfo (D).  
4 For panels A–C: lane 1, APPXL-Y, no enzyme; lanes 2–5, APPXL-Y cleavage under the BER  
5 conditions; lanes 6–9, APPXL-Y cleavage under the NIR conditions; lanes 10–13, AP site cleavage  
6 under the BER conditions; lanes 14–17, AP site cleavage under the NIR conditions; lane 18, AP site  
7 untreated (A) or treated with NaOH (B, C). Panel D: lane 1, APPXL-Y, no enzyme; lanes 2–5, APPXL-  
8 Y cleavage under the BER conditions; lanes 6–9, APPXL-Y cleavage under the NIR conditions;  
9 lanes 10–13, AP site cleavage under the BER conditions. BER and NIR conditions for the individual  
10 AP endonucleases are specified in Supplementary Table S2.  
11  
12  
13  
14

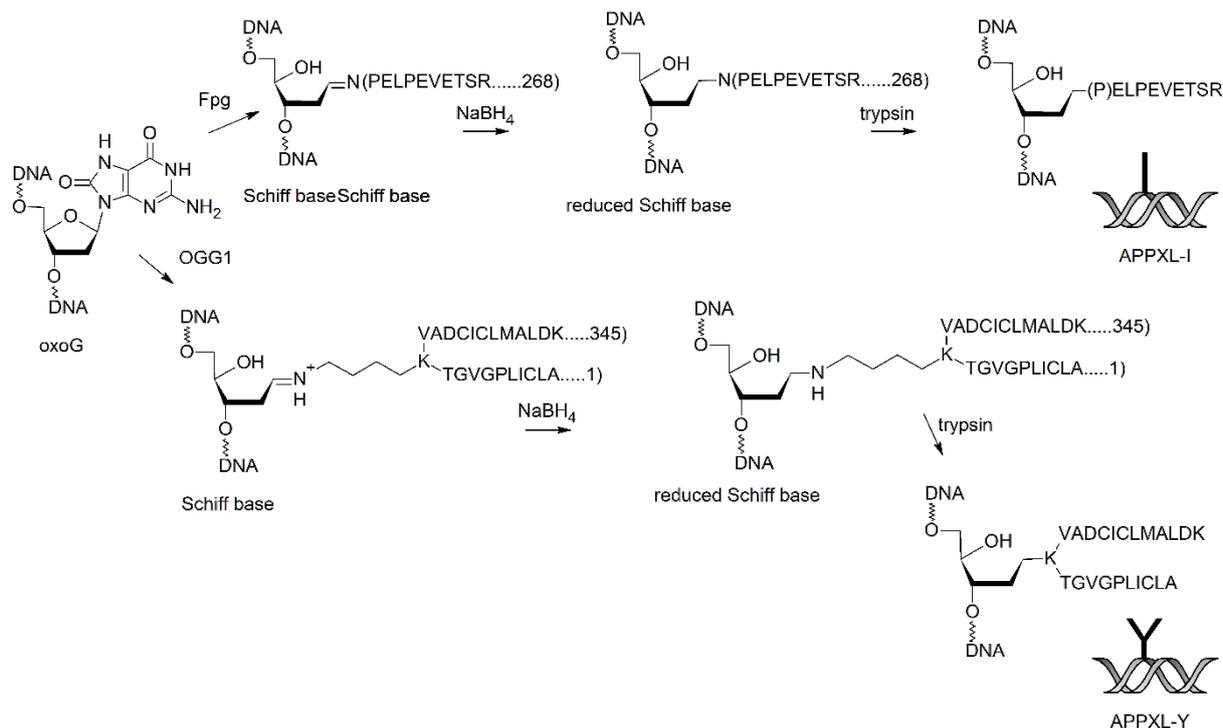
15 **Figure 7.** Base excision repair continuation after the cleavage of AP site–peptide cross-links. A, B,  
16 Extension by KF of the primer with the displacement of the downstream strand bearing a 5'-terminal  
17 cross-link after Nfo cleavage of APPXL-I (A) and APPXL-Y (B). Lane 1, cross-linked strand; lane 2,  
18 APPXL treated with Nfo; lanes 3–5, APPXL treated with Nfo followed by KF extension for the  
19 indicated time; lane 6, size marker corresponding to the unmodified strand. C, Reconstitution of the  
20 full BER cycle of APPXL-I with Nfo, Pol I, and DNA ligase. Lane 1 and 2, unmodified and cross-linked  
21 strands, respectively; lane 3, substrate cleaved by Nfo; lane 4, substrate cleaved by Nfo in the  
22 presence of Pol I without dNTPs; lane 5; substrate cleaved by Nfo and extended by Pol I; lane 6,  
23 substrate cleaved by Nfo, extended by Pol I and ligated. D General scheme of possible BER-  
24 dependent repair of AP site–peptide cross-links.  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60

**Table 1.** Steady-state kinetics of cleavage of APPXL substrates by AP endonucleases.

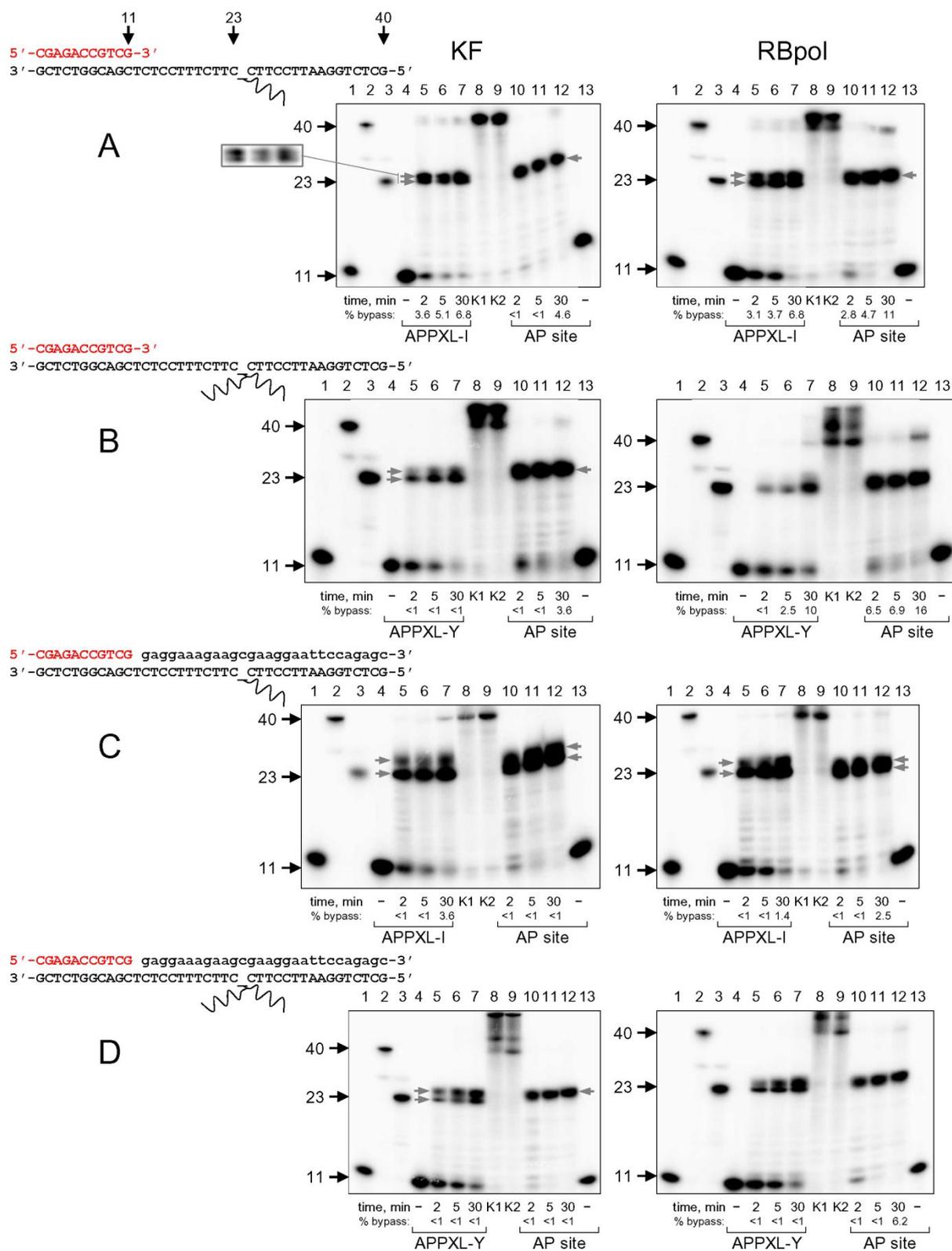
Enzyme	Substrate	Conditions	$K_M$ , nM	$k_{cat}$ , $s^{-1}$ ( $\times 10^{-5}$ )	$k_{cat}/K_M$ , $nM^{-1}\cdot s^{-1}$ ( $\times 10^{-8}$ )
Apn1p	APPXL-I	BER	190 ± 60	2.0 ± 0.3	10 ± 3
Apn1p	APPXL-I	NIR	21 ± 8	1.2 ± 0.1	58 ± 24
Apn1p	APPXL-Y	BER	110 ± 30	47 ± 5	430 ± 130
Apn1p	APPXL-Y	NIR	200 ± 80	65 ± 13	330 ± 150
Apn1p	THF	BER	8.8 ± 3.0	160 ± 20	18600 ± 6700
Apn1p	OHU	NIR	61 ± 21	38 ± 4	620 ± 220
Nfo	APPXL-I	BER	n/s <sup>1</sup>	n/s	5.2 ± 1.0
Nfo	APPXL-I	NIR	130 ± 50	6.5 ± 1.0	50 ± 20
Nfo	APPXL-Y	BER	n/s	n/s	200 ± 30
Nfo	APPXL-Y	NIR	61 ± 17	260 ± 20	4200 ± 1200
Nfo	THF	BER	150 ± 60.0	66 ± 12	430 ± 180
Nfo	OHU	NIR	140 ± 40.0	46 ± 5	320 ± 90
APE1	APPXL-I	BER	n/c <sup>2</sup>	n/c	n/c
APE1	APPXL-I	NIR	n/c	n/c	n/c
APE1	APPXL-Y	BER	n/c	n/c	n/c
APE1	APPXL-Y	NIR	44 ± 19	1.6 ± 0.2	37 ± 16
APE1	THF	BER	21 ± 7	580 ± 70	28200 ± 10000
APE1	OHU	NIR	170 ± 80	19 ± 4	110 ± 50

<sup>1</sup> n/s, not saturated, saturation with substrate could not be achieved;  $k_{cat}/K_M$  determined from the initial phase of reaction velocity vs substrate concentration plots.

<sup>2</sup> n/c, not cleaved, cleavage absent or too low to reliably estimate kinetic constants.

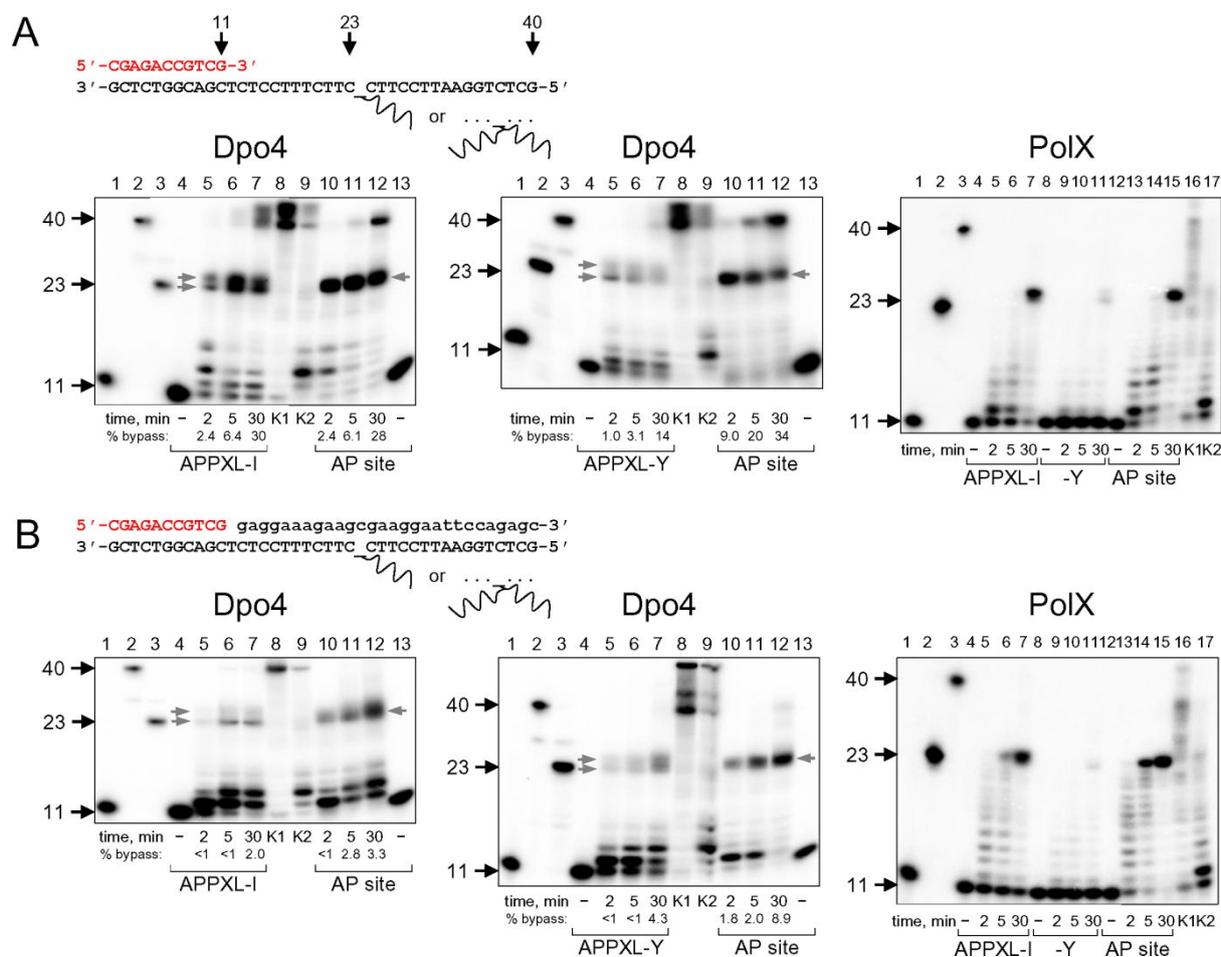


**Figure 1.** Scheme of the DNA–peptide cross-link preparation. 8-OxoG-containing DNA duplexes were treated with bifunctional DNA glycosylases Fpg (top) or OGG1 (bottom) to produce a Schiff base conjugate with the nascent aldehydic AP site through a terminal amino group in case of Fpg or an internal amino group in case of OGG1. The enzyme–DNA intermediates were stabilized with  $\text{NaBH}_4$  to yield covalent DNA-protein cross-links (DPXLs). Then, DPXL were subjected to full trypsinolysis to obtain AP-peptide cross-links.

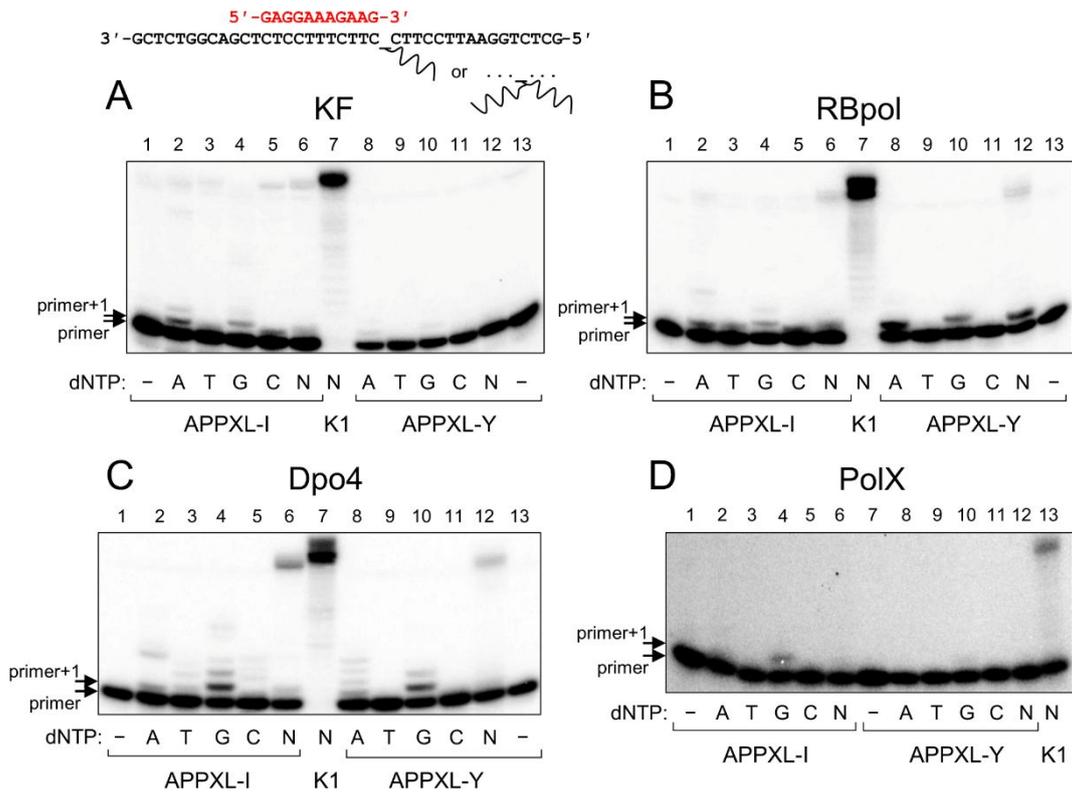


**Figure 2.** Stalling of KF and RBpol by AP-peptide cross-links. A, APPXL-I, primer-template; B, APPXL-Y, primer-template; C, APPXL-I, primer-downstream strand-template; D, APPXL-Y, primer-downstream strand-template. The  $^{32}\text{P}$ -labeled primer is shown in red, the downstream strand is in lowercase. Grey arrows mark the sites of synthesis termination at the lesion. Reaction time and the nature of the lesion are indicated under the gel images. In all panels, lanes 1–3 are size markers

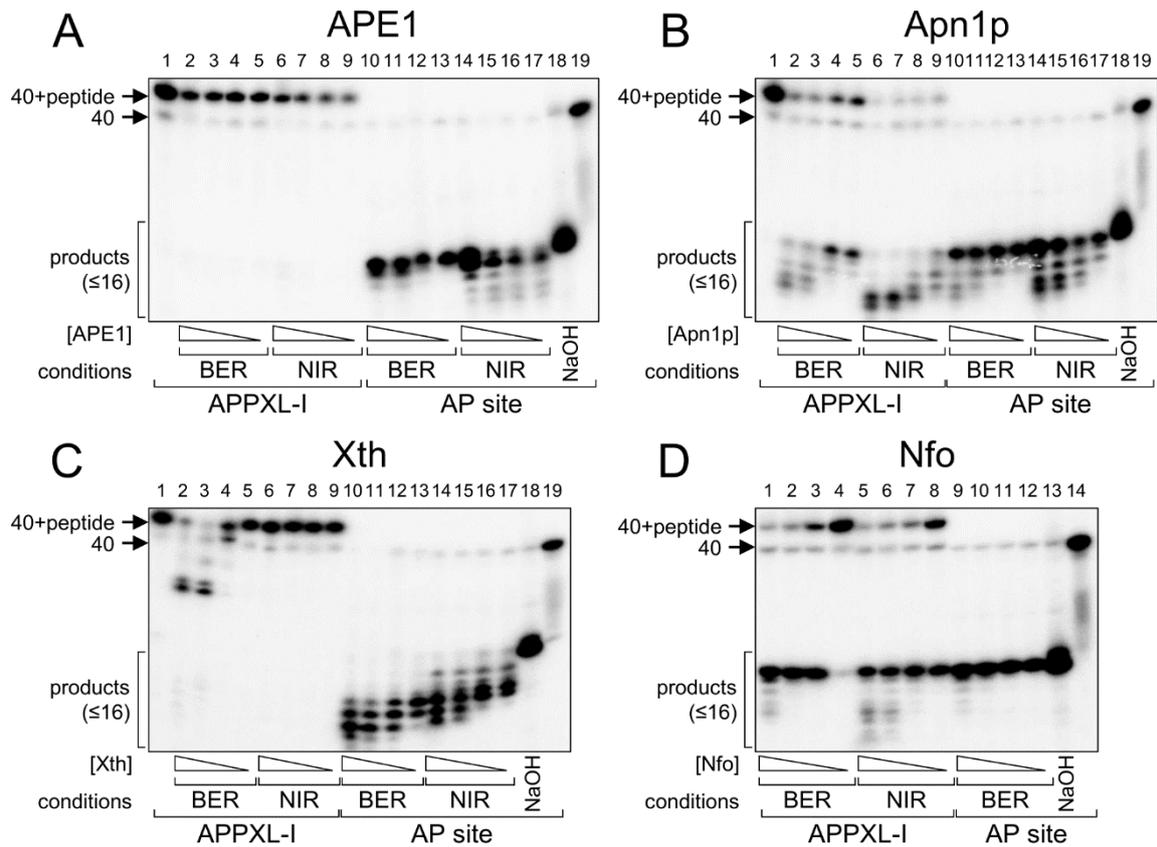
1  
2  
3 corresponding to the primer (11 nt), full-size product (40 nt), and primer extended to the cross-link site  
4 (23 nt). "-", no enzyme added; K1, undamaged primer–template, 30 min; K2, undamaged primer–  
5 template–downstream strand, 30 min. The inset shows KF pause sites at a lower band intensity.  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



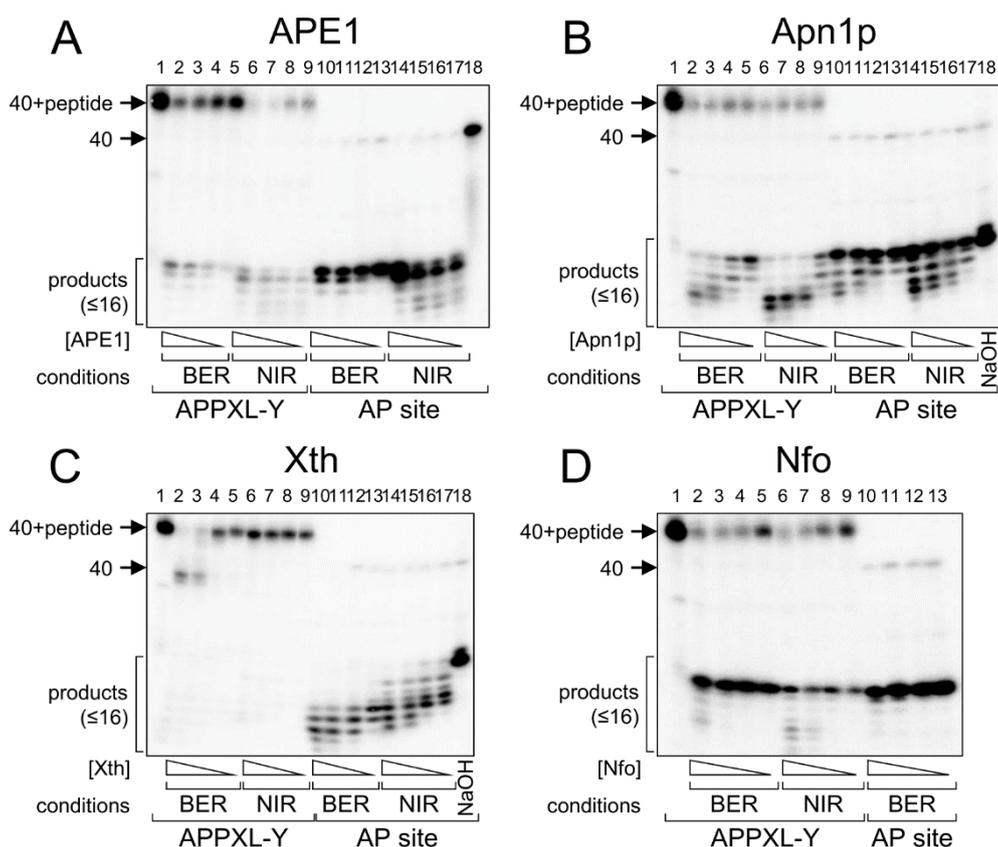
**Figure 3.** Stalling of Dpo4 and PoIX by AP-peptide cross-links. (A), primer-template; (B), primer-downstream strand-template. The nature of the polymerase and the cross-link is indicated over and below the gel images, respectively. The  $^{32}\text{P}$ -labeled primer is shown in red, the downstream strand is in lowercase. Grey arrows mark the sites of synthesis termination at the lesion. Reaction time and the nature of the lesion are indicated under the gel images. In all panels, lanes 1–3 are size markers corresponding to the primer (11 nt), full-size product (40 nt), and primer extended to the cross-link site (23 nt). “–”, no enzyme added; K1, undamaged primer-template, 30 min; K2, undamaged primer-template-downstream strand, 30 min. No bypass was observed for PoIX.



**Figure 4.** Incorporation of dNMPs opposite APPXLs by **KF (A)**, **RBpol (B)**, **Dpo4 (C)** and **PoIX (D)**. The nature of the substrate and dNTP is indicated below the gel images. The <sup>32</sup>P-labeled primer is shown in red. "-", no enzyme or dNTPs added; N, all dNTPs added; K1, undamaged primer–template, all dNTPs, 30 min.

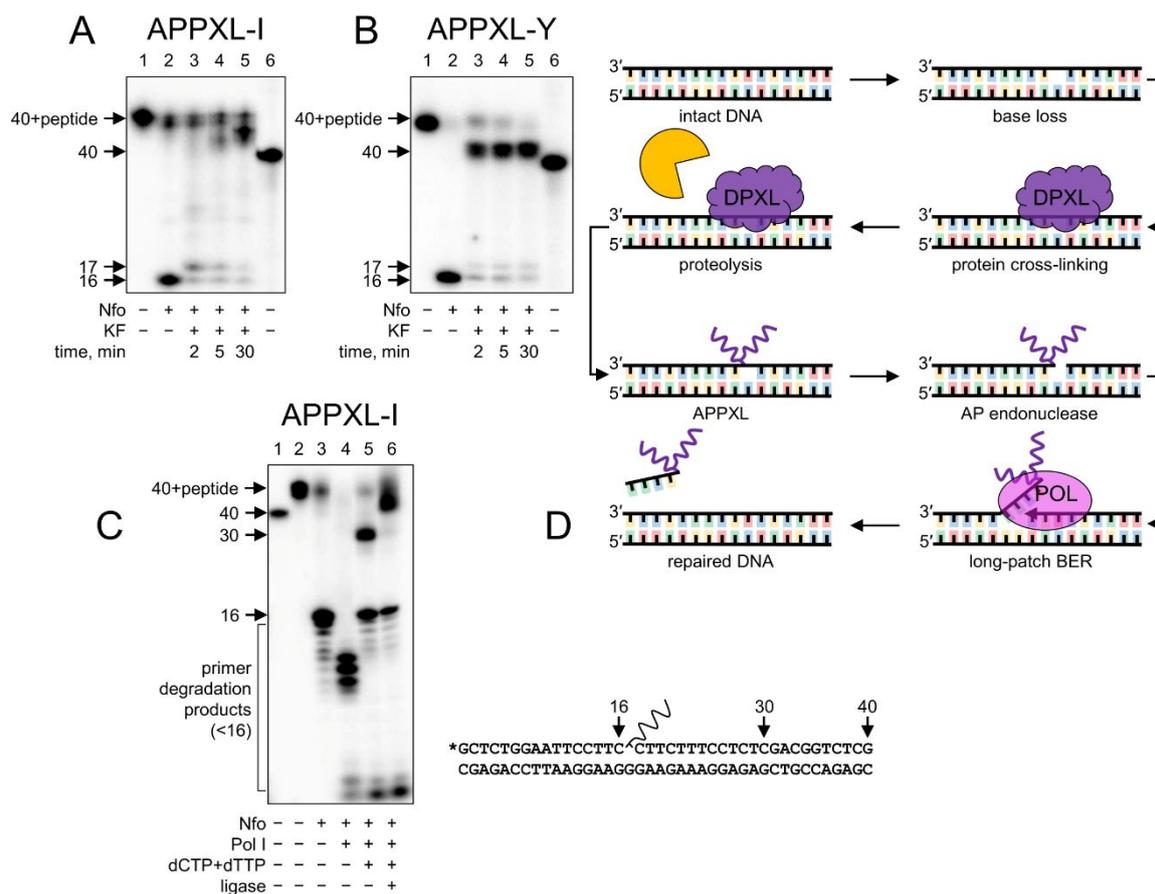


**Figure 5.** Cleavage of APPXL-I by AP endonucleases: APE1 (A), Apn1p (B), Xth (C), and Nfo (D). For panels A–C: lane 1, APPXL-I, no enzyme; lanes 2–5, APPXL-I cleavage under the BER conditions; lanes 6–9, APPXL-I cleavage under the NIR conditions; lanes 10–13, AP site cleavage under the BER conditions; lanes 14–17, AP site cleavage under the NIR conditions; lane 18, AP site treated with NaOH; lane 19, AP site, no enzyme. Panel D: lanes 1–4, APPXL-I cleavage under the BER conditions; lanes 5–8, APPXL-I cleavage under the NIR conditions; lanes 9–12, AP site cleavage under the BER conditions; lane 13, AP site treated with NaOH; lane 14, AP site, no enzyme. BER and NIR conditions for the individual AP endonucleases are specified in Supplementary Table S2.



**Figure 6.** Cleavage of APPXL-Y by AP endonucleases: APE1 (A), Apn1p (B), Xth (C), and Nfo (D).

For panels A–C: lane 1, APPXL-Y, no enzyme; lanes 2–5, APPXL-Y cleavage under the BER conditions; lanes 6–9, APPXL-Y cleavage under the NIR conditions; lanes 10–13, AP site cleavage under the BER conditions; lanes 14–17, AP site cleavage under the NIR conditions; lane 18, AP site untreated (A) or treated with NaOH (B, C). Panel D: lane 1, APPXL-Y, no enzyme; lanes 2–5, APPXL-Y cleavage under the BER conditions; lanes 6–9, APPXL-Y cleavage under the NIR conditions; lanes 10–13, AP site cleavage under the BER conditions. BER and NIR conditions for the individual AP endonucleases are specified in Supplementary Table S2.



**Figure 7.** Base excision repair continuation after the cleavage of AP site-peptide cross-links. A, B, Extension by KF of the primer with the displacement of the downstream strand bearing a 5'-terminal cross-link after Nfo cleavage of APPXL-I (A) and APPXL-Y (B). Lane 1, cross-linked strand; lane 2, APPXL treated with Nfo; lanes 3–5, APPXL treated with Nfo followed by KF extension for the indicated time; lane 6, size marker corresponding to the unmodified strand. C, Reconstitution of the full BER cycle of APPXL-I with Nfo, Pol I, and DNA ligase. Lane 1 and 2, unmodified and cross-linked strands, respectively; lane 3, substrate cleaved by Nfo; lane 4, substrate cleaved by Nfo in the presence of Pol I without dNTPs; lane 5, substrate cleaved by Nfo and extended by Pol I; lane 6, substrate cleaved by Nfo, extended by Pol I and ligated. D General scheme of possible BER-dependent repair of AP site-peptide cross-links.

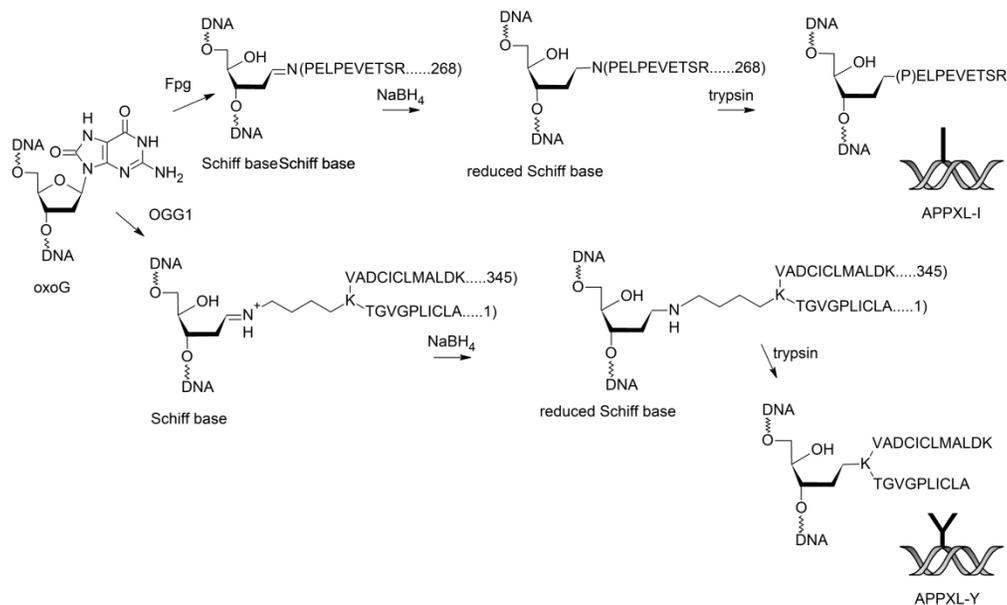
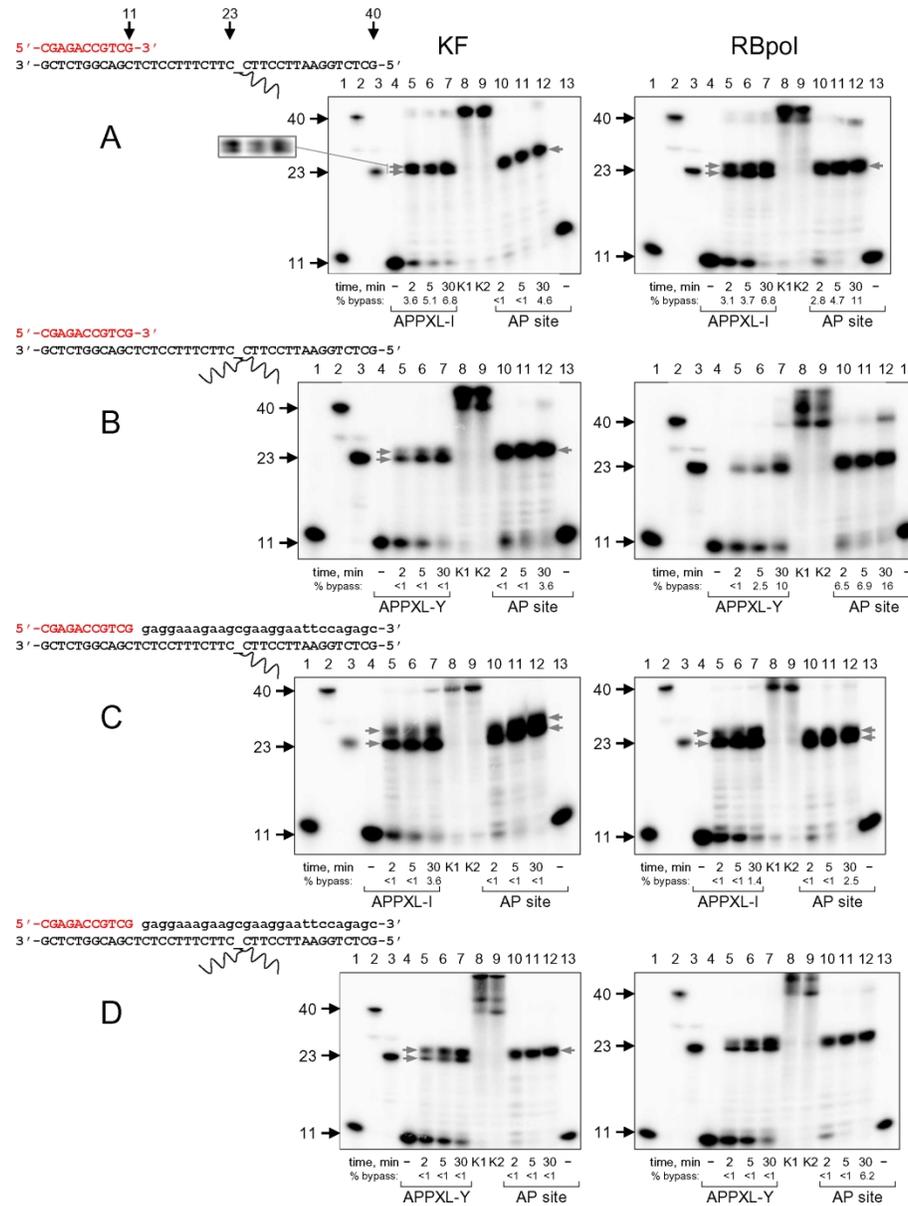


Figure 1. Scheme of the DNA–peptide cross-link preparation. 8-OxoG-containing DNA duplexes were treated with bifunctional DNA glycosylases Fpg (top) or OGG1 (bottom) to produce a Schiff base conjugate with the nascent aldehydic AP site through a terminal amino group in case of Fpg or an internal amino group in case of OGG1. The enzyme–DNA intermediates were stabilized with  $\text{NaBH}_4$  to yield covalent DNA-protein cross-links (DPXLs). Then, DPXL were subjected to full trypsinolysis to obtain AP-peptide cross-links.

695x419mm (150 x 150 DPI)



45 Figure 2. Stalling of KF and RBpol by AP-peptide cross-links. A, APPXL-I, primer-template; B, APPXL-Y,  
46 primer-template; C, APPXL-I, primer-downstream strand-template; D, APPXL-Y, primer-downstream  
47 strand-template. The <sup>32</sup>P-labeled primer is shown in red, the downstream strand is in lowercase. Grey  
48 arrows mark the sites of synthesis termination at the lesion. Reaction time and the nature of the lesion are  
49 indicated under the gel images. In all panels, lanes 1-3 are size markers corresponding to the primer (11  
50 nt), full-size product (40 nt), and primer extended to the cross-link site (23 nt). "-", no enzyme added; K1,  
51 undamaged primer-template, 30 min; K2, undamaged primer-template-downstream strand, 30 min. The  
52 inset shows KF pause sites at a lower band intensity.

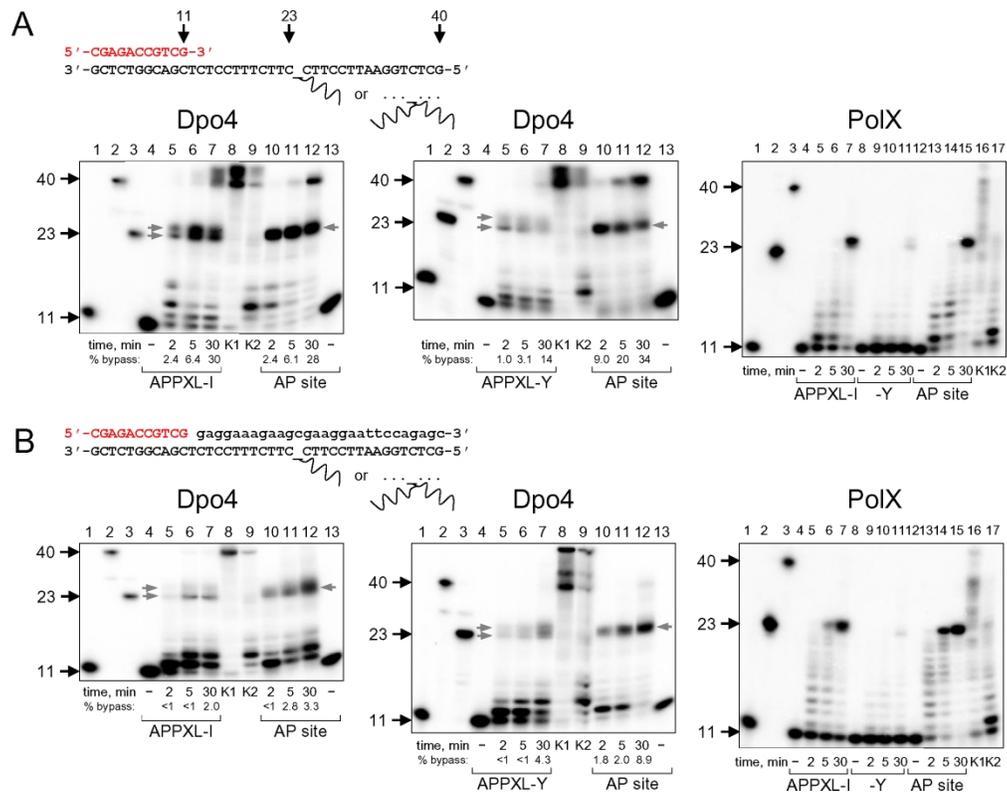


Figure 3. Stalling of Dpo4 and PolX by AP-peptide cross-links. (A), primer-temple; (B), primer-downstream strand-temple. The nature of the polymerase and the cross-link is indicated over and below the gel images, respectively. The 32P-labeled primer is shown in red, the downstream strand is in lowercase. Grey arrows mark the sites of synthesis termination at the lesion. Reaction time and the nature of the lesion are indicated under the gel images. In all panels, lanes 1-3 are size markers corresponding to the primer (11 nt), full-size product (40 nt), and primer extended to the cross-link site (23 nt). "-", no enzyme added; K1, undamaged primer-temple, 30 min; K2, undamaged primer-temple-downstream strand, 30 min. No bypass was observed for PolX.

787x617mm (72 x 72 DPI)

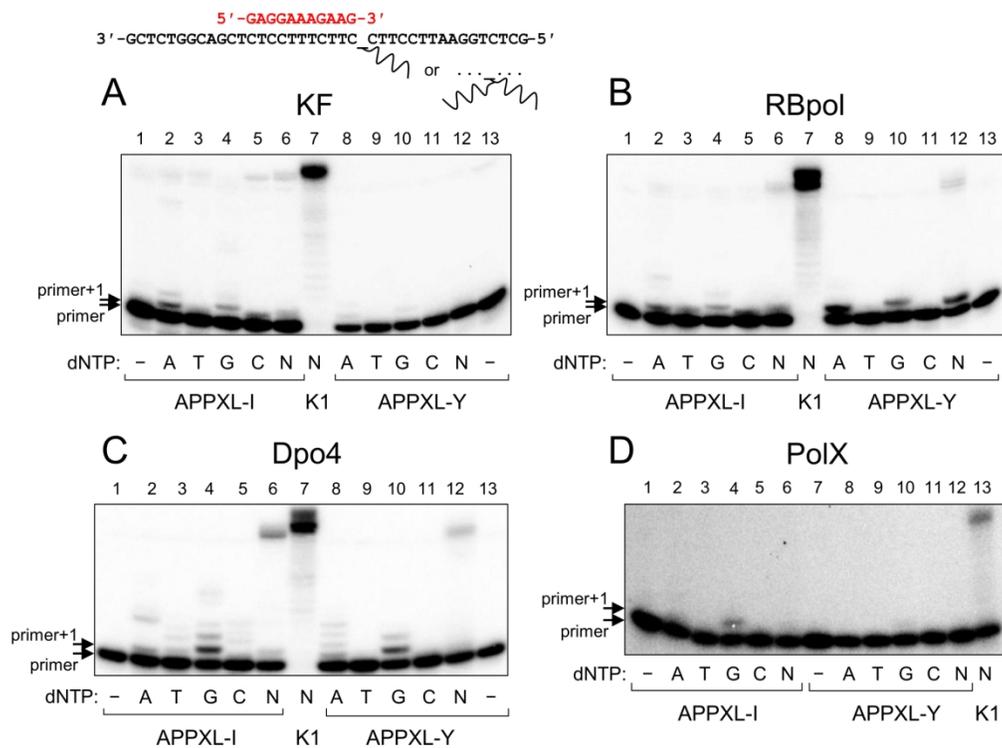


Figure 4. Incorporation of dNMPs opposite APPXLs by KF (A), RBpol (B), Dpo4 (C) and PolX (D). The nature of the substrate and dNTP is indicated below the gel images. The  $^{32}\text{P}$ -labeled primer is shown in red. "-", no enzyme or dNTPs added; N, all dNTPs added; K1, undamaged primer-template, all dNTPs, 30 min.

423x311mm (118 x 118 DPI)

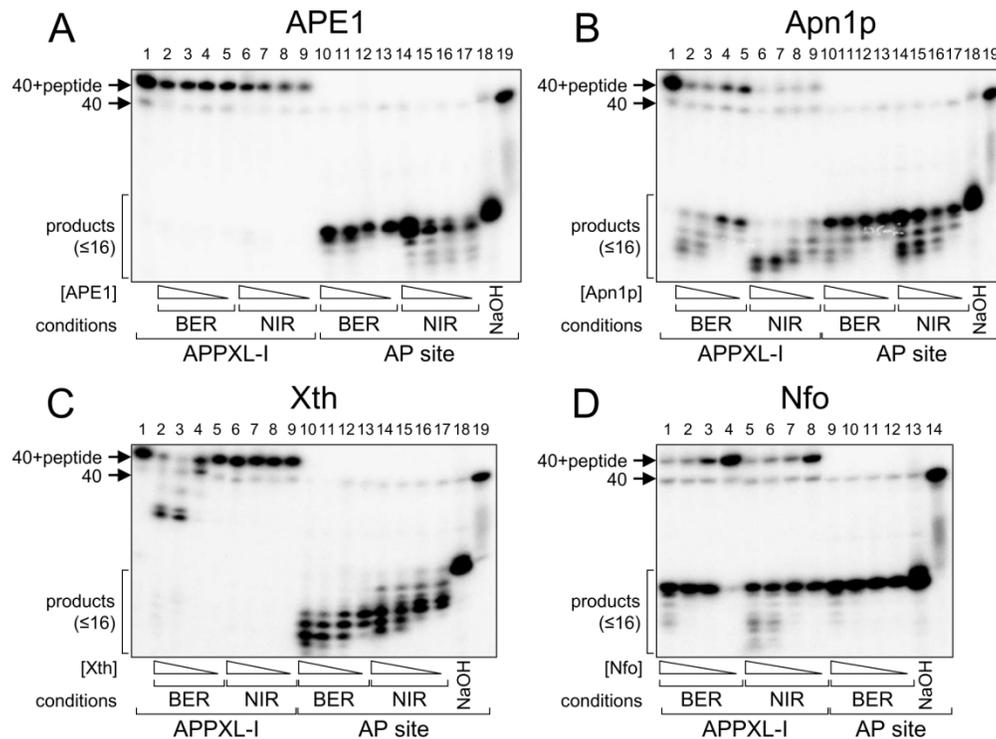


Figure 5. Cleavage of APPXL-I by AP endonucleases: APE1 (A), Apn1p (B), Xth (C), and Nfo (D). For panels A-C: lane 1, APPXL-I, no enzyme; lanes 2-5, APPXL-I cleavage under the BER conditions; lanes 6-9, APPXL-I cleavage under the NIR conditions; lanes 10-13, AP site cleavage under the BER conditions; lanes 14-17, AP site cleavage under the NIR conditions; lane 18, AP site treated with NaOH; lane 19, AP site, no enzyme. Panel D: lanes 1-4, APPXL-I cleavage under the BER conditions; lanes 5-8, APPXL-I cleavage under the NIR conditions; lanes 9-12, AP site cleavage under the BER conditions; lane 13, AP site treated with NaOH; lane 14, AP site, no enzyme. BER and NIR conditions for the individual AP endonucleases are specified in Supplementary Table S2.

394x288mm (118 x 118 DPI)

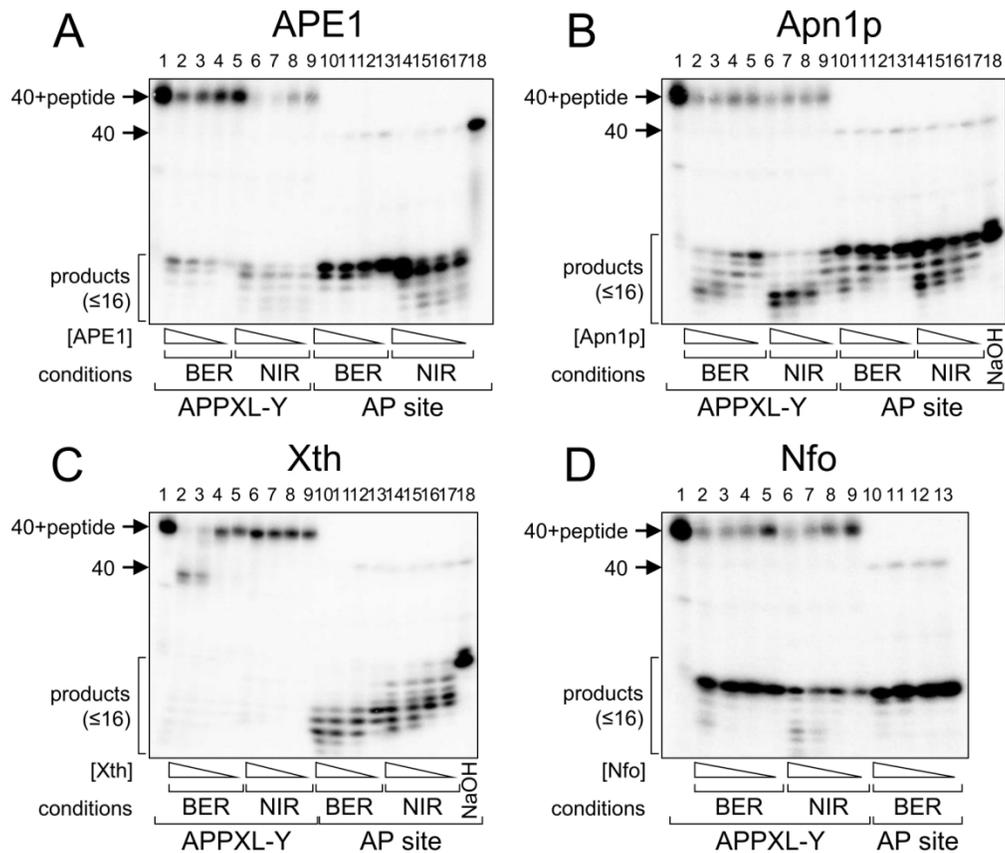


Figure 6. Cleavage of APPXL-Y by AP endonucleases: APE1 (A), Apn1p (B), Xth (C), and Nfo (D). For panels A–C: lane 1, APPXL-Y, no enzyme; lanes 2–5, APPXL-Y cleavage under the BER conditions; lanes 6–9, APPXL-Y cleavage under the NIR conditions; lanes 10–13, AP site cleavage under the BER conditions; lanes 14–17, AP site cleavage under the NIR conditions; lane 18, AP site untreated (A) or treated with NaOH (B, C). Panel D: lane 1, APPXL-Y, no enzyme; lanes 2–5, APPXL-Y cleavage under the BER conditions; lanes 6–9, APPXL-Y cleavage under the NIR conditions; lanes 10–13, AP site cleavage under the BER conditions. BER and NIR conditions for the individual AP endonucleases are specified in Supplementary Table S2.

341x288mm (118 x 118 DPI)

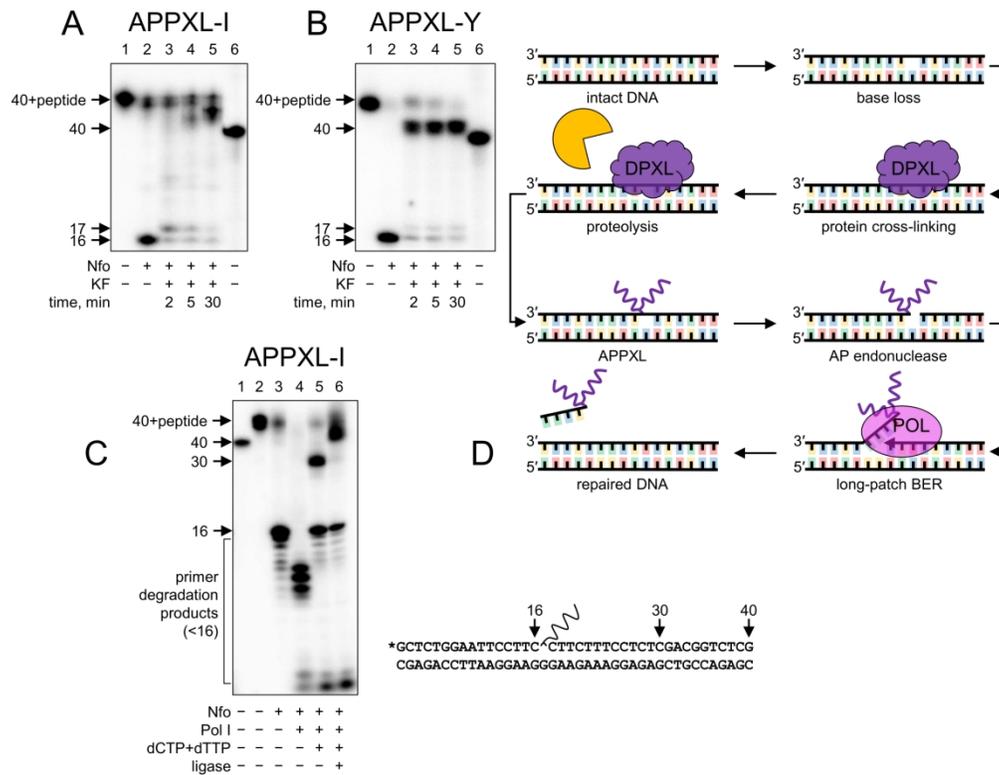


Figure 7. Base excision repair continuation after the cleavage of AP site-peptide cross-links. A, B, Extension by KF of the primer with the displacement of the downstream strand bearing a 5'-terminal cross-link after Nfo cleavage of APPXL-I (A) and APPXL-Y (B). Lane 1, cross-linked strand; lane 2, APPXL treated with Nfo; lanes 3-5, APPXL treated with Nfo followed by KF extension for the indicated time; lane 6, size marker corresponding to the unmodified strand. C, Reconstitution of the full BER cycle of APPXL-I with Nfo, Pol I, and DNA ligase. Lane 1 and 2, unmodified and cross-linked strands, respectively; lane 3, substrate cleaved by Nfo; lane 4, substrate cleaved by Nfo in the presence of Pol I without dNTPs; lane 5, substrate cleaved by Nfo and extended by Pol I; lane 6, substrate cleaved by Nfo, extended by Pol I and ligated. D General scheme of possible BER-dependent repair of AP site-peptide cross-links.

475x364mm (118 x 118 DPI)

## Supplementary materials for

### Abasic site–peptide cross-links are blocking lesions repaired by AP endonucleases

Anna V. Yudkina<sup>1,2,\*</sup>, Nikita A. Bulgakov<sup>1</sup>, Svetlana V. Baranova<sup>1</sup>, Alexander A. Ishchenko<sup>3</sup>, Murat K. Sapparbaev<sup>3</sup>, Vladimir V. Koval<sup>1</sup> and Dmitry O. Zharkov<sup>1,2,\*</sup>

<sup>1</sup> SB RAS Institute of Chemical Biology and Fundamental Medicine, Novosibirsk, 630090, Russia

<sup>2</sup> Department of Natural Sciences, Novosibirsk State University, Novosibirsk, 630090, Russia

<sup>3</sup> Groupe "Mechanisms of DNA Repair and Carcinogenesis", Equipe Labellisée LIGUE 2016, CNRS UMR9019, Université Paris-Saclay, Gustave Roussy Cancer Campus, F-94805 Villejuif, France

\* To whom correspondence should be addressed. Tel: +7 383 363 5187; Fax: +7 383 363 5128; Email: dzharkov@niboch.nsc.ru. Correspondence may also be addressed to Anna Yudkina. Tel: +7 383 363 5188; Fax: +7 383 363 5128; Email: ayudkina@niboch.nsc.ru.

This file contains

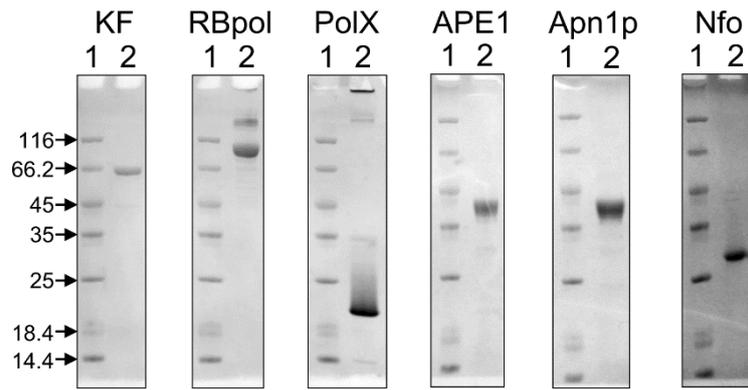
- Supplementary Tables S1–S2
- Supplementary Figures S1–S8
- A list of Supplementary References

**Supplementary Table S1.** Oligonucleotides used in this work.

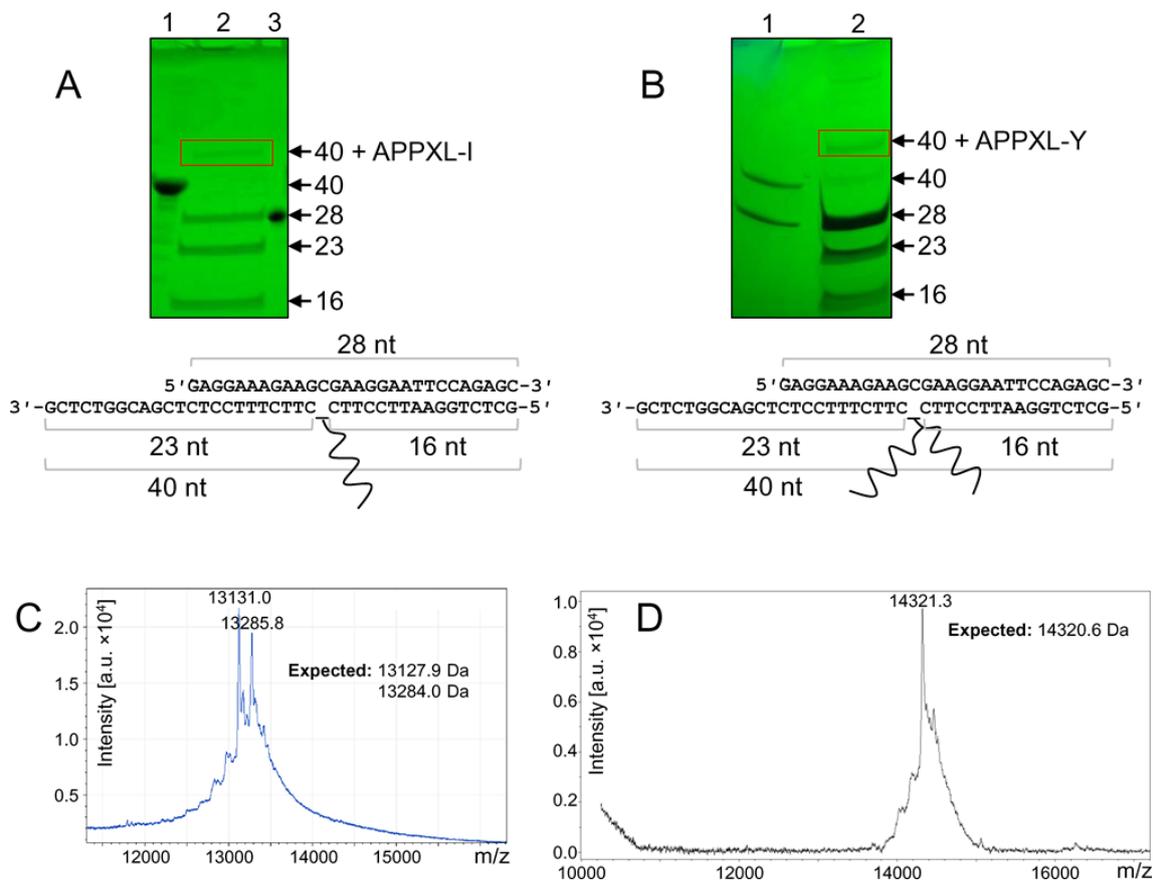
ID	Sequence, 5'→3'
<i>Assembly of substrates for cross-linking and enzyme assays</i>	
40oG	GCTCTGGAATTCCTTCXCTTCTTTTCCTCTCGACGGTCTCG; X = oxoG
40U	GCTCTGGAATTCCTTCXCTTCTTTTCCTCTCGACGGTCTCG; X = U
40G	GCTCTGGAATTCCTTCGCTTCTTTTCCTCTCGACGGTCTCG
40compC	CGAGACCGTCGAGAGGAAAGAAGCGAAGGAATTCAGAGC
40compG	CGAGACCGTCGAGAGGAAAGAAGGGAAGGAATTCAGAGC
28down	GAGGAAAGAAGCGAAGGAATTCAGAGC
28oG	GAGGAAAGAAGXGAAGGAATTCAGAGC; X = oxoG
21comp	GGAATTCCTTCCCTTCTTTCC
<i>AP endonuclease kinetics</i>	
23OHU	TCTCCTTTTCCTCXCTTCCCTCTC; X = OHU
23comp	AGAGGAAAGGAGCGAAGGGAGAG
20THF	GAGTGAGAXGTCAAGTTGGG; X = THF
20comp	CCCAACTTGACGTCTCACTC
<i>Polymerase primers</i>	
16pri	GCTCTGGAATTCCTTC
13pri	GAGAGGAAAGAAG
11pri	CGAGACCGTCG
<i>Electrophoretic mobility markers</i>	
40m	CGAGACCGTCGAGAGGAAAGAAGCGAAGGAATTCAGAGC
23m	CGAGACCGTCGAGAGGAAAGAAG
11m	CGAGACCGTCG

**Supplementary Table S2.** Composition of BER and NIR buffers.

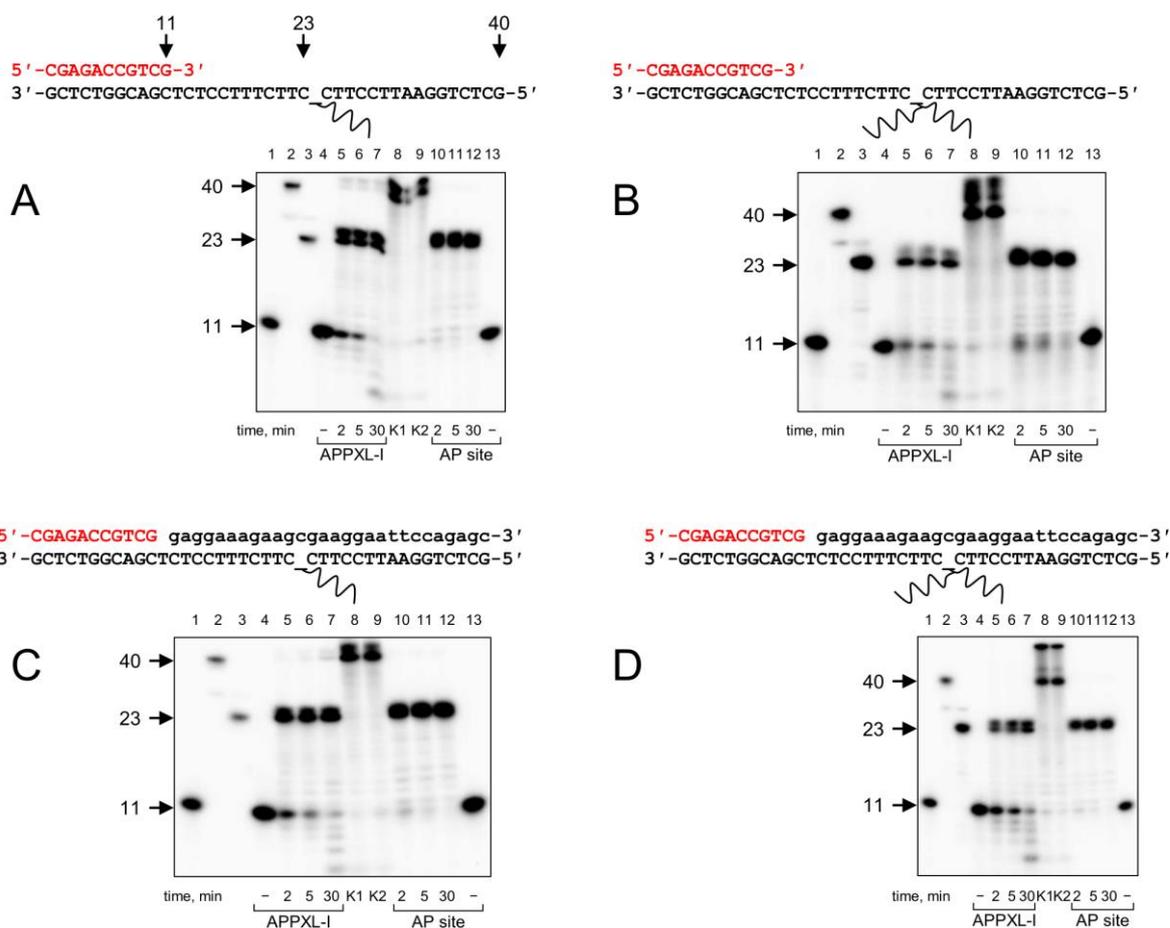
<b>Enzyme</b>	<b>BER buffer (1×)</b>	<b>NIR buffer (1×)</b>
APE1	20 mM HEPES–KOH (pH 8.0) 100 mM KCl 5 mM MgCl <sub>2</sub> 100 µg/ml BSA 1 mM DTT Reference: (1)	20 mM HEPES–KOH (pH 6.8) 25 mM KCl 0.5 mM MgCl <sub>2</sub> 100 µg/ml BSA 1 mM DTT References: (1-3)
Apn1p	20 mM HEPES–KOH (pH 7.6) 100 mM KCl 5 mM MgCl <sub>2</sub> 100 µg/ml BSA 1 mM DTT Reference: (4)	20 mM HEPES–KOH (pH 7.6) 50 mM KCl 5 mM MgCl <sub>2</sub> 100 µg/ml BSA 1 mM DTT References: (4,5)
Nfo	20 mM HEPES–KOH (pH 7.6) 100 mM KCl 100 µg/ml BSA 1 mM DTT Reference: (6)	20 mM HEPES–KOH (pH 7.6) 50 mM KCl 100 µg/ml BSA 1 mM DTT Reference: (6)
Xth	20 mM HEPES–KOH (pH 7.6) 100 mM KCl 5 mM MgCl <sub>2</sub> 100 µg/ml BSA 1 mM DTT Reference: (4)	20 mM HEPES–KOH (pH 7.6) 25 mM KCl 5 mM CaCl <sub>2</sub> 100 µg/ml BSA 1 mM DTT Reference: (7)



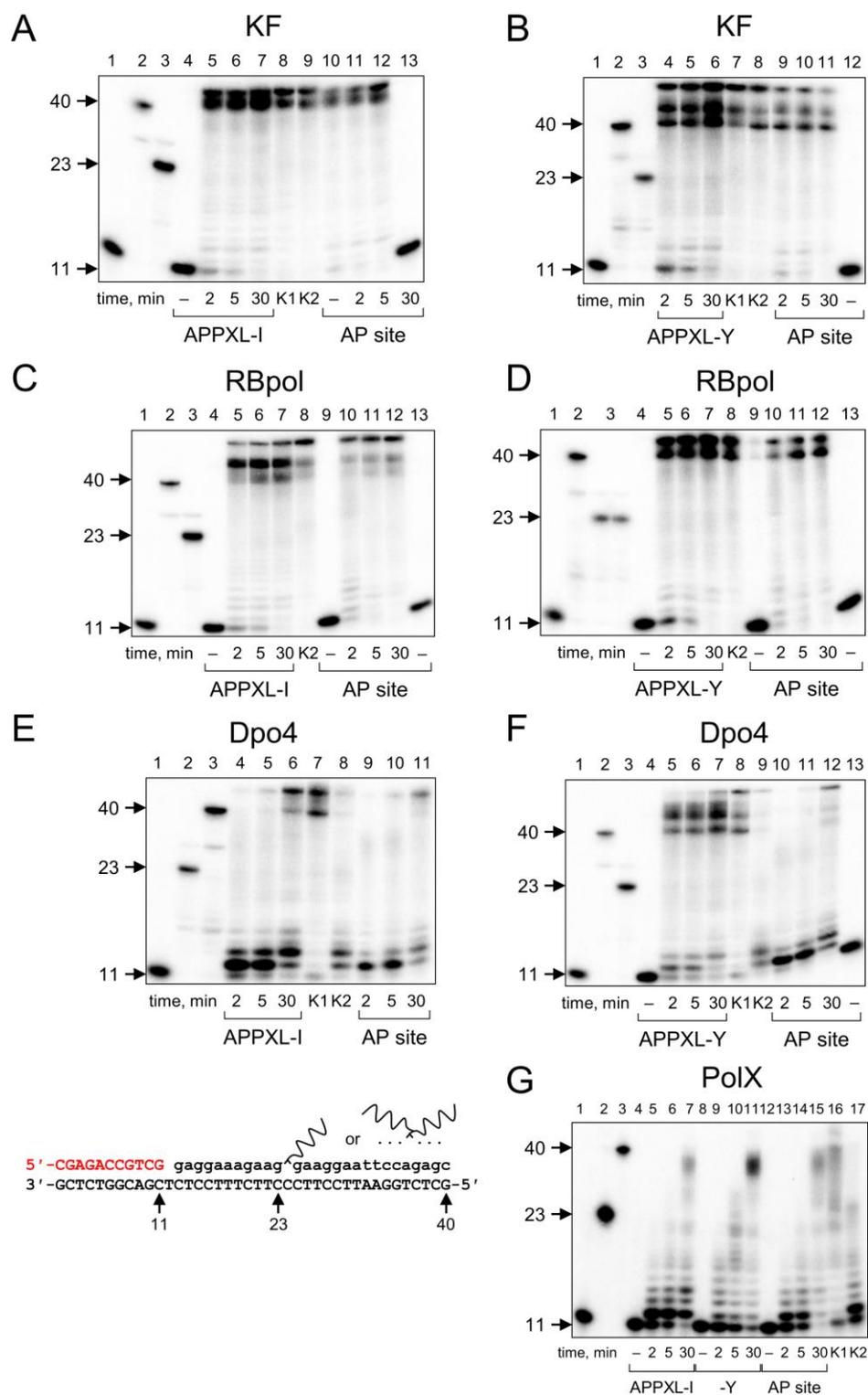
**Supplementary Figure S1.** Images of Coomassie-stained gels of the preparations of KF, RBpol, PolIX, APE1, Apn1p and Nfo used in this work. In all panels, lane 1 is a set of molecular weight markers (sizes indicated next to the leftmost panel), lane 2 is the purified protein.



**Supplementary Figure S2.** Preparation of AP-peptide cross-links. (A), (B), purification of 40-nt adducts APPXL-I (A) and APPXL-Y (B) on a denaturing 20% polyacrylamide gel. Lanes 1 and 3 in (A) and 1 in (B) contain molecular weight markers (40 nt and 28 nt). (C), (D), MALDI mass spectra of APPXL-I (C) and APPXL-Y (D). Mass spectra were obtained on a Bruker Autoflex Speed mass spectrometer equipped with a VSL-337 ND N<sub>2</sub> laser (337 nm) at 1–5 ns pulse duration, positive linear mode within the m/z range 3000–22000, laser frequency 60 Hz, ion source 1 voltage 19 kV, ion source 2 voltage 16.7 kV, lens voltage 8.75 kV, and 5000 laser shots at 200 Hz.

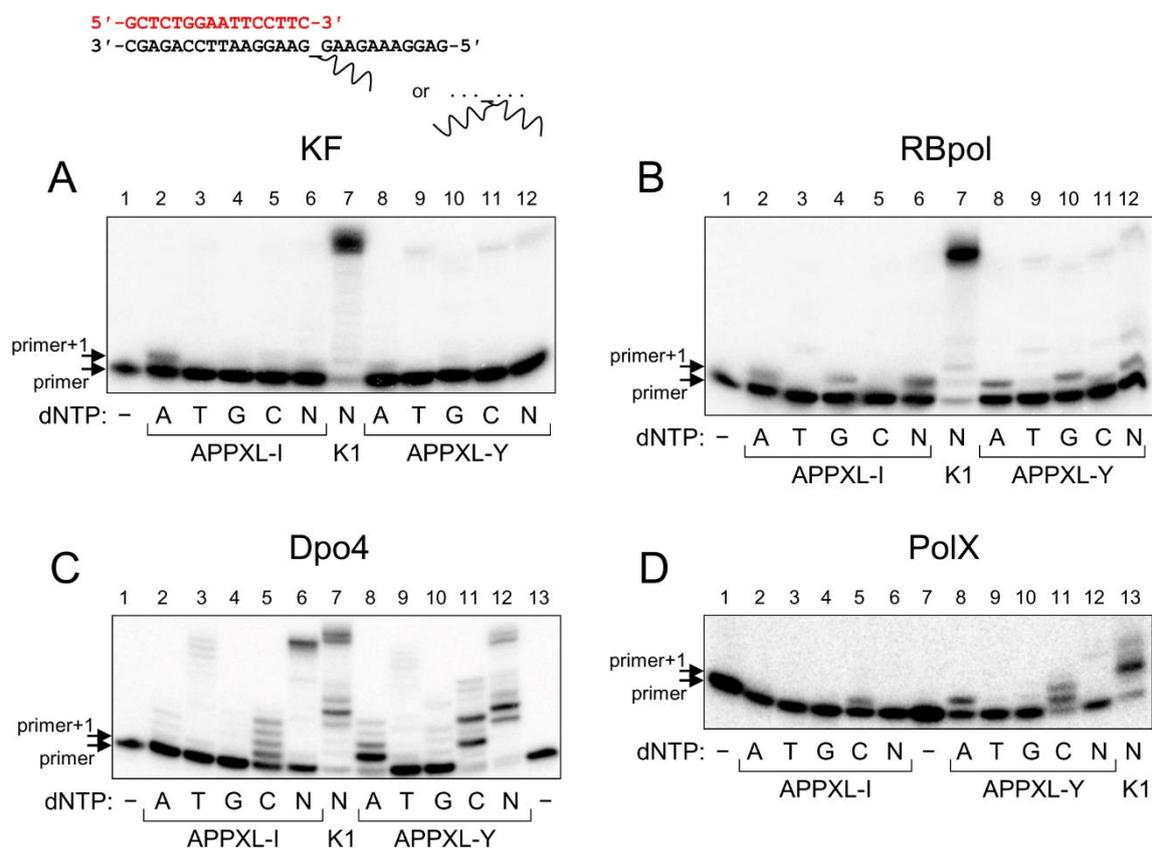


**Supplementary Figure S3.** Stalling of 3'→5' exonuclease-proficient KF by AP-peptide cross-links. A, APPXL-I, primer-*template*; B, APPXL-Y, primer-*template*; C, APPXL-I, primer-*downstream strand*-*template*; D, APPXL-Y, primer-*downstream strand*-*template*. The <sup>32</sup>P-labeled primer is shown in red, the downstream strand is in lowercase. Reaction time and the nature of the lesion are indicated under the gel images. In all panels, lanes 1–3 are size markers corresponding to the primer (11 nt), full-size product (40 nt), and primer extended to the cross-link site (23 nt). “-”, no enzyme added; K1, undamaged primer-*template*, 30 min; K2, undamaged primer-*template*-*downstream strand*, 30 min.

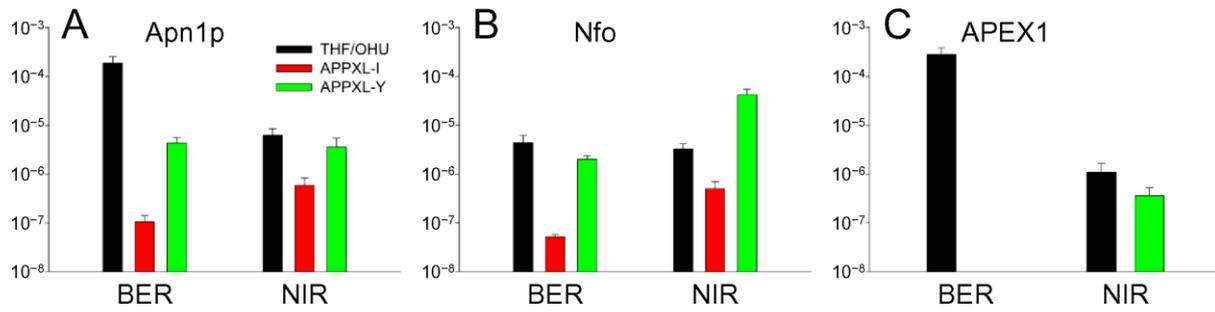


**Supplementary Figure S4.** Displacement of a downstream strand bearing an AP-peptide cross-link by DNA polymerases. A, APPXL-I, KF; B, APPXL-Y, KF; C, APPXL-I, RBpol; D, APPXL-Y, RBpol; E, APPXL-I, Dpo4; F, APPXL-Y, Dpo4; G, APPXL-I and APPXL-Y, PolX. The <sup>32</sup>P-labeled primer is shown in red, the downstream strand is in lowercase. Reaction time and the nature of the lesion are indicated under the gel images. In all panels, lanes 1–3 are size markers corresponding to the primer (11 nt), full-size product (40 nt), and primer extended to the cross-link site (23 nt). “-”, no enzyme

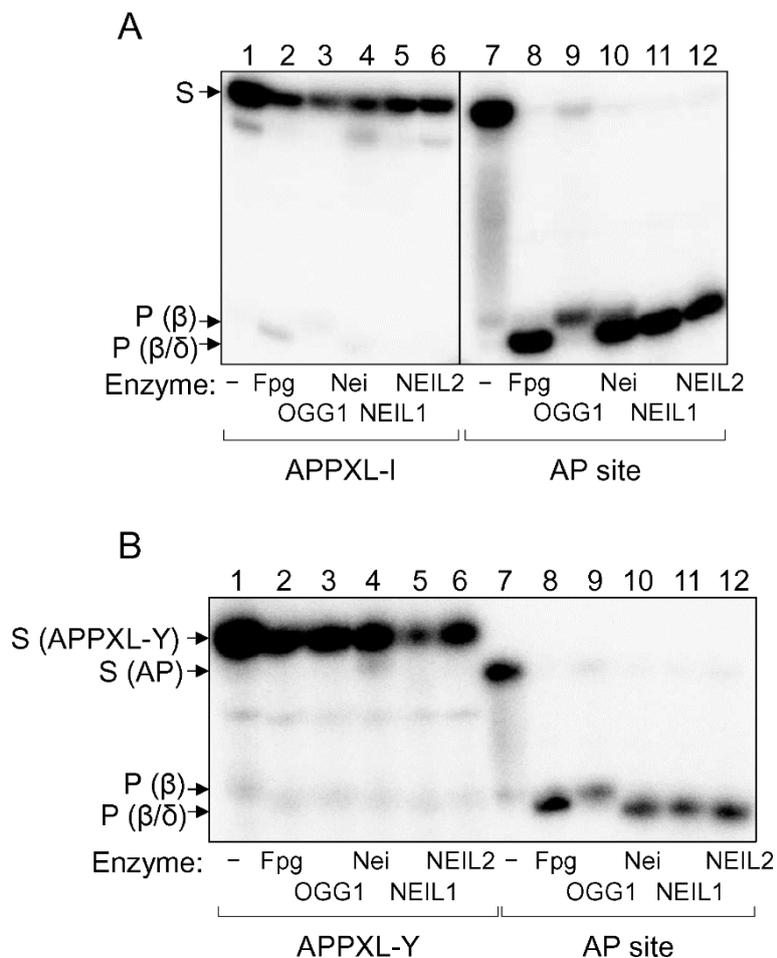
1  
2  
3 added; K1, undamaged primer–template, 30 min; K2, undamaged primer–template–downstream  
4 strand, 30 min.  
5  
6  
7  
8  
9  
10  
11  
12  
13  
14  
15  
16  
17  
18  
19  
20  
21  
22  
23  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60



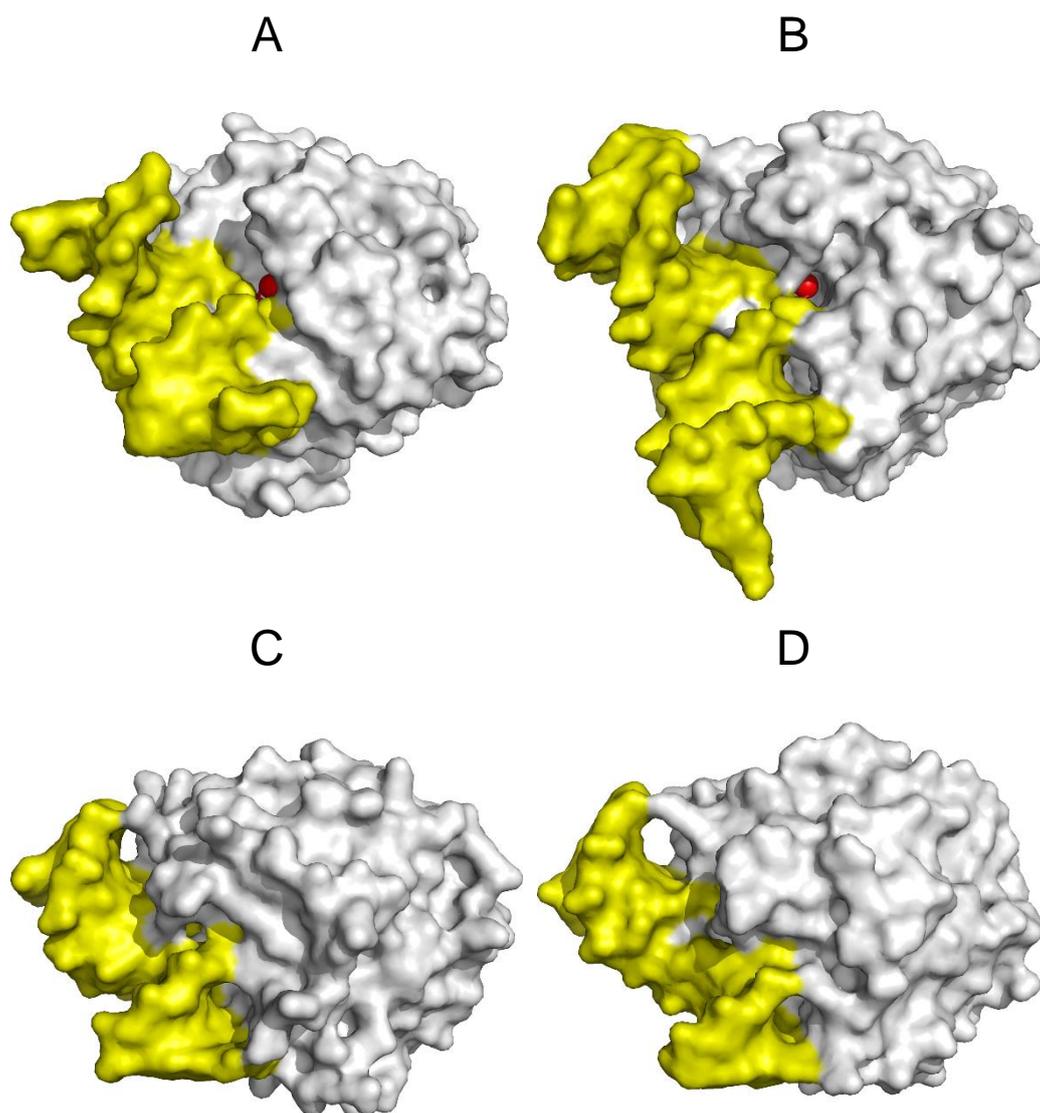
**Supplementary Figure S5.** Incorporation of dNMPs opposite APPXL-I and APPXL-Y by KF (A), RBpol (B), Dpo4 (C) and PolX (D) in a different sequence context. The nature of the DNA and dNTP substrate is indicated below the gel image. The <sup>32</sup>P-labeled primer is shown in red. "-", no enzyme or dNTPs added; N, all dNTPs added; K1, undamaged primer-*template*, all dNTPs, 30 min.



**Supplementary Figure S6.** Graphical summary of the kinetic data from Table 1 showing a comparison of the enzymes' catalytic efficiency ( $k_{cat}/K_M$ ) on different substrates under different conditions. A, Apn1p; B, Nfo; C, APEX1. Black bars correspond to the canonical substrates (THF for the BER conditions, OHU for the NIR conditions), red bars, APPXL-I, green bars, APPXL-Y.



**Supplementary Figure S7.** Resistance of APPXL-I (A) and APPXL-Y (B) to bifunctional DNA glycosylases. The nature of the enzyme and the substrate is indicated below the gel image. In both panels, lanes 1–6 are APPXL substrates, lanes 7–12, are the substrate containing a natural AP site.



**Supplementary Figure S8.** Structures of DNA-bound *E. coli* Nfo (A; 1QUM, Ref. 8), African swine fever virus AP endonuclease (B; 6KI3, Ref. 9), human APEX1 (C; 1DEW, Ref. 10) and *N. meningitides* AP endonuclease (D; 4B5F, Ref. 11). The structures are aligned to present DNA (yellow) in the same orientation. The surfaces are built with a 1.4 Å probe radius. The C1' atom of the AP site is shown as a red ball. The Figure is made in PyMol (Schrödinger, New York, NY).

### Supplementary References

1. Gros, L., Ishchenko, A.A., Ide, H., Elder, R.H. and Saparbaev, M.K. (2004) The major human AP endonuclease (Ape1) is involved in the nucleotide incision repair pathway. *Nucleic Acids Res.*, **32**, 73-81.
2. Daviet, S., Couve-Privat, S., Gros, L., Shinozuka, K., Ide, H., Saparbaev, M. and Ishchenko, A.A. (2007) Major oxidative products of cytosine are substrates for the nucleotide incision repair pathway. *DNA Repair*, **6**, 8-18.
3. Redrejo-Rodríguez, M., Vigouroux, A., Mursalimov, A., Grin, I., Alili, D., Koshenov, Z., Akishev, Z., Maksimenko, A., Bissenbaev, A.K., Matkarimov, B.T. *et al.* (2016) Structural comparison of AP endonucleases from the exonuclease III family reveals new amino acid residues in human AP endonuclease 1 that are involved in incision of damaged DNA. *Biochimie*, **128-129**, 20-33.
4. Ishchenko, A.A., Yang, X., Ramotar, D. and Saparbaev, M. (2005) The 3'→5' exonuclease of Apn1 provides an alternative pathway to repair 7,8-dihydro-8-oxodeoxyguanosine in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.*, **25**, 6380-6390.
5. Ishchenko, A.A., Sanz, G., Privezentzev, C.V., Maksimenko, A.V. and Saparbaev, M. (2003) Characterisation of new substrate specificities of *Escherichia coli* and *Saccharomyces cerevisiae* AP endonucleases. *Nucleic Acids Res.*, **31**, 6344-6353.
6. Ishchenko, A.A., Ide, H., Ramotar, D., Nevinsky, G. and Saparbaev, M. (2004)  $\alpha$ -Anomeric deoxynucleotides, anoxic products of ionizing radiation, are substrates for the endonuclease IV-type AP endonucleases. *Biochemistry*, **43**, 15210-15216.
7. Ischenko, A.A. and Saparbaev, M.K. (2002) Alternative nucleotide incision repair pathway for oxidative DNA damage. *Nature*, **415**, 183-187.
8. Hosfield, D.J., Guan, Y., Haas, B.J., Cunningham, R.P. and Tainer, J.A. (1999) Structure of the DNA repair enzyme endonuclease IV and its DNA complex: Double-nucleotide flipping at abasic sites and three-metal-ion catalysis. *Cell*, **98**, 397-408.
9. Chen, Y., Chen, X., Huang, Q., Shao, Z., Gao, Y., Li, Y., Yang, C., Liu, H., Li, J., Wang, Q. *et al.* (2020) A unique DNA-binding mode of African swine fever virus AP endonuclease. *Cell Discov.*, **6**, 13.
10. Mol, C.D., Izumi, T., Mitra, S. and Tainer, J.A. (2000) DNA-bound structures and mutants reveal abasic DNA binding by APE1 and DNA repair coordination. *Nature*, **403**, 451-456.
11. Lu, D., Silhan, J., MacDonald, J.T., Carpenter, E.P., Jensen, K., Tang, C.M., Baldwin, G.S. and Freemont, P.S. (2012) Structural basis for the recognition and cleavage of abasic DNA in *Neisseria meningitidis*. *Proc. Natl Acad. Sci. U.S.A.*, **109**, 16852-16857.