APOBEC3A efficiently deaminates methylated, but not TET-oxidized, cytosine bases in DNA

Emily K. Schutsky, Christopher S. Nabel, Amy K. F. Davis, Jamie E. DeNizio, and Rahul M. Kohli

SUPPLEMENTARY INFORMATION

Supplementary Figures S1-S7, Supplementary Table S1

Figure S1

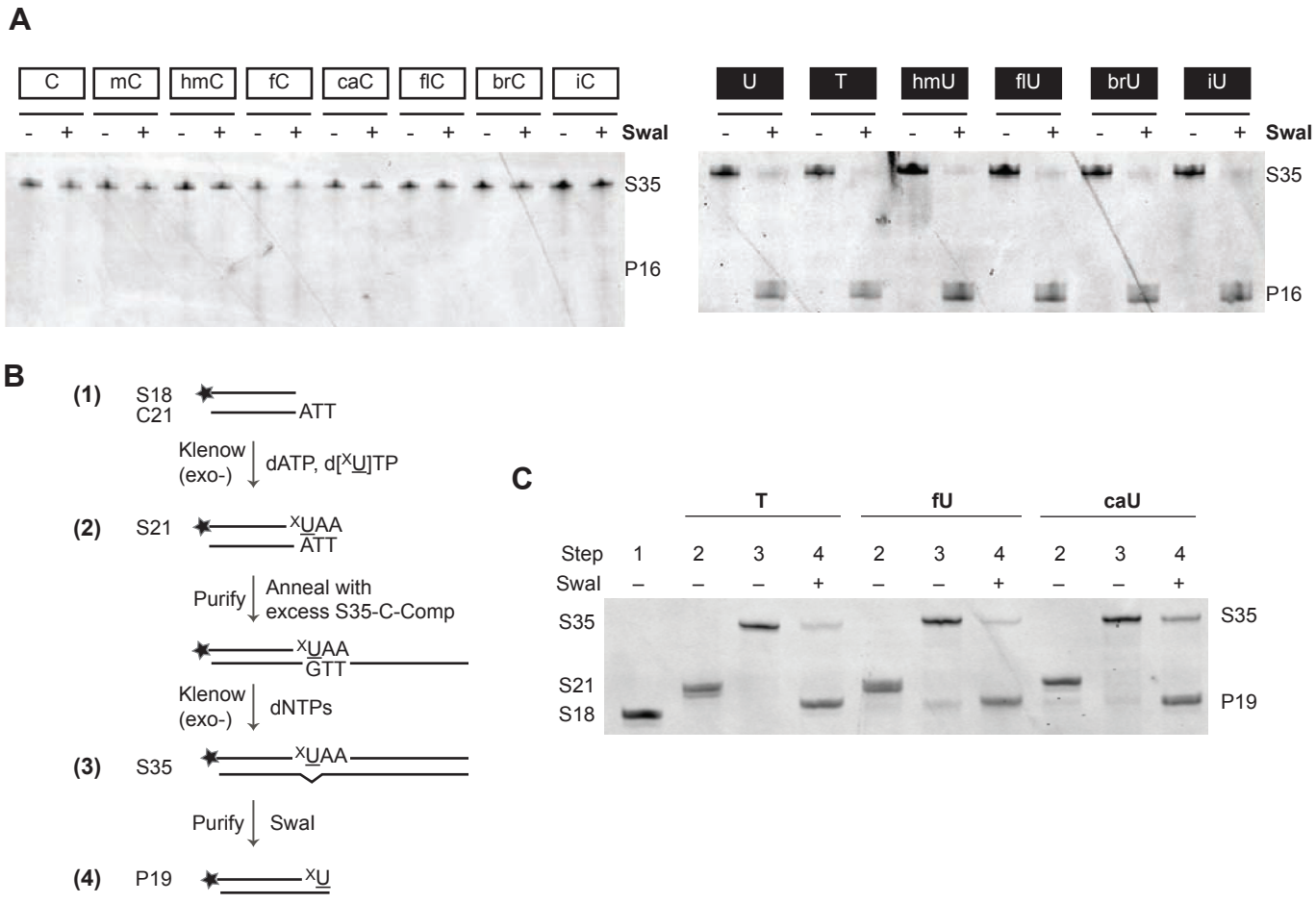
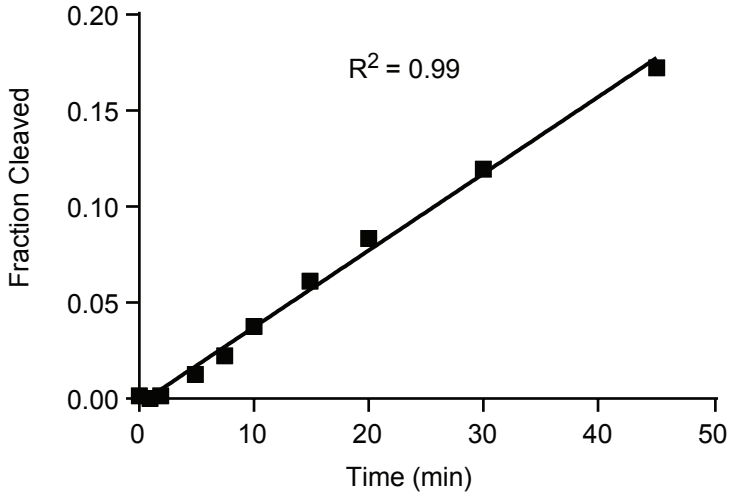


Figure S1. SwaI selectivity for modified uracils (ˣU) and not modified cytosines (ˣC). (A) SwaI has no cleavage activity on ˣC-containing substrates, and effectively cleaves ˣU-containing DNA. (B) Phosphoramidites for 5-formyluracil (fU) and 5-carboxyluracil (caU) were not commercially available to synthesize product controls with standard phosphoramidite chemistry. Therefore, fU- and caU-containing oligos were synthesized chemoenzymatically. A 5'-FAM labeled oligonucleotide S18 was annealed to a 21-mer complement and extended with dATP and either TTP, 5-formyl-dUTP, or 5-carboxyl-dUTP and Klenow(exo-) polymerase. The extended S21 strand was then annealed to excess S35-C-Comp and again subjected to primer extension to generate the 35-mer product control containing T, fU, or caU. (C) The purified oligonucleotides from each chemoenzymatic step are shown. While the corresponding ˣC substrates were not cleaved (see (A)), SwaI is able to cleave the oligonucleotides containing T, fU, and caU. Notably, a 5'-label was required for monitoring steps in the chemoenzymatic synthesis. The SwaI cleavage products for fU and caU controls therefore are associated with 5'-labeled 19-mers (P19) rather than 3'-labeled 16-mers (P16) for all other product controls and for the deaminated substates, which all have 3'-FAM labels. For clarity, the fU and caU product controls are not shown in Fig 2 to avoid confusion between the P16 resulting from cleavage of deaminated fC and caC.

The recently reported structures of SwaI in complex with DNA (PDB 5TGX, 5TH3, Reference 44) provide some insight into the reasons why SwaI is well suited to our assay. In the structures, the two central base pairs of the restriction site (5'–ATTTAAAT–3') are highly distorted and lose Watson-Crick base pairing with the opposite strand; instead, the A bases on opposite strands stack with each other and the T bases are rotated away from the helix. These rotated T bases make specific contacts with SwaI via their Watson-Crick face, while there is space in the vicinity of the 5-position methyl group, with the closest residue located ~3.8 Å away. This latter observation provides a rationale for how the enzyme is permissive to the identity of our variously modified uracil analogs.

Figure S2

A



B

Substrate	[A3A] (nM)	Time (min)
S35-C	0.5	3
S35-mC	3	3
S35-flC	3	3
S35-brC	15	8
S35-iC	50 - 750	25
S35-hmC	150 - 1400	45
S35-fC	200 - 1000	45
S35-caC	200 - 1000	45

C

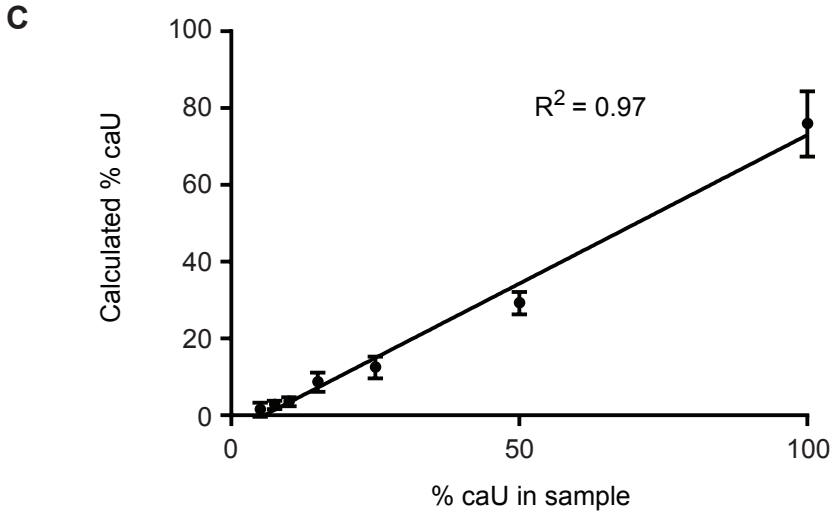


Figure S2. Conditions for kinetic experiments with A3A and caU detection limit. (A) 500 nM radiolabeled S35-C substrate was incubated with 100 pM A3A at 37 ºC. Serial timepoints were taken and terminated by heat inactivation. The samples were subjected to the UDG-based assay and product formation was quantified. Product formation was linear to at least 45 min. (B) Conditions used to for determination of deamination kinetics in Table 1 (main text). Enzyme and time conditions were chosen with target of <20% substrate conversion to product. (C) Limit of detection for caU. Known mixtures of S35 oligonucleotides with caC or caU were generated, treated with SwaI, and the amount of cleavage product was quantified. Shown is the plot of known caU in the sample versus calculated caU from the SwaI digest. This experiment was done in four independent replicates and the linear regression was calculated by GraphPad Prism software. Although caU cleavage is not complete, the detected product is linear from 5 to 100%, with detectable cleavage when the product is as low as 5% of the sample. The kinetic assays were performed with 2 μM substrate, providing a limit of detection of 100 nM. Under the assay conditions tested this translates to a k_{cat} detection limit of 0.002 min^{-1} for A3A acting on caC-containing DNA.

Figure S3

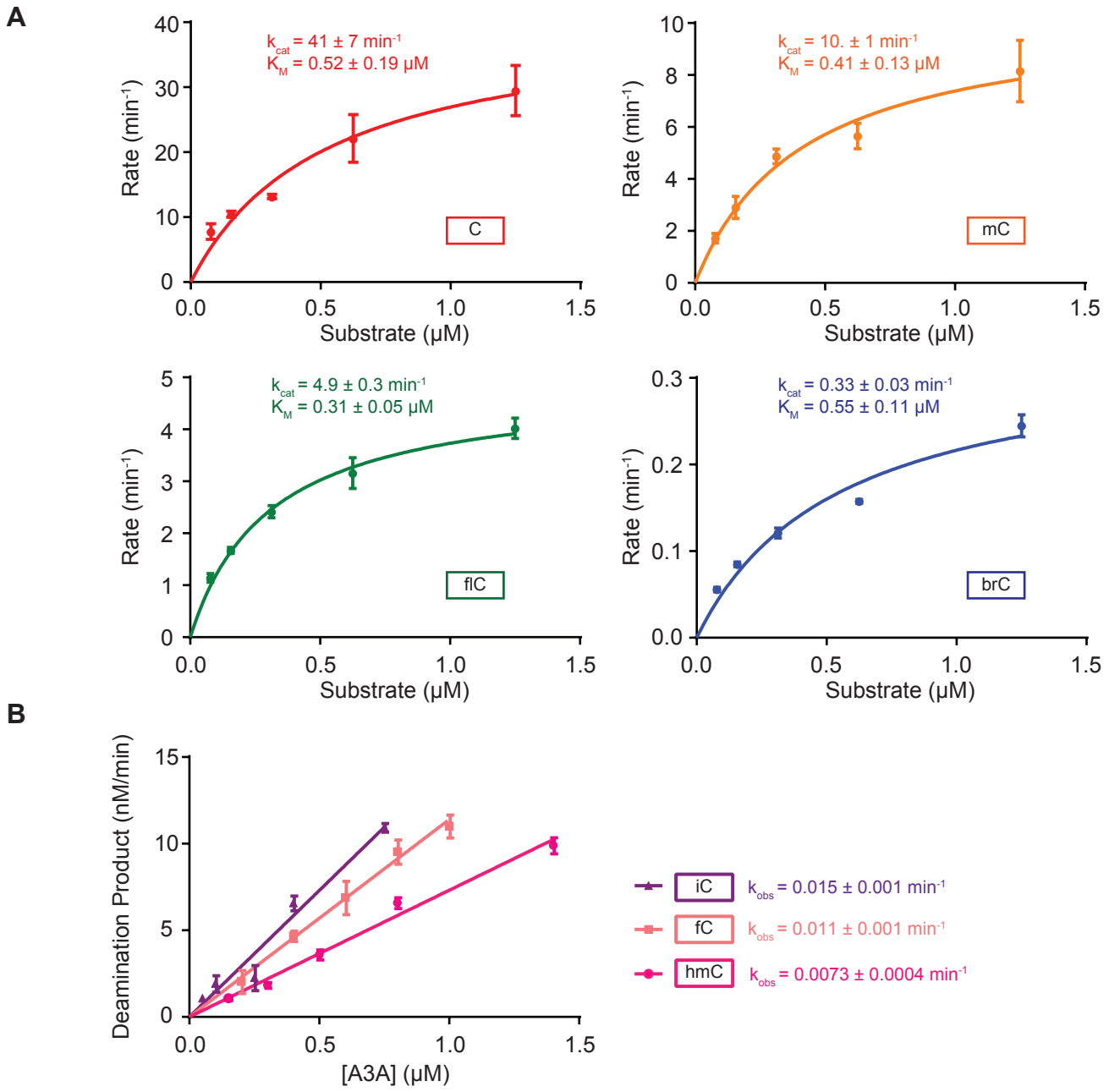


Figure S3. Curves from steady-state kinetic analyses. (A) Rate versus substrate concentration curves for the substrates S35-C, S35-mC, S35-flC, and S35-brC with corresponding kinetic constants shown. Data were fit with Michaelis-Menten parameters using GraphPad Prism software. Error bars indicate standard deviation from three independent replicates. (B) Deamination product versus enzyme concentration curves for S35-iC, S35-fC, and S35-hmC substrates performed at saturating substrate concentrations. Experiments were performed in duplicate (S35-iC) or triplicate (S35-hmC, S35-fC) and fit with linear regression on GraphPad Prism software after appropriate background subtraction from control experiments without A3A treatment. Error bars indicate range for S35-iC experiments and standard deviation from replicate experiments for all other substrates.

Figure S4

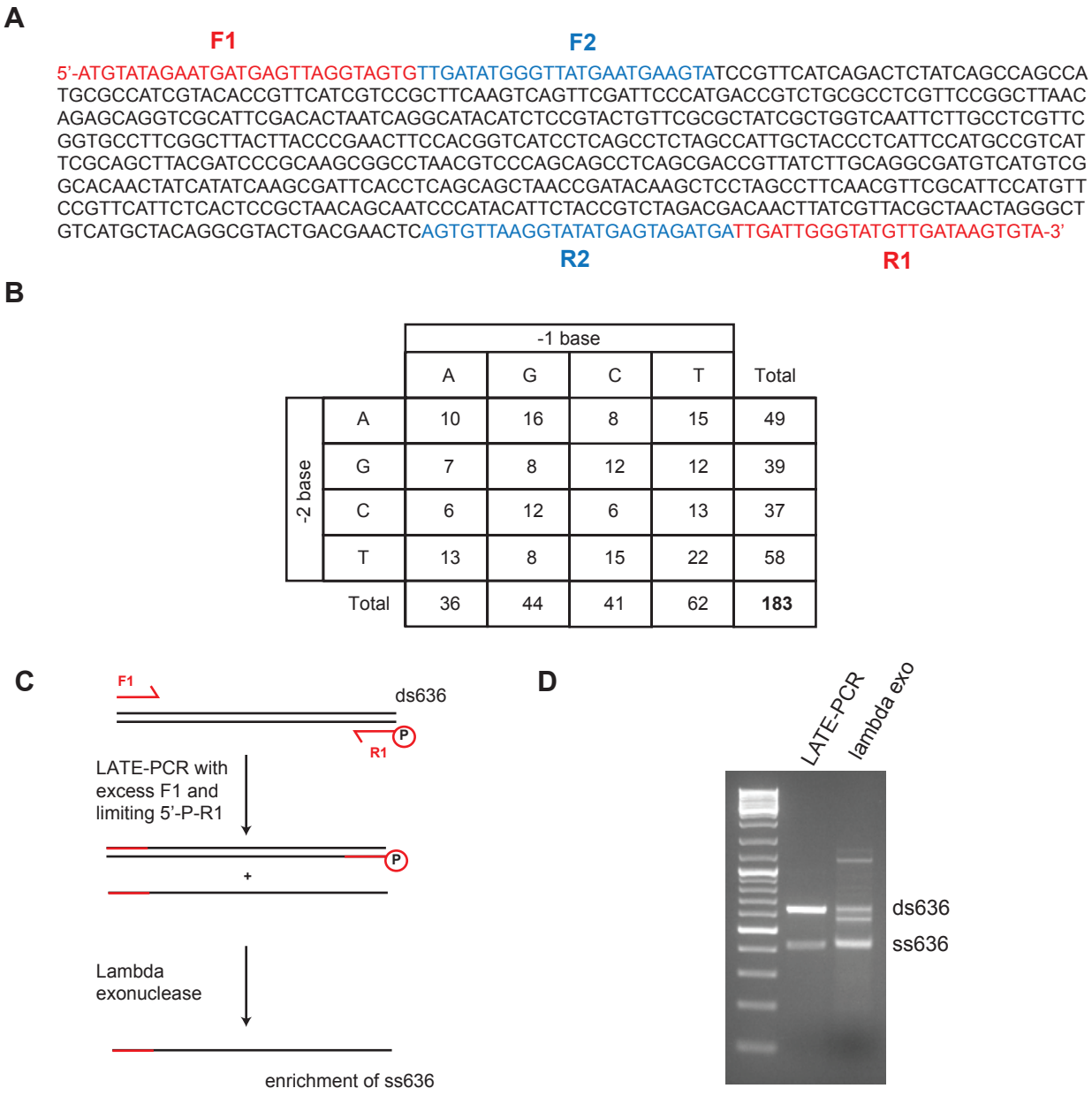


Figure S4. Design and synthesis of long ssDNA substrates. (A) 636mer DNA substrate designed to analyze C and mC deamination. Primer touchdown locations are shown in red (exterior primers; F1 and R1) and blue (interior primers; F2 and R2). (B) Table of sequence contexts and their occurrences in the 636mer substrate. The sequence contains 183 cytosine bases with all NNC triplets represented (range 6-22 times). (C) Schematic of ssDNA 636mer synthesis. LATE-PCR is performed with dATP, dGTP, dTTP and either dCTP or 5-methyl-dCTP. A limiting amount of the phosphorylated primer R1 is used, giving rise to both double-stranded and single-stranded species. The phosphorylated strand is degraded by lambda exonuclease to enrich for the ssDNA substrate. (D) A 2% agarose gel stained with Sybr-Gold, depicting double-stranded and single-stranded species resulting from LATE-PCR and subsequent lambda exonuclease treatment. The ssDNA is gel purified from the lambda exonuclease treated sample and used for further studies.

Figure S5

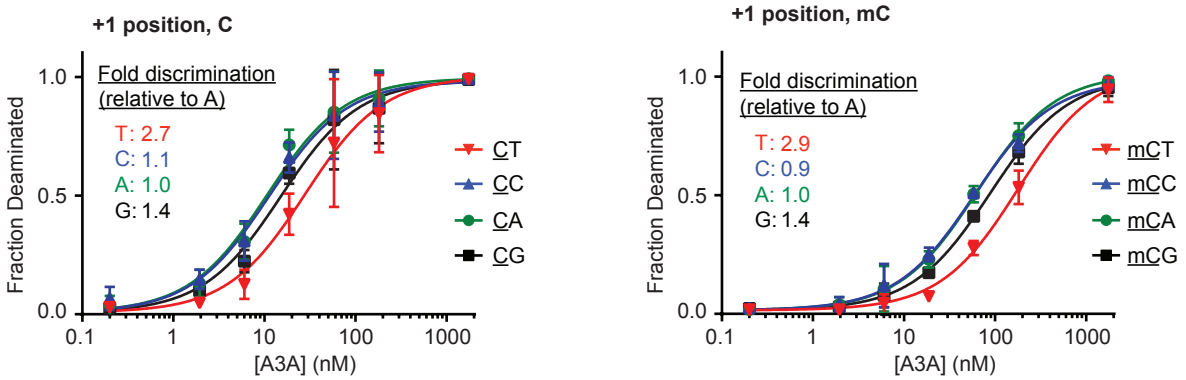


Figure S5. The effect of downstream (+1) position on C and mC deamination. The deamination experiments on long ssDNA substrates described in Figure 3 were analyzed for the impact of the downstream (+1) position on C and mC deamination. Plotted are the fractions of deaminated bases in each context as a function of A3A concentration, shown in identical format to the -2 and -1 position analysis in Figure 3. The EC_{50} is the enzyme concentration to achieve 50% deamination of C or mC bases in each condition and was calculated by fitting each curve to a hyperbolic function. The fold discrimination is the ratio of EC_{50} for a given base, relative to that of A, the most favored +1 base for C deamination.

Figure S6

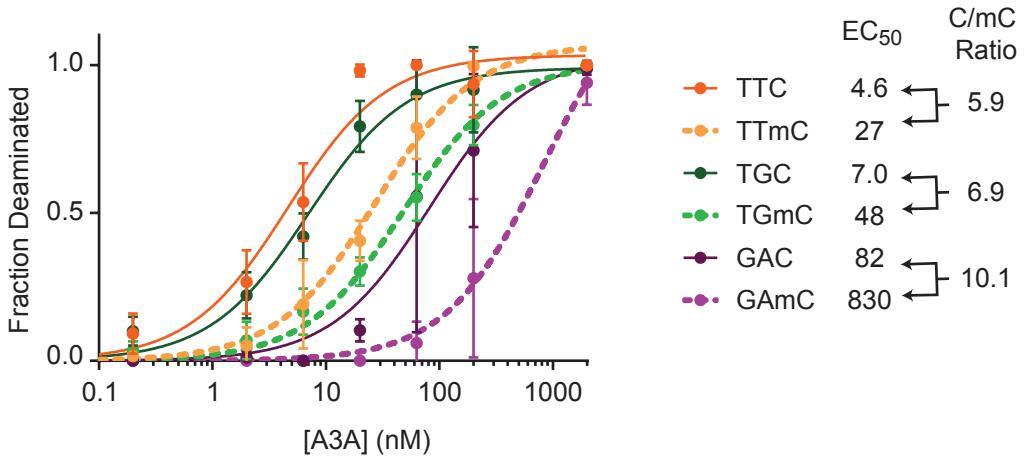


Figure S6. The combined effect of upstream sequence context on C and mC deamination. For the 636-mer ssDNA substrate deamination experiments shown in Figure 3, the cytosine bases in a preferred TTC/mC, intermediate TGC/mC, or disfavored GAC/mC context were analyzed separately for deamination by A3A. Plotted is the fraction of deaminated bases in each condition as a function of A3A concentration. The values for EC_{50}, the enzyme concentration to achieve 50% deamination of all C or mC bases in a particular context, are reported, along with the ratio of C:mC deamination for each sequence context.

Figure S7

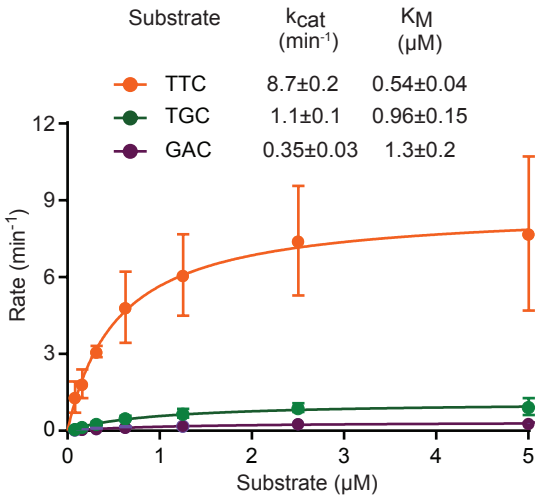


Figure S7. Kinetics of C deamination in varied sequence contexts. Oligonucleotides containing a single cytosine in preferred (TTC), intermediate (TGC), or disfavored (GAC) sequence context were reacted with A3A under steady-state conditions appropriate for each substrate (see Material and Methods). Deamination was analyzed by a UDG-coupled assay and fit to Michaelis-Menten parameters. While the overall k_{cat} values determined by the UDG-coupled assay are lower than those determined by the TDG-coupled assay (Figure 4), the trends are consistent. Importantly, differences in reactivity are predominantly attributable to changes in k_{cat}. While the determined k_{cat} values span a ~25-fold range across the various sequence contexts, the K_M values differ by only 2.2-fold.

Table S1

A

Name	Oligonucleotide Sequence	Theoretical Mass (Da)	Observed Mass (Da)
S35-C	5' TGAGGAATGAAGTTGATT**C**AAATGTGATGAGGTGA 3' - FAM	11548	11547
S35-U	5' TGAGGAATGAAGTTGATT**U**AAATGTGATGAGGTGA 3' - FAM	11549	11549
S35-mC	5' TGAGGAATGAAGTTGATT**mC**AAATGTGATGAGGTGA 3' - FAM	11562	11565
S35-T	5' TGAGGAATGAAGTTGATT**T**AAATGTGATGAGGTGA 3' - FAM	11563	11566
S35-hmC	5' TGAGGAATGAAGTTGATT**hmC**AAATGTGATGAGGTGA 3' - FAM	11578	11582
S35-hmU	5' TGAGGAATGAAGTTGATT**hmU**AAATGTGATGAGGTGA 3' - FAM	11579	11587
S35-fC	5' TGAGGAATGAAGTTGATT**fC**AAATGTGATGAGGTGA 3' - FAM	11577	11573
S35-caC	5' TGAGGAATGAAGTTGATT**caC**AAATGTGATGAGGTGA 3' - FAM	11593	11593
S35-flC	5' TGAGGAATGAAGTTGATT**flC**AAATGTGATGAGGTGA 3' - FAM	11567	11569
S35-flU	5' TGAGGAATGAAGTTGATT**flU**AAATGTGATGAGGTGA 3' - FAM	11568	11571
S35-brC	5' TGAGGAATGAAGTTGATT**brC**AAATGTGATGAGGTGA 3' - FAM	11628	11629
S35-brU	5' TGAGGAATGAAGTTGATT**brU**AAATGTGATGAGGTGA 3' - FAM	11629	11629
S35-iC	5' TGAGGAATGAAGTTGATT**iC**AAATGTGATGAGGTGA 3' - FAM	11675	11680
S35-iU	5' TGAGGAATGAAGTTGATT**iU**AAATGTGATGAGGTGA 3' - FAM	11676	11677

B

S35-TTC	5' TGAGGAATGAAGTTGA*TT***C**GGATGTGATGAGGTGA 3'	11010	11015
S35-TTU	5' TGAGGAATGAAGTTGA*TT***U**GGATGTGATGAGGTGA 3'	11011	11013
S35-TTmC	5' TGAGGAATGAAGTTGA*TT***mC**GGATGTGATGAGGTGA 3'	11024	11025
S35-TTT	5' TGAGGAATGAAGTTGA*TT***T**GGATGTGATGAGGTGA 3'	11025	11028
S35-TGC	5' TGAGGAATGAAGTTGA*TG***C**GGATGTGATGAGGTGA 3'	11035	11037
S35-TGU	5' TGAGGAATGAAGTTGA*TG***U**GGATGTGATGAGGTGA 3'	11036	11040
S35-TGmC	5' TGAGGAATGAAGTTGA*TG***mC**GGATGTGATGAGGTGA 3'	11049	11049
S35-TGT	5' TGAGGAATGAAGTTGA*TG***T**GGATGTGATGAGGTGA 3'	11050	11054
S35-GAC	5' TGAGGAATGAAGTTGA*GA***C**GGATGTGATGAGGTGA 3'	11044	11046
S35-GAU	5' TGAGGAATGAAGTTGA*GA***U**GGATGTGATGAGGTGA 3'	11045	11048
S35-GAmC	5' TGAGGAATGAAGTTGA*GA***mC**GGATGTGATGAGGTGA 3'	11058	11065
S35-GAT	5' TGAGGAATGAAGTTGA*GA***T**GGATGTGATGAGGTGA 3'	11059	11057

C

S35-C-Comp	5' TCACCTCATCACATTTGAATCAACTTCATTCCTCA 3'	10520	10521
S35-TT-Comp	5' TCACCTCATCACATCCGTCTCAACTTCATTCCTCA 3'	10457	10458
S35-TG-Comp	5' TCACCTCATCACATCCGCATCAACTTCATTCCTCA 3'	10466	10467
S35-GA-Comp	5' TCACCTCATCACATCCGAATCAACTTCATTCCTCA 3'	10490	10491

Table S1. Sequences and mass confirmation of substrate and product control oligonucleotides. (A) Oligonucleotides were synthesized using standard phosphoramidite chemistry including a 3'-FAM moiety. The masses of these oligonucleotides were confirmed using MALDI-TOF mass spectrometry. (B) Oligonucleotides used for mC sequence preference studies were ordered from IDT, and mass was confirmed using MALDI-TOF mass spectrometry. A DNA standard (MW=11598.2 Da) was ordered from IDT and read on MALDI as 11601 Da, indicative of mass range accuracy. (C) Complement oligonucleotides used in SwaI or TDG-based assays were ordered from IDT and the reported masses are from ESI-MS analysis performed by IDT.