Quantitative analysis of the location- and sequence-dependent deamination by APOBEC3G using real-time NMR**

Ayako Furukawa, Kenji Sugase, Ryo Morishita, Takashi Nagata, Tsutomu Kodaki, Akifumi Takaori-Kondo, Akihide Ryo, and Masato Katahira

Dedication((optional)

Abstract: Human APOBEC3G (A3G) deaminates the newly synthesized minus strand of human immunodeficiency virus-1 (HIV-1), resulting in the abolition of the HIV-1 infectivity against virus infectivity factor (Vif)-deficient strains. [1-6] *A unique property of A3G is that it deaminates a CCC hot spot located close to the 5' end more effectively than one less close to the 5' end. However, the mechanism of this property is elusive because the deamination process that includes A3G non-specific binding to DNA and sliding along it cannot be analyzed by existing methods using the Michaelis-Menten theory. In this study, we have developed a new real-time NMR method to examine the non-specific binding and sliding processes explicitly, and applied it to the analysis of the deamination by A3G. As a result, the location-dependent deamination can be explained by the difference in catalytic rates depending on the direction of approach of A3G to the target cytidine. Real-time NMR experiments also showed that A3G deaminates CCCC tandem hotspots with little redundancy suggesting that A3G efficiently*

[*] Dr. A. Furukawa, Prof. T. Nagata, Prof. T. Kodaki, Prof. M. Katahira
Institute of Advanced Energy, Graduate School of Energy Science,Kyoto University
Gokasho, Uji, Kyoto 611-0011, Japan
Fax: +81 774 38 3524
E-mail: katahira@iae.kyoto-u.ac.jp
Dr. K. Sugase
Bioorganic Research Institute, Suntory Foundation for Life Sciences
1-1-1 Wakayamadai, Shimamoto, Mishima, Osaka 618-8503, Japan
Dr. R. Morishita
CellFree Sciences Co., Ltd., Ehime University Venture
Matsuyama, Ehime 790-8577, Japan
Prof. A. Takaori
Department of Hematology and Oncology, Graduate School of Medicine, Kyoto University
54 Shogoin-Kawaracho, Sakyo-ku, Kyoto 606-8507, Japan
Prof. A. Ryo, Dr. R. Morishita
Department of Microbiology, Yokohama City University School of Medicine
3-9 Fukuura, Kanazawa-ku, Yokohama 236-0004, Japan

Supporting information for this article is available on the WWW under http://dx.doi.org/10.1002/anie.201xxxxxx.((Please delete if not appropriate))

mutates many CCC hotspots scattered throughout the HIV-1 genome.

A3G possesses two consensus zinc-finger type cytidine deaminase motifs (CD1 and CD2),[6] but only CD2 is catalytically active.[7,8] A3G preferably deaminates the third cytidine of a CCC sequence in single-stranded DNA (ssDNA).[9,10] It was reported that A3G nonspecifically binds to ssDNA and slides on ssDNA over 30 nm (69 nucleotides) without directional preference.[11,12] Interestingly, A3G deaminates CCC hot spots in the location-dependent manner.[10] A 5' to 3' gradient of mutations in HIV RNA, which is transcribed from the minus strand DNA, was observed in vivo. [9,13] This would be due to the 3' to 5' deamination gradient of the minus strand DNA by A3G. The deamination by A3G has been analyzed by various methods, such as gel shift assay,[10,14,15] single molecule FRET,[11] and atomic force microscopy.[12,16] We previously demonstrated that real-time NMR can be utilized to monitor the deamination, and revealed that A3G CD2 deaminates the third cytidine of CC$\underline{C}$ much faster than the second one (C$\underline{C}$C).[17] An advantage of real-time NMR over other methods is that it can directly detect a site-specific deamination reaction in real time. Moreover, this method is sensitive to weak interactions because highly concentrated ssDNA (mM order) can be used for the NMR experiment. The real-time NMR method that we used for the analysis of the deamination has been gaining in popularity and has been used by other groups.[18,19]

Our real-time NMR method monitors the intensity change of the H5-H6 total correlation spectroscopy (TOCSY) peak of the third cytidine of CC$\underline{C}$ in real-time. Using this method, we previously analyzed the deamination of a hot spot that solely exists in ssDNA. To gain an insight into the mechanism underlying the location-dependent deamination by A3G, we further developed this NMR method. Firstly, to determine whether or not real-time NMR has sufficiently high spectral and time resolution to distinguish multiple deamination reactions occurring on an ssDNA, we monitored deamination reactions of ssDNA comprising two CCC hot spots (S_{2CCC} in Table 1; Supporting Information). Obviously, the two hot spots were deaminated at different rates by full-length A3G (Figure 1a). Surprisingly, CD2 alone also deaminated the two hot spots in a location-dependent manner, exhibiting higher activity than the full-length A3G (Figure 1b). The higher activity of A3G CD2 was previously observed by another group.[14] F-tests showed that the differences in the deamination rate between the two hot spots is statistically significant for both the full-length A3G and CD2. The same result was obtained with ssDNA in which the positions of two units of S_{2CCC}, ATTCCCAAT and ATACCCATT were swapped. These results confirm that CD2 alone causes the

location-dependent deamination. Although catalytically inactive CD1 also supposedly contributes to the location-dependent deamination, we assumed that the intrinsic characteristics of CD2 are important for the location-dependent deamination by full-length A3G. We focused on CD2 in the following experiments. Subsequently, we investigated whether or not the real-time NMR method can be used to examine A3G sliding along ssDNA. Since A3G binds to ssDNA but not to dsDNA,[10] intervening dsDNA should block A3G sliding. As shown in Figure 1c and d, the deamination rates of two hot spots in a substrate with short dsDNA between them were almost the same, whereas the rates were location-dependent without dsDNA ($S_{CCCdsCCC}$ and $S_{CCCssCCC}$ in Table 1). This finding clearly indicates that real-time NMR can sense A3G CD2 sliding along ssDNA.

Table 1. Oligonucleotides used in this study

Name	Sequence
S_{2CCC}	ATTCCCAATTTTTTTTATACCCCATTT
$S_{CCCssCCC}$	ATTCCCAATTTTTTTTTTTTTTTTTTTTTATACCCCAT$_{23}$
$S_{CCCdsCCC}$	ATTCCCAATTCGCGAACGCGCAAGCGCATACCCCAT$_{23}$
	GCGCTTGCGCGTTCGCG
$S_{5'CCC}$	TTACCCCAT$_{42}$
S_{mCCC}	T$_{21}$ACCCCAT$_{22}$
$S_{3'CCC}$	T$_{41}$ACCCCATTT
S_{CCCC}	ATTCCCCATT
S_{CCCU}	ATTCCCUATT
S_{CCUC}	ATTCCUCATT
S_{CCCA}	ATTCCCAATT
S_{CCUA}	ATTCCUAATT
S_{CCUU}	ATTCCUUATT

The cytidines that are deaminated are underlined.
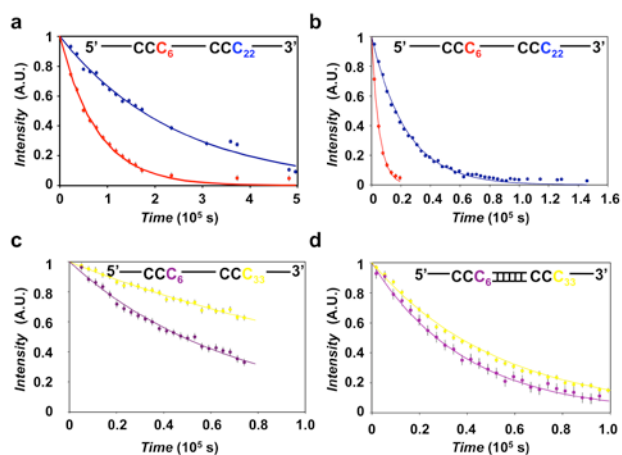
Figure 1. Real-time monitoring of deamination reactions at two CCC hot spots in an ssDNA. (a, b) The deamination reactions of full-length A3G (a) and CD2 (b) were monitored for C6 (red) and C22 (blue) in S_{2CCC}. The deamination rates for full-length A3G are $1.3 \times 10^{-5} \pm 4.0 \times 10^{-7}$ s^{-1} for C6 and $4.1 \times 10^{-6} \pm 7.4 \times 10^{-8}$ s^{-1} for C22, and those for CD2 are $2.0 \times 10^{-4} \pm 7.5 \times 10^{-6}$ s^{-1} for C6 and $4.2 \times 10^{-5} \pm 5.0 \times 10^{-7}$ s-1 for C22. (c, d) The deamination reactions of CD2 were monitored for C6 (purple) and C33 (yellow) in $S_{CCCssCCC}$ (c) and $S_{CCCdsCCC}$ (d).

We confirmed that real-time NMR is sufficiently sensitive to the location-dependent deamination and sliding processes of A3G CD2. Based on the results, we designed experiments and constructed a kinetic model to quantitatively analyze the deamination reaction. Since A3G binds to ssDNA nonspecifically, the binding rate should depend on the length of the ssDNA, and the concentrations of A3G and ssDNA. On the other hand, the sliding duration to reach a hot spot should depend on the length of the ssDNA and the position of a hot spot. Therefore, the kinetics of the deamination can be characterized by performing real-time NMR experiments with different concentrations of ssDNA and A3G CD2, different lengths of ssDNAs, and different positions of hotspots. According to this idea, we developed a theoretical kinetic model for the analysis of real-time NMR data (Supporting Information). The deamination reaction monitored by NMR as the change of the intensity, $I(t)$, is expressed as:

$$I(t) = I_0 \exp(-k_{deami}t) \qquad (1)$$

, where I_0 is the initial intensity, proportional to the concentration of ssDNA, and k_{deami} represents the apparent deamination rate calculated as below:

$$k_{deami} = \frac{[^N S][^N E]}{S_0 K_d}\{k_{cat(3'\to5')}(1-\alpha^n) + k_{cat(5'\to3')}(1-\alpha^{N-n+1})\} \qquad (2)$$

$$[^N S] = \frac{1}{2}\left(-{}^N\beta + S_0 - E_0 + \sqrt{({}^N\beta - S_0 + E_0)^2 + 4E_0\,{}^N\beta}\right) \qquad (3)$$

$$[^N E] = \frac{1}{2}\left(-{}^N\beta - S_0 + E_0 + \sqrt{({}^N\beta + S_0 - E_0)^2 + 4E_0\,{}^N\beta}\right) \qquad (4)$$

$$\alpha = k_s/(k_s + k_{off}), \quad {}^N\beta = \frac{K_d}{2\{N+1-(1-\alpha^{N+1})/(1-\alpha)\}} \qquad (5)$$

, where N and n represent the total number of nucleotides in ssDNA and the position of the reacting cytidine from the 3' end, respectively. In this kinetic model, A3G CD2 binds to any nucleotide of ssDNA at association rate k_{on}, and dissociates from ssDNA at dissociation rate k_{off}. A3G CD2 slides along ssDNA in both directions at sliding rate k_s (Figure 2a). Since the ssDNAs we used are relatively short (49-57 nucleotides), we assumed that A3G CD2 does not change in direction during a single sliding event. Note that the A3G-ssDNA complex structure has not been determined yet, thus how A3G recognizes a hot spot is unknown. On the other hand, the structure of free A3G CD2 has already been determined, and it has a protruding bump that hangs over one side of the catalytic pocket including the catalytic residue Glu259 and the zinc ion (Figure S1a).[14,17,20] We assumed that the accommodation of the target cytidine into the catalytic pocket is interfered with by the bump depending on the direction of approach of A3G CD2 (Figure S1b and c). Then, we incorporated two catalytic rates, $k_{cat(3'\to5')}$ and $k_{cat(5'\to3')}$, into the model, corresponding to the catalytic rates for the cases where A3G CD2 approaches the target cytidine from downstream and upstream, respectively. Since A3G slides along ssDNA without directional preference[12], the location-dependent deamination is difficult to explain without the two different k_{cat} values. These k_{cat} values represent the rates of the ssDNA's entry into the catalytic pocket and its deamination. A global fit of multiple real-time NMR data collected under different conditions, as mentioned above, provides parameters α, K_d ($= k_{off}/k_{on}$), $k_{cat(3'\to5')}$, and $k_{cat(5'\to3')}$. Note that it is difficult to separate α into k_s and k_{off} because either of the sliding or the dissociation process cannot be perturbed experimentally without affecting the other process. The NMR data are fitted to equation (1) using the program GLOVE.[21]

For this analysis, we obtained real-time NMR data for $S_{5'CCC}$, S_{mCCC} and $S_{3'CCC}$ (Table 1) in addition to the aforementioned two hot spots of $S_{CCCssCCC}$. These data fit well to equation (1), yielding $k_{cat(3'\to5')}$ of 68 s^{-1} and $k_{cat(5'\to3')}$ of 14 s^{-1} (Figure 2b). The closer a CCC hot spot is located to the 5' end, the more chance it

has of being detected by A3G approaching from downstream than from upstream; therefore a CCC hot spot located close to the 5' end is deaminated more rapidly than one less close to the 5' end.
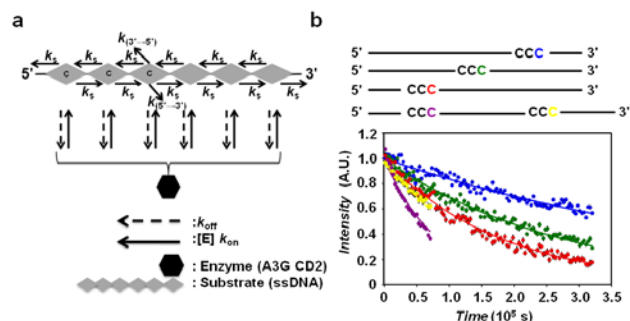

Figure 2. **Quantitative analysis of the location dependent-deamination reaction using the real-time NMR.** (a) A kinetic model of A3G deamination. (b) The real-time NMR data for C6 of $S_{5'CCC}$ (red), C26 of S_{mCCC} (green), C45 of $S_{3'CCC}$ (blue), and both C6 (purple) and C33 (yellow) of $S_{CCCssCCC}$ were fitted to equation (1).

For further applications of our real-time NMR, we analyzed the deamination of CCCC, in which two CCC sequences tandemly overlap (S_{CCCC} in Table 1). The clusters of 3-6 consecutive cytidines are scattered on the HIV genome, and CCCC is the second most abundant after CCC. Unfortunately, NMR peaks overlapped during the deamination reactions, for example, the TOCSY peaks of the fourth cytidines of CCC<u>C</u> and CCU<u>C</u>. Thus, the time course of each deamination reaction could not be monitored separately (Figure 3a and b). However, we characterized the deamination reactions using the NMR spectrum of the same ssDNA whose deamination reactions were fully completed. As a result, the third and fourth cytidines of CCCC were found to be deaminated with equal efficiency by A3G CD2 because the intensities of peaks 2 (CCU<u>C</u>) and 3 (CCC<u>U</u>+CCU<u>U</u>) were nearly identical. This finding is different from that of a previous study by gel shift assay; the third cytidine was deaminated more efficiently than the fourth.[15] This discrepancy could be caused by the difference in the monitoring durations between our NMR method (6 hours) and the gel shift assay (3 minutes). The real-time NMR can monitor multiple reactions directly and simultaneously using a single sample, whereas the gel shift assay cannot. Therefore, our results should provide more accurate information.

The deamination of CCCU into CCUU is redundant and may be unnecessary for abolition of the HIV infectivity. However, the above experiment could not provide information on the deamination efficiency as to CCCU because the NMR peaks of CC<u>U</u>U and CC<u>U</u>C overlapped. Thus, we monitored the deamination reactions as to CCCA (undeaminated site) and CCCU (singly deaminated site) to determine whether or not A3G has a preference for either of the two sequences, S_{CCCA} and S_{CCCU} (Table 1). In this case, the deamination reactions were monitored by one-dimensional ^{1}H NMR to increase the spectral resolution (Figure 3c and d). The results revealed that A3G preferentially deaminates an undeaminated site rather than a singly deaminated site because the intensity of the deaminated product CC<u>U</u>A (peak 4) was higher than that of CC<u>U</u>U (peak 3). A3G would exploit this sequential preference to efficiently deaminate multiple

hotspots scattered on the HIV genome with little redundant deamination of CCCC.
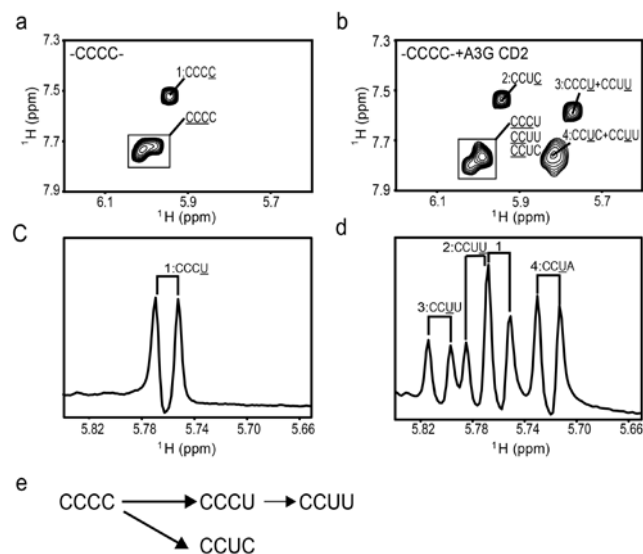
Figure 3. **The deamination reactions as to CCCC of A3G CD2.** (a, b) The TOCSY spectrum of S_{CCCC} (a) and that at 6 hours after the addition of A3G CD2 (b). The peaks are labeled with the sequences of the corresponding substrates or deaminated products in which the assigned residues are underlined (Supporting Information). (c, d) ^{1}H NMR spectrum of an equimolar mixture of S_{CCCA} and S_{CCCU} (c), and that at 40 minutes after the addition of A3G CD2 (d). Peaks 1 to 4 were assigned by measuring NMR spectra of S_{CCUU} and S_{CCUA}. (e) Two routes of the deamination reactions as to CCCC.

In conclusion, we analyzed the real-time NMR data for the location-dependent deamination by A3G CD2 using the newly derived kinetic model. As a result, the location-dependent deamination can be explained by the difference in k_{cat} of CD2 depending on the direction of approach to the target cytidine. We also characterized the deamination reactions as to CCCC using the real-time NMR even though multiple reactions occurred simultaneously. Previously, it was proposed that the location-dependent deamination is caused by two binding orientations on ssDNA and the existence of a deamination dead-region (~30 nucleotides) at the 3' end of the ssDNA.[10,14] However, this model cannot explain the location-dependent deamination observed for a long ssDNA (70 nucleotides)[15] because the contribution of the dead-region to the location-dependency becomes smaller. On the other hand, our model can explain the location-dependent deamination even for such a long ssDNA.

Recently, it was proposed that APOBEC family proteins are involved in epigenesis.[22] They may contribute to the removal of an epigenetic marker, a methyl group of a 5-methyl cytosine, through deamination of a 5-hydroxymethyl cytosine. Our real-time NMR method can also be applied to examine such DNA modifications and provide insights into epigenesis. Furthermore, this method can also be used for the analysis of post-translation modifications of proteins.

[1]   A. M. Sheehy, N. C. Gaddis, J. D. Choi, M. H. Malim, *Nature* **2002**, *418*, 646-650.

[2]   R. S. Harris, M. T. Liddament, *Nat Rev Immunol* **2004**, 4, 868-877.

[3]   D. Lecossier, F. Bouchonnet, F. Clavel, A. J. Hance, *Science* **2003**, *300*, 1112.

[4]   B. Mangeat, P. Turelli, G. Caron, M. Friedli, L. Perrin, D. Trono, *Nature* **2003**, *424*, 99-103.

[5]   H. Zhang, B. Yang, R. J. Pomerantz, C. Zhang, S. C. Arunachalam, L. Gao, *Nature* **2003**, *424*, 94-98.

[6]   R. S. Harris, M. T. Liddament, *Nat Rev Immunol* **2004**, *4*, 868-877.

[7]   G. Hache, M. T. Liddament, R. S. Harris, *J Biol Chem* **2005**, *280*, 10920-10924.

[8]   F. Navarro, B. Bollman, H. Chen, R. Konig, Q. Yu, K. Chiles, N. R. Landau, *Virology* **2005**, *333*, 374-386.

[9]   Q. Yu, R. Konig, S. Pillai, K. Chiles, M. Kearney, S. Palmer, D. Richman, J. M. Coffin, N. R. Landau, *Nat Struct Mol Biol* **2004**, *11*, 435-442.

[10]   L. Chelico, P. Pham, P. Calabrese, M. F. Goodman, *Nat Struct Mol Biol* **2006**, *13*, 392-399.

[11]   L. S. Shlyakhtenko, A. Y. Lushnikov, A. Miyagi, M. Li, R. S. Harris, Y. L. Lyubchenko, *Biochemistry* **2012**, *51*, 6432-6440.

[12]   G. Senavirathne, M. Jaszczur, P. A. Auerbach, T. G. Upton, L. Chelico, M. F. Goodman, D. Rueda, *J Biol Chem* **2012**, *287*, 15826-15835.

[13]   R. Suspène, C. Rusniok, J. P. Vartanian, S. Wain-Hobson, *Nucleic Acids Res* **2006,** *34*, 4677-4684.

[14]   L. G. Holden, C. Prochnow, Y. P. Chang, R. Bransteitter, L. Chelico, U. Sen, R. C. Stevens, M. F. Goodman, X. S. Chen, *Nature* **2008**, *456*, 121-124.

[15]   R. Nowarski, E. Britan-Rosich, T. Shiloach, M. Kotler, *Nat Struct Mol Biol* **2008**, *15*, 1059-1066.

[16]   L. Chelico, C. Prochnow, D. A. Erie, X. S. Chen, M. F. Goodman, *J Biol Chem* **2010**, *285*, 16195-16205.

[17]   A. Furukawa, T. Nagata, A. Matsugami, Y. Habu, R. Sugiyama, F. Hayashi, N. Kobayashi, S. Yokoyama, H. Takaku, M. Katahira, *EMBO J* **2009**, *28*, 440-451.

[18]   S. Harjes, W. C. Solomon, M. Li, K. M. Chen, E. Harjes, R. S. Harris, H. Matsuo, *J Virol* **2013**, *87*, 7008-7014.

[19]   I. J. Byeon, J. Ahn, M. Mitra, C. H. Byeon, K. Hercik, J. Hritz, L. M. Charlton, J. G. Levin, A. M. Gronenborn, *Nature Commun* **2013**, *4*, 1890.

[20]   K. M. Chen, E. Harjes, P. J. Gross, A. Fahmy, Y. Lu, K. Shindo, R. S. Harris, H. Matsuo, *Nature* **2008**, *452*, 116-119.

[21]   K. Sugase, T. Konuma, J. C. Lansing, P. E. Wright, *J Biomol NMR* **2013**, *56*, 275-283.

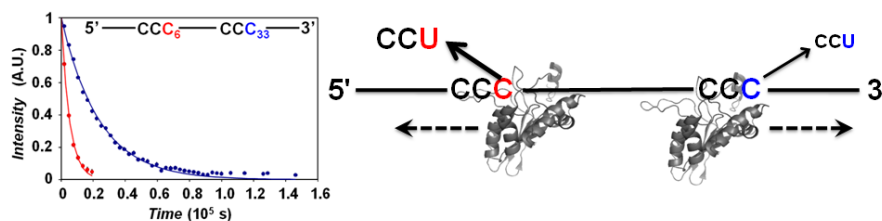[22]   N. Bhutani, D. M. Burns, H. M. Blau, *Cell* **2011**, *146*, 866-872.

APOBEC3G (A3G) efficiently deaminates cytidines located close to the 5' end of single- stranded minus DNA of the HIV-1 genome. However, this mechanism is elusive because no suitable method exists to analyze the deamination process that includes A3G non-specific binding to DNA and sliding along it. Here, we quantitated this process using a newly developed real-time NMR method. As a result, the location-dependent deamination can be explained by two catalytic rates depending on the direction of approach to the target cytidine.

Materials and methods

Preparation of A3G

The cDNAs of full-length and CD2 A3G were inserted into the pEU-E01-GST-ps-Flag expression vector (Cell Free Sciences). Full-length A3G with a GST tag was synthesized using the wheat germ cell-free protein expression system as described previously.[1] The synthesized protein was loaded onto a Glutathione-Sepharose resin column (GE Healthcare), and the GST tag was cleaved with TEV protease on the column. A3G was eluted from the column with PBS (pH 7.4) containing 5 mM DTT, 10 μM ZnCl$_2$, and 0.04 % Briji35. The protein solution was concentrated using an ultrafiltration cartridge (Millipore). For A3G CD2, the protein was expressed and purified as described previously.[2] After purification, the protein was dialyzed against the same buffer as that for full-length A3G and concentrated.

Monitoring of deamination using NMR

Each DNA purchased from Fasmac was dissolved in PBS (pH 7.4) containing 5 mM DTT, 10 μM ZnCl$_2$, and 0.04% Briji35 at a concentration of 220 μM. The molar ratios of A3G:ssDNA were 1:44 (Figure 1a,b), 1:100 (Figure 1c,d and 2), and 1:5 (Figure 3). After the addition of A3G, a series of ^{1}H one- and two-dimensional TOCSY spectra of DNA were recorded with a mixing time of 20 ms at 25 ºC to monitor the deamination in real-time on a Bruker DRX 600 NMR spectrometer equipped with a cryoprobe and a Z-gradient (Bruker Biospin). The experimental error was estimated from the noise level of the TOCSY spectra, and the real-time NMR data were fitted to $I(t) = I_0\exp(-k_{\text{deami}}t)$ to derive an apparent

deamination rate, k_{deami}.

Derivation of a theoretical kinetic model for the analysis of real-time NMR data

A general enzyme reaction is expressed as:

$$E + S \leftrightarrow ES \rightarrow EP \rightarrow E + P$$
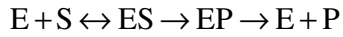
, where E, S, and P represent enzyme, substrate and product, respectively. The enzyme-substrate complex ES completely changes into the enzyme-product complex EP, provided that the enzyme dissociates from the substrate much slower than the catalytic reaction occurs. The enzyme does not rebind to the product. In the case of A3G, however, the deamination reaction does not always occur even if A3G binds to ssDNA because A3G can form non-active complexes with ssDNA.[3, 4] In our kinetic model, A3G binds to any nucleotide of ssDNA at binding rate k_{on}, and dissociates from ssDNA at dissociation rate k_{off}, because A3G exhibits no sequential preference for binding. [5] This model also includes multiple separate interactions between A3G and ssDNA. As explained in the main text, our model assumes that A3G slides along ssDNA in both directions at sliding rate k_{s}, without a change in direction during a single sliding event. The catalytic rate is defined as $k_{\text{cat}(3' \rightarrow 5')}$ when A3G approaches a target cytidine from downstream, whereas it is $k_{\text{cat}(5' \rightarrow 3')}$ when A3G approaches from upstream (Figure 3a). Although A3G cannot deaminate a single cytidine, we first describe a hypothetical case where A3G can deaminate a single cytidine, and the total number of nucleotides in ssDNA is set to 3 for simplicity. Namely, A3G deaminates cytidines in the sequences Cxx xCx, and xxC (x is a nucleotide other than cytidine). The theoretical time courses of the intensities of cytidines located at the first, second, and third positions from the 3' end, $I_i(t)$ (i=1, 2, 3) are expressed as follows, respectively:

$$I_1(t) = I_0 \exp\left\{-\left[\frac{[ES_1]}{S_0}k_{cat(3' \to 5')} + \frac{[ES_{1'}]}{S_0}k_{cat(5' \to 3')}\right]t\right\}$$

$$I_2(t) = I_0 \exp\left\{-\left[\frac{[ES_2]}{S_0}k_{cat(3' \to 5')} + \frac{[ES_{2'}]}{S_0}k_{cat(5' \to 3')}\right]t\right\} \quad \text{(S1)}$$

$$I_3(t) = I_0 \exp\left\{-\left[\frac{[ES_3]}{S_0}k_{cat(3' \to 5')} + \frac{[ES_{3'}]}{S_0}k_{cat(5' \to 3')}\right]t\right\}$$

,where I_0 is the initial intensity, proportional to the total concentration of ssDNA. $[ES_1]$, $[ES_2]$, and $[ES_3]$ are the concentrations of ssDNA bound to A3G under the condition that A3G slides toward the 5' end. $[ES_{1'}]$, $[ES_{2'}]$, and $[ES_{3'}]$ are the same as $[ES_1]$, $[ES_2]$, and $[ES_3]$ except that A3G slides toward the 3' end. Subscripts indicate the positions of the cytidines from the 3' end. $[ES_i]/S_0$ (i=1, 2, 3, 1', 2', 3') represents the fractional population of each A3G:ssDNA complex. When the concentrations of free ssDNA, free A3G and total A3G are defined as [S], [E], and E_0, respectively,

$$[S]+[ES_1]+[ES_2]+[ES_3]+[ES_{1'}]+[ES_{2'}]+[ES_{3'}] = S_0 \quad \text{(S2)}$$

$$[E]+[ES_1]+[ES_2]+[ES_3]+[ES_{1'}]+[ES_{2'}]+[ES_{3'}] = E_0 \quad \text{(S3)}$$

We assume that A3G interacts with the products (Uxx, xUx, and xxU) at the same k_{on} and k_{off} as for the substrates. Then, the concentration of A3G at each position of ssDNA is held in equilibrium even if the cytidine is deaminated. In this case, the concentrations in (6) are derived by solving the following differential equation under the condition that $d\mathbf{C}/dt$=0:

$$\frac{d}{dt}\mathbf{C} = \mathbf{KC}$$

$$\mathbf{C} = \left([S] \quad [ES_1] \quad [ES_2] \quad [ES_3] \quad [ES_{1'}] \quad [ES_{2'}] \quad [ES_{3'}]\right)^T$$

$$\mathbf{K} = \begin{pmatrix} -6[E]k_{on} & k_{off} & k_{off} & k_{off}+k_s & k_{off}+k_s & k_{off} & k_{off} \\ [E]k_{on} & -k_{off}-k_s & 0 & 0 & 0 & 0 & 0 \\ [E]k_{on} & k_s & -k_{off}-k_s & 0 & 0 & 0 & 0 \\ [E]k_{on} & 0 & k_s & -k_{off}-k_s & 0 & 0 & 0 \\ [E]k_{on} & 0 & 0 & 0 & -k_{off}-k_s & k_s & 0 \\ [E]k_{on} & 0 & 0 & 0 & 0 & -k_{off}-k_s & k_s \\ [E]k_{on} & 0 & 0 & 0 & 0 & 0 & -k_{off}-k_s \end{pmatrix}$$

(S4)

Simple mathematical manipulation together with (7) and (8) provides,

$$[ES_1] = [ES_{3'}] = [E][S](1-\alpha)/K_d$$
$$[ES_2] = [ES_{2'}] = [E][S](1+\alpha)(1-\alpha)/K_d$$
$$[ES_3] = [ES_{1'}] = [E][S](1+\alpha+\alpha^2)(1-\alpha)/K_d$$

(S5)

$$[S] = \frac{1}{2}\left(-\beta+S_0-E_0+\sqrt{(\beta-S_0+E_0)^2+4E_0\beta}\right)$$

(S6)

$$[E] = \frac{1}{2}\left(-\beta-S_0+E_0+\sqrt{(\beta+S_0-E_0)^2+4E_0\beta}\right)$$

(S7)

$$\alpha = k_s/(k_{off}+k_s), \qquad \beta = \frac{K_d}{2\{3+1-(1-\alpha^{3+1})/(1-\alpha)\}}$$

(S8)

,where K_d is the dissociation constant ($=k_{off}/k_{on}$). By substituting (10) for (6),

$$I_1(t) = I_0\exp\left\{-\frac{[E][S](1-\alpha)}{S_0K_d}\left[k_{cat(3'\to5')}+k_{cat(5'\to3')}(1+\alpha+\alpha^2)\right]t\right\}$$

$$I_2(t) = I_0\exp\left\{-\frac{[E][S](1-\alpha)}{S_0K_d}\left[k_{cat(3'\to5')}(1+\alpha)+k_{cat(5'\to3')}(1+\alpha)\right]t\right\}$$

(S9)

$$I_3(t) = I_0\exp\left\{-\frac{[E][S](1-\alpha)}{S_0K_d}\left[k_{cat(3'\to5')}(1+\alpha+\alpha^2)+k_{cat(5'\to3')}\right]t\right\}$$

In the case where the total number of nucleotides is N.

$$^NI_n(t) = I_0\exp\left\{-\frac{[^NE][^NS](1-\alpha)}{S_0K_d}\left[k_{cat(3'\to5')}(1+\alpha+\alpha^2+\cdots+\alpha^{n-1})+k_{cat(5'\to3')}(1+\alpha+\alpha^2+\cdots+\alpha^{N-n+1})\right]t\right\}$$

$$= I_0\exp\left\{-\frac{[^NE][^NS]}{S_0K_d}\left[k_{cat(3'\to5')}(1-\alpha^n)+k_{cat(5'\to3')}(1-\alpha^{N-n+1})\right]t\right\}$$

(S10)

$$[^{N}S] = \frac{1}{2}\left(-^{N}\beta + S_0 - E_0 + \sqrt{\left(^{N}\beta - S_0 + E_0\right)^2 + 4E_0\,^{N}\beta}\right)$$

$$[^{N}E] = \frac{1}{2}\left(-^{N}\beta - S_0 + E_0 + \sqrt{\left(^{N}\beta + S_0 - E_0\right)^2 + 4E_0\,^{N}\beta}\right)$$

(S11)

$$^{N}\beta = \frac{K_d}{2\left\{N + 1 - \left(1 - \alpha^{N+1}\right)/(1-\alpha)\right\}}$$

The superscripts N and n indicate the total number of nucleotides of ssDNA and the position of the reacting cytidine from the 3' end of ssDNA, respectively.

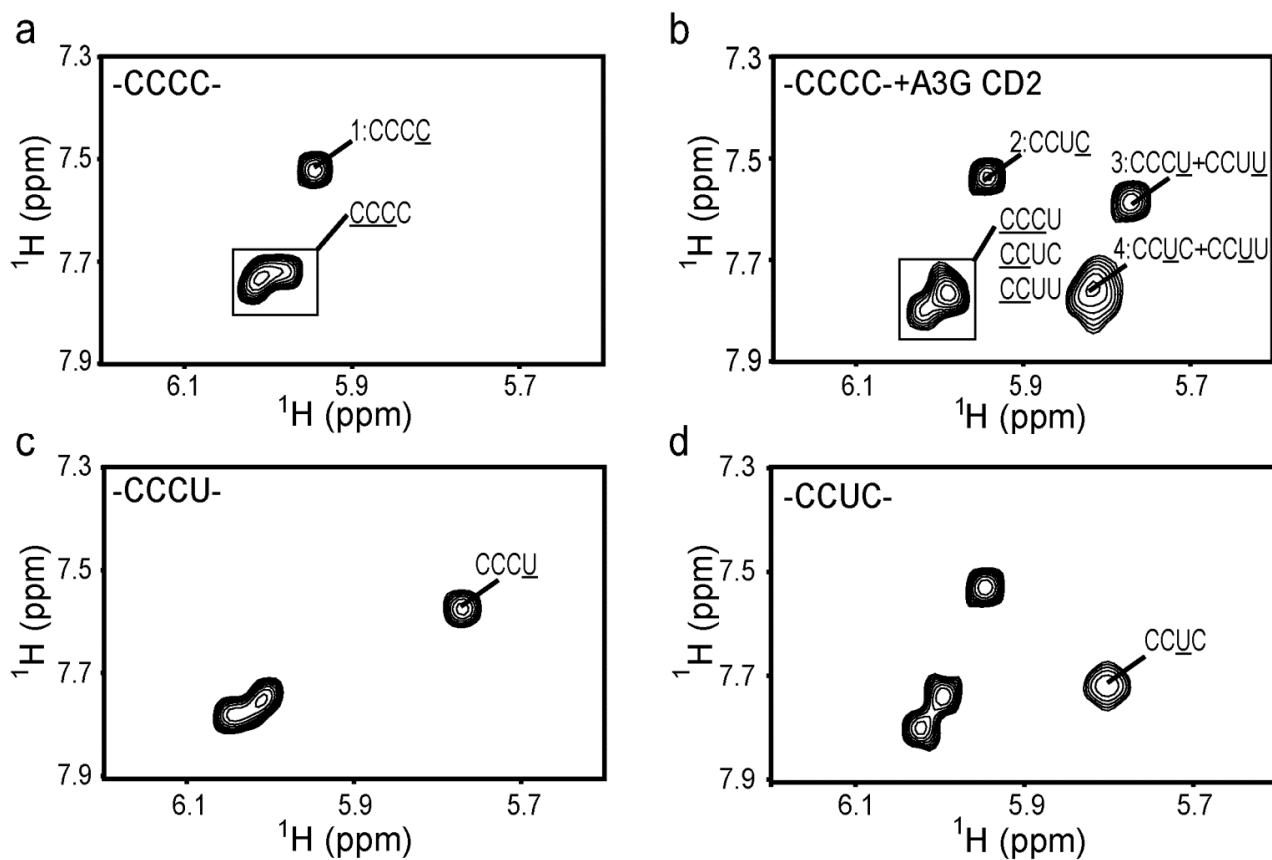**Chemical shift assignments of the cytidines and uridines in Figure 3a and b**

TOCSY spectra of S_{CCCC} were recorded at 0 and 6 hours after the addition of A3G CD2. To assign peaks 1-4, TOCSY spectra of S_{CCCU} and S_{CCUC} were also recorded (Figure S2). Through comparison between Figure S2a and c, peak 1 was assigned to the fourth cytidine of CCC<u>C</u> (corresponding cytidine or uridine is underlined). The intensity of peak 2 did not change any more after 6-hour incubation, and thus the reactions CCCC → CCCU and CCCC → CCUC should have been completed. Therefore, peak 2 was assigned to the fourth cytidine of the CCU<u>C</u>. Through comparison between Figures S2b and c, and between Figure S2b and d, peaks 3 and 4 were assigned to the fourth uridine of CCC<u>U</u> and the third uridine of CC<u>U</u>C, respectively. Peak 4 is broad, and thus it should include the correlation peak for the third uridine of CC<u>U</u>U. Similarly, peak 3 should contain the correlation peak of CCU<u>U</u> because of its structural similarity to CCUC. The absence of broadening for peak 3 implies that the chemical shifts for these two peaks are virtually identical.
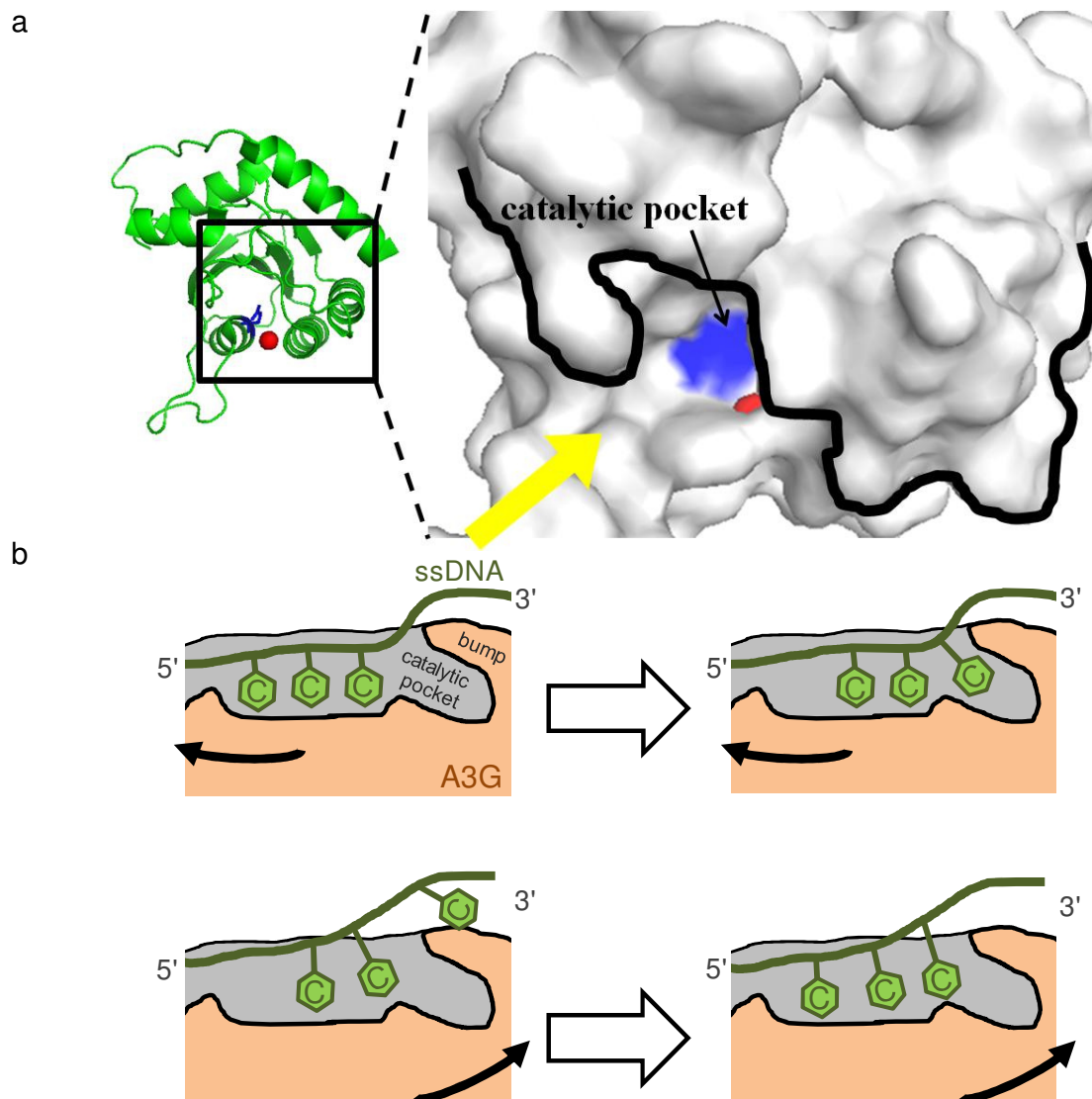
Figure S1. Model for the sliding-direction-dependent binding to the catalytic pocket.

(a) Overall structure of A3G CD2 (PDB:3E1U) and a closed view of the catalytic pocket as a surface

representation. The catalytic pocket includes Glu259 (blue) and a zinc ion (red), which are important for

the deamination reaction. The putative bump is highlightened by a bold line. (b) Schematic diagrams of

sliding-direction-dependnet entry of cytidine into the catalytic pocket. Sliced views at the yellow arrow in

(a) are illustrated. The hollow cavity and the bump in A3G are depicted in gray and orange, respectively. The ssDNA backbone and cytidines are depicted as dark green lines and light green hexagons, respectively. Black arrows indicate the direction of A3G sliding. When A3G approaches a CCC hot spot from downstream (3'), the third cytidine of CCC can be ideally accommodated into the catalytic pocket for the deamination reaction (upper panel). On the other hand, when A3G approaches from upstream (5'), the third cytidine encounters a bump before reaching the pocket. Thus, the cytidine needs to overcome the bump to reach the pocket, causing overrunning of the cytidine. In this case, the cytidine cannot be ideally accommodated in the catalytic pocket (bottom panel), resulting in reduction of the catalytic rate.

Figure S2. The chemical shifts assignments of the deamination products of CCCC.

The H5-H6 TOCSY spectrum of S_{CCCC} (a) and that at six hours after the addition of A3G CD2 (b). The molar ratio of A3G:S_{CCCC} was 1:5. The H5-H6 TOCSY correlation peaks of S_{CCCU} (c) and those of S_{CCUC} (d). The peaks are labeled with the sequence of the corresponding substrate or deaminated product in which the assigned residue is shown in underline.

REFERENCES

[S1] K. Takai, T. Sawasaki, Y. Endo, *Nature Protocols* **2010**, *5*, 227-238.

[S2] A. Furukawa, T. Nagata, A. Matsugami, Y. Habu, R. Sugiyama, F. Hayashi, N. Kobayashi, S. Yokoyama, H. Takaku, M. Katahira, *EMBO J* **2009**, *28*, 440-451.

[S3] L. S. Shlyakhtenko, A. Y. Lushnikov, A. Miyagi, M. Li, R. S. Harris, Y. L. Lyubchenko, *Biochemistry* **2012**, *51*, 6432-6440.

[S4] G. Senavirathne, M. Jaszczur, P. A. Auerbach, T. G. Upton, L. Chelico, M. F. Goodman, D. Rueda, *J Biol Chem* **2012**, *287*, 15826-15835.

[S5] L. Chelico, C. Prochnow, D. A. Erie, X. S. Chen, M. F. Goodman, *J Biol Chem* **2010**, *285*, 16195-16205.