

**Supplemental Table 1.** Statistical analysis of the deamination site preferences for the A3A protein among oligonucleotides with a TC/CT open/stem target and different flanking bases.<sup>a</sup>

a. Preference for flanking bases with TC/CT target in open or stem structure.

target structure \ comparison of flanking base	A vs. T	A vs. G	A vs. C	T vs. G	T vs. C	G vs. C
TC open	ns	****	**	***	*	ns
TC stem	****	ns	ns	****	**	**
CT open	****	ns	****	****	ns	****
CT stem	****	****	**	ns	**	****

b. Preference for open or stem structure with each flanking nucleotide base.

flanking base \ comparison	TC open vs. TC stem	CT open vs. CT stem	TC open vs. CT open	TC stem vs. CT stem
A	****	****	***	****
T	*	****	ns	***
G	****	ns	*	ns
C	**	***	***	***

<sup>a</sup> Significance was analyzed by using a two-way ANOVA and significance of  $p \leq 0.05$ . “ns” = not significant; “\*\*\*\*\*” =  $p \leq 0.0001$ ; “\*\*\*\*” =  $p \leq 0.001$ ; “\*\*\*” =  $p \leq 0.01$ ; “\*\*” =  $p \leq 0.05$ .

**Supplemental Table 2.** Statistical analysis of deamination site preferences of the A3B protein among oligonucleotides with a TC/CT open/stem target and different flanking bases.<sup>a</sup>

a. Preference for flanking bases with a TC/CT target in an open or stem structure.

target structure \ comparison of flanking base	A vs. T	A vs. G	A vs. C	T vs. G	T vs. C	G vs. C
TC open	ns	ns	ns	ns	ns	ns
TC stem	ns	ns	ns	ns	ns	ns
CT open	ns	ns	ns	ns	ns	ns
CT stem	ns	ns	ns	ns	ns	ns

b. Preference for an open or stem structure with each flanking nucleotide base.

flanking base \ comparison	TC open vs. TC stem	CT open vs. CT stem	TC open vs. CT open	TC stem vs. CT stem
A	ns	ns	ns	ns
T	ns	*	ns	ns
G	ns	ns	ns	ns
C	ns	ns	ns	ns

<sup>a</sup> Significance was analyzed with a two-way ANOVA and significance of  $p \leq 0.05$ . “ns” = not significant; “\*\*\*\*\*” =  $p \leq 0.0001$ ; “\*\*\*\*” =  $p \leq 0.001$ ; “\*\*\*” =  $p \leq 0.01$ ; “\*\*” =  $p \leq 0.05$ .

**Supplemental Table 3.** Statistical analysis of the deamination preferences of the A3C protein among oligonucleotides with a TC/CT open/stem target site and different flanking bases.<sup>a</sup>

a. Preference for flanking bases with a TC/CT target site in an open or stem structure.

target structure \ comparison of flanking base	A vs. T	A vs. G	A vs. C	T vs. G	T vs. C	G vs. C
TC open	ns	ns	ns	ns	ns	ns
TC stem	ns	ns	ns	ns	ns	ns
CT open	ns	ns	ns	ns	ns	ns
CT stem	ns	ns	ns	ns	ns	ns

b. Preference for an open or stem structure with each flanking nucleotide base.

flanking base \ comparison	TC open vs. TC stem	CT open vs. CT stem	TC open vs. CT open	TC stem vs. CT stem
A	ns	ns	ns	ns
T	ns	ns	ns	ns
G	ns	ns	ns	ns
C	ns	ns	ns	ns

<sup>a</sup> Significance was analyzed by using a two-way ANOVA with significance of  $p \leq 0.05$ . “ns” = not significant; “\*\*\*\*\*” =  $p \leq 0.0001$ ; “\*\*\*\*” =  $p \leq 0.001$ ; “\*\*\*” =  $p \leq 0.01$ ; “\*\*” =  $p \leq 0.05$ .

**Supplemental Table 4.** Statistical analysis of the deamination preferences of the A3F protein among oligonucleotides with a TC/CT open/stem target site and different flanking bases.<sup>a</sup>

a. Preference for flanking bases with a TC/CT target site in open or stem structure.

target structure \ comparison of flanking base	A vs. T	A vs. G	A vs. C	T vs. G	T vs. C	G vs. C
TC open	**	ns	ns	ns	*	ns
TC stem	ns	ns	ns	ns	ns	ns
CT open	*	ns	ns	***	ns	**
CT stem	***	**	ns	ns	*	ns

b. Preference for an open or stem structure with each flanking nucleotide base.

flanking base \ comparison	TC open vs. TC stem	CT open vs. CT stem	TC open vs. CT open	TC stem vs. CT stem
A	ns	ns	ns	ns
T	*	***	ns	**
G	ns	ns	ns	ns
C	ns	*	ns	ns

<sup>a</sup> Significance was analyzed by using a two-way ANOVA and significance of  $p \leq 0.05$ . “ns” = not significant; “\*\*\*\*\*” =  $p \leq 0.0001$ ; “\*\*\*\*” =  $p \leq 0.001$ ; “\*\*\*” =  $p \leq 0.01$ ; “\*\*” =  $p \leq 0.05$ .

**Supplemental Table 5.** Statistical analysis of deamination site preferences of the A3G protein among oligonucleotides with a TC/CT open/stem target and different flanking bases. <sup>a</sup>

a. Preference for flanking bases with a TC/CT target in an open or stem structure.

target structure \ comparison of flanking base	A vs. T	A vs. G	A vs. C	T vs. G	T vs. C	G vs. C
TC open	ns	**	***	*	****	****
TC stem	ns	ns	****	*	***	****
CT open	***	ns	****	****	ns	****
CT stem	****	*	ns	*	****	***

b. Preference for an open or stem structure with each flanking nucleotide base.

flanking base \ comparison	TC open vs. TC stem	CT open vs. CT stem	TC open vs. CT open	TC stem vs. CT stem
A	ns	***	ns	*
T	ns	****	***	*
G	ns	*	ns	ns
C	ns	ns	ns	ns

<sup>a</sup> Significance was analyzed by using a two-way ANOVA and significance of  $p \leq 0.05$ . “ns” = not significant; “\*\*\*\*\*” =  $p \leq 0.0001$ ; “\*\*\*\*” =  $p \leq 0.001$ ; “\*\*\*” =  $p \leq 0.01$ ; “\*\*” =  $p \leq 0.05$ .

**Supplemental Table 6.** Statistical analysis of the deamination site preferences of the A3H protein among oligonucleotides with a TC/CT open/stem target and different flanking bases. <sup>a</sup>

a. Preference for flanking bases with a TC/CT target in an open or stem structure.

target structure \ comparison of flanking base	A vs. T	A vs. G	A vs. C	T vs. G	T vs. C	G vs. C
TC open	ns	****	*	****	ns	***
TC stem	****	ns	ns	****	****	ns
CT open	****	ns	****	****	ns	****
CT stem	ns	ns	ns	ns	ns	ns

b. Preference for an open or stem structure with each flanking nucleotide base.

flanking base \ comparison	TC open vs. TC stem	CT open vs. CT stem	TC open vs. CT open	TC stem vs. CT stem
A	****	**	***	ns
T	****	****	ns	****
G	****	**	ns	ns
C	****	****	ns	ns

<sup>a</sup> Significance was analyzed by using a two-way ANOVA and significance of  $p \leq 0.05$ . “ns” = not significant; “\*\*\*\*\*” =  $p \leq 0.0001$ ; “\*\*\*\*” =  $p \leq 0.001$ ; “\*\*\*” =  $p \leq 0.01$ ; “\*\*” =  $p \leq 0.05$ .

**Supplemental Table 7.** Characteristics of A3F and A3G deamination hotspots in the HIV-1 BRU *pol* gene region spanning nucleotides 2574-3301 in the proviral DNA.<sup>a</sup>

a. Contexts and DNA secondary structure of the minus-strand DNA.

number	amino acid mutations	contexts	secondary structure	<u>TC</u> (-1 position is T)
1	D237N	AT <u>C</u> A	11-loop	✓
2	E233K	TT <u>C</u> A	5-loop	✓
3	E224K	TT <u>C</u> TT	open	✓
4	G213E	GT <u>C</u> CC	half stem, C in stem	✓
5	G213E	TC <u>C</u> CC	half stem, C in stem	
6	E204K	CT <u>C</u> CT	19-loop	✓
7	E203K	CT <u>C</u> T A	19-loop	✓
8	G196R	CC <u>C</u> T A	open	
9	E194K	TT <u>C</u> T A	open	✓
10	D192N	GT <u>C</u> A	open	✓
11	D186N	AT <u>C</u> A	bulge	✓
12	D185N	AT <u>C</u> C	stem	✓
13	E169K	CT <u>C</u> T A	3-loop	✓
14	G155E	AT <u>C</u> CT	16-loop	✓
15	G152E	AT <u>C</u> CC	open	✓
16	E122K	TT <u>C</u> A	open	✓
17	D113N	AT <u>C</u> A	open	✓

b. Percentage of different DNA secondary structures of the A3F and A3G deamination hotspots involving TC.

secondary structure	percentage
open	41%
stem	6%
bulge	6%
loop	35%
half stem	12%

c. Percentage of each nucleotide at the -2 or +1 position of the A3F and A3G deamination open hotspot involving TC.

nucleotide	-2 position	+1 position
A	33%	67%
T	50%	17%
C	0%	17%
G	17%	0%

<sup>a</sup> Data from Mohammadzadeh et al., 2019. The target cytidine is underlined.

**Supplemental Table 8.** Characteristics of A3F and A3G deamination hotspots within the HIV-1 NL4-3 protease gene region spanning nucleotides 2250-2631 in the proviral DNA.<sup>a</sup>

a. Contexts and DNA secondary structure.

number	amino acid number	overexpression	contexts	secondary structure	TC (-1 position is T)
1	90	A3G	GT <u>C</u> A	8-loop	✓
2	87	A3G	TT <u>C</u> TT	open	✓
3	86	A3G	TT <u>C</u> CA	open	✓
4	78	A3G	TC <u>C</u> TA	14-loop	
5	73	A3G	AC <u>C</u> T	stem	
6	65	A3G	TT <u>C</u> TA	half stem, C in stem	✓
7	60	A3G	AT <u>C</u> A	open	✓
8	57	A3G	GT <u>C</u> TT	open	✓
9	52	A3G	AC <u>C</u> T	stem	
10	51	A3G	CT <u>C</u> CA	22-loop	✓
11	51	A3G	TC <u>C</u> A	22-loop	
12	48	A3G	CC <u>C</u> CTA	22-loop	
13	48	A3G	CCC <u>C</u> TA		
14	48	A3G	CCC <u>C</u> TA		
15	46	A3G	AT <u>C</u> A	22-loop	✓
16	42	A3G	TT <u>C</u> CA	bulge	✓
17	42	A3G	TC <u>C</u> A	bulge	
18	41	A3G or A3F	AT <u>C</u> T	stem	✓
19	36	A3F	TT <u>C</u> AT	open	✓
20	34	A3G or A3F	TT <u>C</u> TA	open	✓
21	30	A3G or A3F	AT <u>C</u> AT	open	✓
22	21	A3F	TT <u>C</u> CT	open	✓
23	20	A3G	TC <u>C</u> TT	open	

b. Percentage of different DNA secondary structures of A3F and A3G deamination hotspots involving TC.

secondary structure	protein involved		
	total	A3F	A3G
open	57%	80%	50%
stem	7%	20%	8%
bulge	7%	0%	8%
loop	21%	0%	25%
half stem	7%	0%	8%

c. Percentage of each nucleotide at the -2 or +1 position of A3F and A3G deamination open hotspots involving TC.

nucleotide	-2 position			+1 position		
	total	A3F	A3G	total	A3F	A3G
A	25%	25%	33%	38%	50%	33%
T	63%	75%	50%	38%	25%	50%
C	0%	0%	0%	25%	25%	17%
G	13%	0%	17%	0%	0%	0%

d. Percentage of TC or CC target sites among A3F and A3G deamination hotspots.

	total	A3F	A3G
<u>TC</u> target	61%	100%	57%
<u>CC</u> target	39%	0	43%

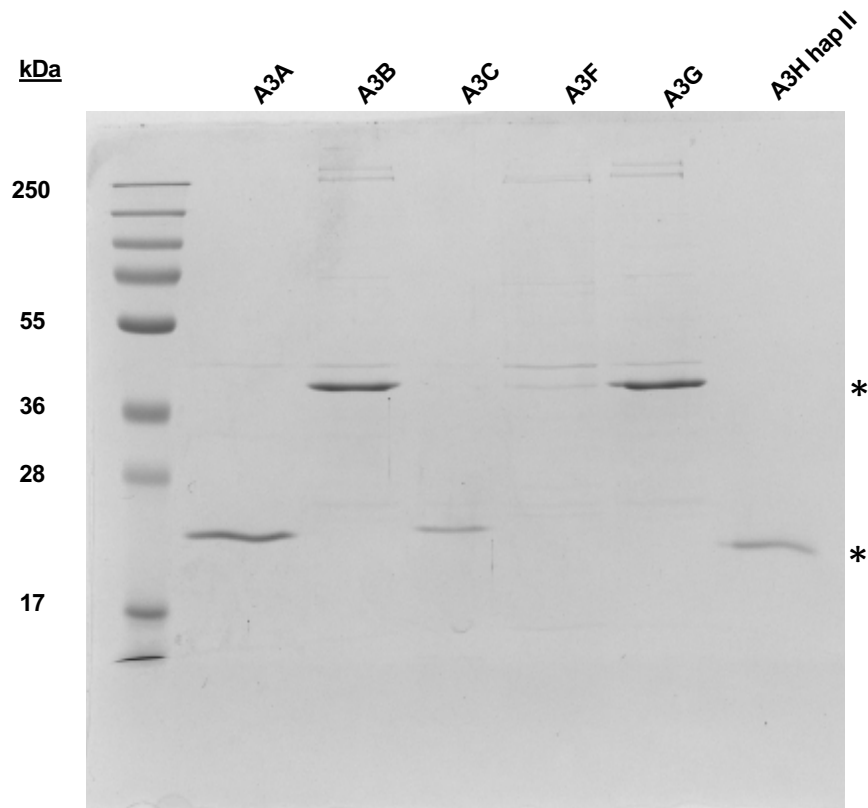
<sup>a</sup> Data from Ara, Love, & Chelico, 2014. The target cytidine is underlined.

**Supplemental Table 9.** Effect of the nucleotide at -2 position on APOBEC3-induced C-to-T mutations at the TC target. <sup>a</sup>

APOBEC3	sample source	cell culture	patients
	A3G	C > T > A > G	C > T > A > G
	A3F	T > A > C > G	A > T > C > G
	A3D	A > T > C > G	/
	A3H	T > C > A > G	/
	A3G & A3F	C > T > A > G	/
	A3G & A3H	C > T > A > G	/
	A3F & A3H	/	T > A > C > G

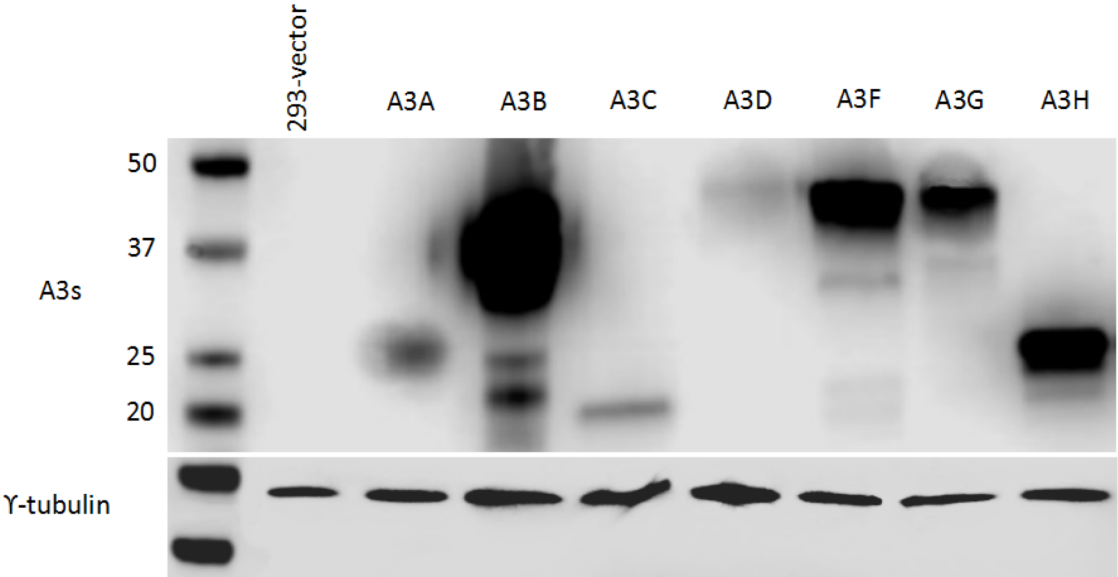
<sup>a</sup> Data from Desimmi et al., 2016. The “/” = not determined; “>” = larger. The target cytidine is underlined.

## Supplemental Figure 1



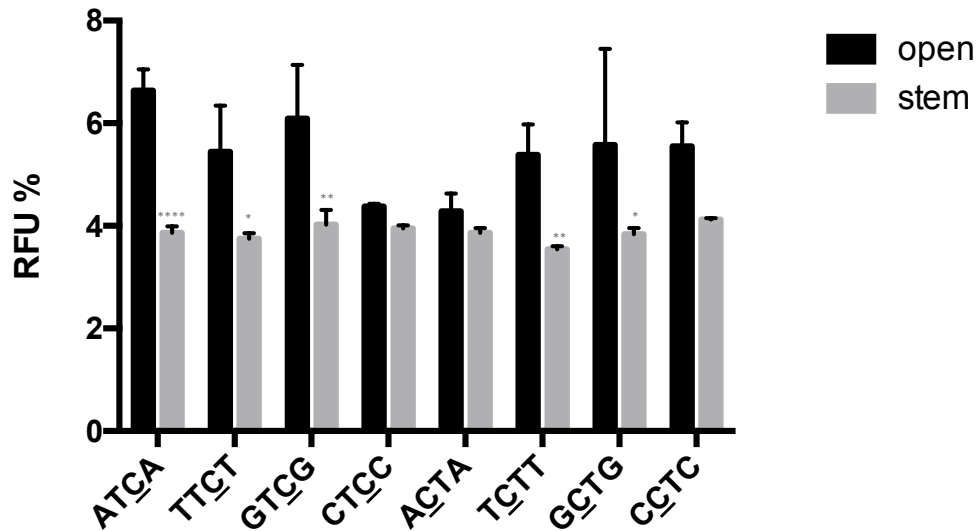
**Supplemental Figure 1. Analysis of APOBEC3 proteins.** Recombinant baculovirus was used for expression of APOBEC3 proteins. GST-A3A, GST-A3B, GST-A3C, GST-A3F, GST-A3G and GST-A3H expression plasmids were constructed by using the pAcG2T or pFAST-bac1 vector (BD Biosciences). Recombinant baculovirus was used to infect *Sf9* cells, with harvesting after 72 h postinfection. Cells were lysed, and the GST-tagged proteins were purified to obtain protein that was cleaved from the GST tag. The asterisks (\*) denote the expression of the different A3 proteins.

Supplemental Figure 2



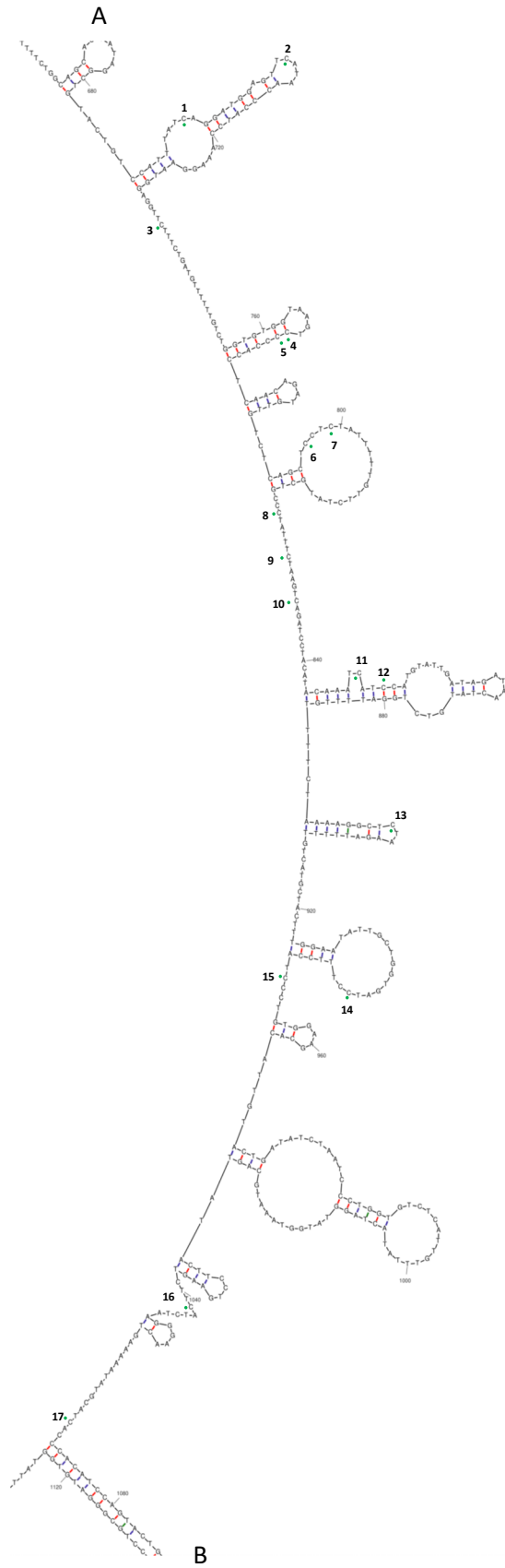
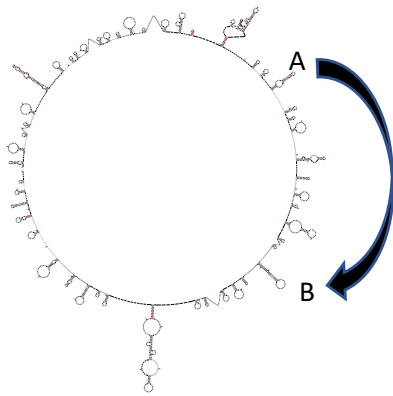
**Supplemental Figure 2. Confirmation of APOBEC3 protein expression by immunoblot analysis.** A3A and A3B expressing cell lines were obtained by transiently transfecting HEK 293 cells as described in the Materials and Methods. A3C, A3D, A3F, A3G and A3H expressing cell lines were generated by transfecting HEK 293 cells and selecting for stable expressors. The parental cell line was transfected with an empty vector (293-vector), and was used as a negative control. Y-tubulin was used as loading control. The image is representative of 2 independent experiments.





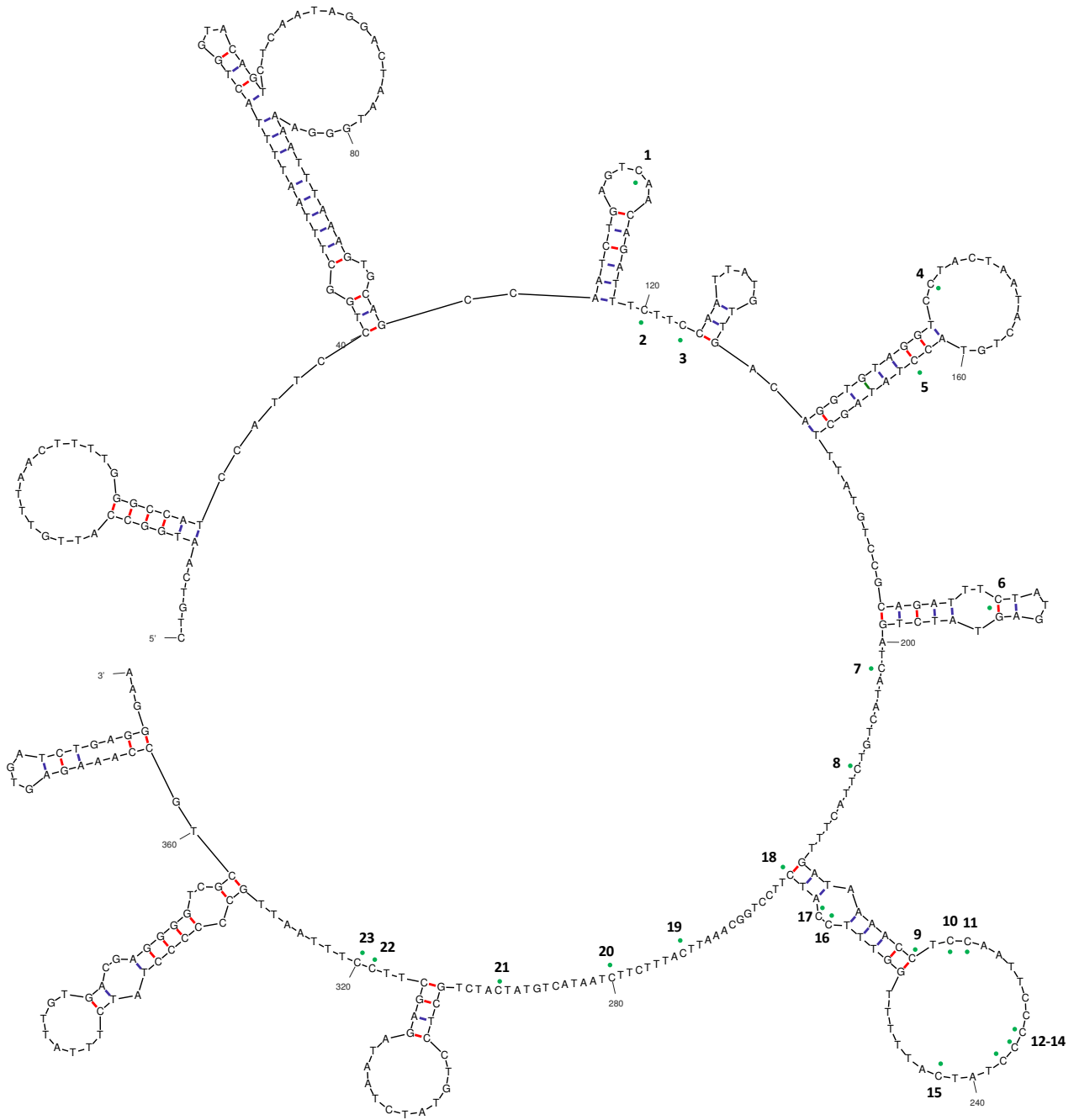
**Supplemental Figure 3. Comparison of the background fluorescence of the open and stem oligonucleotides in order confirm predicted DNA secondary structures.** The XTUX, XTCX and XCTX open and stem oligonucleotides were treated with UDG in the absence of APOBEC3 proteins. The relative fluorescence units (RFU) was determined, with the RFU of the XTUX oligonucleotide being set at 100%. The RFU measurements for the XTCX and XCTX oligonucleotides were normalized to that of the XTUX oligonucleotide. Data is presented as the average  $\pm$  standard deviation of 3 independent experiments. The asterisks above any bar of a stem oligonucleotide indicates a significant difference between the stem and open oligonucleotide pair. Significance was analyzed by using a one-way ANOVA with significance defined as a  $p \leq 0.05$ ; “\*\*\*\*” =  $p \leq 0.0001$ ; “\*\*” =  $p \leq 0.01$ ; “\*” =  $p \leq 0.05$ .

## Supplemental Figure 4



**Supplemental Figure 4. A3F/A3G deamination hotspots within the HIV-1 BRU *pol* gene region.** Sequencing data was previously published (Mohammadzadeh et al., 2019). Sequence of the minus-strand DNA is shown. The DNA secondary structure (*i.e.*, *pol* 1792-3536) was predicted by using the Mfold program. The sequenced area, labeled A to B (nucleotides 2574-3301 of the complete proviral DNA), is shown. Deaminated C nucleotides are denoted by a number (1-17) and a green circle adjacent to the C nucleotide. Additional details regarding each deaminated C nucleotide is provided in Supplemental Table 7.

## Supplemental Figure 5



**Supplemental Figure 5. A3F/A3G deamination hotspots within the HIV-1 NL4-3 protease gene.** Sequence data of the HIV-1 NL4-3 protease gene (nucleotides 2250-2631 from the proviral DNA) was previously published (Ara, Love, & Chelico, 2014). The minus-strand DNA sequence is shown. The DNA secondary structure was predicted by using the Mfold program. Deaminated C nucleotides are denoted by a number (1-23) and a green circle adjacent to the C nucleotide. Additional details regarding each deaminated C nucleotide is provided in Supplemental Table 8.