Supporting Information

High-Resolution Mapping of Amino Acid Residues in DNA-Protein Cross-Links Enabled by Ribonucleotide-Containing DNA

Jin Tang[†], Wenxin Zhao[†], Nathan G. Hendricks[‡], Linlin Zhao[†§]*

[†]Department of Chemistry and [§]Environmental Toxicology Graduate Program, University of California, Riverside, Riverside, California 92521, United States

[‡]Proteomics Core, Institute for Integrative Genome Biology, University of California, Riverside, Riverside, CA 92521, United States

Corresponding Author: E-mail: linlin.zhao@ucr.edu. Tel: +1 951-827-9081

Table of Content

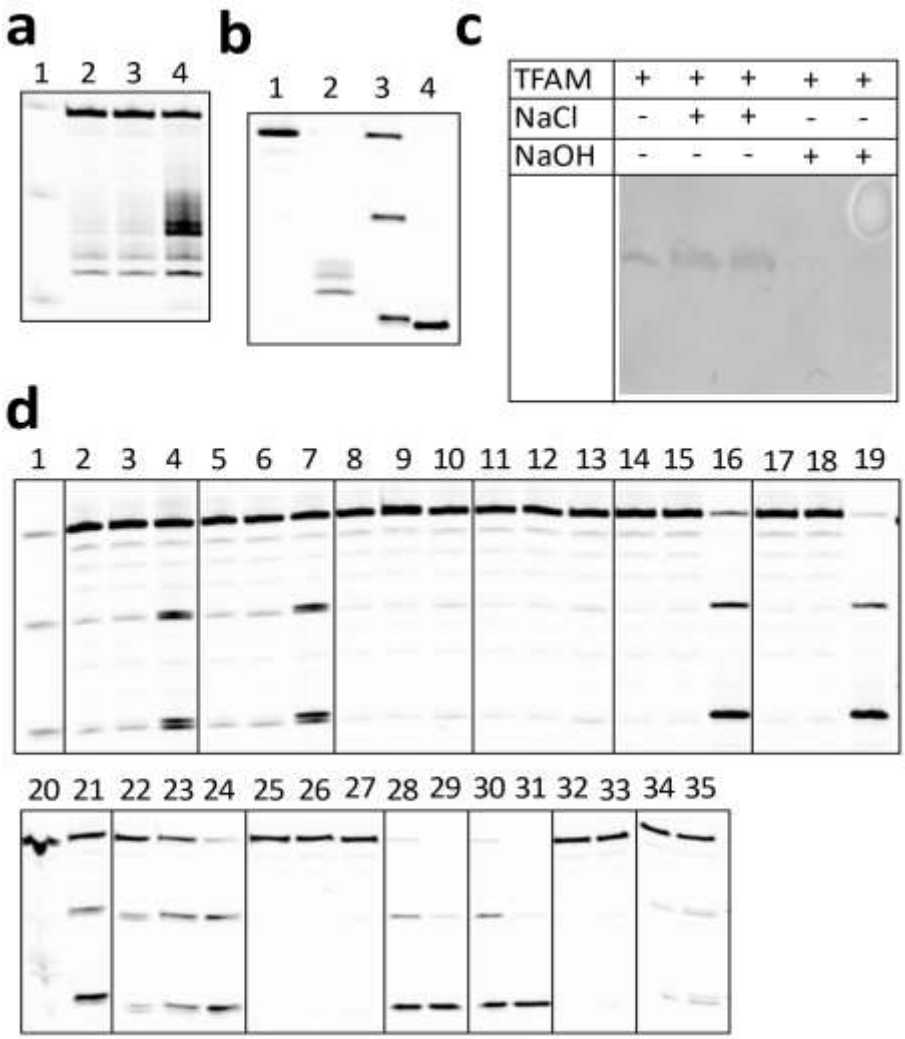

Figure S1. Optimization of ribonucleotide cleavage conditions. (a) D4 cleavage reaction with RNase HII. The reaction was performed in 50 mM Tris pH 8.0, 50 mM NaCl, 1 mM DTT, 0.1 mg/mL BSA, 5% glycerol, and 10 mM $MgCl_2$. And 1 uM double-strand D4 was incubated with 0.1 uM RNase HII. The incubation time point of 1, 3 and 21 hours were taken (lanes 2, 3, and 4, respectively). Lane 1 is the strand break marker cleaved at two ribonucleosides. The sample was prepared by incubating D4 with 0.3 M NaOH at 55°C for 10 min. (b) D4 cleavage reaction with 0.3 M NaOH at 55°C. Lane 1 is D4; lane 2 is the strand break of D4 at the AP site; lane 3 is the strand break at two ribonucleoside sites maker by incubating D4 at 0.3 M NaOH, 55°C for 10 min. Lane 4 is the product of D4 incubated at 0.3 M NaOH for 2 hours. (c) Incubation of TFAM at 55°C with either 0.3 M NaCl or 0.3 M NaOH, two replicates. (d) The reaction conditions to cleave the ribonucleotides on D4. Reaction conditions are listed in Table S2. Lane 20 is D4. Lane 1 and Lane 21 are the strand break standards made by incubating D4 with 0.3 M NaOH for 10 min.

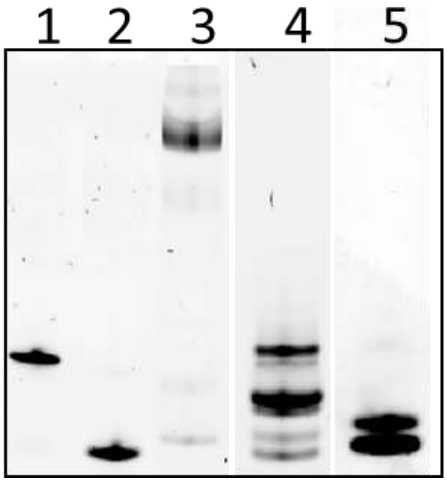

Figure S2. Preparation of TFAM-DNA cross-links monitored by gel electrophoresis. Lane 1 is D3. Lane 2 is a strand break marker at the AP site derived from D2. Lane 3 is D2 DPC generated in the presence of $NaCNBH_3$. Lane 4 is trypsin digested D2 DPC. Lane 5 is the product of trypsin digested D2 DPC after Cleave R reaction.

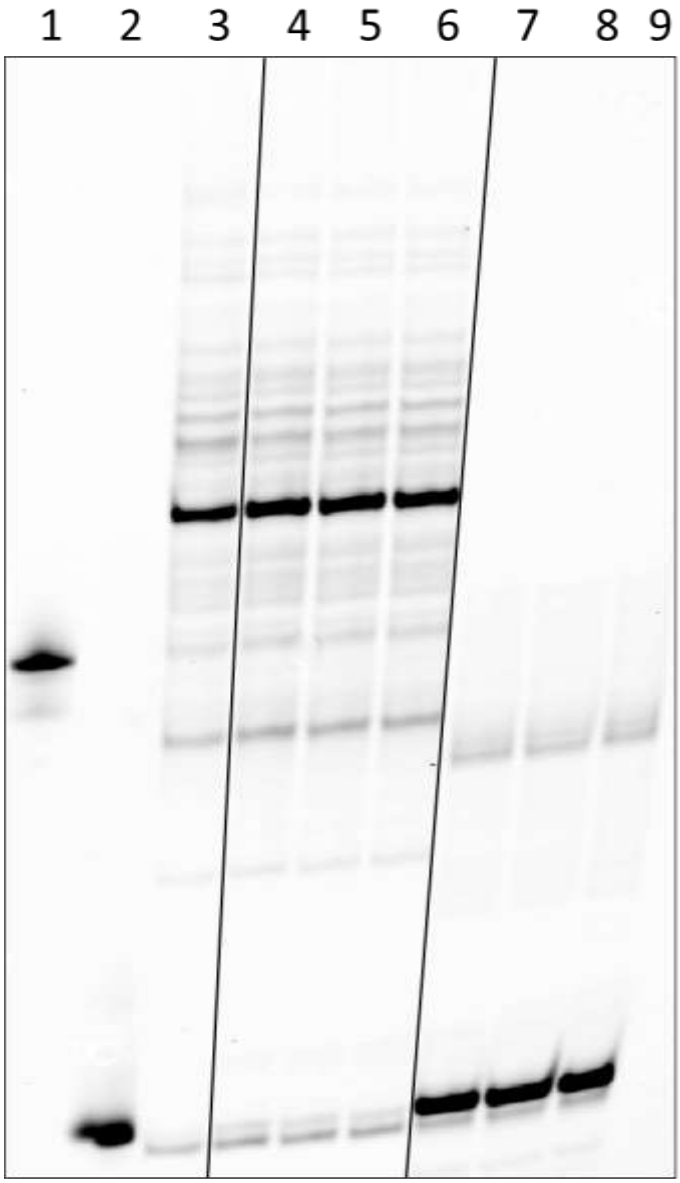

Figure S3. Stability of D2-peptide cross-links from trypsin digestion of D2-TFAM DPC under Cleave R condition and NaOH treatment. Lane 1, D2. Lane 2, cleaved single-stranded breaks at AP sites from D2 after NaOH treatment. Lane 3, trypsin-digested D2-TFAM DPC without treatment. Lane 4-6 (triplicate) treatment of digested D2-TFAM DPC under Cleave R condition for 20 h. Lane 7-9 (triplicate) treatment of digested D2-TFAM DPC with 0.3 M NaOH at 55°C for 2 hours.

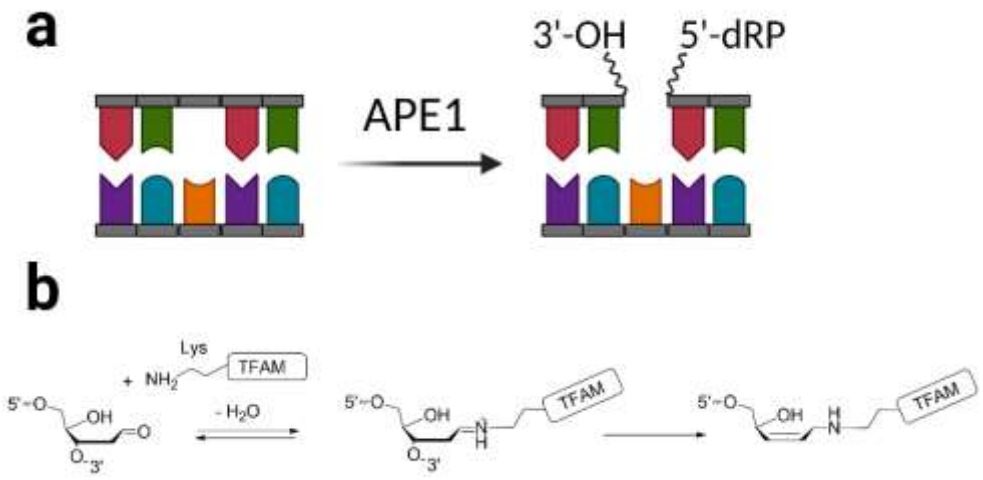

Figure S4. (a) APE1 cleaves the DNA backbone 5' of the AP lesion to yield a 3'-OH terminal and a 5'-deoxyribose phosphate residue.[1] (b) TFAM(peptide)-AP-DNA cross-links undergo β-elimination before reductive amination by $NaBH_3CN$ to afford a mass adduct of 102.

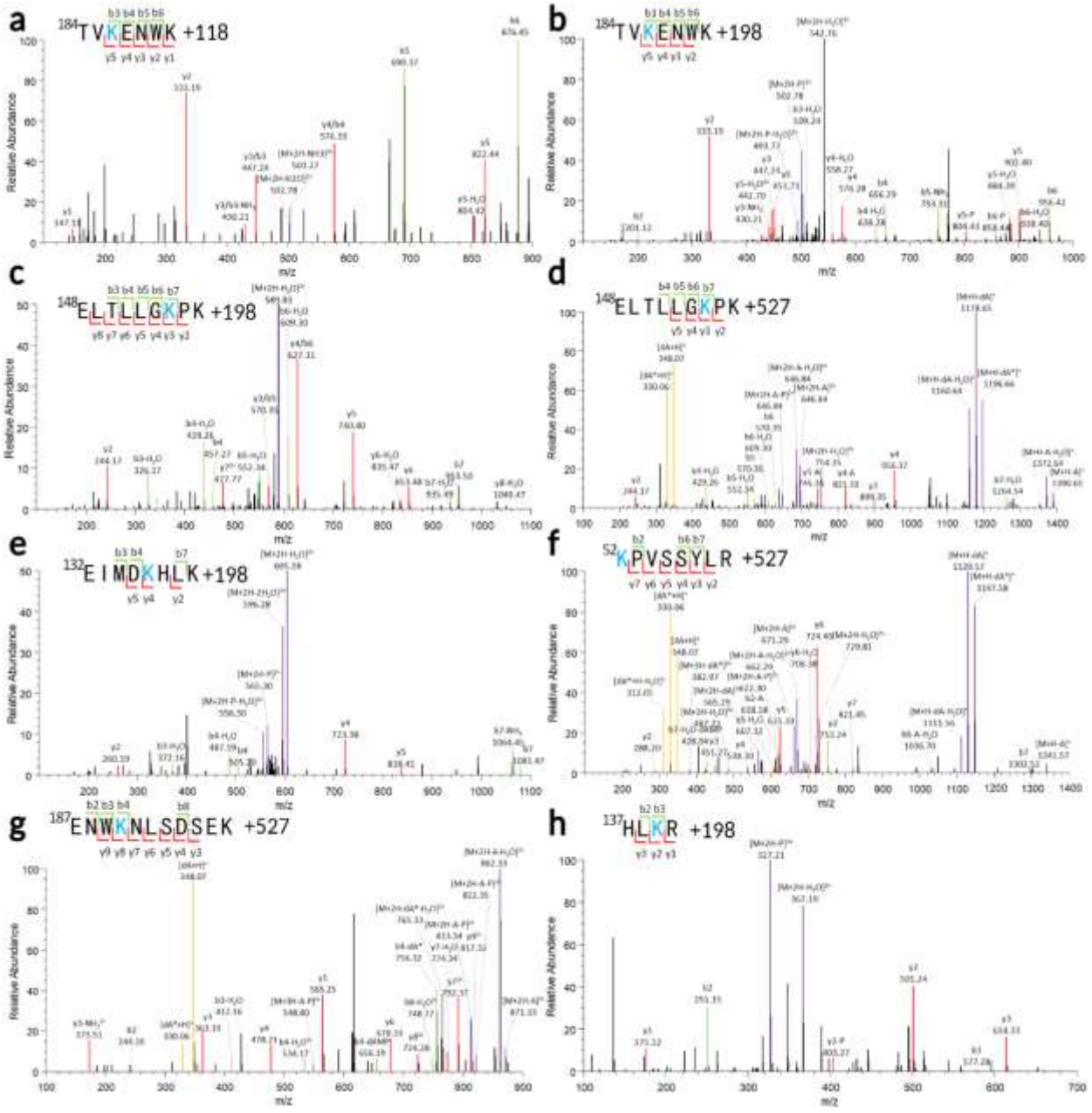

Figure S5. The MS2 spectra of the crosslinks identified in this study, part 1.

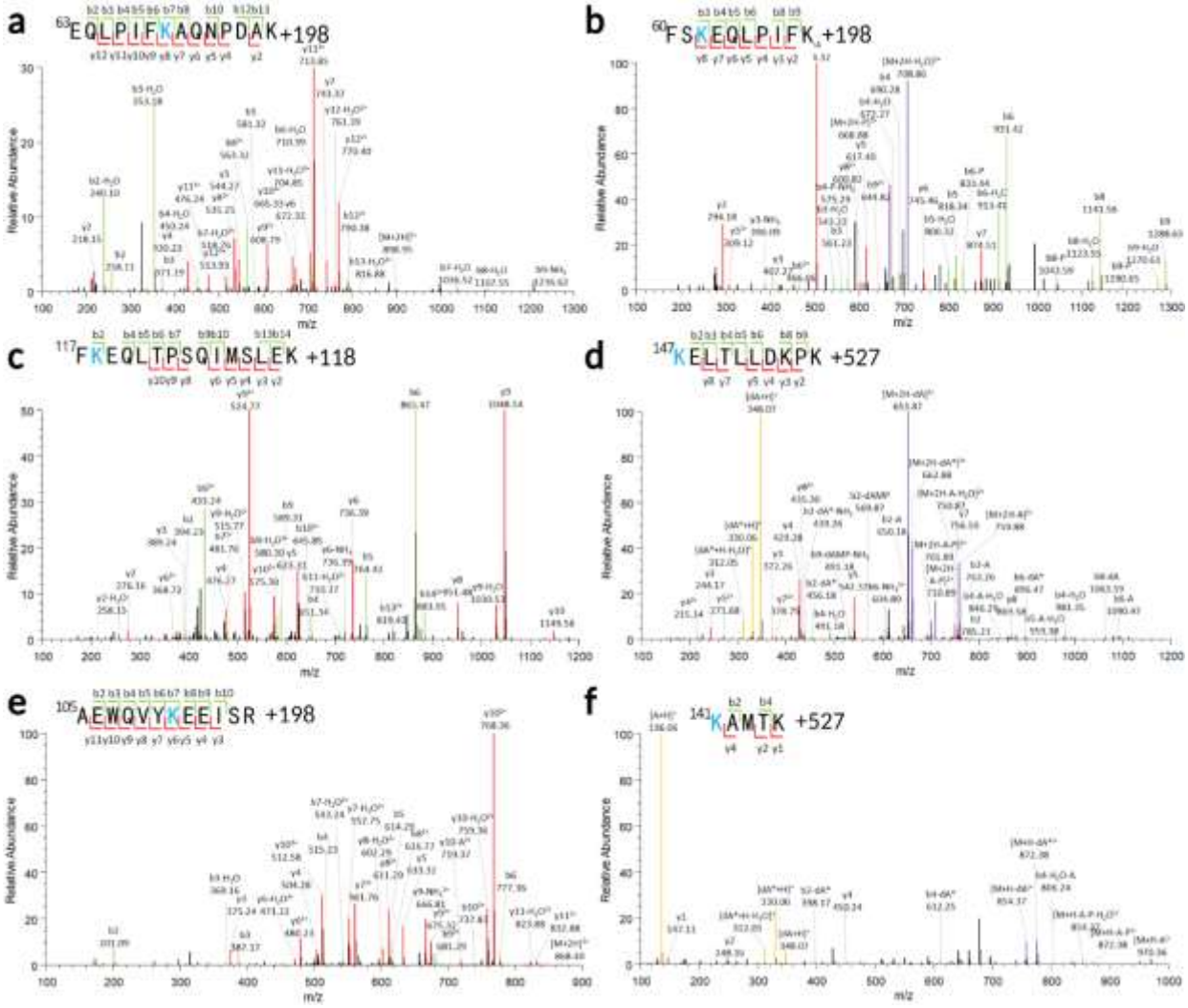

Figure S6. MS2 spectra of the crosslinks identified in this study, part 2.

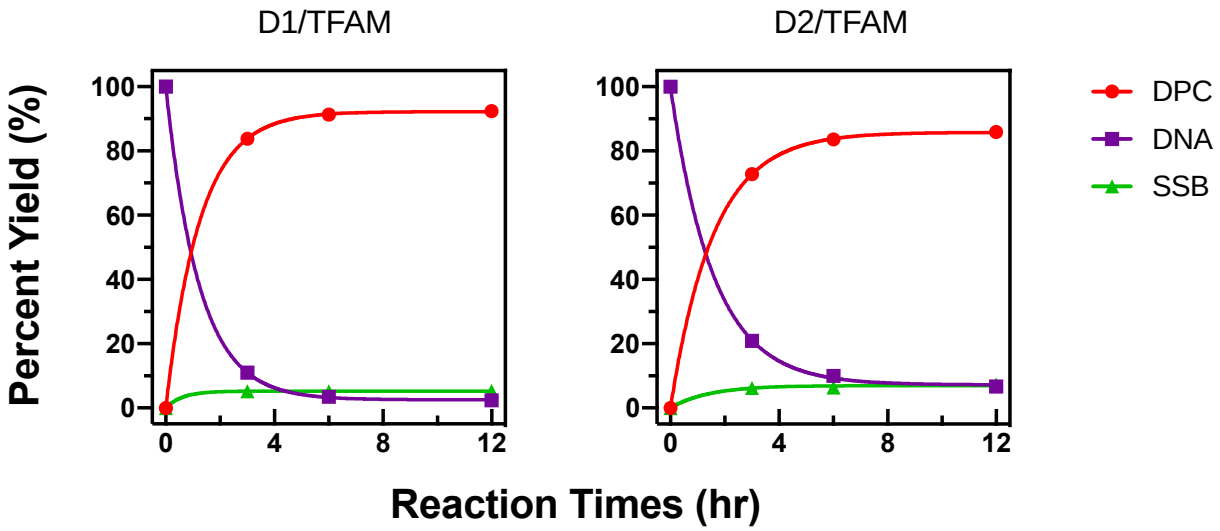

Figure S7. Quantification of gel analysis of TFAM:AP-ODN reactions shown in Figure 2e. D1 is a double-strand (ds) DNA with an AP modification located on one strand, as illustrated in Figure 2b. D2 contains the identical nucleotide sequences to D1, except that it has two neighboring (of the AP lesion) deoxynucleotides substituted with ribonucleotides. The percent yield of each species was fit to a single exponential equation to obtain the apparent formation rate (k_{fo}) of DNA-protein cross-links (DPC) and the cleavage rate (k_{cl}) of AP-ODN. For reactions with D1 and TFAM, the percent yield of DPC is 92%, k_{fo} is 0.79 hr^{-1}, and k_{cl} is 0.81 hr^{-1}. For reactions with D2 and TFAM, the percent yield of DPC is 86%, k_{fo} is 0.63 hr^{-1}, and k_{cl} is 0.63 hr^{-1}.

Table S1. Sequences of the DNA substrates used in this study. For simplicity, substrates D1-D5 denote dsDNA annealed to a complementary strand. X indicates the AP site, and U is deoxyuridine. The underlined nucleotides denote ribonucleotides.

Name	**Sequence (5'-3')**
D1	FAM - TAA CAG TCA CCC CCC XAC TAA C
D2	FAM - TAA CAG TCA CCC CCrC XrAC TAA C
D3	FAM - TAA CAG TCA CCC CCrC UrAC TAA C
D4	FAM - TAA CAG TCA CCC CrCC UArC TAA C
D5	FAM - TAA CAG TCA CCC CrCC XArC TAA C
Complementary strand	GTT AGT TGG GGG GTG ACT GTT A

Table S2. Cleavage reaction conditions in Figure S1d

Incubation condition	Length of incubation (h)	Lane number
pH=10, [K+]=0.03 M, [Mg2+]=5 mM, T=55°C	1	2
	2	3
	18	4
pH=10, [K+]=0.01 M, [Mg2+]=5 mM, T=55°C	1	5
	2	6
	18	7
pH=10, [K+]=0.03 M, [Mg2+]=5 mM, T=37°C	1	8
	2	9
	18	10
pH=10, [K+]=0.01 M, [Mg2+]=5 mM, T=37°C	1	11
	2	12
	18	13
pH=10, [K+]=0.03 M, T=55°C	1	14
	2	15
	18	16
pH=10, [K+]=0.01 M, T=55°C	1	17
	2	18
	18	19
pH=13.4, [K+]=0.001 M, T= room temperature	2	22
	8	23
	24	24
pH=12, [K+]=0.001M, [Mg2+]=15 mM, T=37°C	2	25
	8	26
	24	27
pH=10, glycine-NaOH buffer, [K+]=0.001 M, [Mg2+]=15 mM, T=55°C	8	28
	24	29
pH=10, glycine-NaOH buffer, [Mg2+]=15 mM, T=55°C	8	30
	24	31
pH=10, NaOH solution, [K+]=0.02 M, [Mg2+]=5 mM, T=55°C	8	32
	24	33
pH=10, NaOH solution, [K+]=0.02 M, T=55°C	8	34
	24	35

Table S3. Identified lysine residues of TFAM cross-linked with the AP lesion

Ion m/z, charge state	Peak area	Starting residue	Crosslinked peptide	Crosslinked residue	Charge	m/z	mass adduct	Δm [ppm]	Note
492.8825, z=3	12327082	52	KPVSSYLR	K52	3	492.8825	527.0818	2.37	
738.8200, z=2	5119750	52	KPVSSYLR	K52	2	738.8200	527.0818	2.6	
478.9149, z=3	14292681	60	FSKEQLPIFK	K62	3	478.9149	198.0293	0.73	
717.8683, z=2	26719839	60	FSKEQLPIFK	K62	2	717.8683	198.0293	0.54	
599.6343, z=3	23513831	63	EQLPIFKAQNPDAK	K69	3	599.6343	198.0293	2.37	
898.9478, z=2	18528305	63	EQLPIFKAQNPDAK	K69	2	898.9478	198.0293	2.67	
709.6531, z=3	1.01E+08	63	EQLPIFKAQNPDAK	K69	3	709.6531	527.0818	0.17	C13
1063.9760, z=2	48499082	63	EQLPIFKAQNPDAK	K69	2	1063.9760	527.0818	0.42	C13
552.2747, z=3	87499742	70	AQNPDAKTTELIR	K76	3	552.2747	198.0293	2.25	
827.9083, z=2	1.1E+08	70	AQNPDAKTTELIR	K76	2	827.9083	198.0293	2.44	
661.9597, z=3	4.65E+08	70	AQNPDAKTTELIR	K76	3	661.9597	527.0818	3.13	
992.4349, z=2	3.04E+08	70	AQNPDAKTTELIR	K76	2	992.4349	527.0818	2.39	
552.6153, z=3	23470686	105	AEWQVYKEEISR	K111	3	552.6153	118.0630	1.65	
828.4192, z=2	18982627	105	AEWQVYKEEISR	K111	2	828.4192	118.0630	1.83	
579.2705, z=3	47453109	105	AEWQVYKEEISR	K111	3	579.2705	198.0293	1.16	
868.4026, z=2	62176994	105	AEWQVYKEEISR	K111	2	868.4026	198.0293	2.04	
688.9551, z=3	1.57E+08	105	AEWQVYKEEISR	K111	3	688.9551	527.0818	1.61	
1032.9289, z=2	92505104	105	AEWQVYKEEISR	K111	2	1032.9289	527.0818	1.76	

638.3371, z=3	33482050	117	FKEQLTPSQIM^{OX}SLEK	K118	3	638.3371	118.0630	1.07	
774.6785, z=3	55473176	117	FKEQLTPSQIM^{OX}SLEK	K118	3	774.6785	527.0818	3.2	
1161.5138, z=2	7752968	117	FKEQLTPSQIM^{OX}SLEK	K118	2	1161.5138	527.0818	3.18	
404.5302, z=3	46823451	132	EIMDKHLK	K136	3	404.5302	198.0293	0.67	
606.2918, z=2	52592837	132	EIMDKHLK	K136	2	606.2918	198.0293	0.35	
409.8621, z=3	33493714	132	EIM^{OX}DKHLK	K136	3	409.8621	198.0293	1.32	
614.2893, z=2	27462866	132	EIM^{OX}DKHLK	K136	2	614.2893	198.0293	1.42	
519.5467, z=3	1.56E+08	132	EIM^{OX}DKHLK	K136	3	519.5467	527.0818	1.87	
778.8163, z=2	1.11E+08	132	EIM^{OX}DKHLK	K136	2	778.8163	527.0818	2.08	
514.2148, z=3	3.13E+08	132	EIMDKHLK	K136	3	514.2148	527.0818	1.39	
770.8183, z=2	2.35E+08	132	EIMDKHLK	K136	2	770.8183	527.0818	1.39	
376.1973, z=2	27329832	137	HLKR	K139	2	376.1973	198.0293	0.77	
553.2132, z=2	1.19E+08	141	KAMTK	K141	2	553.2132	527.0818	3.38	
1105.4186, z=1	2947804	141	KAMTK	K141	1	1105.4186	527.0818	3.4	
631.2629, z=2	21301011	140	RKAMTK	K141	2	631.2629	527.0818	1.6	
551.9396, z=3	13746344	147	KELTLLGKPK	K147	3	551.9396	527.0818	1.16	
827.4062, z=2	5595916	147	KELTLLGKPK	K147	2	827.4062	527.0818	2.03	
558.8481, z=2	8293783	148	ELTLLGKPK	K154	2	558.8481	118.0630	0.82	
598.8319, z=2	1.68E+08	148	ELTLLGKPK	K154	2	598.8319	198.0293	1.85	
1196.6560, z=1	9477711	148	ELTLLGKPK	K154	1	1196.6560	198.0293	1.87	

509.2416, z=3	1.61E+08	148	ELTLLGKPK	K154	3	509.2416	527.0818	1.89	
763.3587, z=2	4.59E+08	148	ELTLLGKPK	K154	2	763.3587	527.0818	2.17	
511.783, z=2	15554123	184	TVKENWK	K186	2	511.7830	118.0630	6.33	
483.1944, z=3	5931178	184	TVKENW^{OX}+16K	K186	3	483.1944	527.0818	1.49	
551.7642, z=2	3.41E+08	184	TVKENWK	K186	2	551.7642	198.0293	2.34	
1102.5, z=1	23504853	184	TVKENWK	K186	1	1102.5213	198.0293	3	
559.7613, z=2	1.16E+08	184	TVKENW^{OX}+16K	K186	2	559.7613	198.0293	1.68	
477.8629, z=3	3.4E+08	184	TVKENWK	K186	3	477.8629	527.0818	1.78	
716.7882, z=2	1.15E+09	184	TVKENWK	K186	2	716.7882	527.0818	-6.78	C13
503.7857, z=2	37482775	184	TVKENWK	K186	2	503.7857	102.0681	6.73	
626.2455, z=3	4434088	187	ENWKNLSDSEK	K190	3	626.2455	527.0818	3.11	
938.8653, z=2	6038440	187	ENWKNLSDSEK	K190	2	938.8653	527.0818	4.14	
421.7237, z=2	46868679	228	KDLLR	K228	2	421.7237	198.0293	1.52	
842.4403, z=1	3479367	228	KDLLR	K228	1	842.4403	198.0293	2.39	
586.2503, z=2	1.61E+08	228	KDLLR	K228	2	586.2503	527.0818	1.69	
1171.4922, z=1	5252689	228	KDLLR	K228	1	1171.4922	527.0818	1.2	

The crosslinked amino acid residues are in blue, and oxidative modifications are denoted by OX in red. Methionine oxidation has the mass adduct of 16; tryptophan oxidation has the mass shift of +4, +16, +20, and +32,[2] labeled the mass adduct of tryptophan oxidation.

References

(1) López, D. J.; Rodríguez, J. A.; Bañuelos, S. Molecular Mechanisms Regulating the DNA Repair Protein APE1: A Focus on Its Flexible N-Terminal Tail Domain. *Int. J. Mol. Sci.* **2021**, *22* (12), 6308. https://doi.org/10.3390/ijms22126308.

(2) Lam, X. M.; Lai, W. G.; Chan, E. K.; Ling, V.; Hsu, C. C. Site-Specific Tryptophan Oxidation Induced by Autocatalytic Reaction of Polysorbate 20 in Protein Formulation. *Pharm. Res.* **2011**, *28* (10), 2543–2555. https://doi.org/10.1007/s11095-011-0482-x.