Selective Inhibition of APOBEC3 Enzymes by Single-Stranded DNAs Containing 2’-Deoxyzebularine

Fareeda M. Barzak¹, Stefan Harjes¹, Maksim V. Kvach¹, Harikrishnan M. Kurup¹, Geoffrey B. Jameson¹,², Vyacheslav V. Filichev¹,²*, Elena Harjes¹,²*

¹ School of Fundamental Sciences, Massey University, Private Bag 11 222, Palmerston North 4442, New Zealand; ² Maurice Wilkins Centre for Molecular Biodiscovery, Auckland 1142, New Zealand.

*E-mail: e.harjes@massey.ac.nz, v.filichev@massey.ac.nz

Table of Contents
1 Synthesis of oligodeoxynucleotides containing 2’-deoxyzebularine (dZ) .............................................S2
  1.1 General methods .................................................................................................................................S2
  1.2 Synthesis of dZ-modified oligos ........................................................................................................S2
2 Protein expression and purification .........................................................................................................S4
3 Kinetic characterization of deamination activity using NMR-based assay ...........................................S8
4 Evaluation of inhibitors in NMR-based assay ......................................................................................S18
References ..................................................................................................................................................S20
1 Synthesis of oligodeoxynucleotides containing 2'-deoxyzebularine (dZ)

1.1 General methods
High-resolution electrospray mass spectra were recorded on a ThermoScientific Q Exactive Focus Hybrid Quadrupole-Orbitrap mass spectrometer. Ions generated by ESI were detected in negative ion mode. Total ion count (TIC) was recorded in centroid mode over the m/z range of 500-3,000 and analyzed using ThermoXcalibur Qual Browser. ‘Saltless buffer’ used for oligonucleotide desalting consists of 10 µM Tris-HCl (pH 8.0), 1 µM EDTA and 0.001% w/v NaN₃.

1.2 Synthesis of dZ-modified oligos
2’-Deoxyzebularine phosphoramidite was prepared as described previously. ¹

Oligonucleotide syntheses were carried out on a MerMade-4 DNA/RNA synthesizer (BioAutomation) on a 5 µmol scale using the manufacturer’s standard protocol. Coupling time of 2’-deoxyzebularine phosphoramidite was increased to 5 min. The final detritylation step was omitted and DMT-ON oligonucleotides were cleaved from the solid support and deprotected with conc. ammonia solution (1.0 mL) at room temperature for 24 h. After filtering, an aq. solution of 0.3 M LiClO₄ (0.5 mL) was added and oligonucleotides were precipitated with acetone (14 mL). The DMT-ON oligonucleotides were isolated by reversed-phase HPLC on 250/10 mm, 5 µm, 300 Å C18 column (Phenomenex) in a gradient of CH₃CN (0→60% for 15 min, 4.6 mL/min) in 0.1 M TEAA buffer (pH 7.0) with detection at 260 nm. DMT-ON oligonucleotides were freeze-dried and manually detritylated with 80% aq. AcOH (2 mL) during 20 min at room temperature. 3 M AcONa solution (0.5 mL) was then added and oligonucleotides were precipitated with 2-propanol (11 mL). DMT-OFF oligonucleotides were purified by RP-HPLC on 250/4.6 mm, 5 µm, 300 Å C18 column (Phenomenex) in a gradient of CH₂CN (0→25% for 20 min, 1.3 mL/min) in 0.1 M TEAA buffer (pH 7.0) (Chart S1). Fractions containing desired oligonucleotides were combined, freeze-dried, dissolved in milli-Q water (1.5 mL) and desalted on a NAP-25 column (GE Healthcare) against ‘saltless buffer’. Pure products were quantified by measuring absorbance at 260 nm, analyzed by ESI-MS (Table S1) and concentrated by freeze-drying. TdZA-oligo was synthesized previously. ¹
Table S1. Modified oligonucleotides and purified using protocols described above.

| Name         | DNA sequence 5'→3'       | ESI-MS                |
|--------------|--------------------------|-----------------------|
| CCdZ-oligo   | ATTCCdZAATT              | 2945.5377 / 2945.5294 |
| dZCC-oligo   | ATTdZCCAATT              | 2945.5217 / 2945.5294 |

Chart S1. Reverse-phase HPLC analytical profiles of purified CCdZ and dZCC oligos.
2 Protein expression and purification

Proteins were prepared as described previously\(^1\)(Chart S2). The A3B C-terminal domain (residues 187 to 378, UniProt ID Q9UH17) was cloned into pET24a vector (Novagen) to produce A3B\(_{\text{CTD}}\) proteins with a non-cleavable C-terminal His\(_6\)-tag (LEHHHHHH), adapted from a previously described study.\(^2\) Derivative constructs previously reported were used in this study, A3B\(_{\text{CTD}}\)-QM-ΔL3-AL1swap (termed A3A-mimic), and A3B\(_{\text{CTD}}\)-DM expressed in the \(E.\ coli\) strain C41(DE3)pLysS (Lucigen). \(E.\ coli\) culture was grown at 37 °C in LB medium. Once mid-log growth phase was established the culture was supplemented with 100 µM zinc chloride, before inducing protein expression by the addition of isopropyl \(\beta\)-D-1-thiogalactopyranoside (IPTG) at a final concentration of 0.5 µM and incubating overnight at 18 °C.

GST-fused A3G\(_{\text{CTD}}\)(191-384, NM_021822, wt) was purified as previously described.\(^3\) The N-terminal glutathione S-transferase (GST)-fused A3G\(_{\text{CTD}}\) was expressed in \(E.\ coli\) BL21(DE3) cells overnight at 17°C. After harvesting, the cells were resuspended in 50 mM sodium phosphate buffer (pH 7.4) and lysed by sonication. After ultracentrifugation at 25,000 \(g\) for 10 min, the supernatant was added to glutathione (GSH)-Sepharose, which was subsequently washed. For kinetic analysis, the GST fusion protein was eluted from the Sepharose matrix with 100 mM GSH in phosphate buffer. By using filtration at 4,000 \(g\), the buffer was changed to a solution containing 75 mM sodium phosphate and 75 mM citrate, at pH 5.5.
A3G<sub>CTD</sub>  ------EILRHSMDPPTFNNEPWRGRHETYLCYEVENMHDNWVLLQQRPGFCN
A3A           MEASPASGPHLMDPHIFTSNFNNNG---IGRHKTYLCEVERLNGTSKVMDQHRGFLKN
A3B<sub>CTD</sub>  ------EILRHLMTDPFTFDNDFLRLRV&LRYLCEVERLNGTWMQHRGFLKN
A3B<sub>CTD</sub>-DM  ------EILRLMDPDTSNFNNDPLVLRRQTYLCEVERLNGTWMQHRGFLCN
A3A-mimic  ------EILRHLMDPPTSNFNNG---IGRHKTYLCEVERLNGTSVKMDQHMFLCN

A3G<sub>CTD</sub>  QAPKHGFLEGRHAELCFLDVIFPKLILDDQYTVCTFSTWSPCFS---CAQEMAKFIKN
A3A           QAKNLLCGFYGRHAELRFLDLVPSLQPAQTYRTVIFSWSFSWGCAGEVRAFLQEN
A3B<sub>CTD</sub>  EAKNLLCGFYGRHAELRFLDLVPSLQPAQTYRTVIFSWSFSWGCAGEVRAFLQEN
A3B<sub>CTD</sub>-DM  EAKNLLCGFYGRHAELRFLDLVPSLQPAQTYRTVIFSWSFSWGCAGEVRAFLQEN
A3A-mimic  E----------SGRHAELRFLDLVPSLQPAQTYRTVIFSWSFSWGCAGEVRAFLQEN

A3G<sub>CTD</sub>  KHVSLCIFTARIYDDQGRCQEGLRTLAEAGAKISMTYSEFKHCWTFDVHQCPFPQW
A3A           THVRLRIFAARIYDYDPLYKEALQMLRDAGAQVSMTYDEFYCWTFVRQGCPCFWQ
A3B<sub>CTD</sub>  THVRLRMKAARIYDDPLYKEALQMLRDAGAQVSMTYDEFYCWTFVRQGCPCFWQ
A3B<sub>CTD</sub>-DM  THVRLRMKAARIYDDPLYKEALQMLRDAGAQVSMTYDEFYCWTFVRQGCPCFWQ
A3A-mimic  THVRLRMKAARIYDDPLYKEALQMLRDAGAQVSMTYDEFYCWTFVRQGCPCFWQ

A3G<sub>CTD</sub>  DGLDEHSQDSLGRAILQNPQEN
A3A           DGLDEHSQALSGLRAILQNPQEN
A3B<sub>CTD</sub>  DGLDEHSQALSGLRAILQNPQEN
A3B<sub>CTD</sub>-DM  DGLDEHSQALSGLRAILQNPQEN
A3A-mimic  DGLDEHSQALSGLRAILQNPQEN

**Chart S2.** A3B<sub>CTD</sub> and A3G<sub>CTD</sub> are C-terminal domains of A3B and A3G. A3B<sub>CTD</sub>-DM (double mutant): L230K and F308K; A3A-mimic = A3B<sub>CTD</sub>-QM-∆L3-AL1 (Quadra Mutant: F200S, W228S, L230K, and F308K; loop3 deleted: Ala-242 to Tyr-250 replaced by Ser; loop 1 swapped with A3A). Green shaded are identical amino acids between A3A and A3B<sub>CTD</sub>.**
Chart S3. Surface diagrams (A and B) showing active site (A) and 180°-rotated backside (B) of A3B_{CTD} (pdb-code: 5cqh). C and D are corresponding cartoon diagrams.

Two A3B_{CTD}-DM mutations (L230K and F308K) are in blue and are located on the backside from the active site. Active site, zinc-coordinating residues and zinc ion, are in magenta.
Chart S4. pH influence on rate of deamination by A3A-mimic and A3B<sub>CTD</sub>-DM on TCA-oligo assessed using $^1$H-NMR activity assay in 50 mM citrate-phosphate buffer at indicated pH, supplemented with 200 mM NaCl, 2 mM β-mercaptoethanol, 200 µM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS) and 10 % deuterium oxide. pH 5.5 was selected for A3A-mimic and pH 7.5 for A3B<sub>CTD</sub>-DM.
3 Kinetic characterization of deamination activity using NMR-based assay

Kinetic characterization of A3 protein variants (A3B<sub>CTD</sub>-QM-ΔL3-AL1swap, A3B<sub>CTD</sub>-DM, and A3G<sub>CTD</sub>) on preferred dC in CCC-, dZZC-, CCdZ-oligos was made using an established NMR-based assay<sup>3</sup> (see Table S2 for description of oligos used). Measurements were acquired on a 700-MHz Bruker NMR spectrometer equipped with a 1.7-mm cryoprobe at 298 K. A series of <sup>1</sup>H NMR spectra were recorded of the substrates at concentrations ranging from 50 µM to 750 µM (or until saturation) with 50 nM of A3B<sub>CTD</sub>-QM-ΔL3-AL1swap, 2 µM of A3B<sub>CTD</sub>-DM, or 850 nM A3G<sub>CTD</sub> in buffer containing 10% deuterium oxide; 50 mM citrate-phosphate, 200 mM NaCl, 2 mM β-mercaptoethanol, 200 µM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS); pH 5.5 for A3B<sub>CTD</sub>-QM-ΔL3-AL1swap, pH 7.5 for A3B<sub>CTD</sub>-DM or 50 mM sodium phosphate, 1 mM citrate, 100 mM NaCl, 2 mM β-mercaptoethanol, 50 µM DSS, pH 6.0 for A3G<sub>CTD</sub>. See explanation below for the choice of pH used in this study.

<sup>1</sup>H NMR spectra of oligonucleotide standards containing 2'-deoxyuridine (dU) were conducted to determine where the H-5 proton doublet signals of dU products appear (Chart S5). The H-5 proton doublet signal of dU at the 5' position in dUCC-oligo appears between 5.79 to 5.76 ppm, signals for dU in the middle position of CdUC appear between 5.76 to 5.74 ppm, dU at the 3' position in CCdU appears between 5.68 to 5.66 ppm. The H-5 proton doublet signal of dU of the first product (CCdU or dUCC) were baselined and integrated, followed by converting integrated signal area to substrate concentration and plotting versus time of the reaction. The data were then fitted by linear regression to determine the initial speed of the reaction as a function of substrate concentration (Charts S8, S10, S12, S14, S16). The data as double reciprocal plots (Charts S9, S11, S13, S15 and S17) were then fitted with linear regression to determine $K_m$ and $k_{cat}$ (see explanation below). Uncertainties of $K_m$ and $k_{cat}$ were calculated using error-propagation method.

In ideal situation, it should be possible to obtain $K_m$ and $k_{cat}$ values by direct fitting of the substrate/initial rate data. For A3 enzymes, the analysis of product formation over long time is complicated by several factors: i) instability of enzymes during long experiments and ii) when substrate concentration decreases the enzyme starts to deaminate other cytosines in the CCdU- and dUCC-motifs. Using NMR assay the initial rate of deamination for the preferred cytosine (the one that is deaminated first) can be obtained. By performing experiments at different substrate concentrations, the data obtained that can be fitted with
a straight line in a double-reciprocal plot, which is much more meaningful to analyse than
fitting into any other curve shape.

$K_m$ and $K_i$ (see next section on evaluation of inhibitors) values are calculated values obtained
from a linear fit of the data acquired in experiments at various substrate or inhibitor
concentrations using Lineweaver-Burk or Dixon plots, respectively. So errors for $K_m$ and $K_i$
can be only calculated based on the fit of experimental data. One should emphasise that
each point on these graphs is an individual experiment where product formation/substrate
disappearance was monitored over time at a given substrate and/or inhibitor
concentrations using the same enzyme concentration. Each kinetic experiment takes 1.5 –
2.5 hours on 700 MHz NMR. For reliability of $K_m$ and $K_i$ values, it is much more meaningful to
perform experiments at various substrate or inhibitor concentrations and thus get more
data points in Lineweaver-Burk or Dixon plots than performing the same experiment
multiple times
For proper evaluation of $K_m$ and $K_i$ values, the use of the same but suboptimal pH, for
example pH = 7 between optimal pH of A3A-mimic and A3Bctd, is not feasible in NMR assay:
reaction speed of A3A-mimic and A3Bctd is at least halved (Chart S4 in Supporting
Information) and for A3Gctd is 10 times lower in comparison with reaction speed at optimal
pH.³ For A3Gctd, the enzyme cannot be saturated with the CCC-oligo (substrate) at pH 7
even at 10 mM substrate concentration.³ This means that it is unfeasible not only to reliably
detect enzyme activity on modified substrates and to measure residual enzyme activity in
the presence of inhibitors but also to kinetically characterise substrates and inhibitors at pH
= 7. One should note that for the wild-type enzymes the discrimination pattern for
deamination of 5’-C versus 3’-C in the CCC-motif is prevalent in physiological pH around 7,⁴
which means that our inhibitors would be still selective on wild-type enzymes at this pH.
Table S2. Oligonucleotides designed and used in study

| Name           | DNA sequence, 5’-3’ | Rational                                                                                                                                 |
|----------------|--------------------|----------------------------------------------------------------------------------------------------------------------------------------|
| TCA-oligo      | ATTTCAATT          | Used as preferred substrate sequence motif of A3A and A3B<sub>CTD</sub>                                                            |
| CCC-oligo      | ATTCCCAATT         | The 3<sup>rd</sup> dC (from 5’ to 3’ ends) is the preferred substrate of A3G<sub>CTD</sub> and preferentially deaminated by A3A. The 1<sup>st</sup> dC is preferentially deaminated by A3B<sub>CTD</sub>. |
| TdZA-oligo     | ATTTdZATTT         | Non-selective inhibitor of A3A and A3B<sub>CTD</sub>                                                                                   |
| CCdZ-oligo     | ATTCdZAATT         | Modified oligonucleotide to compare with the substrate (CC<sub>C</sub>) and study inhibition of A3A-mimic. Determine if it is a substrate of A3B<sub>CTD</sub> (the 1<sup>st</sup> dC). Inhibitor of A3G<sub>CTD</sub>. |
| dZCC-oligo     | ATtdZCCAATT        | Modified oligonucleotide to compare with the substrate (CC<sub>C</sub>) and study inhibition of A3B<sub>CTD</sub>. Determine if it is a substrate of A3A and A3G<sub>CTD</sub> (the 2<sup>nd</sup> dC in the sequence). |

Chart S5. <sup>1</sup>H-NMR spectra of 5'-ATTCCCAATT sequence and possible products of dC deamination in pH 5.5 buffer (50 mM citrate-phosphate, 200 mM NaCl, 2 mM β-mercaptoethanol, 200 μM 4,4-dimethyl-4-silapentane-1-sulfonic acid (DSS); pH 5.5).
**Chart S6.** Two plausible stem-loop structures predicted by m-Fold for CCC-oligo, both with unfavourable positive Gibbs free energy of formation.

**Chart S7.** CD spectra of the CCC-oligo (10 nM) at 20 °C in 10 mM Li-cacodylate buffer at indicated pH supplemented with 100 mM NaCl. CD spectra were recorded using a Chirascan CD spectrophotometer (150 W Xe arc) from Applied Photophysics with a Quantum Northwest TC125 temperature controller. CD spectra (average of at least 3 scans) were recorded between 220 and 350 nm with 1 nm intervals, 120 nm/min scan rate followed by subtraction of a background spectrum (buffer only). Stem-loop DNA structures usually have distinctive features of DNA duplexes with a peak at 280 nm with positive ellipticity and a
peak at 250 nm with negative ellipticity\textsuperscript{5} (not observed here). We also did not see any distinctive patterns of other non-canonical DNA secondary structures\textsuperscript{5} including an i-motif formed in C-rich regions. m-Fold cannot predict such structures.

**Chart S8.** Speed of deamination catalyzed by A3A-mimic (A3B\textsubscript{CTD-QM-ΔL3-AL1swap}, 50 nM) as a function of substrate (5'-ATTCC\underline{C}AATT) concentration. Bold, underlined C is deaminated.

**Chart S9.** Double reciprocal plot of inverse speed of deamination catalyzed by A3A-mimic (A3B\textsubscript{CTD-QM-ΔL3-AL1swap}) as a function of the inverse substrate concentration (5'-\underline{ATTCCCAATT}). Bold, underlined C is deaminated.
**Chart S10.** Speed of deamination catalyzed by A3B<sub>CTD</sub>-DM (2 µM) as a function of substrate (5'-ATT<sub>C</sub>CCAATT) concentration. Bold, underlined C is deaminated.

**Chart S11.** Double reciprocal plot of inverse speed of deamination catalyzed by A3B<sub>CTD</sub>-DM as a function of the inverse substrate concentration (5'-ATT<sub>C</sub>CCAATT). Bold, underlined C is deaminated.
**Chart S12.** Speed of deamination catalyzed by A3A-mimic (A3B<sub>CTD</sub>-QM-ΔL3-AL1swap, 50 nM) as a function of substrate (5'-ATT<sub>dZ</sub>C<sub>C</sub>AATT) concentration. Bold, underlined C is deaminated.

**Chart S13.** Double reciprocal plot of inverse speed of deamination catalyzed A3A-mimic (A3B<sub>CTD</sub>-QM-ΔL3-AL1swap, 50 nM) as a function of the inverse substrate concentration (5'-ATTdZCCCAAATT). Bold, underlined C is deaminated.
**Chart S14.** Speed of deamination catalyzed by A3B_{CTD}-DM (2 µM) as a function of substrate (5'-ATT\textbf{C}CdZAATT) concentration. Bold, underlined C is deaminated.

**Chart S15.** Double reciprocal plot of inverse speed of deamination catalyzed by A3B_{CTD}-DM as a function of the inverse substrate concentration (5'-ATT\textbf{C}CdZAATT). Bold, underlined C is deaminated.
**Chart S16.** Speed of deamination catalyzed by A3G$_{CTD}$ (850 nM) as a function of substrate (5'-ATT$dZC$C$AATT$) concentration.

**Chart S17.** Double reciprocal plot of inverse speed of deamination catalyzed by A3G$_{CTD}$ as a function of the inverse substrate concentration (5'-ATT$dZC$C$AATT$). Bold, underlined C is deaminated.
Chart S18. $^1$H NMR spectrum of CCdZ-oligo (350 µM) in the presence of A3A-mimic (50 nM) at different time points showing that no changes are detected for CCdZ-oligo. Uridine product peaks should appear in the boxed region.
4. Evaluation of inhibitors in NMR-based assay

The NMR-based inhibition assay described previously was conducted to determine the deaminase activity of A3 variants in the presence of dZ-containing oligodeoxynucleotides. NMR spectra were recorded using 350 µM 5'-ATTTCATT T as a substrate for A3A-mimic and A3B_{CTD–DM} with varying concentrations of dZ-oligos ranging from 5 µM to 100 µM in the presence of 50 nM of A3B_{CTD–QM–ΔL3–AL1swap} (A3A-mimic) or 2 µM of A3B_{CTD–DM} in activity assay buffer as mentioned previously at 298 K.

Integration of the H-5 proton doublet signal of the cytosine (between 5.92 to 5.88 ppm, Chart S19) is converted to substrate concentration and plotted versus time of the reaction (Chart S20). This plot was then fitted with linear regression to determine the speed of the reaction in the presence of inhibitor. The Dixon plot of inverse speed versus inverse inhibitor concentration (Figures 3C and D in the main text) was then fitted with linear regression to derive the inhibition constant ($K_i$). Uncertainty of $K_i$ was calculated using error-propagation method.

Chart S19. $^1$H NMR spectra of TCA-oligo in the presence of A3A-mimic in the absence (left) and presence of inhibitor (CCdZ-oligo, right) at different time intervals. $^1$H NMR spectra of the substrate and the product are provided for reference.
Chart S20. Substrate concentration (TCA-oligo) plotted versus time of the reaction in the presence of A3A-mimic in the absence (green dots) and presence of inhibitor (CCdZ-oligo, 50 µM, blue dots) showing linear regression.
References

[1] Kvach, M. V., Barzak, F. M., Harjes, S., Schares, H. A. M., Jameson, G. B., Ayoub, A. M., Moorthy, R., Aihara, H., Harris, R. S., Filichev, V. V., Harki, D. A., and Harjes, E. (2019) Inhibiting APOBEC3 Activity with Single-Stranded DNA Containing 2’-Deoxyzebularine Analogues, *Biochemistry* 58, 391-400.

[2] Shi, K., Carpenter, M. A., Banerjee, S., Shaban, N. M., Kurahashi, K., Salamango, D. J., McCann, J. L., Starrett, G. J., Duffy, J. V., Demir, O., Amaro, R. E., Harki, D. A., Harris, R. S., and Aihara, H. (2017) Structural basis for targeted DNA cytosine deamination and mutagenesis by APOBEC3A and APOBEC3B, *Nat. Struct. Mol. Biol.* 24, 131-139.

[3] Harjes, S., Solomon, W. C., Li, M., Chen, K.-M., Harjes, E., Harris, R. S., and Matsuo, H. (2013) Impact of H216 on the DNA binding and catalytic activities of the HIV restriction factor APOBEC3G, *J. Virol.* 87, 7008-7014.

[4] Byeon, I.-J. L., Byeon, C.-H., Wu, T., Mitra, M., Singer, D., Levin, J. G., and Gronenborn, A. M. (2016) Nuclear Magnetic Resonance Structure of the APOBEC3B Catalytic Domain: Structural Basis for Substrate Binding and DNA Deaminase Activity, *Biochemistry* 55, 2944-2959.

[5] Kypr, J., Kejnovska, I., Renciuk, D., and Vorlickova, M. (2009) Circular dichroism and conformational polymorphism of DNA, *Nucleic Acids Res.* 37, 1713-1725.