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, Akitumi Takaorikondo, Akihide Ryo, and Masato Katahira*

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

A3G possesses two consensus zinc-finger type cytidine deaminase motifs (CD1 and CD2),[7,8] but only CD2 is catalytically active.[7,9] 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,[11,14] 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 CCCC much faster than the second one (CC+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 CCCC 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_{XCCC}, ATTCCCAAT and ATACCCATTT 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 SCCCssCCC 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

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>SCCC</td>
<td>ATCCGGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTATCCGGATTTT</td>
</tr>
<tr>
<td>SCCCdsCCC</td>
<td>ATCCGGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTATCCGGATTTTTCGCCGCGCTGTTGCGG</td>
</tr>
<tr>
<td>SCCCssCCC</td>
<td>ATCCGGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTATCCGGATTTTTCGCCGCGCTGTTGCGG</td>
</tr>
<tr>
<td>SCCC</td>
<td>ATCCGGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTATCCGGATTTT</td>
</tr>
<tr>
<td>SCCC</td>
<td>ATCCGGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTATCCGGATTTT</td>
</tr>
<tr>
<td>SCCC</td>
<td>ATCCGGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTATCCGGATTTT</td>
</tr>
<tr>
<td>SCCC</td>
<td>ATCCGGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTATCCGGATTTT</td>
</tr>
<tr>
<td>SCCC</td>
<td>ATCCGGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTATCCGGATTTT</td>
</tr>
<tr>
<td>SCCC</td>
<td>ATCCGGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTATCCGGATTTT</td>
</tr>
<tr>
<td>SCCC</td>
<td>ATCCGGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTATCCGGATTTT</td>
</tr>
<tr>
<td>SCCC</td>
<td>ATCCGGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTATCCGGATTTT</td>
</tr>
<tr>
<td>SCCC</td>
<td>ATCCGGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTATCCGGATTTT</td>
</tr>
</tbody>
</table>

The cytidines that are deaminated are underlined.

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:

\[
I(t) = I_0 \exp(-k_{\text{total}} t)
\]

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

\[
k_{\text{total}} = \frac{1}{S_k K_k} [S] [E]
\]

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

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

\[
a = k_a (k_c + k_{\text{off}}) \frac{\beta}{2(N+1-|1-a|)}
\]

\[
b = K_k \frac{\beta}{2(N+1-|1-a|)}
\]

A global fit of multiple real-time NMR data collected under different conditions, as mentioned above, provides parameters \( a, K_k \) (\( = k_{\text{off}}/k_{\text{on}} \)), \( k_{\text{cat}(3'\rightarrow5')} \), and \( k_{\text{cat}(5'\rightarrow3')} \), into the model, corresponding to the catalytic rates for the cases where A3G CD2 approaches the target cytidine from upstream and downstream, respectively. Since A3G slides along ssDNA without directional preference\[22\], the location-dependent deamination is difficult to explain without the two different \( k_{\text{cat}} \) values. These \( k_{\text{cat}} \) 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 \( a, K_k \) (\( = k_{\text{off}}/k_{\text{on}} \)), \( k_{\text{cat}(3'\rightarrow5')} \), and \( k_{\text{cat}(5'\rightarrow3')} \). Note that it is difficult to separate \( a \) into \( k_{\text{on}} \) and \( k_{\text{off}} \) 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 SCCC, SCCCdsCCC and SCCCssCCC (Table 1) in addition to the aforementioned two hot spots of SCCCssCCC. These data fit well to equation (1), yielding \( k_{\text{cat}(3'\rightarrow5')} \) of 68 s\(^{-1}\) and \( k_{\text{cat}(5'\rightarrow3')} \) 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 CCC (red), C26 of S CCC (green), C45 of S CCC (blue), and both C6 (purple) and C33 (yellow) of S CCC 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 CCC 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 CCCC and CCCC. 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 (CCUC) and 3 (CCUC+[CCUC]) 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. 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 CCCC 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 CCUU because the NMR peaks of CCCC and CCCC overlapped. Thus, we monitored the deamination reactions as to CCCC (undeaminated site) and CCCC (singly deaminated site) to determine whether or not A3G has a preference for either of the two sequences, S CCC and S CCC (Table 1). In this case, the deamination reactions were monitored by one-dimensional 1H 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 CCU (peak 4) was higher than that of CCU (peak 3). A3G would exploit this sequential preference to efficiently deaminate multiple hotspots scattered on the HIV genome with little redundant deamination of 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 kcat 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. However, this model cannot explain the location-dependent deamination observed for a long ssDNA (70 nucleotides) 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. 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.

Received: ((will be filled in by the editorial staff))
Published online on ((will be filled in by the editorial staff))
Quantitative analysis of the location- and sequence-dependent deamination by APOBEC3G using real-time NMR.

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.
Supporting Information

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. 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 ZnCl₂, 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. 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 ZnCl₂, 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 ¹H 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
Derivation of a theoretical kinetic model for the analysis of real-time NMR data

A general enzyme reaction is expressed as:

\[ \text{E} + \text{S} \leftrightarrow \text{ES} \rightarrow \text{EP} \rightarrow \text{E} + \text{P} \]

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 \( k_{\text{on}} \), and dissociates from ssDNA at dissociation rate \( k_{\text{off}} \), 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 \( k_{\text{cat}(3'\rightarrow5')} \) when A3G approaches a target cytidine from downstream, whereas it is \( k_{\text{cat}(5'\rightarrow3')} \) 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:
\[
I_1(t) = I_0 \exp\left\{ -\frac{[\text{ES}_1]}{S_0} k_{\text{cat}(3'\rightarrow5')} + \frac{[\text{ES}_1]}{S_0} k_{\text{cat}(5'\rightarrow3')} \frac{t}{k_1} \right\}
\]

\[
I_2(t) = I_0 \exp\left\{ -\frac{[\text{ES}_2]}{S_0} k_{\text{cat}(3'\rightarrow5')} + \frac{[\text{ES}_2]}{S_0} k_{\text{cat}(5'\rightarrow3')} \frac{t}{k_2} \right\}
\]

\[
I_3(t) = I_0 \exp\left\{ -\frac{[\text{ES}_3]}{S_0} k_{\text{cat}(3'\rightarrow5')} + \frac{[\text{ES}_3]}{S_0} k_{\text{cat}(5'\rightarrow3')} \frac{t}{k_3} \right\}
\]

(S1)

, where \( I_0 \) is the initial intensity, proportional to the total concentration of ssDNA. \([\text{ES}_1], [\text{ES}_2], \text{ and } [\text{ES}_3] \) are the concentrations of ssDNA bound to A3G under the condition that A3G slides toward the 5’ end. \([\text{ES}_1'], [\text{ES}_2'], \text{ and } [\text{ES}_3'] \) are the same as \([\text{ES}_1], [\text{ES}_2], \text{ and } [\text{ES}_3] \) except that A3G slides toward the 3’ end. Subscripts indicate the positions of the cytidines from the 3’ end. \([\text{ES}_i]/S_0 (i=1, 2, 3, 1', 2', 3') \) 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], \text{ and } E_0, \text{ respectively},
\]

\[
[S] + [\text{ES}_1] + [\text{ES}_2] + [\text{ES}_3] + [\text{ES}_1'] + [\text{ES}_2'] + [\text{ES}_3'] = S_0
\]

(S2)

\[
[E] + [\text{ES}_1] + [\text{ES}_2] + [\text{ES}_3] + [\text{ES}_1'] + [\text{ES}_2'] + [\text{ES}_3'] = E_0
\]

(S3)

We assume that A3G interacts with the products (Uxx, xUx, and xxU) at the same \( k_{\text{on}} \) and \( k_{\text{off}} \) 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 \( dC/dt=0 \):
\[ \frac{d}{dt} C = KC \]

\[ C = \begin{bmatrix} [S] & [ES_1] & [ES_2] & [ES_3] \end{bmatrix}^T \]

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

(S4)

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

\[
[ES_1] = [ES_2] = [E][S](1 - \alpha)/K_d
[ES_2] = [ES_3] = [E][S](1 + \alpha)(1 - \alpha)/K_d
[ES_3] = [ES_4] = [E][S](1 + \alpha^2)(1 - \alpha)/K_d
\]

(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)

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

(S8)

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

\[
I_i(t) = I_0 \exp \left\{ -\frac{[E][S](1 - \alpha)}{S_0 K_d} \left[ k_{\text{cat}(3\rightarrow5)} + k_{\text{cat}(5\rightarrow3)} (1 + \alpha + \alpha^2) \right] \right\}
\]

\[
I_2(t) = I_0 \exp \left\{ -\frac{[E][S](1 - \alpha)}{S_0 K_d} \left[ k_{\text{cat}(3\rightarrow5)} (1 + \alpha) + k_{\text{cat}(5\rightarrow3)} (1 + \alpha) \right] \right\}
\]

\[
I_3(t) = I_0 \exp \left\{ -\frac{[E][S](1 - \alpha)}{S_0 K_d} \left[ k_{\text{cat}(3\rightarrow5)} (1 + \alpha + \alpha^2) + k_{\text{cat}(5\rightarrow3)} \right] \right\}
\]

(S9)

In the case where the total number of nucleotides is \(N\).

\[
^N I_n(t) = I_0 \exp \left\{ -\frac{[E]^n[S](1 - \alpha)}{S_0 K_d} \left[ k_{\text{cat}(3\rightarrow5)} (1 + \alpha + \alpha^2 + \cdots + \alpha^{n-1}) + k_{\text{cat}(5\rightarrow3)} (1 + \alpha + \alpha^2 + \cdots + \alpha^{N-n+1}) \right] \right\}
\]

\[
= I_0 \exp \left\{ -\frac{[E]^n[S]}{S_0 K_d} \left[ k_{\text{cat}(3\rightarrow5)} (1 - \alpha^n) + k_{\text{cat}(5\rightarrow3)} (1 - \alpha^{N-n+1}) \right] \right\}
\]

(S10)
\[ [^N S] = \frac{1}{2} \left( -[^N \beta + S_0 - E_0 + \sqrt{[^N \beta - S_0 + E_0]^2 + 4E_0[^N \beta]} \right) \]
\[ [^N E] = \frac{1}{2} \left( -[^N \beta - S_0 + E_0 + \sqrt{[^N \beta + S_0 - E_0]^2 + 4E_0[^N \beta]} \right) \]
\[ ^N \beta = \frac{K_d}{2[N + 1 - (1 - \alpha^{N+1}]/(1 - \alpha)]} \]

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 SCCCC were recorded at 0 and 6 hours after the addition of A3G CD2. To assign peaks 1-4, TOCSY spectra of SCCCCU and SCCUC were also recorded (Figure S2). Through comparison between Figure S2a and c, peak 1 was assigned to the fourth cytidine of CCCC (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 \( \rightarrow \) CCCU and CCCC \( \rightarrow \) CCUC should have been completed. Therefore, peak 2 was assigned to the fourth cytidine of the CCUCC. Through comparison between Figures S2b and c, and between Figure S2b and d, peaks 3 and 4 were assigned to the fourth uridine of CCCU and the third uridine of CCUC, respectively. Peak 4 is broad, and thus it should include the correlation peak for the third uridine of CCU. Similarly, peak 3 should contain the correlation peak of CCUU 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


