Supporting Information for:

Correlation of clustered abasic DNA lesion processing with structural and dynamic DNA helix distortion

Emmanuelle Bignon[1,2], Hugo Gattuso[3,4], Christophe Morell[2], Francois Dehez[4,] Alexandros G. Georgakilas[5,*], Antonio Monari[3,4,*], Elise Dumont[1,*]

[1]Univ Lyon, Ens de Lyon, CNRS, Université Lyon 1, Laboratoire de Chimie UMR 5182, F-69342, Lyon, France, [2]Institut des Sciences Analytiques, Université de Lyon 1 and CNRS, Villeurbanne France, [3]Université de Lorraine -Nancy, Theory-Modeling-Simulation SRSMC, Vandoeuvre-lès-Nancy, France, [4]CNRS, Theory-Modeling-Simulation SRSMC, Vandoeuvre-lès-Nancy, [5]Physics Department, School of Applied Mathematical and Physical Sciences, National Technical University of Athens (NTUA), Zografou 15780, Athens, Greece

* To whom correspondence should be addressed: elise.dumont@ens-lyon.fr
*Correspondence may also be addressed to: antonio.monari@univ-lorraine.fr
*Correspondence may also be addressed to: alexg@mail.ntua.gr

The authors wish it to be known that, in their opinion, the first 2 authors should be regarded as joint First Authors

Tautomerism of the abasic site

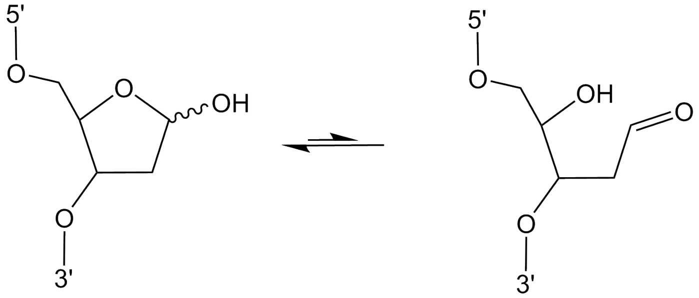

Figure S1 Illustration of the 99:1 equilibrium between the open and closed form

Hydrogen bond between the ejected base and DNA backbone

Molecular dynamics setup

All classical molecular dynamics simulations were performed using the Amber12 suite of programs [1], using ff99bsc0 [2] and gaff [3] force fields. Since no experimental structure is available for sequences used by Georgakilas et al. [4], we built the corresponding 23bp oligonucleotides with the nab module. RESP charges [5] and parameters were generated for the AP site according to the standard Antechamber protocol (see table S1). Then, the system was placed in an octahedral TIP3P [6] water box of 12A buffer, with 44 sodium cations to ensure the medium's neutrality. For each sequence, 400ns classical molecular dynamics were performed. A 10000 steps minimization, including 5000 steps of steepest descent, was first performed in order to adapt our system to the force field. Then, temperature was increased from 0K to 300K in a 20ps themalization run. Langevin thermostat with a 1.0 ps^{-1} collision frequency γln was used to keep it constant during the remaining of the simulation. Then, a first 100ps equilibration run in NPT was performed, followed by a second one in NVT conditions. Finally, a 400ns production was done in the NPT thermodynamic ensemble.

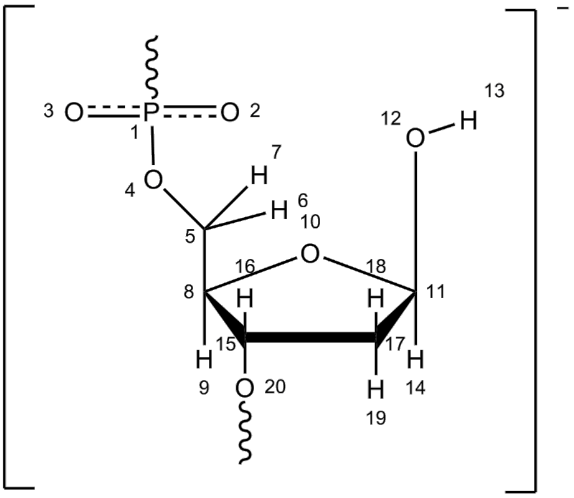

Figure S2: Numbering convention for the constituting atoms of the AP site

	Name	Type	Charge
1	P	P	1.1659
2	OP2	O2	-0.7761
3	OP1	O2	-0.7761
4	O5’	OS	-0.4954
5	C5’	CI	-0.0069
6	H5’	H1	0.0754
7	H5”	H1	0.0754
8	C4’	CT	0.2098
9	H4’	H1	0.0046
10	O4’	OS	-0.4058
11	C1’	CT	0.3679
12	O1	OH	-0.5959
13	H1	HO	0.3761
14	H1’	HC	0.02
15	C3’	CT	0.3259
11	H3’	HC	0.0562
12	C2’	CT	-0.286
13	H2’	HC	0.0941
14	H2”	HC	0.0941
15	O3’	OS	-0.5232

Table S1: Atomic charges (in a.u.) and type specification for the AP site see Figure S3 for the numbering

Minor groove occupation of the etrahelical AP site

The minor groove occupation was estimated considering the diehedral angle between the C4'-C5' (on the base at n+2) and O4'-C1' (on the AP site) see Figure S4. A positive value for the angle indicates minor groove occupancy.

		% minor groove occupancy
seq0	Ap12	92.9
	Ap35	99.9
seq+1	Ap13	17.4
	Ap35	14.4
seq-1	Ap11	18.1
	Ap35	7.9
seq+3	Ap15	76.4
	Ap35	16.0
seq-3	Ap9	92.5
	Ap35	73.4
seq+5	Ap17	28.0
	Ap35	21.5
seq-5	Ap7	39.1
	Ap35	41.1
seqC	Ap12	6.4

Table S2 Minor groove occupancy of the extrahelical AP sites

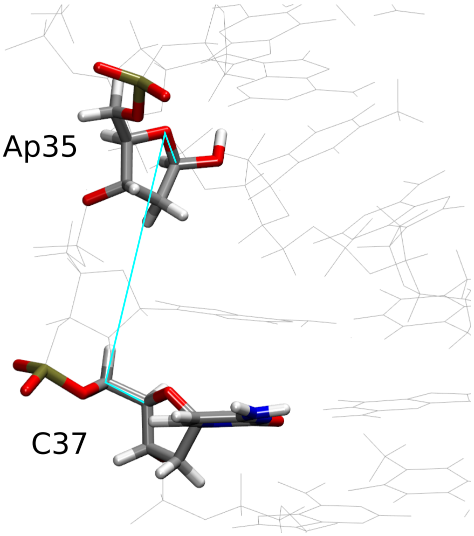

Figure S3 representation of the atoms used to calculated the dihedral angle and estimate minor groove occupancy.

Crucial distances reported in Figure 2-4 in Main text.

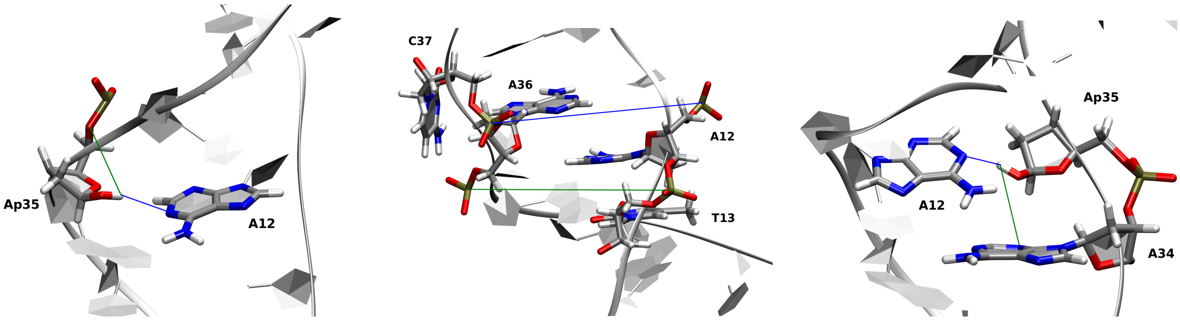

Figure S4 Cartoon representation reported in the tree panels of Figure 2 c) in the main text

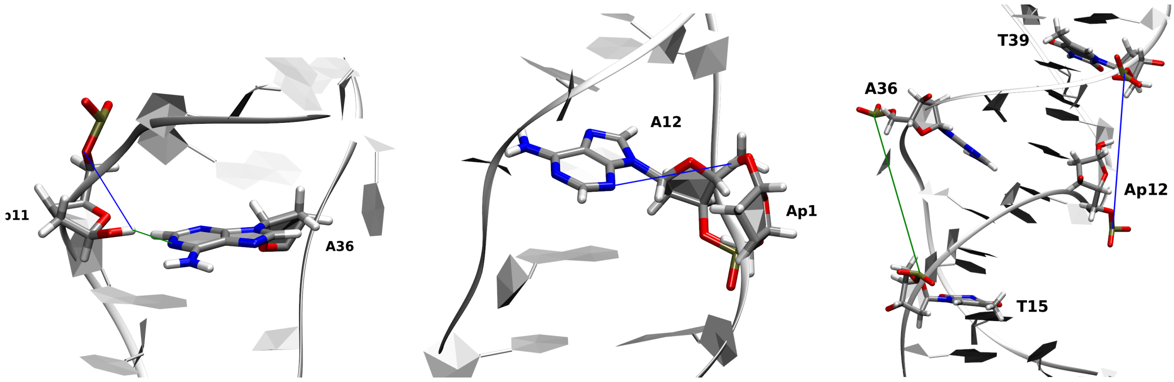

Figure S5 Cartoon representation reported in the tree panels of Figure 3 c) in the main text

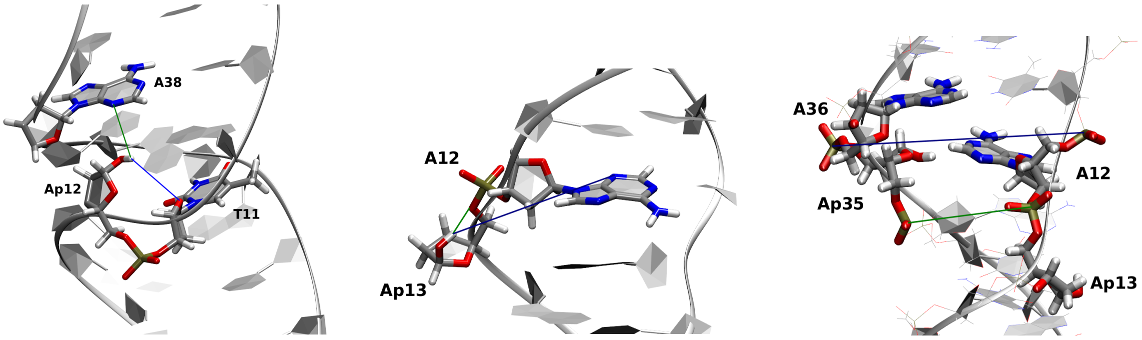

Figure S6 Cartoon representation reported in the tree panels of Figure 4 c) in the main text

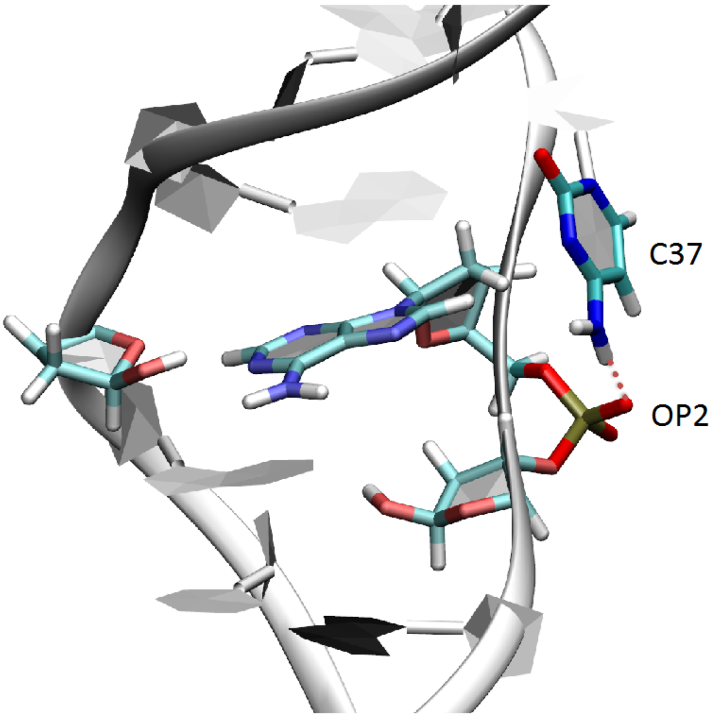

Figure S7 Representative MD snapshot showing the occurence of an HB between the ejected C37 pyrimidine and the backbone oxygen. The latter locks C37 in an extrahelical position.

Choice of the specific oligonucleotide sequence

The choice for the specific oligonucleotide sequence was based on the existent results for DNA AP cluster processing for both enzymes Nfo and APE1 as described in the original publications. Additional reasons for this was that the specific sequence as searched initially by Dr. Georgakilas did not form any hairpin loops and it was quite stable in solution even after the induction of the AP sites. For the control DNA sequence, current calculations agreeing with the older ones, gave an approximate melting temperature of ~51.7 ^{o}C rising to ~59.2 ^{o}C for 50 Mm Na^+ salt adjusted [1]. For more details on Tm calculations please visit :
http://biotools.nubic.northwestern.edu/OligoCalc.html

In addition using BLAST, a sequence alignment analysis revealed similarity for this oligonucleotide sequence with several primates like Macaca and Human chromosomes 1, 6, 14. Some details are shown below:

BLASTN 2.4.0+ Sequence alignment

http://blast.ncbi.nlm.nih.gov/Blast.cgi?CMD=Get&RID=TM4RT4FT015

5′-AGAGGATATGTXTGTATGGAGAG-3′
3′- TCTCCTATACATACATACCTCTC-5’

TM4RT4FT015 (Expires on 07-29 22:13 pm)
Query ID **lcl|Query_129577**
Molecule type: nucleic acid
Query Length: 23

Homo sapiens mitochondrial pyruvate carrier 1 (MPC1), RefSeqGene on chromosome 6

Sequence ID: ref|NG_032888.1|Length: 25095Number of Matches: 1

Related Information

Map Viewer-aligned genomic context
Range 1: 16249 to 16268GenBankGraphics

Alignment statistics for match #1

Score	**Expect**	**Identities**	**Gaps**	**Strand**
37.4 bits(40)	2.1	20/20(100%)	0/20(0%)	Plus/Minus

```
Query  1      TCTCCTATACATACATACCT  20
              ||||||||||||||||||||
Sbjct  16268  TCTCCTATACATACATACCT  16249
```

Homo sapiens AT-rich interaction domain 4B (ARID4B), RefSeqGene on chromosome 1

Sequence ID: ref|NG_029484.1|Length: 168323Number of Matches: 1

Related Information

Map Viewer-aligned genomic context
Range 1: 115262 to 115281GenBankGraphicsNext MatchPrevious Match

Alignment statistics for match #1

Score	Expect	Identities	Gaps	Strand
37.4 bits(40)	2.1	20/20(100%)	0/20(0%)	Plus/Minus

```
Query   3       TCCTATACATACATACCTCT   22
                ||||||||||||||||||||
Sbjct   115281  TCCTATACATACATACCTCT   115262
```

Homo sapiens FOSMID clone ABC8-42662600P1 from chromosome 4, complete sequence

Sequence ID: gb|AC234090.4|Length: 39078Number of Matches: 1

Related Information

Map Viewer-aligned genomic context
Range 1: 18162 to 18181GenBankGraphicsNext MatchPrevious Match

Alignment statistics for match #1

Score	Expect	Identities	Gaps	Strand
37.4 bits(40)	2.1	20/20(100%)	0/20(0%)	Plus/Plus

```
Query   1      TCTCCTATACATACATACCT   20
               ||||||||||||||||||||
Sbjct   18162  TCTCCTATACATACATACCT   18181
```

Homo sapiens FOSMID clone ABC11-49614700O23 from chromosome 14, complete sequence

Sequence ID: gb|AC231261.2|Length: 39862Number of Matches: 1

Related Information

Map Viewer-aligned genomic context

Range 1: 19814 to 19833GenBankGraphics

Score	Expect	Identities	Gaps	Strand
37.4 bits(40)	2.1	20/20(100%)	0/20(0%)	Plus/Plus

```
Query  1      TCTCCTATACATACATACCT  20
              ||||||||||||||||||||
Sbjct  19814  TCTCCTATACATACATACCT  19833
```

Homo sapiens interleukin 1 receptor accessory protein like 1 (IL1RAPL1), RefSeqGene on chromosome X

Sequence ID: ref|NG_008292.1|Length: 1375337Number of Matches: 1

Related Information

Map Viewer-aligned genomic context

Range 1: 561228 to 561247GenBankGraphics

Alignment statistics for match #1

Score	Expect	Identities	Gaps	Strand
37.4 bits(40)	2.1	20/20(100%)	0/20(0%)	Plus/Minus

```
Query  1       TCTCCTATACATACATACCT  20
               ||||||||||||||||||||
Sbjct  561247  TCTCCTATACATACATACCT  561228
```

References

1. Case DA, Berryman JT, Betz RM *et al.* AMBER 2015.
2. Pérez A, Marchán I, Svozil D *et al.* Refinement of the AMBER Force Field for Nucleic Acids: Improving the Description of alphaγConformers. *Biophys J* 2007;**92**:3817–29.
3. Wang J, Wolf RM, Caldwell JW *et al.* Development and testing of a general amber force field. *J Comput Chem* 2004;**25**:1157–74.
4. Georgakilas AG, Bennett P V, Sutherland BM. High efficiency detection of bi-stranded abasic clusters in gamma-irradiated DNA by putrescine. *Nucleic Acids Res* 2002;**30**:2800–8.
5. Wang J, Cieplak P, Kollman PA. How well does a restrained electrostatic potential (RESP) model perform in calculating conformational energies of organic and biological molecules? *J Comput Chem* 2000;**21**:1049–74.
6. Mark P, Nilsson L. Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J Phys Chem A* 2001;**105**:9954–60.