Supporting Information

DRONE: Direct tracking of DNA cytidine deamination and other DNA modifying activities

Tomoaki Sasaki, Shalley N. Kudalkar, Nicole Bertoletti, and Karen S. Anderson*
Department of Pharmacology, Yale University
333 Cedar St. SHM B-350 New Haven, CT 06520

*Corresponding author
Karen S. Anderson
Email: karen.anderson@yale.edu

Contents

Table S1: Oligonucleotide sequences used in this study

Oligonucleotide	Sequence
APOBEC3G Substrate	5'-TAGAAAGGGAGACCCAAAGAGGAAAGGTGA – 3'
APOBEC3G Product	5'-TAGAAAGGGAGACCU*AAAGAGGAAAGGTGA – 3'
Unmethylated C DNA	5'-CCTATGCGCATCAGTTTTCTGATGCGCATAGG-3'
Methylated C DNA	5'- CCTATGmC**GCATCAGTTTTCTGATGCGCATAGG – 3'
Hydroxymethylated C DNA	5'- CCTATGhmC***GCATCAGTTTTCTGATGCGCATAGG – 3'
T DNA	5'- CCTATGTGCATCAGTTTTCTGATGCGCATAGG – 3'
DNMT3B Substrate	5'-CCTATGCGmCATCAGTTTTCTGATGmCGmCATAGG-3'
FAM APOBEC Substrate	5'- /6-FAM****/ TAA GAG GAA AGA CCC AAA GAA GAG GAA - 3'
22mer Substrate	5'-TTAGTGATTTCAGAGAGAGGAT-3'
22mer Product	5'-TTAGTGATTTUAGAGAGAGGAT-3'
35mer Substrate	5'- AATTAGAAAGGGAGACCCAAAGAGGAAAGGTGAAT-3'
35mer Product	5'- AATTAGAAAGGGAGACCUAAAGAGGAAAGGTGAAT-3'
43mer Substrate	5'-ATTATTATTATTCAAATGGATTTATTTATTTATTTATTTATTT-3'
43mer Product	5'-ATTATTATTATTUAAATGGATTTATTTATTTATTTATTTATTT-3'
APOBEC3B 30mer Substrate	5'-TAGAAAGGGAGTATCAAAGAGGAAAGGTGA-3
PIK3CA Substrate	5'-TGCTCAGTGATTTCAGAGAGAGGAT-3'

*U represents an internal deoxyuridine residue
**mC representes a methylated cytosine
***hmC represents a hydroxymethylated cytosine
**** 6-FAM represents a 6-carboxyflourescein label at the 5' end of the oligonucleotide

Table S2: Chromatography conditions for pilot APOBEC oligonucleotide separation (Figure 1B)

Flow Rate	0.3 ml/min
Mobile phase A	15 mM TEA, 400 mM HFIP
Mobile phase B	50% Buffer A, 50% Methanol
Gradient	40-45% B in 10 min
Injection	1 μL
Temperature	60°C
Detector	Agilent 1260 Infinity II Diode Array Detector WR
Detection Method	UV @ 260 nm, 2.5 Hz
Sample	10 μM each of APOBEC3G substrate and APOBEC3G product oligonucleotides in water

Table S3: Chromatography conditions for APOBEC3G enzyme turnover experiment (Figure 2)

Flow Rate	0.3 ml/min
Mobile phase A	15 mM TEA, 400 mM HFIP
Mobile phase B	50% Buffer A, 50% Methanol
Gradient	40-45% B in 10 min
Injection	10 μL
Temperature	60°C
Detector	Agilent 1260 Infinity II Diode Array Detector WR
Detection Method	UV @ 260 nm, 2.5 Hz
Sample	APOBEC3G quenched timepoints processed as described in the Experimental Methods

Table S4: Chromatography conditions for epigenetic DNA modification separation (Figure 3B)

Flow Rate	0.3 ml/min
Mobile phase A	15 mM TEA, 400 mM HFIP
Mobile phase B	50% Buffer A, 50% Methanol
Gradient	40-50% B in 10 min
Injection	1 μL
Temperature	60°C
Detector	Agilent 1260 Infinity II Diode Array Detector WR
Detection Method	UV @ 260 nm, 2.5Hz
Sample	10 μM each of TET/DNMT3B substrate and TET/DNMT3B product oligonucleotides in water

Table S5: Chromatography conditions for APOBEC methylated oligonucleotide separation (Figure 3B)

Flow Rate	0.3 ml/min
Mobile phase A	15 mM TEA, 400 mM HFIP
Mobile phase B	50% Buffer A, 50% Methanol
Gradient	40-45% B in 10 min
Injection	1 μL
Temperature	60°C
Detector	Agilent 1260 Infinity II Diode Array Detector WR
Detection Method	UV @ 260 nm, 2.5 Hz
Sample	10 μM each of methylated C DNA substrate and T DNA product oligonucleotides in water

Table S6: Chromatography conditions for DNMT3B enzyme turnover experiment (Figure 3C)

Flow Rate	0.3 ml/min
Mobile phase A	15 mM TEA, 400 mM HFIP
Mobile phase B	50% Buffer A, 50% Methanol
Gradient	40-50% B in 10 min
Injection	10 μL
Temperature	60°C
Detector	Agilent 1260 Infinity II Diode Array Detector WR
Detection Method	UV @ 260 nm, 2.5 Hz
Sample	DNMT3B quenched timepoints processed as described in the Experimental Methods

Table S7: Chromatography conditions for FAM modified oligonucleotide separation (Figure 4B)

Flow Rate	0.3 ml/min
Mobile phase A	15 mM TEA, 400 mM HFIP
Mobile phase B	50% Buffer A, 50% Methanol
Gradient	40-42% B in 35 min
Injection	1 μL
Temperature	60°C
Detector	Agilent 1260 Infinity II FLD Spectra
Detection Method	Ex 245 nm, Em 520 nm, PMT 10, response time 0.5 s, step size 10 nm
Sample	10 μM each of FAM substrate and product oligonucleotides in water

Table S8: Chromatography conditions for APOBEC3G turnover of FAM modified oligonucleotides (Figure 4C)

Flow Rate	0.3 ml/min
Mobile phase A	15 mM TEA, 400 mM HFIP
Mobile phase B	50% Buffer A, 50% Methanol
Gradient	40-42% B in 35 min
Injection	10 μL
Temperature	60°C
Detector	Agilent 1260 Infinity II FLD Spectra
Detection Method	Ex 245 nm, Em 520 nm, PMT 10, response time 0.5 s, step size 10 nm
Sample	APOBEC3G quenched timepoints of fluorescent oligonucleotides processed as described in the Experimental Methods

Table S9: Chromatography conditions for *PIK3CA* substrate turnover (Figure 5)

Flow Rate	0.2 ml/min
Mobile phase A	15 mM TEA, 400 mM HFIP
Mobile phase B	50% Buffer A, 50% Methanol
Gradient	40-50% B in 55 min
Injection	10 μL
Temperature	60°C
Detector	Agilent 1260 Infinity II Diode Array Detector WR
Detection Method	UV @ 260 nm, 2.5 Hz
Sample	APOBEC3B quenched timepoints with PIK3CA substrate oligonucleotide processed as described in the Experimental Methods

Table S10: Chromatography conditions for nanaomycin inhibition experiment (Figure 4A)

Flow Rate	0.3 ml/min
Mobile phase A	15 mM TEA, 400 mM HFIP
Mobile phase B	50% Buffer A, 50% Methanol
Gradient	48-50% B in 20 min
Injection	10 μL
Temperature	60°C
Detector	Agilent 1260 Infinity II Diode Array Detector WR
Detection Method	UV @ 260 nm, 2.5 Hz
Sample	Nanaomycin inhibition timepoints processed as described in the Experimental Methods

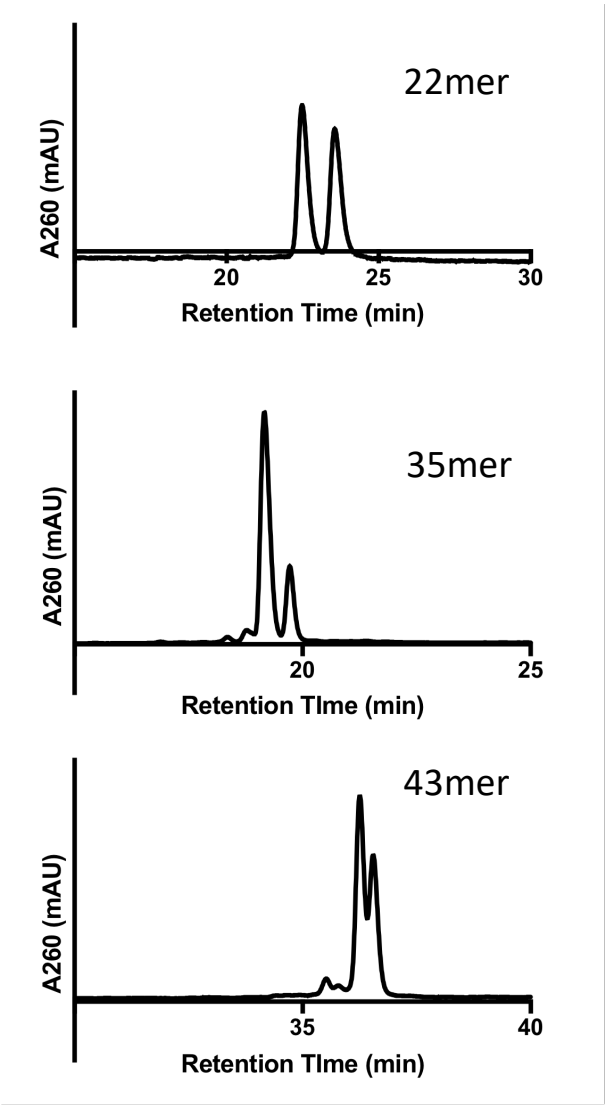

Figure S1: Separation of oligonucleotides of various lengths ranging from 22-43 bases. Baseline resolution was attained for 22mer and 35mer reactant and product oligonucleotides, whereas 43mer oligonucleotide mix was not baseline resolved with a tandem column setup. 22mer gradient was 38-42% B over 30 minutes at 0.3ml/min (single column), while 35mer and 43mer gradient was 40-50% B over 55 minutes at 0.2ml/min (tandem column format)

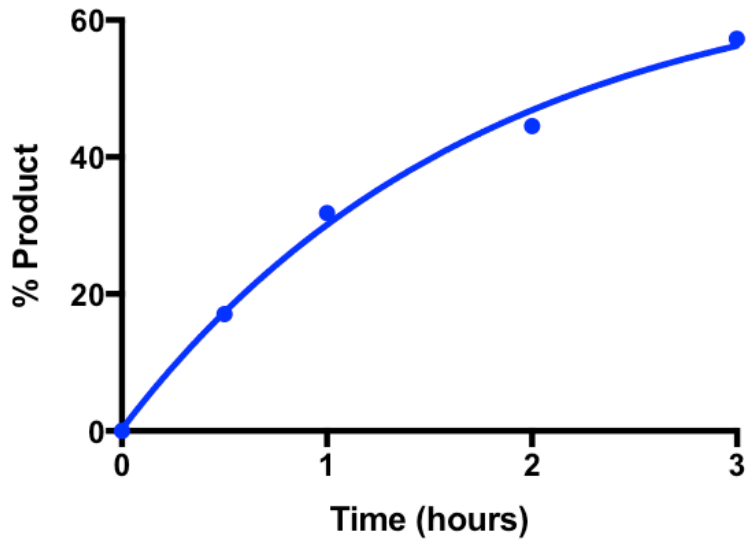

Figure S2: APOBEC3G deamination timecourse observed through the conventional UDG cytidine deamianse assay. Deamination experiments were conducted under single turnover conditions with [APOBEC3G] of 5 μM and [substrate] of 1 μM in a buffer consisting of 50 mM Tris pH 8.0, 100 mM NaCl, 0.1% Triton X-100, and 1 mM DTT. Reaction timepoints were collected by quenching with EDTA at a final concentration of ~50mM and heating to 95°C for 5 minutes. Following quenching, each sample was post-processed with 5 units of UDG and subsequent addition of NaOH at a final concentration of 0.2N. Samples were then loaded on a denaturing polyacrylamide gel, exposed to a phosphor screen, and analyzed by densitometry to calculate % product.

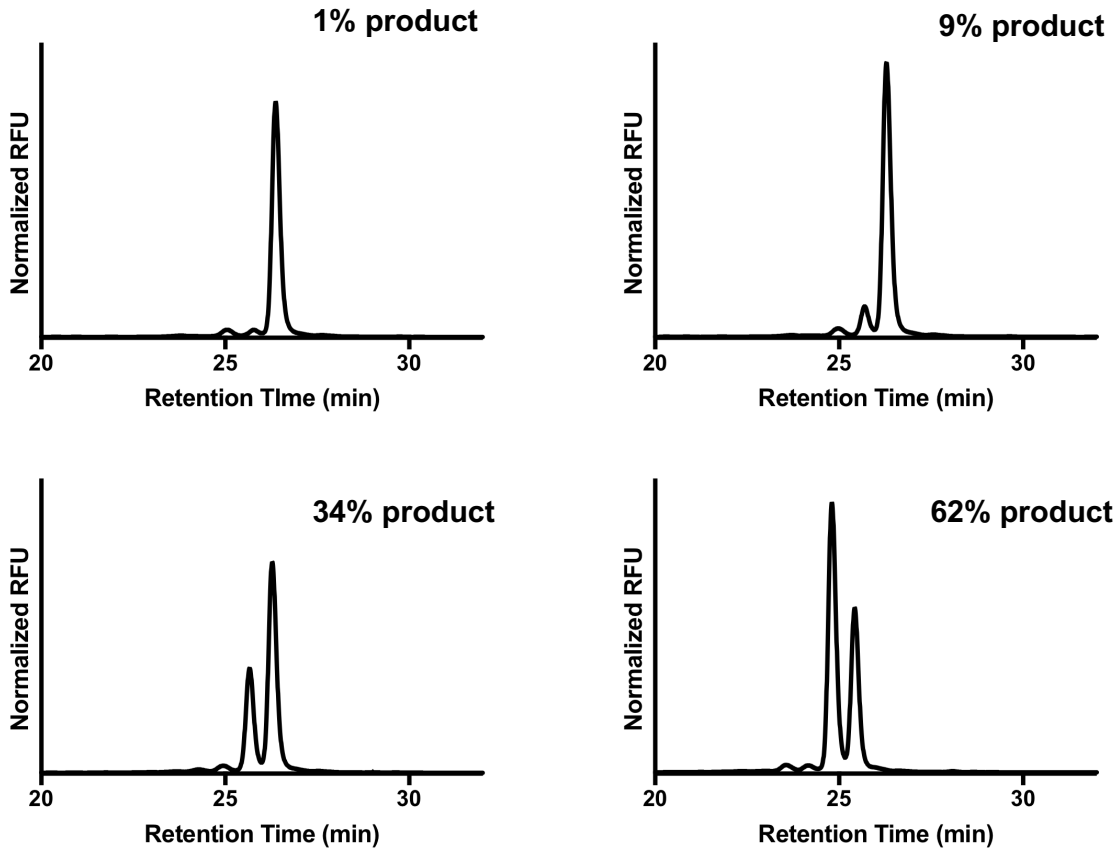

Figure S3: Separation of various ratios of APOBEC3G 30mer substrate and product oligonucleotides. 1 µl of 10 µM various reactant product oligonucleotide mix was injected in the UHPLC and % product was determined by the AUC ratios of substrate and product peaks via fluorescent detector (FLD). Gradient performed was 40-50%B over 55 min at a flow rate of 0.2 ml/min (tandem column format).

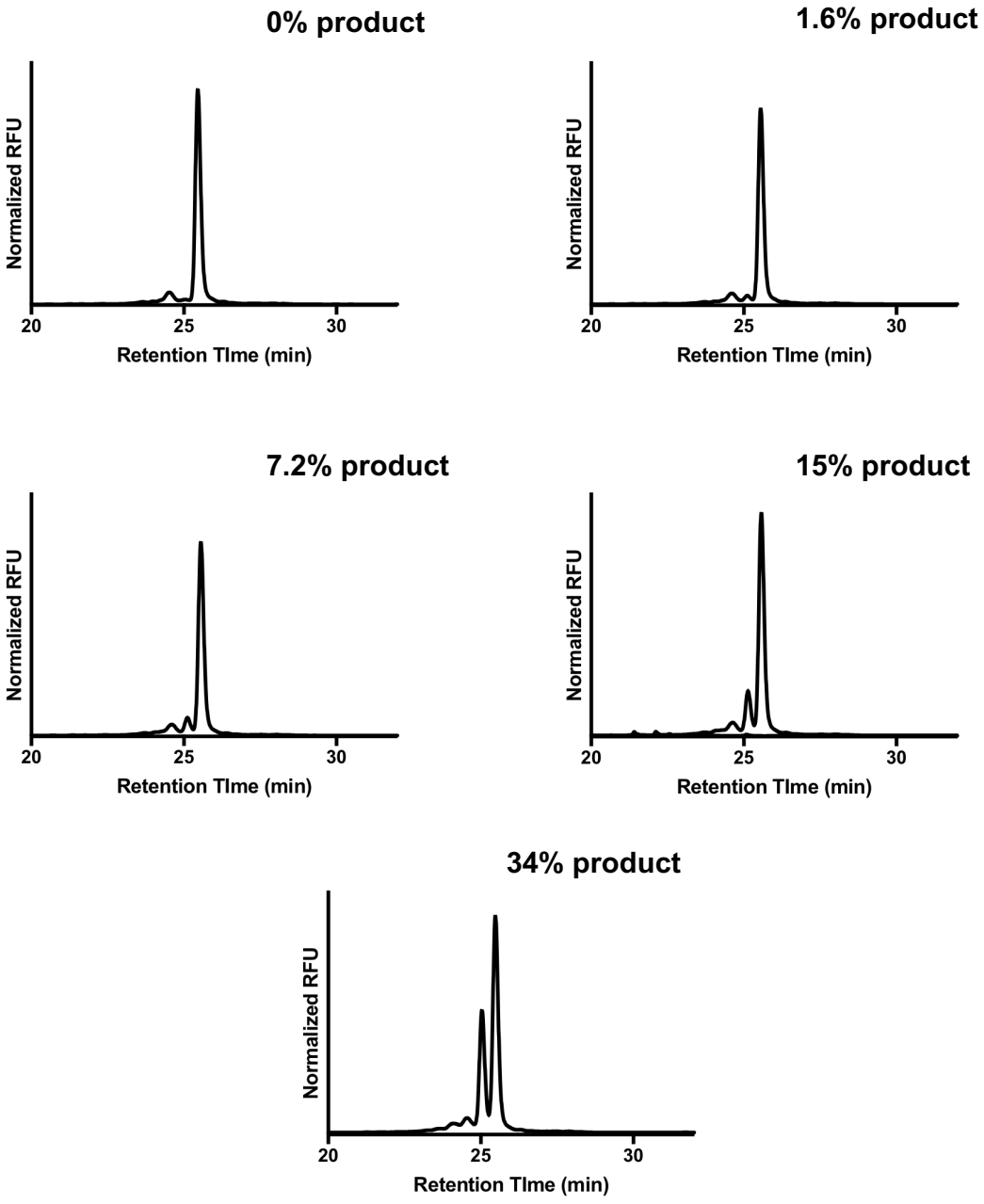

Figure S4: Enzymatic reactions with APOBEC3G corresponding to various % product. After quenching and dialyzing, 10 µl of each reaction timepoint was injected into the UHPLC and % product was determined by the AUC ratios of substrate and product peaks via FLD. 0%, 1.6%, 7.2%, 15%, and 34% product correspond to A3G enzymatic timepoints quenched at 0 hour, 1 minute, 6 minutes, 20 minutes, and 50 minutes, respectively. Gradient performed was 40-50%B over 55 min at a flow rate of 0.2 ml/min (tandem column format).

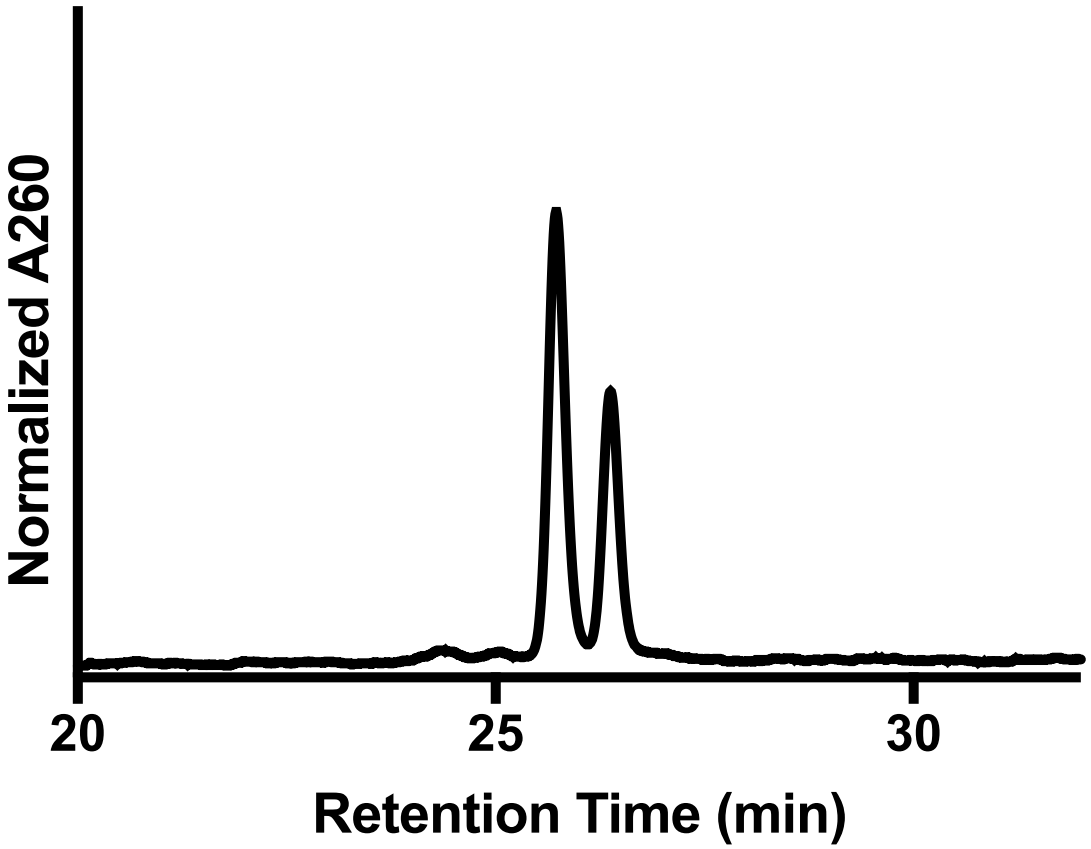

Figure S5: Restoration of baseline resolution of the 30mer APOBEC3B substrate and product oligonucleotides using a tandem column set-up. Gradient performed was 40-50%B over 55 min at a flow rate of 0.2 ml/min (tandem column format).

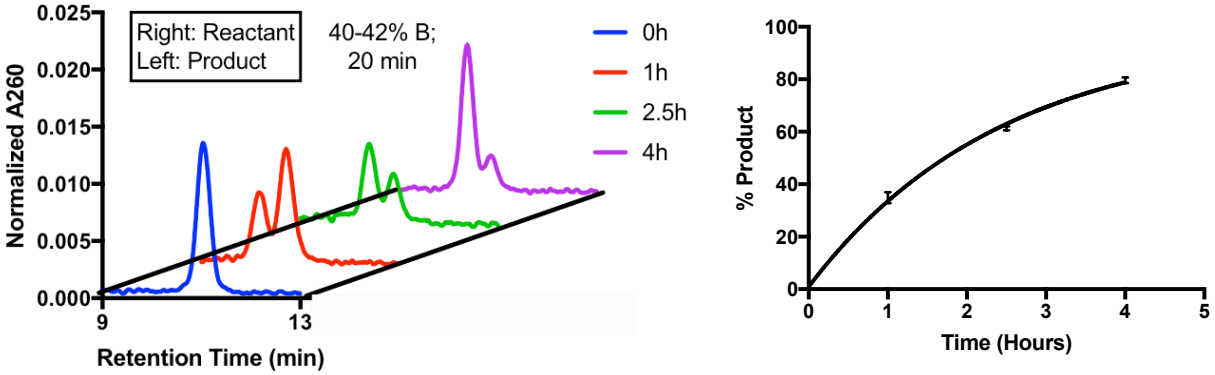

Figure S6: APOBEC3B deamination timecourse. Recombinant N-terminal MBP tagged APOBEC3B was generated for APOBEC3B deamination experiments as described in the methods. Deamination experiments were conducted under single turnover conditions with [MBP-APOBEC3B] of 2 μM and [substrate] of 1 μM. Substrate oligonucleotide sequence is 5'-TAGAAAGGGAGTATCAAAGAGGAAAGGTGA-3'. Gradient conducted for this experiment was 40-42% B over 20 minutes at a flow rate of 0.3 ml/min.

Experimental Methods

Materials and reagents

All DNA oligonucleotides used in this study were ordered HPLC or PAGE purified (Integrated DNA Technologies, Coralville, IA). A full list of oligonucleotides used in this study is summarized in Table S1.

For enzyme reactions and DNA extractions, a pre-mixed solution of 25:24:1 phenol-chloroform-isoamyl alcohol mixture was used (Cat. # 77617, Sigma-Aldrich, St. Louis, MO).

For UHPLC separations, all experiments were conducted on the Agilent 1290 Infinity II LC System (Agilent Technologies, Santa Clara, CA). Individual gradients were performed using the InfinityLab Poroshell 120 HPH-C18, 2.1 x 100 mm, 1.9 µm diameter (Part # 695675-702, Agilent Technologies, Santa Clara, CA). Absorbance detection was conducted on the 1260 DAD WR (Part # G7115A, Agilent Technologies, Santa Clara, CA). Fluorescence detection was conducted on the 1260 FLD Spectra (Part # G7121B, Agilent Technologies, Santa Clara, CA). Reagents used for mobile phase include trimethylamine HPLC grade (Cat. #O4884-100, Fisher Scientific, Hampton, NH), hexafluoroisopropanol (Cat. #105228, Sigma-Aldrich, St. Louis, MO), and methanol (Cat. #MX0488, Merck Millipore, Burlington, MA).

Protein expression and purification – APOBEC3G and APOBEC3B

Recombinant FL APOBEC3G was expressed and purified using a baculovirus expression system. The cDNA construct cloned into pFastBacI was kindly donated from the Levin laboratory and protein expression was conducted as reported previously.[1] For protein purification, the harvested cells were first resuspended and lysed in a buffer containing 25mM HEPES, 500mM NaCl, 10% Glycerol, 0.1% Triton X-100, 1mM EDTA, 10mM $CaCl_2$ supplemented with 5mM β-mercaptoethanol, 100ug/ml DNase I, 40ug/ml RNase A, protease inhibitor tablet. Clarified lysates were then bound to Glutathione Sepharose 3B resin slurry, washed in two steps using high (1M NaCl) and mid (500mM NaCl) salt wash buffer, and were eluted using a buffer consisting of 50mM Tris pH 8, 200mM NaCl, 5% Glycerol, 40mM reduced glutathione, 2mM $CaCl_2$, and 5mM β-mercaptoethanol. The GST-tagged protein was then subject to enterokinase cleavage (6 units rEK, 12 hours RT) after buffer exchanging to cleavage buffer (50mM NaCl, 20mM Tris-HCl pH7.4, 2mM CaCl2, and 0.1% Triton X-100). Enterokinase was removed from the cleavage reaction using the rEK Capture Kit (Merck Millipore, Burlington, MA) and the cleaved protein was then bound to Glutathine Sepharose 4B resin (GE Healthcare, Chicago, IL). Flowthrough and wash fractions were collected containing pure APOBEC3G, buffer exchanged to final storage buffer 50mM Tris pH 8, 200mM NaCl, 10% glycerol, 10uM $ZnCl_2$, 0.1% Triton X-100, and 2mM dithiothreitol (DTT).

The expression plasmid for MBP tagged APOBEC3B was kindly donated from Xiaojiang Chen's lab. Expression and purification was conducted as described previously with the exception of the purification involving a one-step amylose resin purification protocol.[2]

The final protein was stored in a storage buffer containing 50 mM Tris, 200 mM NaCl, 10% Glycerol, 0.1% Triton X-100., and 2 mM DTT.

Protein expression and purification – DNMT3B Catalytic Domain

The plasmids pET11 and pET28a containing the P561-E853 fragment coding sequences of the human gene DNMT3B and the E177-L386 fragment DNMT3L, respectively, were obtained from ARVYS (Trumbull, CT USA).

The pET-based plasmids containing the coding sequences of DNMT3B with a N-terminal His10-tag containing a TEV (Tobacco Etch Virus) protease cleavage site, and DNMT3L with a N-terminal FLAG tag, were co-transformed into Rosetta 2 (DE3) pLysS competent cells. The transformed bacteria cells were grown overnight in 100 mL of LB medium containing 100 µg/mL of ampicillin and 30 µg/mL of kanamycin at 37°C. Subsequently, 15 mL of the overnight culture were transferred in 1 L of the aforementioned medium and allowed to grow at 37 °C until an OD_{600} of 0.3 was reached and the culture was cooled to 16° C. When the culture reached the OD_{600} of 0.6 the cells were induced with 0.5 mM IPTG. The bacteria cells were harvested after 4 hours by centrifugation and stored at -80° C overnight before proceeding with the purification.

The cell pellet was resuspended in a buffer containing 50 mM Tris, 500 mM NaCl, 0.5 mM TCEP, 10 mM Imidazole, 10 mM $MgSO_4$, 0.5% (v/v) Triton X-100, 10% (v/v) glycerol and cOmplete Protease Inhibitor Cocktail Tablet (Roche, Germany) adjusted at a pH of 7.4. The cells were disrupted with a high-pressure homogenizer and the

obtained lysate was centrifuged at 30000 g for 1h at 4°C. The supernatant was applied to a Ni-NTA column, and a washing step was applied with a buffer containing 50 mM Tris, 500 mM NaCl, 0.5 mM TCEP, 20 mM Imidazole, 10 mM $MgSO_4$, 5 mM ATP and 10% (v/v) glycerol to remove non-specifically bound protein. The target protein was eluted using a linear gradient of imidazole from 20 mM to 600 mM. TEV protease was added to the eluted protein mixture to cleave the N-terminal His8-tag and it was dialyzed overnight at 4°C to reduce the imidazole concentration in the sample (50 mM Tris, 500 mM NaCl, 0.5 mM TCEP and 10% (v/v) glycerol adjusted at a pH of 7.4). A second Ni-NTA column was used for separation of the TEV protease from the DNMT3B/3L protein complex in which the target protein complex was collected from the flowthrough of the column. Furthermore, we conducted size exclusion chromatography purification step (Superdex 200 16/60) with a running buffer comprising 50 mM Tris, 300 mM NaCl, 0.5 mM TCEP and 5% (v/v) glycerol. We then added 100% glycerol to attain a final concentration of glycerol of 20% in the purified protein preparation. The protein solution was flash-frozen in liquid nitrogen and stored at -80°C.

Resolution of substrate and product oligonucleotides

All resolution experiments for substrate and product oligonucleotides including the separation of 30mer APOBEC3G oligonucleotides, epigenetically modified oligonucleotides, and fluorescent oligonucleotides were conducted by first preparing 1:1 mixes of each oligonucleotide at a final concentration of 10 µM each in double distilled water. 1 µl of the resulting mixture solution (10 picomoles DNA) was injected for UHPLC analysis.

Enzymatic reaction timepoints and sample clean-up

All enzymatic reactions were conducted with a final oligonucleotide concentration of 1 µM and a final protein concentration of 5 µM at 37°C at a total volume of 20 µl. For APOBEC reactions, 5 µM enzyme and 1 µM substrate oligonucleotide were incubated in a buffer consisting of 50mM Tris pH 8.0, 100mM NaCl, 0.1% Triton X-100, and 1mM DTT. For DNMT3B reactions, final enzyme and oligonucleotides concentrations were 5 and 1 µM, respectively. These reactions were initiated with 50 µM of SAM and were incubated in a buffer of 50mM Tris pH 8.0, 50mM NaCl, and 0.1% beta-mercaptoethanol. Each timepoint was collected by quenching with an equal volume of 25:24:1 phenol-chloroform-isoamyl alcohol mixture. Following termination of reactions, the individual timepoints were centrifuged for 1 minute at 16,000 x g RCF, and the top aqueous layer was isolated. The resulting samples were then dialyzed using the Slide-A-Lyzer™ MINI dialysis device against 4L of double distilled water over three hours (1 L water 1 hour, 1 L water 1 hour, 2 L water 1 hour). For dialysis of multiple timepoints and/or samples, they were collectively dialyzed in the same container given the negligible reaction volume in comparison to the dialysate volume. Subsequently, the samples were dried down with a speed vacuum centrifuge before reconstitution with 20 µl water. A total of 10 µL (10 picomoles DNA) was injected into the UHPLC for analysis.

Inhibition experiment with Nanaomycin A

4 mM nanaomycin A prepared in 100% DMSO was kindly provided by the Bosenberg group at Yale University. For inhibition experiments, 0.5 µl of nanaomycin A dilution

series or 100% DMSO was incubated with 17.5 µl DNMT3B-buffer mixture for 15 minutes on ice to allow for maximal binding before addition of 2 µl of 10 µM substrate to initiate the reaction at 37°C. The reaction was subsequently quenched after 16 hours and cleaned up as described previously for UHPLC analysis.

Fluorescent detection of oligonucleotides

5' FAM-labeled oligonucleotides were reconstituted in TE buffer (10 mM Tris pH 8.0, 1 mM EDTA) at a concentration of 100 µM. Stock oligonucleotides were then stored in amber tubes for light protection. Subsequent dilutions were performed in water prior to injection in UHPLC. For fluorescent detection using the FLD on the UHPLC, an excitation wavelength of 245 nm and an emission wavelength of 520 nm were used. Specific chromatography conditions are listed in Tables S7-S8.

Chromatography conditions

Prior to each run, the column was equilibrated in starting conditions for 10 minutes. Detailed chromatography conditions are listed in Tables S2-S10.

Normalizing signals across samples

In order to normalize signal across various samples, we employed a methodology to integrate the chromatogram to extract out the area under the curve (AUC). For a given chromatogram, we first select out the regions containing the peaks of interest in terms of retention time. The values of each time coordinate is normalized by dividing by the

total AUC to yield the normalized A260 or RFU. The normalized signal is then plotted for comparison with other timepoints.

Quantifying % product and extraction of observed rate

To quantify % product turned over, we employ the formula, % product = (AUC product/AUC reactant + AUC product) * 100. % product for each timepoint is calculated using this method to generate product formation over time. An observed single turnover rate constant k_{obs} was extracted by fitting to a single exponential curve, *Percent product* = $A(1-e^{-kobs*time})$, in which A is the maximum % product.